

**Mutational Studies on 17 $\beta$ -HSD14, Serial Synchrotron X-ray  
Crystallography, Solubility Enhancement using Cyclodextrins  
and Fragment-Based Drug Discovery  
Multiple Blocks to Pave the Road of Drug Design**

**Dissertation**

zur

Erlangung des Doktorgrades  
der Naturwissenschaften  
(Dr. rer. nat.)

dem

Fachbereich Pharmazie der  
Philipps-Universität Marburg  
vorgelegt

von

**Mohammed Badran**

Aus

Ashrafiah (Damaskus)

Marburg, 2020

Erstgutachter: Prof. Dr. Gerhard Klebe  
Institut für Pharmazeutische Chemie  
Philipps-Universität Marburg

Zweitgutachter: Prof. Dr. Andreas Heine  
Institut für Pharmazeutische Chemie  
Philipps-Universität Marburg

Eingereicht am  
Tag der mündlichen Prüfung am  
Hochschulkennziffer: 1180

Die Untersuchungen der vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. Gerhard Klebe am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von November 2016 bis September 2020 angefertigt.



## Summary

The first topic of this thesis (Chapter 2) presents a mutational study performed on 17 $\beta$ -hydroxysteroid dehydrogenase type 14 (17 $\beta$ -HSD14) S205 variant. Five different mutations were done with respect to five amino acids which are believed to have an essential role in the enzyme activity and assembly. The five variants are: His93Ala, Gln148Ala, Lys158Ala, Tyr253Ala and Cys255Ala. The mutated amino acids are located in the active site of the enzyme (His93, Gln148 and Lys158) or on a flexible loop of the enzyme, which is located above the active site (Tyr253 and Cys255). X-ray crystallography is the method utilized in this study to obtain a 3D crystal structure of each variant. A non-steroidal potent 17 $\beta$ -HSD14 inhibitor (inhibitor **1**) has been crystallized in complex with each variant, that has been used to verify the binding capability of the mutated enzyme. Enzymatic assays have been performed with each variant to compare the activity of each one. Estrogen (estradiol) and androgen (5-diol) have been used as a substrate in the enzyme kinetics assay with NAD<sup>+</sup> as a cofactor.

The second part of this thesis (Chapter 3) is focused on a new crystal sample holder (the Roadrunner I chip) which is used in Serial Synchrotron X-ray Crystallography (SSX). The Roadrunner I chip is a micro-patterned sample holder from single crystalline silicon (waiver technology) with micropores. The aim of using the Roadrunner I chip is to have a sample holder that can present hundreds to thousands of crystals to the high intensity PETRA III beam line P11 (DESY – Hamburg) without interfering with the diffraction pattern. In this study, Thermolysin (TLN) is the protein used to test the limit of this new method. Thermolysin crystals were grown, washed, soaked and frozen at cryogenic temperature without removing them from the chip. Data sets were collected of TLN crystals while they are located on the chip. The experimental part of this study was performed at *Deutsches Elektronen-Synchrotron* (DESY), Hamburg at PETRA III P11 beamline in collaboration with associated laboratories at the facility.

The third part of this thesis (Chapter 4) discusses cyclodextrins (CDs) and their ability to enhance hydrophobic compounds solubility in aqueous solutions. The targeted protein in this study is 17 $\beta$ -HSD14. Many compounds were assembled for this study, such as a fluorine-compound library, hydrophobic drugs and sex hormones. The aim of this study is to obtain a compound that binds to the enzyme by introducing it as a compound/CD complex. Most of the

compounds used in this study have already been tested with 17 $\beta$ -HSD14 without the use of CDs, but due to their low solubility it was not possible to introduce them in crystallization samples of the enzyme. The data obtained from this study show the effect of the compound/CD complex, as it is introduced to the enzyme via co-crystallization method.

The fourth part of this thesis (Chapter 5) focuses on a fragment screening. A 96-fragment library is screened against trypsin using X-ray crystallography. This study focuses on the difference of hits and partial hits obtained from the fragment screening. Fragment screening has been performed on two trypsin crystal form (trigonal and orthorhombic). The data obtained from this study show the different results from each screen and how the crystal form and the fragment delivery method influence the hit ratio. Many aspects were considered in this study, such as the difference in electron density, volume of the binding pocket, anomalous peaks and water channels.

## Zusammenfassung

Das erste Kapitel dieser Dissertation (Kapitel 2) präsentiert eine Mutationsstudie, die an der 17 $\beta$ -Hydroxysteroid-Dehydrogenase Typ 14 (17 $\beta$ -HSD14) S205-Variante durchgeführt wurde. Dafür wurden fünf Aminosäuren mutiert, von denen man annimmt, dass sie eine wesentliche Rolle bei der Enzymaktivität und dem strukturellen Aufbau spielen. Die fünf mutierten Aminosäuren sind: His93Ala, Gln148Ala, Lys158Ala, Tyr253Ala und Cys255Ala. Alle mutierten Aminosäuren befinden sich im aktiven Zentrum des Enzyms (His93, Gln148 und Lys158) oder auf der flexiblen Schleife, die sich über dem aktiven Zentrum befindet (Tyr253 und Cys255). Die Röntgenkristallographie ist die Methode der Wahl, die in dieser Studie verwendet wird, um die Kristallstruktur jeder Variante zu erhalten. Ein nicht-steroidal potenter 17 $\beta$ -HSD14-Inhibitor (Inhibitor 1) wurde in einem Komplex mit dem Wildtyp kristallisiert, der zur Überprüfung der Bindungsfähigkeit des mutierten Enzyms verwendet wurde. Mit jeder Variante wurde ein enzymatischer Assay durchgeführt, um die Aktivität der einzelnen Varianten zu vergleichen. Östrogen (Östradiol) und Androgen (5-Diol) wurden als Substrat im Enzymkinetik-Assay mit NAD<sup>+</sup> als Co-Faktor verwendet.

Der zweite Teil der Arbeit (Kapitel 3) konzentriert sich auf einen neuen Kristallprobenhalter (The Roadrunner I Chip), der in der seriellen Röntgenkristallographie (SSX) eingesetzt wird. Der Roadrunner I-Chip ist ein mikrostrukturierter Probenhalter aus monokristallinem Silizium (Waiver-Technologie) mit Mikroporen. Das Ziel der Verwendung des Roadrunner I-Chips ist es, einen Probenhalter zu haben, der Hunderte bis Tausende von Kristallen dem hochintensiven Röntgenstrahl präsentieren kann, ohne das Beugungsmuster zu stören. In dieser Studie wurde Thermolysin (TLN) als Protein verwendet, um die Grenzen dieser neuen Methode zu testen. Thermolysin-Kristalle wurden bei kryogener Temperatur gezüchtet, gewaschen, in Ligandlösungen eingeweicht und eingefroren, ohne sie vom Chip zu entfernen. Nachdem alle Kristallhandhabungsschritte durchgeführt worden waren, wurden Datensätze von TLN-Kristallen gesammelt, während sie auf dem Chip lokalisiert sind. Der experimentelle Teil dieser Studie wurde am Deutschen Elektronen-Synchrotron (DESY), Hamburg, an der PETRA III P11-Strahllinie in Zusammenarbeit mit assoziierten Labors der Anlage durchgeführt.

Der dritte Teil dieser Arbeit (Kapitel 4) befasst sich mit Cyclodextrinen (CDs) und ihrer Fähigkeit, die Löslichkeit hydrophober Verbindungen in wässrigen Lösungen zu verbessern. Das Zielprotein in dieser Studie ist wieder 17 $\beta$ -HSD14. Viele Verbindungen wurden für diese Studie zusammengestellt, wie z.B. eine Fluorverbindungsbibliothek, hydrophobe Medikamente und Sexualhormone. Das Ziel dieser Studie war es, eine Verbindung zu erhalten, die an das Enzym bindet, indem es als Verbindung/CD-Komplex eingeführt wird. Die meisten der in dieser Studie verwendeten Verbindungen wurden bereits mit 17 $\beta$ -HSD14 ohne die Verwendung von CDs getestet, aber aufgrund ihrer geringen Löslichkeit war es nicht möglich, sie in das Enzym einzubringen. Die aus dieser Studie gewonnenen Daten zeigen die Wirkung des Verbindungs-CD/Komplexes, wie er über die Co-Kristallisationsmethode in das Enzym eingebracht wird.

Der vierte Teil dieser Arbeit (Kapitel 5) konzentriert sich auf das Fragmentscreening. Eine Bibliothek von 96 Fragmenten wurde mittels Röntgenkristallographie gegen Trypsin gescreent. Diese Studie konzentriert sich auf den Unterschied zwischen den durch das Fragmentscreening erhaltenen Treffern und Teiltreffern. Das Fragmentscreening wurde an zwei Kristallformen (trigonal und orthorhombisch) durchgeführt. Die aus dieser Studie gewonnenen Daten zeigen den Unterschied zwischen den einzelnen Screeningergebnissen und den Einfluss der Packung, der Kristallsymmetrie und der Fragmentabgabe auf die Trefferquote. Viele Aspekte wurden in dieser Studie berücksichtigt, wie z.B. der Unterschied in der Elektronendichte, das Volumen der Bindungstasche, anomale Peaks und Wasserkanäle.



# Contents

<b>Summary.....</b>	<b>V</b>
<b>Zusammenfassung.....</b>	<b>VII</b>
<b>Contents.....</b>	<b>IX</b>
<b>Abbreviations.....</b>	<b>XIV</b>
<b>Chapter 1. Introduction.....</b>	<b>1</b>
1.1. 17 $\beta$ -hydroxysteroid dehydrogenase family.....	3
1.2. Serial Synchrotron X-ray crystallography.....	4
1.3. Cyclodextrins.....	7
1.4. Fragment screening.....	9
1.5. Trypsin.....	12
<b>Chapter 2. Mutational Study.....</b>	<b>15</b>
“Mutational and structural studies uncover crucial amino acids determining activity and stability of 17 $\beta$ -HSD14”	
2.1. Introduction.....	17
2.2. Materials and methods.....	20
2.2.1. Preparation of the mutant proteins.....	20
2.2.2. Expression and purification of 17 $\beta$ -HSD14 variants.....	21
2.2.3. 17 $\beta$ -HSD14 kinetics.....	24
2.2.4. Michaelis-Menten kinetic constant.....	26
2.2.5. Native PAGE electrophoresis.....	26
2.2.6. Protein crystallization.....	27

2.2.7.	Data collection and processing.....	29
2.2.8.	Crystal structure determination and refinement.....	29
2.2.9.	Mass spectrometry under denaturing and native conditions.....	30
2.2.10.	Volume calculation of the binding pocket.....	30
2.3.	Results.....	30
2.3.1.	Site directed mutagenesis, expression, purification and characterization.....	30
2.3.2.	Enzyme kinetics of 17 $\beta$ -HSD14 wild type and variants.....	33
2.3.3.	Comparison of crystal structures.....	34
2.3.4.	Volume comparison of the binding pocket.....	37
2.4.	Discussion.....	39
2.4.1.	His93Ala and Gln158Ala variants.....	39
2.4.2.	Lys158Ala variant.....	42
2.4.3.	Tyr253Ala variant.....	43
2.4.4.	Cys255Ala variant.....	47
2.5.	Conclusion.....	50

## **Chapter 3. Serial Synchrotron X-ray Crystallography.....53**

“New sample holder: The Roadrunner I chip”

3.1.	Introduction.....	55
3.2.	Materials and methods.....	57
3.2.1.	Preparation of the protein sample.....	57
3.2.2.	3D printed crystal plates.....	57
3.2.3.	Crystallization condition and protein buffer.....	58
3.2.4.	Setting up the plates.....	59
3.2.5.	Crystal handling on the chip.....	60
3.2.5.1.	Hydration stream.....	60
3.2.5.2.	Incubation with isopropanol.....	61
3.2.5.3.	Soaking on the chip.....	61
3.2.6.	Data collection and processing.....	63

3.2.7.	Crystal structure determination and refinement.....	64
3.3.	Results.....	64
3.3.1.	Soaking outcome.....	64
3.3.2.	Obtained data sets.....	65
3.4.	Discussion.....	69
3.4.1.	Overall Roadrunner usage.....	69
3.4.2.	Comparison of TLN structures.....	71
3.5.	Conclusion.....	73

## **Chapter 4. Solubility Study.....75**

“Solubility enhancement of 17 $\beta$ -HSD14 potential inhibitors using cyclodextrins”

4.1.	Introduction.....	77
4.2.	Materials and methods.....	78
4.2.1.	Fluorine-compound library.....	78
4.2.2.	Hydrophobic drugs.....	79
4.2.3.	Sex hormones.....	83
4.2.4.	Cyclodextrin derivatives.....	84
4.2.5.	Expression and purification of 17 $\beta$ -HSD14.....	85
4.2.6.	Crystallization protocol.....	85
4.2.7.	Data collection and processing.....	86
4.2.8.	Crystal structure determination and refinement.....	86
4.2.9.	Inhibition assay of 17 $\beta$ -HSD14.....	87
4.3.	Results and discussion.....	87
4.3.1.	Compounds solubility with cyclodextrins.....	87
4.3.1.1.	The fluorine-compound library solubility.....	87
4.3.1.2.	Hydrophobic drugs solubility.....	88
4.3.1.3.	Sex hormones solubility.....	88
4.3.2.	Crystals quality and stability with cyclodextrins.....	89
4.3.3.	The fluorine-compound library.....	91
4.3.3.1.	Hits from the fluorine-compound library.....	91

4.3.3.1.1.	Compound FB20.....	92
4.3.3.1.2.	Compound bad25.....	97
4.3.3.1.3.	Compound bad26.....	98
4.3.3.1.4.	Compound bad22.....	100
4.3.3.2.	The fluorine-compound library partial hits.....	101
4.3.4.	The hydrophobic drugs.....	102
4.3.5.	Sex hormones.....	102
4.3.6.	Water channels and cyclodextrin.....	103
4.3.7.	Crystallographic tables.....	105
4.4.	Conclusion.....	107
4.5.	Appendix.....	108

## **Chapter 5. Fragment-Based Drug Discovery.....113**

“Fragment screening against trypsin crystallized in trigonal and orthorhombic space group”

5.1.	Introduction.....	115
5.2.	Materials and methods.....	119
5.2.1.	The 96-fragment library.....	119
5.2.2.	Protein sample.....	119
5.2.3.	Competitive fragment screening on two crystal forms.....	120
5.2.4.	Crystallization protocol.....	121
5.2.4.1.	Trigonal crystal form (soaking).....	121
5.2.4.2.	Orthorhombic crystal form (Soaking/Co-crystallization).....	122
5.2.4.3.	Benzamidine replacement.....	123
5.2.5.	Data collection and processing.....	124
5.2.6.	Crystal structure determination and refinement.....	125
5.2.7.	Volume calculation of the binding pocket.....	125
5.3.	Results and discussion.....	127
5.3.1.	The 96-fragment library hits.....	127
5.3.1.1.	Fragment J2.....	129
5.3.1.2.	Fragment J21.....	131

5.3.1.3.	Fragment J29.....	133
5.3.1.4.	Fragment J85.....	134
5.3.1.5.	Fragment J24.....	136
5.3.2.	Crystallographic tables.....	138
5.3.3.	Volume of the binding pocket.....	142
5.3.4.	Anomalous map.....	144
5.3.5.	Water channels.....	147
5.3.6.	Comparison of crystal structures.....	148
5.4.	Conclusion.....	152
5.6.	Appendix.....	153
<b>References.....</b>		<b>165</b>
<b>Acknowledgments.....</b>		<b>181</b>
<b>Curriculum Vitae.....</b>		<b>185</b>
<b>Erklärung.....</b>		<b>188</b>

# Abbreviations

17 $\beta$ -HSD14	17 beta-HydroxySteroid Dehydrogenase type 14
5-diol	5-androstene-3 $\beta$ , 17 $\beta$ -diol
Å	Ångstrom (1 Å = 10 <sup>-10</sup> m)
BEN	benzamidine HCl
<i>B</i> -factor	Debye-Waller factor
°C	degree Celsius
CD	cyclodextrin
Da	Dalton
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
E1	Estrone
E2	Estradiol
<i>e</i> <sup>-</sup>	Electron
EMA	European Medicine Agency
FBDD	Fragment-based drug discovery
FBLD	Fragment-based lead discovery
FDA	U.S. Food and Drug Administration
$\lambda$	wavelength
g	gram
GLO	glucose
GOL	glycerol
H-bond	hydrogen bond
IPTG	Isopropyl $\beta$ - d-1-thiogalactopyranoside
IM	Intramuscular
IPA	isopropanol
IV	Intravenous
K	Kelvin

$K_i$	Dissociation constant at equilibrium for an inhibitor
kGy	kiloGray
MW	Molecular Weight
NAD(H) <sup>+</sup>	Nicotinamide adenine dinucleotide
NADP(H) <sup>+</sup>	Nicotinamide adenine dinucleotide phosphate
NMR	Nuclear Magnetic Resonance
NSAIDs	Nonsteroidal anti-inflammatory drugs
OD	Optical Density
PAGE	polyacrylamide gel electrophoresis
PDB	Protein data bank
PEG	Polyethylene glycol
pET	Polyethylene Terephthalate
RMSD	root-mean-square deviation
rpm	Round Per Minute
SAXS	Small-angle X-ray scattering
SDR	Short-chain Dehydrogenase/Reductase
SEC	Size Exclusion Chromatography
SSX	Serial X-ray crystallography
T	Testosterone
TCEP	tris (2-carboxyethyl) phosphine
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TLN	Thermolysin
Tris	Tris(hydroxymethyl)aminomethane
TPCK	L-(tosylamido-2-phenyl) ethyl chloromethyl ketone
$V_{max}$	maximum velocity at saturating substrate concentration
W	water molecule

## Amino acids

Ala	Alanine	Leu	Leucine
Arg	Arginine	Lys	Lysine
Asn	Asparagine	Met	Methionine
Asp	Aspartic acid	Phe	Phenylalanine
Cys	Cysteine	Pro	Proline
Gln	Glutamine	Ser	Serine
Glu	Glutamic acid	Thr	Threonine
Gly	Glycine	Trp	Tryptophan
His	Histidine	Tyr	Tyrosine
Ile	Isoleucine	Val	Valine







# 1

## Chapter 1

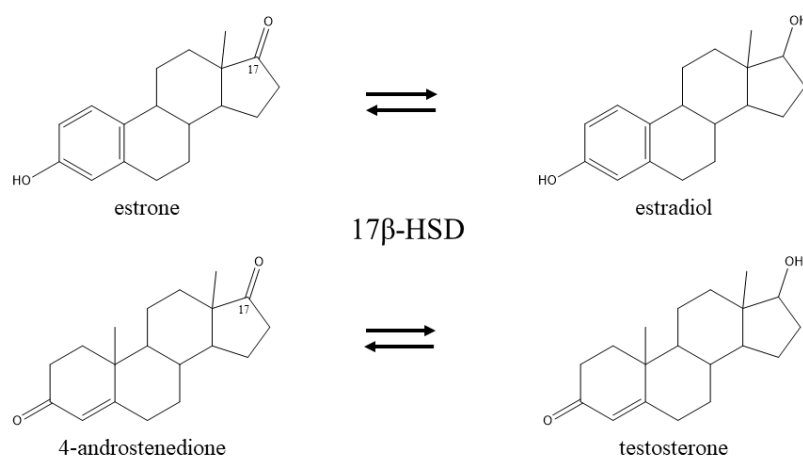
---

### Introduction



## 1.1. 17 $\beta$ -hydroxysteroid dehydrogenase family

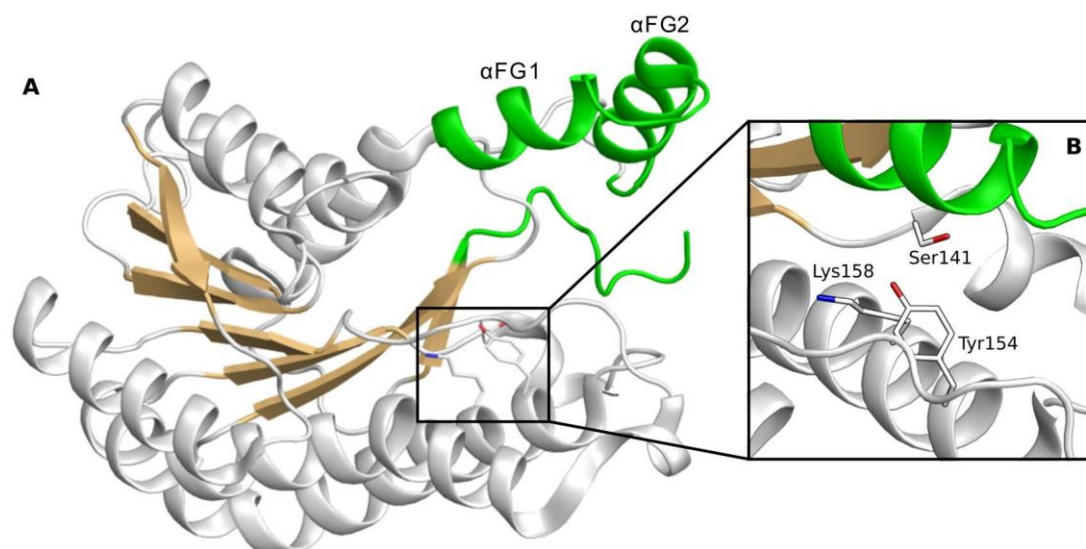
Hydroxysteroid dehydrogenases (HSD) are steroid hormones that act through specific receptors, which are involved in the activation of gene transcription. HSD convert steroids at position 3,5,11,17 and 20, which makes them responsible for the regulation of steroid hormones action [1,2]. 17 $\beta$ -hydroxysteroid dehydrogenase (17 $\beta$ -HSD) has a high potency to estrogens and androgens, that converts them at position 17.



**Figure 1.** Schematic showing the oxidoreduction at C17 oxo/ $\beta$ -hydroxyl groups of androgen and estrogen hormones.

As seen in **Figure 1**, the 17 $\beta$ -HSDs family plays a key role in the formation and degradation of sex hormones, in which, all intracellular availability and potential activation of sex hormones is regulated through this family [3–31].

The interest in 17 $\beta$ -HSDs appeared in the 1980s, with 14 different mammalian 17 $\beta$ -HSDs have been characterized up to now. All members of this family share the same conserved Rossmann-fold domain established by an  $\alpha/\beta$  sandwich folding pattern that is arranged in 6 to 7 central  $\beta$ -sheets, and flanked by 3  $\alpha$ -helices at both sides (**Figure 2**) [32–34].



**Figure 2:** (A) Ribbon representation of the 17 $\beta$ -HSD14 apoenzyme. The conserved Rossmann-fold domain is established by  $\alpha$ -helices (white) and  $\beta$ -sheets (beige). The loops giving special shape to the binding pocket and the variable C-terminal tail are colored in green. (B) Close-up view of the substrate binding pocket. The catalytic amino acids are shown as stick model. All structural representations were prepared with PyMOL [35]. This figure was adopted from Nicole Bertolotti's thesis [36].

Due to the enzyme expression in multiple organs in the human body, there is a growing interest to characterize this protein and reach a therapeutic target that corresponds to the functionality of the steroid hormones.

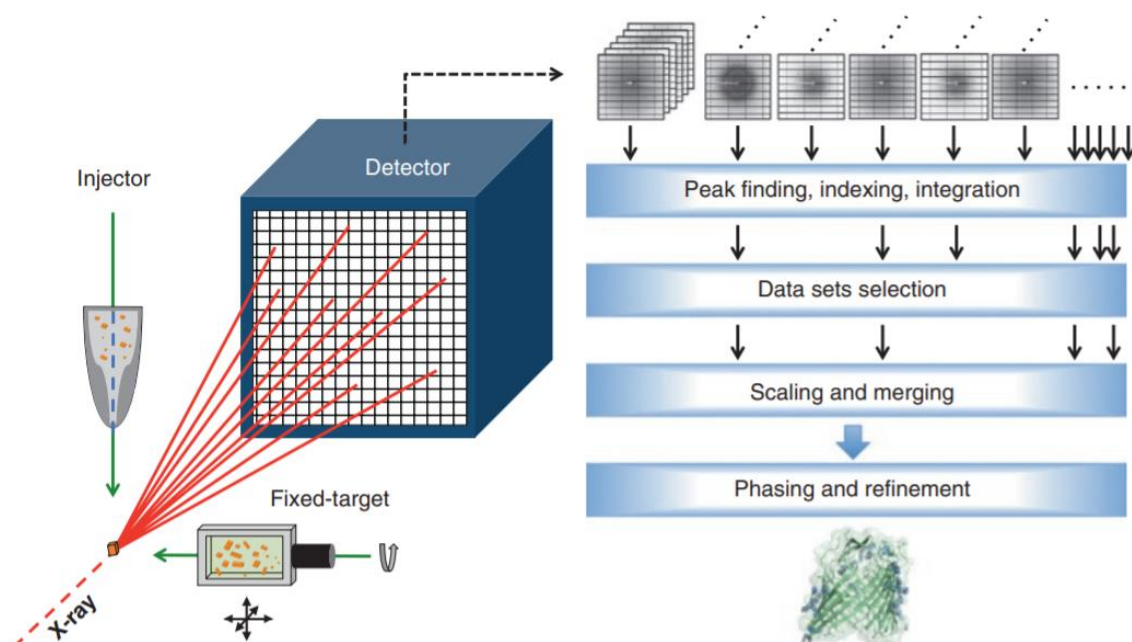
## 1.2. Serial Synchrotron X-ray crystallography

Since the beginning of crystallographic studies, collecting data sets from small-sized crystals has always been a challenge, because of their weak diffraction power that limits the number of signals collected before the damageable effects of the X-ray radiation become significant. Recently, there has been a huge surge in serial X-ray crystallography (SSX) method, due to its usefulness in obtaining data sets from crystals on the nanometre to micrometre scale [37–49]. Collection and merging of data sets from multiple crystals are not something new, in fact, that was state of the art before the advances of cryo-cooling [50]. In previous times, crystallographers have always faced a major challenge, which is the need to optimize the

growth of crystals to a large size to obtain a sufficient diffraction to resolve a protein structure either from one crystal or from multiple crystals of the same protein and merge them [37]. SSX came with a solution to many problems faced before, of course with the technological advancement and the foundation of microfocus X-ray beams, the collection of data sets of small crystal has been facilitated. The reason that many researchers focused on tiny crystals was their advantages over larger crystals. If a crystal has a higher mosaicity, a smaller crystal from the same macromolecule would have a less total number of mosaic blocks, so the broadness of the Bragg peaks will be reduced leading to an improved overall crystal quality. Another main advantage of smaller crystals is the higher tendency of a crystal to bind a ligand during soaking experiment and the higher successful rate of freezing crystals in cryogenic temperature compared to their equivalents that possess a bigger size [51,52]. Another major bottle neck facing crystallographers is radiation damage, which put a limitation on the number of datasets that could be collected from the same crystal. There are multiple types of radiation damage which has a devastating effect on the data sets collected. For example, generation of free radicals, degradation of overall data quality or breaking of chemical bonds which is unfortunately, only visible after obtaining an electron density [53–56]. Therefore, macromolecular crystallography is usually carried out at cryogenic temperature (100 K) to minimize radiation damage [57,58]. Furthermore, measuring at cryogenic temperature can extend the crystal lifetime by two orders compared to its equivalent at room temperature [59–61]. However, collecting data sets at room temperature is considered more convenient, rapid and less consumable.

In SSX, there is a high need to find a good combination of crystal size, X-ray beam size and the velocity of the stream to have most of the crystal illuminated by X-ray while it passes through the beam. From a practical point of view, a certain minimum stream size is needed to avoid clogging. In SSX with fixed-target methods, many scientific institutes and research facilities have been working relentlessly to develop sample holders in which crystals can be grown, buffers can be exchanged and data sets can be collected without removing the crystals from these sample holders [62–64]. The sample holders are designed to develop a system with low interference with the X-ray diffraction pattern and the requirement of less crystallization medium around crystals. The goniometer or scanning stage, which holds the fixed-target samples, allows full control of crystal characterization and data collection. Crystals can be located with a grid scan, and rotation data can be collected for each selected crystal

subsequently [38,65]. Alternatively, the whole sample could be scanned either with still images or with oscillation images [66,67]

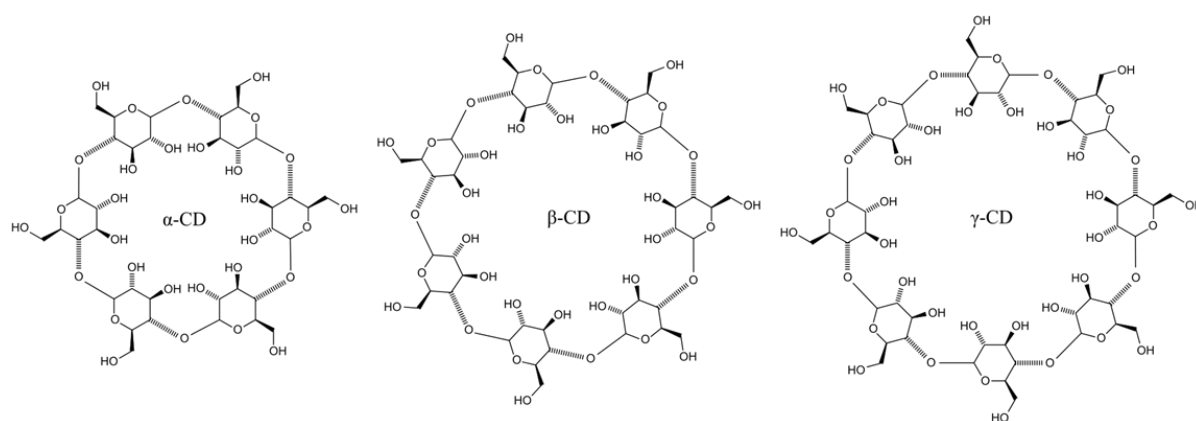


**Figure 3.** Serial synchrotron X-ray crystallography, starting with a crystal ending with a structure. Figure adopted from [68].



### 1.3. Cyclodextrins

Cyclodextrins (CDs) were discovered approximately 100 years ago and the first CD complex was registered in 1953. Cyclodextrins (CDs) are a family of cyclic oligosaccharides, consisting of multiple glucose subunits forming a ring by  $\alpha$ -1,4 glycosidic bonds. Cyclodextrins result from the cyclomaltodextrin glucanotransferase (E.C. 2.4.1.19; CGTase) catalyzed degradation of starch [69,70]. Nowadays, there are multiple uses of CDs in day-to-daily life, such as food additives, drugs, lubricants and many chemical products. The most common CDs come with 6, 7 and 8 D-glucopyranosyl units ( $\alpha$ -,  $\beta$ - and  $\gamma$ -cyclodextrin respectively) as shown in **Figure 4**. The glucose subunits all have the  ${}^4C_1$  (chair) conformation. The three main CDs also share similar structures (bond length and orientations), with an exception for the glucose subunit numbers. Their shape looks like a bowl (cone) stiffed with hydrogen bonding between the 2-OH and 3-OH groups around the outer rim, making a hydrophobic central cavity and hydrophilic outer surface. The strength of the hydrogen bonds is dependent on the number of glucose units in CD,  $\alpha$ -cyclodextrin <  $\beta$ -cyclodextrin <  $\gamma$ -cyclodextrin [71–78].



**Figure 4.** 2D representation of the most common cyclodextrins.

Due to the relatively large number of hydrogen donors and acceptors, CDs are not able to permeate lipophilic membranes. The main focus of this study is the usage of CDs in enhancing compound solubility. CDs are used as agents to increase aqueous solubility of poorly soluble drugs; thus, increasing their bioavailability and stability. There are approximately 30 drugs containing cyclodextrin derivatives (**Table 1**) that have been approved by the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA).

Due to the CDs molecular structure and shape of CDs, they own a unique ability to act as containers for other molecules by locking (entrapping) the hydrophobic part inside the lipophilic internal cavity. However, when in aqueous solution, no covalent bonds are broken or formed during the molecule/CD complex formation [79–85]. The resulting complex hides most of the hydrophobic functionality of the bound molecule by locking the hydrophobic moiety inside the cavity. The hydrophilic hydroxyl groups located on the outer surface are exposed to the outer environment, leading to a water-soluble complex. Furthermore, the entrapping of the hydrophobic moiety inside the cavity would result in the hydrophilic moieties of a compound to point to the outside environment, which would increase the compound's solubility [86–91].

**Table 1.** *Molecular weight and solubility of the most common cyclodextrins.*

Cyclodextrin	MW (Dalton)	Solubility in water (mg/ml) at 25°C
$\alpha$ -cyclodextrin	972	45
$\beta$ - cyclodextrin	1135	5
$\gamma$ - cyclodextrin	1297	80
Methyl- $\beta$ - cyclodextrin	1312	350
2-hydroxypropyl- $\beta$ - cyclodextrin	1400	300
2-hydroxypropyl- $\gamma$ - cyclodextrin	1576	400

These data are based on entries in pharmacopoeias and health organizations in USA, Europe and Japan.

As seen in **Table 1**, the more subunits and moieties are added to the CD molecule, the larger water solubility will be. However, there are other factors that determine the degree of dissociation of the complexes, such as the various environmental factors (dilution, temperature and pH) [92–95]. This concept has made cyclodextrin most intriguing for many drug designers. The fact that a candidate drug physiochemical property, such as solubility, bioavailability, stability and affinity to its receptor are posted while gaining no negative side effects, makes it a favourable additive in the solubility enhancement study performed on 17 $\beta$ -HSD14.

## 1.4. Fragment screening

In a human body there are about 25,000 genes that are protein coding, which means there are at least 25,000 proteins in the human body. In a recent study by Brandon et al. [96] there are around 42 million protein molecules in a simple cell. With this tremendous amount of proteins, finding an efficient and fast screening method for drug candidates was necessary. Drug design of lead compounds that have an affinity to a protein; thus, halting or posting its function is required. The main aim of drug discovery is to find lead compounds that have specificity and potency to their targeted proteins, so drugs for new or old diseases are synthesized accordingly.

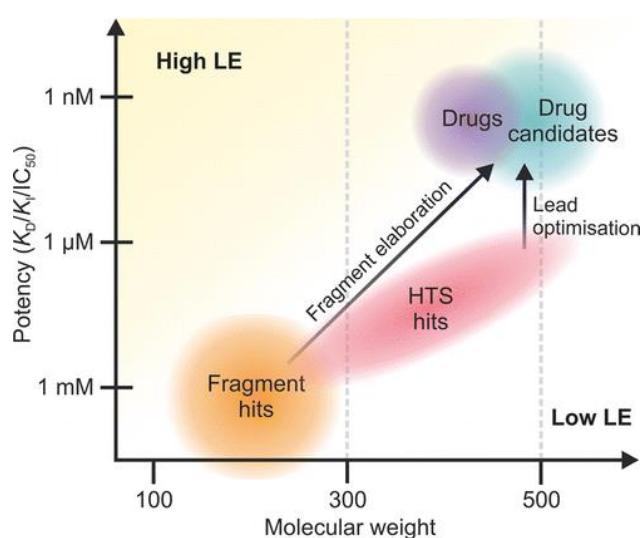
Fragment screening offers a great advantage due to its ability of introducing various moieties to the targeted protein, which is the initial step to reach a final drug based on the results obtained by such a screen. Furthermore, the small size of the fragments can easily reach the protein active sites and travel easier through the protein water channels [97,98]. A successful hit from a fragment screen would provide a promising candidate for a follow-up drug to be developed based on the structure and moieties of the binding fragment. Unlike the fragment used in the screening, a bigger size compound consisting of extra moieties based on the original fragment used or even multiple combined fragments together could provide a more potent ligand with enhanced properties.

The aim of fragment screening is to identify chemical moieties that specifically interact with a particular protein target. Recently, fragment screening has gained a lot of popularity. Multiple designated moieties can be introduced to a single protein to identify which one would form an interaction with that protein. This means, decreasing the time for drug discovery and screening various amounts of protein through this process. Fragments are small chemical entities which become increasingly favourable starting points for drug discovery [99–113]. A standard definition of a fragment suggests that fragments are chemical entities that possess a small size and a low molecular weight ( $MW \leq 300$  Da) [98,105,114].

Fragment-based drug discovery (FBDD) or Fragment-based screening (FBS) has come a long way recently. Even with the High-Throughput Screening (HTS) method being widely used, but the difference between HTS and FBS is the molecule size which ranges from 300-500 Da in the HTS libraries making FBS more favourable nowadays. The bigger molecule size in HTS hinders the development of the pharmacokinetics and pharmacodynamics properties of the

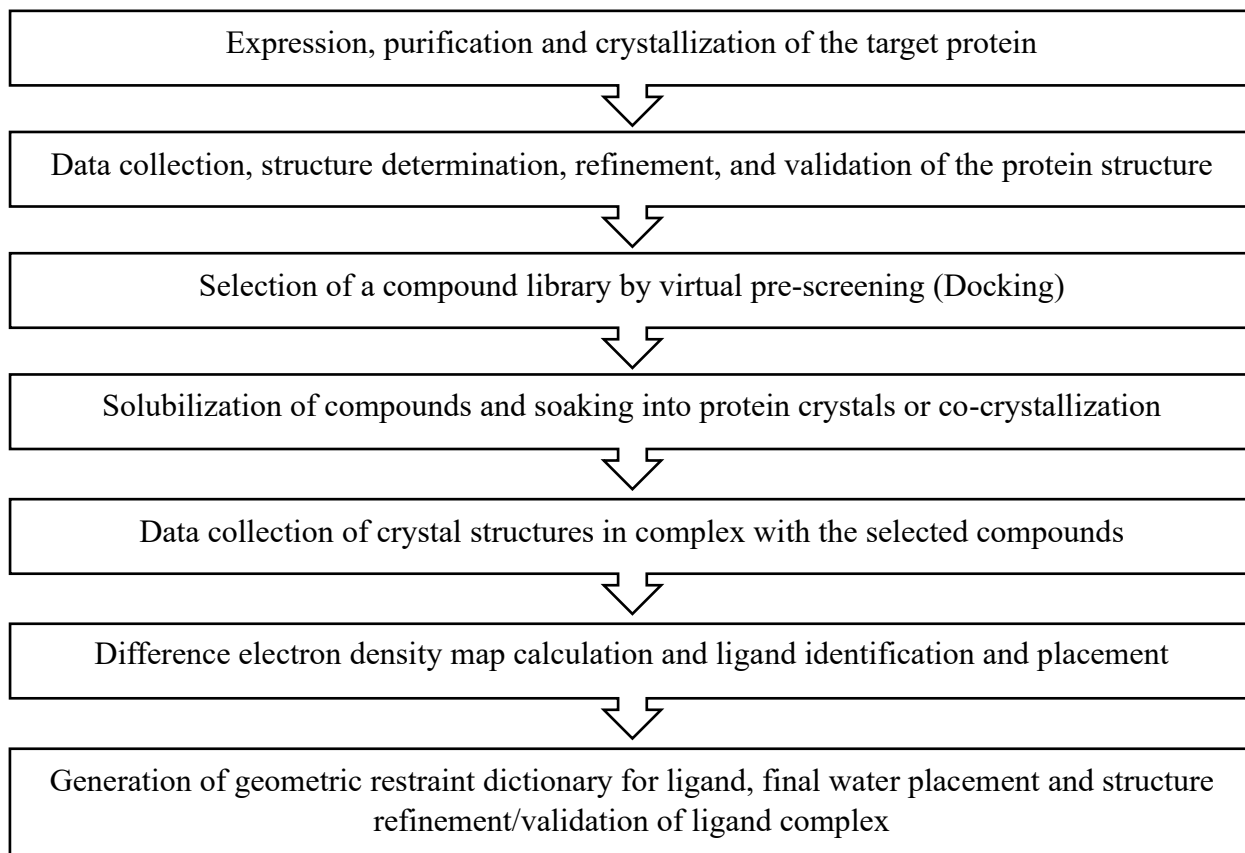
selected molecules because they are less manoeuvrable around the active site and the water channels passing through a crystal of the targeted protein [97,98,115].

There is no doubt that in present days, X-ray crystallography is the dominant method to determine the structure of macromolecules and it is quite sufficient when it comes to fragment screening. Mainly, with all the automization that has been developed to make it less time and resource consumable. The main reason for the effectiveness of this technique is the undeniable result obtained by resolving the 3D structure of the macromolecules which confirms or denies the presence of a fragment in the active site or any other location throughout the protein structure.



**Figure 5.** Comparison of molecular mass vs potency of leads developed from conventional HTS and FBDD approaches. The dashed lines show Lipinski's Rule of Five (500 Da molecular weight cut-off and the Rule of Three 300 Da molecular weight limit). This figure is taken from [111].

In any fragment screening, there are standard steps that need to be met which are organized in the sketch below.

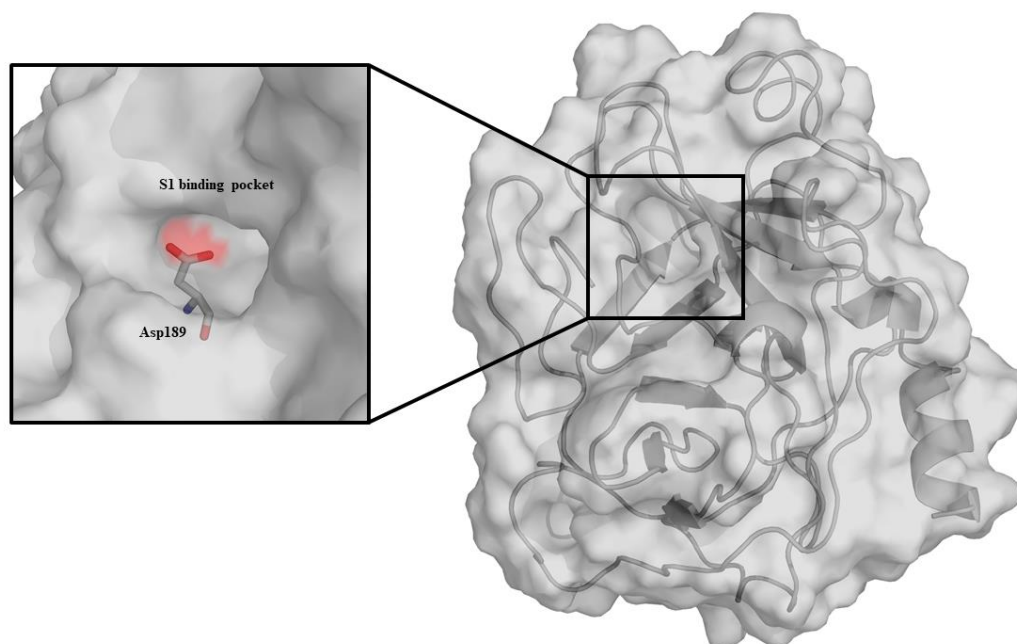


## 1.5. Trypsin

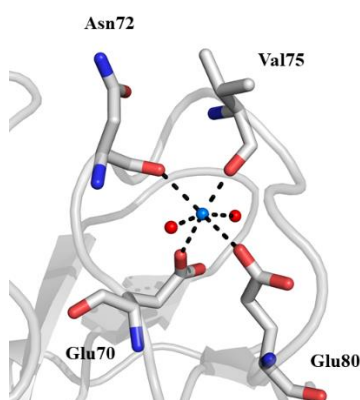
Trypsin (EC 3.4.21.4) was first identified by Kühne in 1876 [116]. It is a well-known pancreatic enzyme and a member of the Protease of mixed nucleophile superfamily A (PA clan). There is a pre-existing availability of trypsin from bovine pancreas making it easy to purify by crystallization [117]. Since many years, trypsin has been associated with the proteolytic activity that cleaved peptide bonds after Lys and Arg [118,119]. Throughout the 80s and 90s, the role of the amino acids present in the enzyme were identified and afterwards, trypsin was linked with hereditary pancreatitis. The enzyme has many functions such as tissue dissociation when mixed with enzymes like elastase and collagenase, cell harvest via trypsinization, protein studies “*in vitro*”, mitochondria isolation, subculturing cells, tryptic mapping, cleavage protein fusion and glycopeptides generation from pure glycoproteins [120–131]. As a general structural description, trypsin is a globular protein of 24 KDa, with a total of 220 residues. The protein is composed of 13 beta-strands, six of which form a beta-barrel structure. There are four regions of alpha helix and six disulfide bridges as well as the Ca<sup>2+</sup> binding loop that extends from Glu70 to Glu80.

Serine proteases evolved a similar catalytic device characterized by the His, Asp and Ser triad (His57, Asp102 and Ser195 in trypsin). These three amino acids are often referred to as a catalytic triad. The serine in the triad is much more reactive than other serines in the protein. The serine hydroxyl is normally protonated at neutral pH, but in the enzyme Ser195 is hydrogen bonded to His57, which is further hydrogen bonded to Asp102 [132–135].

Residues 189-195 and 214-220 form the primary substrate binding pocket called S1 binding pocket. There are residues that form two loops near the S1 binding pocket, residues 185-188 and 221-224, which are called L1 and L2, respectively. There is a high favourability towards basic residues. Residue Asp189 in trypsin lies at the bottom of the S1 binding pocket and forms a salt bridge with the positively charged group at the end of the substrate lysine and arginine side chain, and determines many of the enzyme chemical properties [136–138].



**Figure 6.** overall structure of trypsin and the S1 binding pocket where Asp189 is located at the back end of the pocket.



**Figure 7.** Octahedral coordination of calcium ion (blue sphere) in trypsin by Glu70, Glu80, two water molecules (red spheres) and two carbonyl group from the backbone of Asn72 and Val75.





# 2

## Chapter 2

---

# Mutational Study

Mutational and structural studies uncover crucial amino acids determining activity and stability of 17 $\beta$ -HSD14

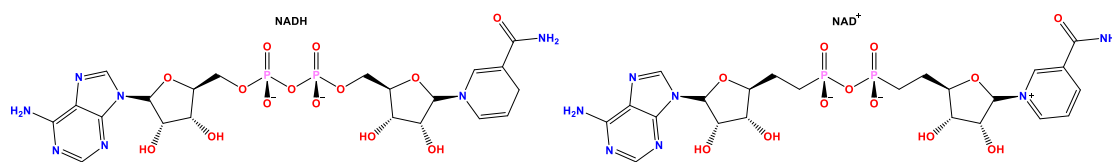
### Introductory remarks

This chapter closely relates to a publication which appeared in the *Journal of Steroid Biochemistry and Molecular Biology* [139]. The fluorescence-based assay was designed and performed by Dr. Sandrine Marchais-Oberwinkler in collaboration with the author of the thesis. The expression and the purification of the 17 $\beta$ -HSD14 variants, the crystallization study and the elucidation of the variant crystal structures established and performed by Dr. Nicole Bertoletti with the author of this thesis. Furthermore, the author significantly contributed to the writing of the manuscript in collaboration with Dr. Sandrine Marchais-Oberwinkler.



## 2.1. Introduction

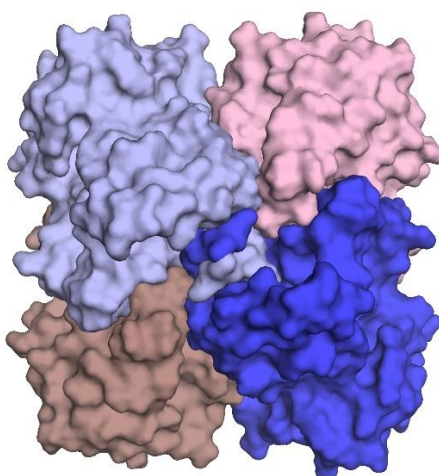
17 $\beta$ -Hydroxysteroid dehydrogenase type 14 (17 $\beta$ -HSD14), which is also known as retSDR3, DHRS10 or SDR47C19, is one of the last members of the short-chain dehydrogenase (SDR) family to be identified [3,140,141]. The purpose of many studies of this enzyme is to figure out its main role in the human body which is yet to be discovered, but throughout previous and current research, it was established that the main substrates for this enzyme are : Estradiol (E2), 5-androstene-3 $\beta$ ,17 $\beta$ -diol (5-diol) and testosterone (T) as they have been tested on the enzyme *in vitro* [140]. There are two variants that occur naturally (T205 and S205) that have been already characterized [142]. The enzyme has a clear role in oxidizing of the hydroxyl group at position 17 of the substrates in the presence of the cofactor NAD<sup>+</sup> [140]. Nevertheless, the main role of this protein still has to be proven by *in vivo* studies.



**Figure 1.** Representation of the cofactor NAD<sup>+</sup> and the reduced form NADH

Until now, 14 different 17 $\beta$ -HSD subtypes have been reported with 17 $\beta$ -HSD14 being the last, two of these subtypes are 17 $\beta$ -HSD1 and 17 $\beta$ -HSD2 which were most previously characterized subunits from the 17 $\beta$ -HSD tree, that predominantly catalyze the oxidation and reduction of estrogens and androgens. Inhibitors of these two enzymes have already been reported [143–148]. The gene of 17 $\beta$ -HSD14 was initially isolated from the retina [149]. and expressed as well in cancer breast tissue [150]. In later immunological studies by Sivik et al. [151], it was demonstrated that 17 $\beta$ -HSD14 has a wide distribution throughout the human body and is located in different organs such as the heart, eye, adrenal glands, testis, liver, esophagus, rectum, skeletal muscles and salivary glands. In another study by Lukacik et al. [140], which was more conservative toward the distribution of the enzyme in the human body, revealing that the mRNA coding this enzyme is mainly expressed in the brain, placenta and liver. The enzyme is located in the cytoplasm as well and assembled in solution as a homo tetramer [142]. The first crystal structure of the wild type has been recently determined as an apo enzyme (PDB ID: 5ICS) [142]. The holoenzyme structure has been determined as well (PDB ID: 5JS6 and 5JSF) and the holoenzyme in complex with both the cofactor and non-steroidal inhibitors (PDB ID:

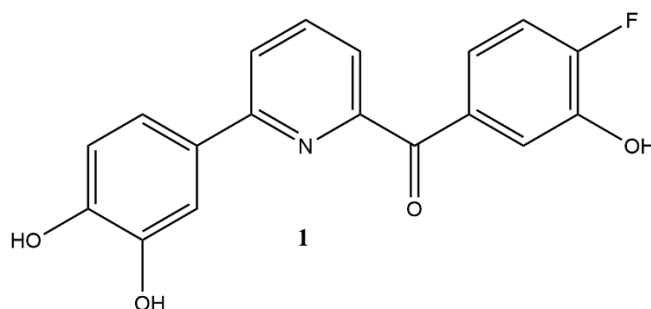
5ICM) [142]. After investigating the crystal structure obtained using X-ray crystallography, it is obvious that the enzyme is a tetramer formed by interactions of two identical monomers to form a dimer that joins with another identical dimer in an organised way (as seen in the crystal structure), in which the flexible loop from each monomer is attached to the adjacent monomer in a hook shape. This dimer formed is linked to the other dimer with multiple amino acid interactions on the surface of the enzyme. The crystal structure shows a tetramer formed by four identical monomers joined together and the flexible loop from each monomer adopts a closed conformation.



**Figure 2.** Surface representation of 17 $\beta$ -HSD14 wild type variant S205 (PDB ID: 6G4L). In the centre of the figure the flexible loops of each monomer are visible in a different colour as they are hooked together forming a dimer (blue and light blue) and joined by a similar dimer (pink and brown) that has the same attachment, but located on the back of the molecule. Both dimers are linked through surface amino acid interactions.

The active site of 17 $\beta$ -HSD14 is characterized by three catalytic amino acids also known as the catalytic triad Ser141, Tyr154 and Lys158. This catalytic triad is shared with other types of 17 $\beta$ -HSD14s and they are involved in interactions between the cofactor and the substrate, so they are responsible for the enzymatic activity of the protein. There are other amino acids beside the ones from the catalytic triad which have a major role in the activity or the assembly of the enzyme, Gln148, Tyr253 and Cys255 which are not present in other types of human 17 $\beta$ -HSDs. Unfortunately, all attempts to obtain a crystal structure with suspected substrates estradiol (E2) and 5-androstene-3 $\beta$ ,17 $\beta$ -diol (5-diol) were unsuccessful, but it is believed that Ser141, Tyr154 and O17 from estradiol (E2) form a trigonal H-bond network similar to the one

observed in 17 $\beta$ -HSD1 [152]. In the active site of 17 $\beta$ -HSD14, Lys158 stabilizes the cofactor NAD<sup>+</sup> by an H-bond interaction with the two hydroxyl groups of the ribose ring of the nicotinamide moiety [139]. In the case of 17 $\beta$ -HSD1, it was suggested that Lys158 is interacting with catalytic amino acid Tyr155 and as for 17 $\beta$ -HSD14, the interaction is formed with the catalytic Tyr154 via electrostatic interactions and acts as a catalytic base. This leads to the lowering of the pK<sub>a</sub> value of Tyr155 in 17 $\beta$ -HSD1 [153]. As suggested by a study done by our working group [139], Lys158 is involved directly in the activity of the enzyme after applying multiple fluorescence assays on the 17 $\beta$ -HSD14 variant Lys158Ala, which showed an absence of activity compared to the wild type, while using both estradiol (E2), 5-androstene-3 $\beta$ ,17 $\beta$ -diol (5-diol) as a substrate and NAD<sup>+</sup> as a cofactor in the activity assay. Cys255 is located on the flexible loop of the enzyme, as seen in the crystal structure, this amino acid has a role in the interface between the two dimers. Cys255 forms a disulfide bridge with the equivalent Cys255' from the adjacent monomer mate. Due to the hook-like shape of the flexible loop, and how the two flexible loops from both adjacent monomer mates are located, they are able to link together like a ribbon. The placement of Cys255 and Cys255' from both adjacent monomers is at the top of the hook. It is believed that the disulfide bridge formed by both has a major role in the stability of the tetrameric assembly of the enzyme. However, a study done by our working group [139] shows the role of the disulfide bridge is irrelevant neither for the activity nor in the stabilization of the dimer or the tetramer. Tyr253 is also located at the flexible loop of 17 $\beta$ -HSD14. To better understand the involvement of this amino acid, the enzyme should be studied as a dimer because of how Tyr253 is placed on the flexible loop, which points toward the binding pocket of the adjacent monomer mate. The crystal structure with a non-steroidal inhibitor **1** (PDB ID: 5ICM) which is a highly potent 17 $\beta$ -HSD14 inhibitor, shows that Tyr253' from the adjacent monomer has an interaction with the fluorine moiety of the inhibitor **1** binding in the active site of the enzyme.



**Figure 3.** Highly potent 17 $\beta$ -HSD14 inhibitor (inhibitor **1**).

Gln148 is located at the rim of the active site of the enzyme acting as a gate keeper. The positioning of this amino acid of the rim has a crucial part in obtaining a binding compound inside the active site. His93 is placed behind Gln148 in the inner pocket of the active site and they both interact with each other. Although located in the binding pocket and its entrance, both His93 and Gln148 do not share any interactions with inhibitor **1**. To get a better understanding of the role of the enzyme and its metabolic participation in the human body, His93, Gln148, Lys158, Tyr253 and Cys255 were investigated via a mutagenesis study. In order to observe the affinity for each mutated variant using inhibitor **1** was used as a standard to check the difference of binding mode in the active site or if there will be any affinity at all. All of these amino acids were replaced with an alanine, each variant contains one mutated amino acid which led to five different mutated variants His93Ala, Gln148Ala, Lys158Ala, Tyr253Ala and Cys255Ala. The variants were all expressed and purified in our lab with addition of NAD<sup>+</sup> as cofactor. Then the protein is crystallized by the vapour diffusion method with the presence of inhibitor **1** with each variant. Five crystal structures were determined using X-ray crystallography as a method of choice. Each one of the five variants were further characterized by enzyme kinetics. The mutagenesis study was carried out on the allelic variant S205 [142].

## 2.2. Materials and methods

### 2.2.1. Preparation of the mutant proteins

The five 17 $\beta$ -HSD14 variants, His93Ala, Gln148Ala, Lys158, Tyr253 and Cys255Ala, were all generated by mutating the p11-HSD17B14 vector (sequenced according to NCBI data bank entry Q9BPX1). The Quick-Change Lightning Multi Site-Directed Mutagenesis kit (Agilent Technologies, La Jolla, CA, USA) was used. The sequence was coded for the wild type variant S205 and T205, but it was the wild type variant S205 that was used in this study, although some kinetics were done using the wild type variant T205, but all the structural studies were done with the S205 variant. The forward primer 5'-CGGACCCGCGTGCGACCATTCGTGAAGGTATGC-3' and the reverse primer 5'-GCATACCTTCACGAATGGTCGCACGCGGGTCCG-3', were obtained from Udo Oppermann, Structural Genomics Consortium (SGC) Oxford. The mutation was introduced to the plasmid using the manufacturer instructions and it was later checked by Sanger sequencing

to ensure that both the mutation and its site are correct. The primers that have been utilized for the generation of the variants are listed in **Table 1**.

**Table 1.** Internal primers used for the site-directed mutagenesis.

Mutation	Direction	Nucleotide sequence of primer
Lys158Ala	Forward	G TAT GTG GCG ACC <u>GCA</u> GGT GCG GTG ACC GC
	Reverse	GC GGT CAC CGC ACC <u>TGC</u> GGT CGC CAC ATA C
His93Ala	Forward	AT AAT GCG GGT CAT <u>GCC</u> CCG CCG CCG CAG CGT C
	Reverse	G ACG CTG CGG CGG CGG <u>GGC</u> ATG ACC CGC ATT AT
Gln148Ala	Forward	TG GGT GCG ATT GGT <u>GCG</u> GCG CAG GCG GTG CCG T
	Reverse	A CGG CAC CGC CTG CGC <u>CGC</u> ACC AAT CGC ACC CA
Cys255Ala	Forward	CTG GGC TAT GGT <u>GCC</u> AAA GCG AGC CGT AGC
	Reverse	GCT ACG GCT CGC TTT <u>GGC</u> ACC ATA GCC CAG
Tyr253Ala	Forward	GCG GAA CTG GGC <u>GCT</u> GGT TGC AAA GCG AGC C
	Reverse	G GCT CGC TTT GCA ACC <u>AGC</u> GCC CAG TTC CGC

The mutant codons of the primer pairs are underlined.

The plasmid sequences were further investigated to ensure a successful introduction of the desired replacement of amino acid (Eurofins Genomics, Ebersberg).

### 2.2.2. Expression and purification of 17 $\beta$ -HSD14 variants

All of the variants shown in this study are from the human wild type variant S205 of 17 $\beta$ -HSD14 and they were expressed in *E. coli* BL21 (DE3) pLysS competent cells (Agilent Technologies) as pET- based vector, p11 Torontal (SGC). These *E. coli* BL21 (DE3) pLysS are chemically competent cells designed for applications that require high-level expression of non-toxic recombinant proteins from high copy number, T7 promoter-based expression systems [154], containing the coding sequence of the human HSD17B14 variant S205 with N-terminal 6His-tag. The competent cells were removed from -80°C and put on ice to prevent a heat shock. The competent cell solution was waited to liquefy slowly while on ice (approximately 20 min).

2  $\mu\text{L}$  of p11-17 human  $\beta\text{HSD}$  type 14 variant S205 is added to 25  $\mu\text{L}$  of the competent cells and stirred gently without any force or shaking then incubated for 30 minutes. The mixture is kept all the time on ice to prevent the decaying of the competent cells. After incubation a heat shock is applied for 30 seconds at a temperature of  $42^{\circ}\text{C}$  using a thermomixer, followed by returning the sample immediately on ice. This procedure allows the plasmid to diffuse through the pores at the surface of the competent cells that were extended by the heat shock [154]. An additional 973  $\mu\text{L}$  of Lysogeny Broth medium (LB-medium) is added to the mixture enlarging the volume to 1 mL. The sample is transferred to a thermomixer and incubated for 45 minutes at  $37^{\circ}\text{C}$  at 350 rpm. After incubation, the sample is centrifuged for one minute at 13000 rpm. 900  $\mu\text{L}$  of the supernatant is removed and the centrifuged cells are resuspended gently by stirring in the remaining 100  $\mu\text{L}$ . Under a sterile hatch, the cells are transferred to an agar plate and spread on top of it and moved to a plate incubator at a temperature of  $37^{\circ}\text{C}$  overnight. One colony of the grown culture on the agar plate is transferred to a flask containing 90 mL of Terrific Broth medium (TB-medium) containing 100  $\mu\text{g}/\text{mL}$  of ampicillin with an additional 10 mL of potassium salt solution, containing potassium dihydrogen phosphate ( $\text{KH}_2\text{PO}_4$ ) 23.1 mg/mL and dipotassium phosphate ( $\text{K}_2\text{HPO}_4$ ) 125.4 mg/mL which stabilizes the pH of the culture. The culture is incubated overnight (at least 12h) at  $37^{\circ}\text{C}$  in a shaking incubator. The 100 mL culture is divided equally on four big 5-liter glass containers containing 900mL of TB-medium, 100 mL potassium salt solution and 100  $\mu\text{g}/\text{mL}$  ampicillin, with 25 mL of the culture is added to each container. The containers are placed in a shaker at  $37^{\circ}\text{C}$  and the optical density  $\text{OD}_{600}$  is checked regularly (ideally, every 20 minutes) using the spectrometer NanoDrop<sup>TM</sup> (Thermo Fisher SCIENTIFIC) and freshly prepared TB-medium as a blank until the  $\text{OD}_{600}$  reaches 0.5. The temperature is decreased to  $15^{\circ}\text{C}$  in the shaker, when the  $\text{OD}_{600}$  reaches 1. A volume of 1 mL IPTG (isopropyl- $\beta$ -D-thiogalactopyranosid) solution 120 mg/mL dissolved in distilled water is added to each container. The cells are left to grow overnight (at least 16h). The cell culture is centrifuged for 15 minutes at 10000 rpm at  $4^{\circ}\text{C}$  and the cells are harvested and the supernatant is disposed. The cells are stored at  $-80^{\circ}\text{C}$ . It is worth mentioning that the cells should not be immediately lysed for purification after harvesting since the final yield of the protein would be cut by half or even less, although the reason for that is yet unknown. After freezing, the cell pellets are transferred to ice and resuspended in lysing buffer (LB) containing 25 mM Tris, 500 mM NaCl, 5mM imidazole, 250 mM glucose, TCEP (Tris 2-carboxyethyl phosphine hydrochloride) 0.25 mM and 1 mM  $\text{NAD}^+$ , after adjusting the pH to 8. 5% w/v triton X is added to the final volume of the LB as well as one tablet of complete protease inhibitor cocktail (Roche) is added to each 100 mL of the lysing buffer. The cells are mixed



and resuspended with LB-medium in a glass dancer until a more viscous texture is reached. The resuspended pellets are lysed using an Emulsiflex. At least four cycles should be performed. The procedure must be done while the pellets are on ice and there should be a small break between each cycle to prevent the pellets from heating up which will lead to the denaturation of the protein. Another 0.5 mM of NAD<sup>+</sup> is added for a further stabilization of the protein. The protein is centrifuged at 10000 rpm for 2 h, the supernatant is transferred to a sterile container and the precipitant is discarded. Using the ÄKTA, the supernatant is run through a HisTrap<sup>TM</sup> Fast Flow Crude (GE Health care) for one or two (more preferable) runs with 2 mL/min (flow through speed) maximum. The HisTrap<sup>TM</sup> is washed with Tris BB (Binding Buffer) pH 8 containing 0.25 mM Tris, 500 mM NaCl, 250 mM glucose, 0.25 mM TCEP and 0.5 mM NAD<sup>+</sup> with 2 mL/min and under a pressure of 0.5 mPa to eliminate non-specific bound proteins until a plateau on the ÄKTA is reached. Then the column is washed with Tris WB (Washing Buffer) containing 50 mM Tris, 1.5 M NaCl, 0.5 mM NAD<sup>+</sup> at 2 mL/min and under 0.5 mPa to eliminate non-specifically bound DNA until a plateau is reached again. The last washing step is performed with Tris BB to expel the Tris WB from the column. The volume of Tris BB buffer used depends on the capacity of the column, but should be at least five times the volume of the column. First, the column is eluted with a ratio of 7% Tris EB (Eluting Buffer) containing 25 mM Tris, 500 mM NaCl, 300 mM imidazole, 250 mM glucose, 0.5 mM TCEP and 0.5 mM NAD<sup>+</sup> and 93% Tris BB to elute non-specifically bound proteins until a plateau is reached, then the column is eluted with 100% Tris EB and the flow through is collected. 10 mL of TEV (Tobacco Etch Virus) at 1 mg/mL is added to the sample and transferred to 4 L of dialysis buffer containing 0.25 mM Tris, 500 mM NaCl, 250 mM glucose, 0.25 mM TCEP and 0.5 mM NAD<sup>+</sup>. The sample is left in the dialysis buffer overnight (at least 12 h) at 4°C. The sample is locked inside a porous membrane (10-20 kDa) while in the dialysis buffer. After dialysis, the sample is applied to a HisTrap<sup>TM</sup> and washed with Tris BB. A gradient of 5% Tris EB to 95% Tris BB is applied to obtain the weakly binding targeted protein inside the HisTrap column in the flow through. Last step is injecting the protein sample into a HiLoad Superdex 200 column (GE Healthcare) after concentrating it to 5 mL using VIVA spin. The peak sample is collected and frozen at -80°C after adding 0.5 mM NAD<sup>+</sup> for more stability of the sample.

### 2.2.3. 17 $\beta$ -HSD14 kinetics

The activity of the 17 $\beta$ -HSD14 wild type is decreased by the long storage time. Enzymatic activity is measured using fresh sample of the enzyme as fast as possible after purification (less than a week). The same is done as well for determining the kinetics of the variants to ensure all measurements for all variants are done at equal conditions. The enzymatic activity was evaluated by fluorometric assay using a protocol developed in our lab [142]. Fluorescence spectra were recorded on a Tecan Sapphire 2 machine. Before starting the enzymatic assay, the concentration of the protein is measured immediately before the assay (1 mg/mL, 3.5  $\mu$ M in the assay) and NAD<sup>+</sup> is added to reach 1.2 mM in the protein buffer. The substrate in the enzymatic activity assay is E2 and 5-diol (32  $\mu$ M) in 100 mM phosphate buffer, pH 8, at 25°C. A separate measurement was done for each substrate. The substrates were first dissolved in DMSO and resuspended in the phosphate buffer reaching a final concentration of 0.05% DMSO in the phosphate buffer. The sample containing the protein and NAD<sup>+</sup> is mixed with the phosphate buffer containing the substrate in a 96-well plate making the volume up to 200  $\mu$ L in each well (180  $\mu$ L of phosphate buffer and 20  $\mu$ L of protein buffer). The fluorescence is produced due to the formation of NADH that is driven by the oxidation of E2 and 5-diol at position 17 in the presence of NAD<sup>+</sup>. NADH fluorescence emission was measured at 496 nm following excitation at 340 nm. The slit width for excitation was 7 nm while that for emission was 15 nm. A linear relationship between product formation and reaction time was obtained. The slope of these progress curves, representing the velocity of the NADH formation, was calculated by linear regression, and revealed the initial velocity. For each new batch of protein, to ensure the purity of the protein sample and the absence of any parasitic activity, a validation experiment was run without E2 as a control. This linear relationship was obtained by keeping the same concentration of protein but with a different substrate concentration with a series of dilutions ranging from 0.1 mM to 7.5 mM in the phosphate buffer.

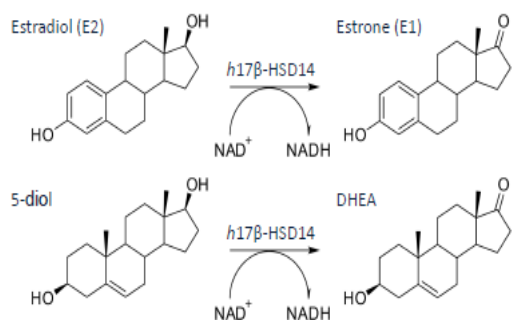
The dilution method is furthermore explained in **Table 2**.

**Table 2.** *The dilution series of the substrate and final concentration in the assay*

<b>Buffer Nr.</b>	<b>Concentration of stock solution (E2 or 5-diol) <sup>a</sup></b>	<b>Volume ratio (substrate + phosphate buffer) <sup>b</sup></b>	<b>Substrate concentration in phosphate buffer</b>	<b>Substrate concentration in assay <sup>c</sup></b>
10	7.5 mM	0.05% + 99.95%	37.5 $\mu$ M	33.7 $\mu$ M
9	5 mM	0.05% + 99.95%	25 $\mu$ M	22.5 $\mu$ M
8	3 mM	0.05% + 99.95%	15 $\mu$ M	13.5 $\mu$ M
7	2 mM	0.05% + 99.95%	10 $\mu$ M	9 $\mu$ M
6	1.5 mM	0.05% + 99.95%	7.5 $\mu$ M	6.75 $\mu$ M
5	1 mM	0.05% + 99.95%	5.25 $\mu$ M	5.625 $\mu$ M
4	0.75 mM	0.05% + 99.95%	3.75 $\mu$ M	3.375 $\mu$ M
3	0.5 mM	0.05% + 99.95%	2.5 $\mu$ M	2.25 $\mu$ M
2	0.3 mM	0.05% + 99.95%	1.5 $\mu$ M	1.35 $\mu$ M
1	0.2 mM	0.05% + 99.95%	1 $\mu$ M	0.9 $\mu$ M
0	0.1 mM	0.05% + 99.95%	0.5 $\mu$ M	0.45 $\mu$ M

<sup>a</sup> The concentration of substrate (E2 or 5-diol) in DMSO. <sup>b</sup> Ratio of substrate dissolved in DMSO to phosphate buffer. <sup>c</sup> The final concentration of substrate in the enzymatic assay after adding the protein sample ratio 1:9 of protein buffer and phosphate-substrate buffer, respectively.

All of the 17 $\beta$ -HSD14 variants were measured under the same conditions, except for the Gln148Ala variant because the reaction in the enzymatic assay was so fast, it was diluted 80-fold to a concentration of 45 nM to be able to determine the  $K_m$ .



**Figure 4.** The oxidation of the OH-group at C17 of both substrates (E2 and 5-diol). The reaction is triggered by the presence of NAD<sup>+</sup>. This mechanism is the driving force for the assay. The cofactor is reduced to NADH which gives the florescent emission in the enzymatic assay.

#### 2.2.4. Michaelis-Menten kinetic constant

To determine both kinetic parameters  $K_m$  and  $V_{max}$ , the substrate concentration was varied between 12% and 250% of the estimated  $K_m$  value. Multiple assay cycles took place in order to achieve an accurate statistic that solidifies the result of other similar runs. The assay for each variant was done in 96-well plates in triplicate (3 wells for each substrate concentration). At least five independent measurements were performed on each variant ( $n = 5-7$ ). The kinetic constants  $K_m$  ( $\mu M$ ) and  $V_{max}$  ( $nM \cdot sec^{-1}$ ) were calculated from the initial velocities by directly fitting the curves in Lineweaver-Burk plots. The  $k_{cat}$  values ( $min^{-1}$ ) were calculated from the  $V_{max}$  values. The fitting and data analysis was performed using GraphPad Prism 7 [155].

#### 2.2.5. Native PAGE electrophoresis

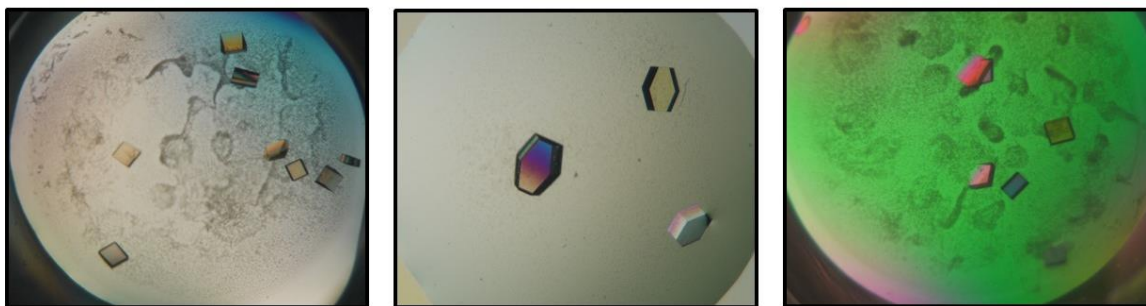
Native PAGE electrophoresis is a method used to determine the net charge, size and shape of the protein sample investigated. In this study it is used to confirm the tetrameric assembly of the variant sample, so both the wild type and the variants are included in this experiment. The stock solutions prepared are acrylamide solution (30 grams acrylamide and 0.8 grams of bis-acrylamide in 100 mL of distilled water) stored at 4°C, separating buffer (1.5 M Tris HCl dissolved in distilled water, pH = 8.8), stacking gel buffer (0.5 M Tris HCl dissolved in distilled water, pH = 6.8) and polymerizing solution (10% ammonium persulfate in distilled water and TEMED). The sample buffer is prepared by adding 15.5 mL of 1 M Tris HCl solution (dissolved in distilled water) pH=6.8, 2.5 mL of 1% bromophenol blue solution, 7 mL distilled water and 25 mL glycerol. This procedure is sufficient to make five gels with an average size of 10\*6 cm

and 2mm thickness. In the case of 17 $\beta$ -HSD14 variant samples, there is no need of adjusting the pH of the sample buffer because the protein buffer has a close pH range. To prepare the electrophoresis buffer, in a 1 L container 3 grams Tris base and 14.4 grams of glycerol are mixed and the volume is adjusted to 1 L with a final pH of 8.3. As for preparing the gel, at first the separating gel, 7.5 mL acrylamide solution, 7.5 mL separating gel buffer, 14.85 mL distilled water, 150  $\mu$ L polymerizing solution and 15  $\mu$ L TEMED were added in a flask and stirred gently for an even mixing and poured into a mold (normal SDS gel mold). 60-70% of the volume should be filled and it should rest for at least 30 minutes to solidify. A layer of isopropanol is applied on top of the gel in the mold to prevent the gel from drying or the formation of any air bubbles on the exposed separating gel's edge. While the separating gel is solidifying, the stacking gel solution should be prepared. 1.5 mL acrylamide solution, 3 mL stacking gel buffer, 7.4 mL distilled water and 100  $\mu$ L polymerizing solution were added in a flask and mixed gently. The isopropanol layer is poured out and the stacking gel is added on top, a plastic separator is added on top before the stacking gel solidifies to define a pocket in which the sample would be loaded into. The gel is gently transferred to an electrophoresis machine and filled with the electrophoresis buffer until the gel is completely soaked. The samples are loaded in the gel with 10  $\mu$ L volume from each and a protein ladder (Thermo Fisher) in one pocket is used as a reference for comparison. The gel electrophoresis is started with protein sample running toward the anode. At last, the gel is stained with Coomassie blue dye. This protocol is based on an established procedure [156] with some modifications to suit the protein sample characterization.

#### 2.2.6. Protein crystallization

Before crystallization of the protein, the activity and the concentration of the enzyme have to be guaranteed for every new protein batch prepared. Activity is checked with enzymatic activity assay (fluorometric assay) and concentration should be 10 mg/mL (35  $\mu$ M) in the protein buffer that contains 500 mM NaCl, 25mM Tris HCl, 250 mM glucose, 0.25 mM TCEP and 0.5 mM NAD<sup>+</sup> at pH=8. The protein was crystallized using hanging drop and sitting drop methods. Although, the crystallization buffers are the same for each variant, a variety of dilutions were used to obtain crystals for each variant. Co-crystallization is the method of choice. A stock of 100 mM of inhibitor **1** was dissolved in DMSO and a final volume of 5% of DMSO in the protein sample making the final concentration of inhibitor **1** in the sample 5mM.

The protein sample was incubated with inhibitor **1** for at least 20 minutes at room temperature. The mixture was centrifuged at 12000 g for 15 minutes to ensure all non-dissolved particles of inhibitor **1** have precipitated. The protein-inhibitor **1** sample drop was placed in the crystal plate and additionally each crystallization buffer was added to the protein drop with a ratio of 1:1. 3  $\mu\text{L}$  each of the protein sample and crystallization buffer was used to make the final volume of the drop 6  $\mu\text{L}$ . The crystallization buffer contains 10-25% (w/v) PEG6000, 0.1 M HEPES pH=7 and 5% DMSO (v/v). Each well in the crystal plate is filled with 800  $\mu\text{L}$  of crystallization buffer. A range of PEG6000 is used in the crystallization buffer 10-25% (w/v) and that is where the difference in the crystallization condition between each variant is noticed. The crystal plates are kept in a controlled cooling area at 18°C. During the crystal formation, any slight imbalance or change with the surrounding temperature ( $\pm 0.5^\circ\text{C}$ ) could affect the formation of the crystals or even, the crystals could form in a different PEG6000 concentration and that is the main reason why a variation of PEG6000 ratio was used in each well. It was seen in many cases that the crystals of the same protein variant are obtained from a different concentration of PEG6000. The hanging drop plates are sealed using a semi-viscous silicone gel placed around the rim of the well to fix the glass slides and the sitting drop plates are sealed by transparent tape (Hampton research). It takes between 10-20 days until a measurable protein crystal is grown. A cryo-protectant buffer is needed before the crystals are frozen in liquid nitrogen. The cryo-protectant buffer contains the same condition as the one the crystal is grown in with an additional 20% glucose v/v and inhibitor **1** (5 mM final concentration) to prevent the washing out of the inhibitor.



**Figure 5.** Crystals of 17 $\beta$ -HSD14 variants.

All of the crystals above are obtained using co-crystallization method in sitting drop plates. The size, shape, fragility and resolution of crystals is dependent on the thermal stability of the environment and at which PEG6000 concentration range they formed in. The crystals that form

in a higher ratio of PEG6000 tend to have less water content and a better resolution in most of the cases compared to their equals from a lower ratio of PEG6000.

### 2.2.7. Data collection and processing

For this study, all data were collecting at Helmholtz-Zentrum Berlin (Berlin-Adlershof, Germany) [157] at BESSY II at beam line MX14.1 and MX14.2. The data collection was done under cryogenic temperature at 100 K (-173°C) and wavelength 0.9184Å on a silicon Pilatus 6M pixel detector at atmospheric pressure. Data sets collected were indexed, processed and scaled using XDSAPP [158]. Details of the data collection and processing are given in **Table 5**.

### 2.2.8. Crystal structure determination and refinement

Crystal structures were determined by molecular replacement using the program PHASER MR [159] from the CCP4 suite [160]. The structure of the wild type 17β-HSD14 S205 variant was used as a search model for molecular replacement (PDB ID: 5ICM) after removing all non-protein molecules and hydrogen atoms using FCONV [161]. In the refinement, a 5% subset of the reflections was used for the  $R_{\text{free}}$  calculations and they were consequently omitted from the refinement. The model was built in COOT [162] and the refinement was done using PHENIX.refine version 1.10.1-2155 [163]. The restraints of the ligands were generated using the Grade Web Server [164] or by eLBOW from PHENIX [163] which was also used for energetically minimization and restraint generation. SMILES codes were obtained from Molinspiration v2014 [165]. First refinement step was performed with default parameters. XYZ coordinates, occupancies and individual B-factors were alternated with structural adaption in COOT [162]. A series of repeated refinement steps were applied until the electron density is well explained by the built model. His93Ala (PDB ID: 6HNO), Gln148Ala (PDB ID: 6FFB), Tyr253Ala (PDB ID: 6G4L) and Cys255Ala (PDB ID: 6GBT) were isotropically refined, while Lys158Ala (PDB ID: 6H0M) was anisotropically refined using PHENIX.refine [163]. Water molecules in all data sets were isotropic refined and hydrogen atoms were included in the final model at calculated positions. Refinement statistics are included in **Table 5**.

### 2.2.9. Mass spectrometry under denaturing and native conditions

The molecular weight of the mutated monomeric proteins was determined by electrospray/ionization (ESI) under denaturing conditions after passing through an RP-HPLC (RP4H or RP-C4) column and analyzed on an Orbitrap Velos Pro from Thermofischer Scientific. The molecular weight of the intact folded protein samples was determined by native mass spectrometry using a TOF-Mass spectrometer (Synapt G2 Si, Waters).

### 2.2.10. Volume calculation of the binding pocket

The volume of the binding pocket was calculated using FCONV [161], a program devolved in our working group. The first step of the program is to identify the residues surrounding the binding pocket. In the case of 17 $\beta$ -HSD14, the protein has to be studied as a dimer because of the residues from the adjacent monomer mate interacting with the inhibitor in the binding pocket. All amino acids clustered for this study should have an interaction with the inhibitor in the binding pocket or forming the interior surface of the same binding pocket. The amino acids that are engaged in this study are clustered into nine continuous chains surrounding the binding pocket: Thr15 – Ala23, Lys41 – Arg48, Cys61 – Gln65, Val85 – Gln97, Glu112 – Thr122, Asn138 – Gly144, Ile146 – Ala151, Ile182 – Gly216 and Thr246 – Lys256. The chain Thr246 – Lys256 is the one identified from the adjacent monomer mate.

## 2.3. Results

### 2.3.1. Site directed mutagenesis, expression, purification and characterization

The five selected amino acids His93, Gln148, Lys158, Tyr253 and Cys255 were individually mutated to alanine. Alanine is frequently selected to functionally probe the importance of residues as it usually maintains the spatial folding without introducing novel specific interactions. All the variants of 17 $\beta$ -HSD14 were applied to the wild type S205 variant. The molecular mass of the variants and the wild-type protein was determined under denaturing conditions. The observed masses, shown in **Table 3**, were consistent with the expected values.



**Table 3.** Molecular mass of the wild-type and the mutants as monomer and as tetramer

Protein	Monomer mass in Da <sup>a</sup> (calculated mass <sup>b</sup> )	Tetramer mass in Da <sup>c</sup>
Wild-type S205 variant	28641 (28641.78)	114500
Lys158Ala	28584 (28584.68)	115170
His93Ala	28576 (28575.71)	nd
Gln148Ala	28585 (28584.72)	115144
Cys255Ala	28609 (28609.72)	nd
Tyr253Ala	28549 (28549.68)	nd

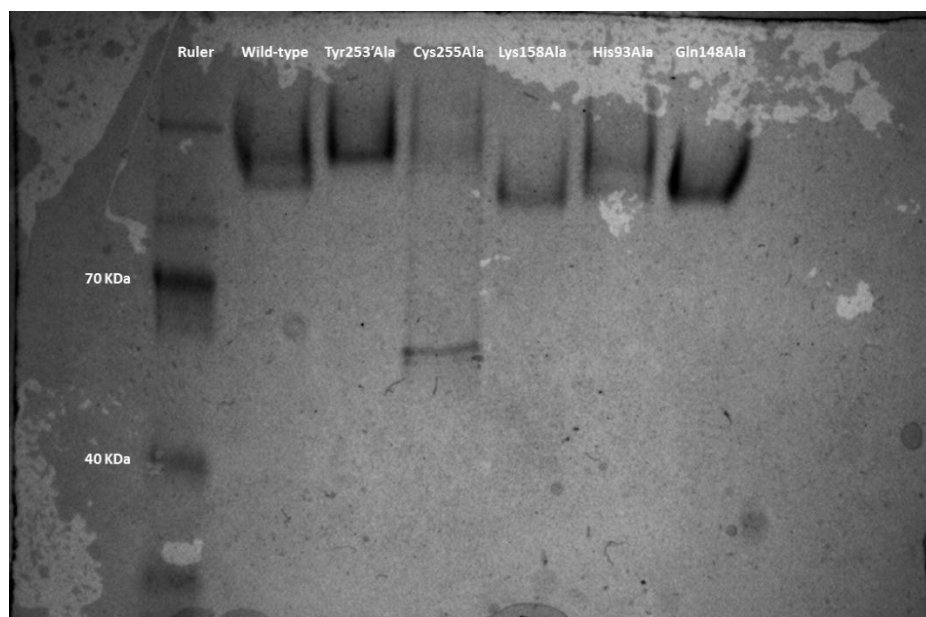
<sup>a</sup> The molecular masses of the monomers were determined by ESI mass spectrometry

<sup>b</sup> Calculated masses using the web Server *expasy* [166]

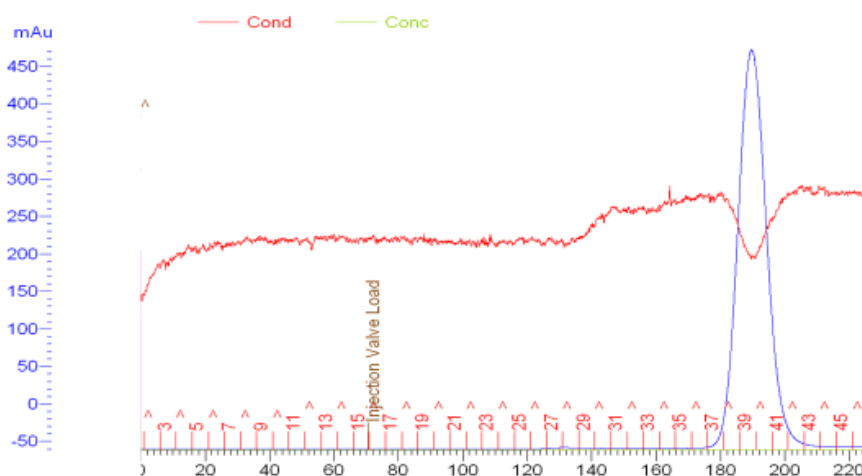
<sup>c</sup> The molecular masses of the tetramers were determined by TOF-mass spectrometry

nd: not determined

The assembly of all studied protein variants was further investigated using size-exclusion chromatography (**Figure 7**), native mass spectrometry and native PAGE gels (**Figure 6**). Although the mass was determined for the wild type (**Table 3**) and all the variants, the native mass determinations **Table 3** were only successful for the wild type (S205 variant) and for both Lys158Ala and Gln148Ala variants confirming that these species also form a homotetramer in the gas phase under the applied MS conditions. As for the His93Ala, Tyr253Ala and Cys255Ala variants, no interpretable mass peak could be obtained under the conditions of the native mass spectrometry experiment. Regarding the size-exclusion chromatography, the column used was SEC200 (HiLoad™ 26/200 Superdex™ 200pg GE Life Science). All protein samples (wild type and variants) eluted after addition of the same eluent volume and with comparable chromatographs (**Figure 7**). The final protein yield after a successful purification of the wild type and the variants ranges from 7 to 13 mg for each litre of cell culture utilized.



**Figure 6.** Native PAGE of the wild-type and the mutants



**Figure 7.** The diagram of the eluted protein peak. The highest peak shows the eluted protein in the flow through after using 100% of elution buffer. the ÄKTA machine was used in this experiment. The sample used in this figure is the wild type (S205 variant). X axis (volume mL) and Y axis (UV emission)

The native PAGE gel analysis showed a similar mobility for all studied variants as for the wild type confirming their tetrameric assembly except for the Cys255Ala variant. Under the conditions of the native gel, a preponderant band was observed at the molecular weight of the

dimer for Cys225Ala and might suggest that for this variant the equilibrium is shifted predominantly toward the dimer [139].

### 2.3.2. Enzyme kinetics of 17 $\beta$ -HSD14 wild type and variants

In a previous study [142], the wild type enzyme S205 variant showed a higher efficacy in the presence of E2 (143 nM sec<sup>-1</sup> mg<sup>-1</sup> of substrate converted by the protein) compared to the efficacy with 5-diol (68 nM sec<sup>-1</sup> mg<sup>-1</sup> of substrate converted by the protein). Although, in the case of the two variants Lys158Ala and Tyr253Ala, no matter which substrate was used both variants have very low or no catalytic activity (between 0.1 and 12 nM sec<sup>-1</sup> mg<sup>-1</sup> of substrate converted by the protein). Unfortunately, due to the absence of activity, no kinetic parameters could be determined for these two variants. The Cys255Ala variant showed a similar turnover ( $k_{cat}$ ) as the wild type for both substrates (E2 and 5-diol), shown in **Table 4**. The  $K_m$  value, however, differed slightly for E2 exhibiting a five-fold higher  $K_m$  and for 5-diol a 25-fold increase. On average, the kinetic profile of this variant was similar to the wild type for both substrates. Obviously, the mutated residue Cys255 did not have a pronounced impact on the kinetic properties of the enzyme. In case of the His93Ala variant, a three to four-fold faster initial velocity was observed for E2 and 5-diol and for Gln148Ala variant a 20-fold increase was observed with E2 but not for 5-diol. The initial velocity of the His93Ala variant was found to be slightly higher than the wild type with E2 (four times). This result matched well with a slightly increased turnover of the mutant. The Gln148Ala variant also catalyzes the reaction faster in presence of E2, and the higher initial velocity correlated well with the improved turnover (0.51 min<sup>-1</sup>) compared to the wild type (0.02 min<sup>-1</sup>). The faster turnover of this mutated form was not observed with 5-diol, thus this variant clearly discriminated between both substrates exhibiting a nine-fold and a 22-fold higher  $k_{cat}$ , respectively. In presence of 5-diol nearly the same  $k_{cat}$  was observed for E2 and 5-diol (4-fold and 2-fold, respectively). The catalytic efficiency ( $k_{cat}/K_m$ ), which allows the comparison of the two substrates, E2 and 5-diol for one enzyme, highlights that for the variant Gln148Ala E2 is more efficient than 5-diol by a factor of five.

**Table 4.** Enzyme kinetics of 17 $\beta$ -HSD14 wild type and variants.

	Estradiol				5-diol			
	Initial velocity <sup>b</sup>	$K_m^c$	$V_{max}^d$	$k_{cat}^e$	Initial velocity <sup>b</sup>	$K_m^c$	$V_{max}^d$	$k_{cat}^e$
<b>Wild-type<sup>f</sup></b>	143 $\pm$ 16	6.2 $\pm$ 1.4	1.3 $\pm$ 0.1	0.02	68 $\pm$ 9.0	6.6 $\pm$ 1.4	1 $\pm$ 0.4	0.02
<b>Lys158Ala<sup>f</sup></b>	5.6 $\pm$ 5.0	nd	nd	nd	12 $\pm$ 2.2	nd	nd	nd
<b>His93Ala<sup>f</sup></b>	541 $\pm$ 35	6.2 $\pm$ 1.9	12.0 $\pm$ 2.4	0.21	214 $\pm$ 28	0.3 $\pm$ 0.1	3.5 $\pm$ 0.2	0.06
<b>Gln148Ala<sup>f</sup></b>	2741 $\pm$ 158	ns	nd	nd	135 $\pm$ 4	2.1 $\pm$ 0.9	2.5 $\pm$ 0.3	0.04
<b>Gln148Ala<sup>g</sup></b>	nd	4.0 $\pm$ 0.4	0.4 $\pm$ 0.1	0.51	nd	nd	nd	nd
<b>Cys255Ala<sup>f</sup></b>	155 $\pm$ 20	1.3 $\pm$ 0.3	2.8 $\pm$ 0.3	0.05	66 $\pm$ 8	0.25 $\pm$ 0.03	1.1 $\pm$ 0.4	0.02
<b>Tyr253Ala<sup>f</sup></b>	5.4 $\pm$ 2.0	nd	nd	nd	0.1 $\pm$ 2.2	nd	nd	nd

Mean values and given standard deviations were calculated based on 5–7 measurements. <sup>b</sup> Given in nM.sec<sup>-1</sup>.mg<sup>-1</sup> protein, [substrate]= 32  $\mu$ M. <sup>c</sup> Given in  $\mu$ M. <sup>d</sup> Given in nM.sec<sup>-1</sup>. <sup>e</sup> Given in min<sup>-1</sup>. <sup>f</sup> Protein concentration: 3.5  $\mu$ M, 1 mg/mL. <sup>g</sup> Protein concentration: 45.5 nM, 0.013 mg/mL. nd: not determined, ns: no saturation.

### 2.3.3. Comparison of crystal structures

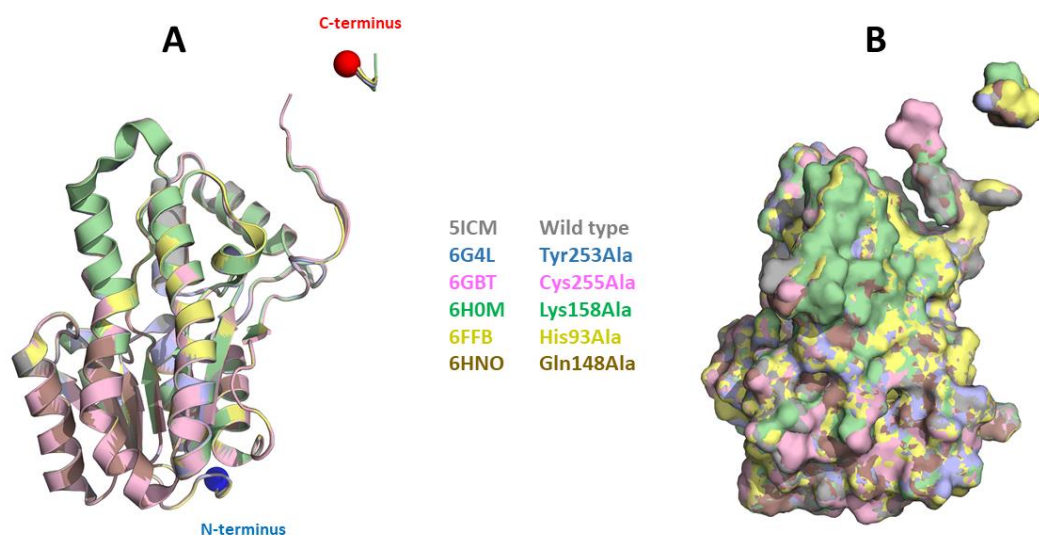
Details of the data collection, processing, diffraction data, refinement and other parameters related to 17 $\beta$ -HSD14 crystal structures variants are listed in **Table 5**.

**Table 5.** Data collection and refinement statistics for all crystal structures

Variant PDB code	Tyr253Ala 6G4L	Cys255Ala 6GBT	Lys158Ala 6H0M	Gln148Ala 6FFB	His93Ala 6HNO
<u>Data collection and processing</u>					
Space group	I 4 2 2	I 4 2 2	I 4 2 2	I 4 2 2	I 4 2 2
Unit cell	91.5 91.5 132.2	91.6 91.6 134.1	91.6 91.6 131.1	91.6 91.6 132.5	91.4 91.4 132.7
Matthews coefficient <sup>a</sup>	2.4	2.4	2.4	2.4	2.4
Solvent content (%)	49.4	49.9	49.0	49.4	49.2
<u>Diffraction data</u> <sup>b</sup>					
Resolution range	50-1.44 (1.53-1.44)	50-2.10 (2.22-2.10)	50-1.25 (1.33-1.25)	50-1.65 (1.75-1.65)	50-1.68 (1.78-1.68)
Unique reflections	50078 (7713)	17086 (2653)	76893 (12249)	33549 (5353)	31576 (5144)
R(I) <sub>sym</sub> (%) <sup>c</sup>	5.5(50.3)	8.7(47.7)	4.5(49.5)	9.3(50.4)	5.8(50.2)
Wilson B factor	23.4	31.6	18.1	18.2	20.7
completeness	99.4(96.3)	99.4(97.4)	99.9(99.7)	97.7(98.0)	97.5(99.7)
redundancy	8.5(8.2)	8.6(9.0)	10.1(9.5)	8.6(8.8)	8.2(8.7)
<1/σ(I)>	20.6(3.2)	20.8(4.8)	27.1(4.1)	14.0(2.9)	23.2(4.0)
<u>Refinement</u>					
Resolution range	46.21-1.44	37.83-2.10	46.07- 1.25	45.34-1.65	46.32-1.68
Reflections used in refinement (work/free)	50076 (47572/2504)	17086 (16231/855)	76893 (73048/3845)	33549 (31871/1678)	31576 (29997/1579)
Final R value for all reflections (work <sup>d</sup> /free <sup>e</sup> ) (%)	15.6/17.9	14.9/19.0	12.6/14.7	14.9/17.5	16.8/20.5
Protein residues	253	255	257	253	257
Water molecules	218	159	239	214	191
rmsd from ideality: bond length (Å)	0.008	0.009	0.007	0.008	0.009
rmsd from ideality: bond angle (°)	1.005	0.886	1.008	0.998	1.038
Ramachandran most favored(%) <sup>f</sup>	97.6	97.7	98.1	98.4	98.0
Ramachandran additionally allowed (%) <sub>f</sub>	2.4	2.3	1.9	1.6	2.0
Mean B factor protein (Å <sup>2</sup> ) <sup>g</sup>	18.9	25.6	15.5	21.1	28.2
Mean B factor ligand (cofactor/ligand) (Å <sup>2</sup> ) <sup>g</sup>	14.7/18.9	23.7/27.1	11.4/-	15.5/35.1	28.9/-
Mean B factor water molecules (Å <sup>2</sup> ) <sup>g</sup>	29.2	34.0	27.2	32.2	35.6

<sup>a</sup> Calculated with Matthews\_coef program from CCP4 suite version 6.4.0. <sup>b</sup> Values in parenthesis describe the highest resolution shell. <sup>c</sup>  $R(I)_{\text{sym}} = [\sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)] \times 100$ , in which  $I(h)$  is the mean of the  $I(h)$  observation of reflection  $h$ . <sup>d</sup>  $R_{\text{work}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$ . <sup>e</sup>  $R_{\text{free}}$  was calculated as shown for  $R_{\text{work}}$  but on refinement-excluded 5% of data. <sup>f</sup> Calculated with PROCHECK [167]. <sup>g</sup> Mean B factors were calculated with MOLEMAN [168].

All protein variants were co-crystallized with inhibitor **1** to check the difference of the binding mode in the various variants. Crystal structures were obtained for all 17 $\beta$ -HSD14 variants, holo-enzyme for His93Ala and Lys158Ala as a ternary complex with the cofactor and in complex with cofactor and inhibitor **1** for variants Gln148Ala, Tyr253Ala and Cys255Ala.



**Figure 8.** Superimposition of the crystal structures of the different 17 $\beta$ -HSD14 mutated variants. **A.** Cartoon representation. **B.** Surface representation. The cofactor and inhibitor **1** are omitted for clarity. The PDB IDs of the corresponding 3D structures are indicated by the colours used to represent the individual protein.

As seen in **Figure 8**, there are no structural changes and all the variant structures exhibit the same overall geometry. In the case of Cys255Ala variant (PDB ID: 6GBT) shown in light pink in **Figure 8**, one extra residue is visible in the electron density at the end of the flexible loop (before the C-terminus) which is Arg259. The C-terminus was not visible in the electron density for the Cys255Ala variant, unlike in other variant structures. To get a better understanding of the difference between the variant structures, the maximum RMSD (maxRMSD) between the C $\alpha$  atoms of two structures was calculated to obtain insight into the local variation between the mutated variants (**Table 6**). Each maximum RMSD value was calculated for each variant with each other as pairs.

**Table 6.** Maximum RMSD between C $\alpha$  carbons for pairs of the different structures

MaxRMSD (Å)	Structures		
0.23	Gln148Ala	+	Lys158Ala
0.24	Cys255Ala	+	Lys158Ala
0.22	His93Ala	+	Cys255Ala
0.21	His93Ala	+	Lys158Ala
0.21	Gln148Ala	+	Cys255Ala
0.18	His93Ala	+	Tyr253Ala
0.18	His93Ala	+	Gln148Ala
0.17	Tyr253Ala	+	Lys158Ala
0.18	Tyr253Ala	+	Cys255Ala
0.17	Tyr253Ala	+	Gln148Ala

Fitting was performed using the McLachlan algorithm as implemented in the program Profit [169].

All maximum RMSD values obtained are rather low and very similar. These values highlight the similarity of all mutants. Nonetheless, a slight variation can be observed with maxRMSD values between pairs with Tyr253Ala. This could be due to the location of this residue on the flexible loop near the C-terminus. The flexible loop in 17 $\beta$ -HSD14 acts as a gate keeper that regulates the compound accessibility to the active site. Residue Tyr253 is pointing toward the active site and mutating it to Alanine gives better accessibility and higher exchange and movement rate in the active site compared to the wild type.

#### 2.3.4. Volume comparison of the binding pocket

The volume of the binding pocket was calculated for each variant structure using FCONV program [161]. This study was done to observe any variation in the volume of the active site cavity especially after mutating amino acids that are located around or inside the binding pocket. The values of volume of the binding pocket are shown in **Table 7**.

**Table 7.** *Volume of the binding pocket in the wild type and different mutated variants*

<b>Protein</b>	<b>Volume of the binding pocket (Å<sup>3</sup>)<sup>a</sup></b>
Wild type	2619.2
His93Ala	2365.8
Gln148Ala	2171.2
Lys158Ala	2359.4
Tyr253Ala	1824.9
Cys255Ala	2514.3

<sup>a</sup> The volume of the binding pocket is measured using FCONV [161].

There are slight differences seen in the volume of the binding pocket (**Table 7**). The largest volume variation is observed for Gln148Ala and Tyr253Ala replacements. This variation might be due to the removal of the side chain residue Tyr253' located on the flexible loop on the adjacent monomer mate, although all calculations were done with respect to the dimer form but the final value of the volume of the binding pocket calculated is for one monomer and this change will affect the cavity in both adjacent monomer mates. In the FCONV calculations, all the amino acids located along the rim of the binding pocket or interacting with the ligand or the cofactor are considered essential in the calculation. By superimposing the binding pockets of all mutants, they all fit identically and there are no significant changes in the arrangement or the positioning of the amino acids that form the binding pocket are observed. The only difference is observed for the Tyr253Ala mutation.

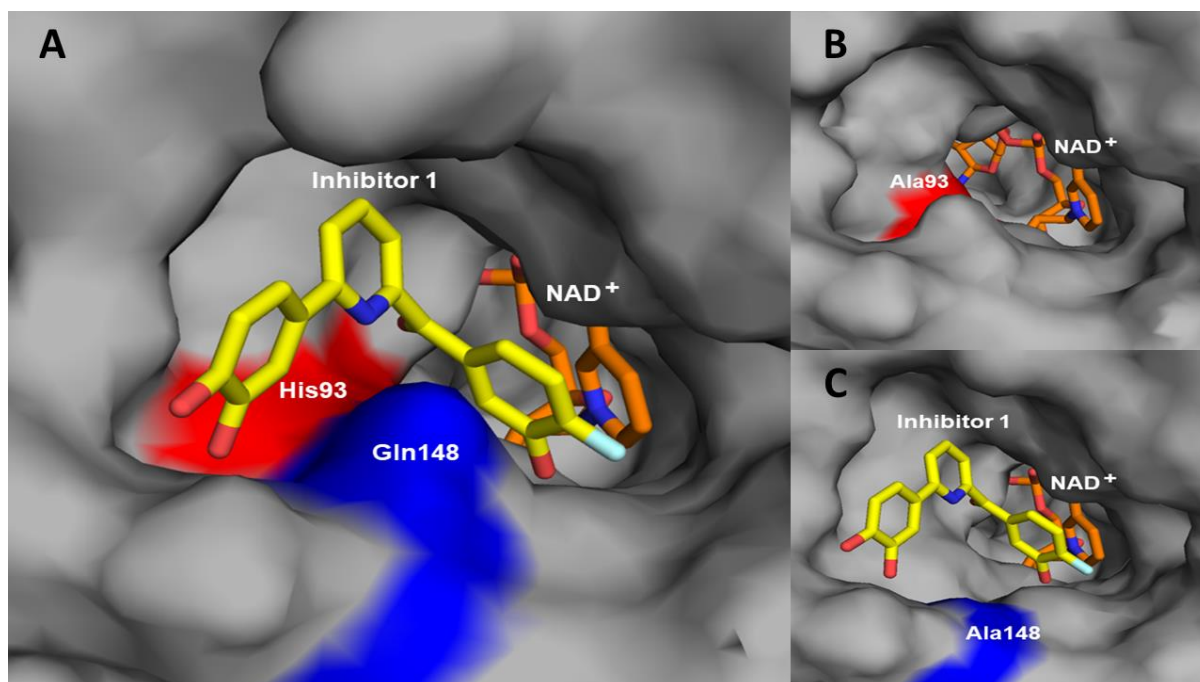


## 2.4. Discussion

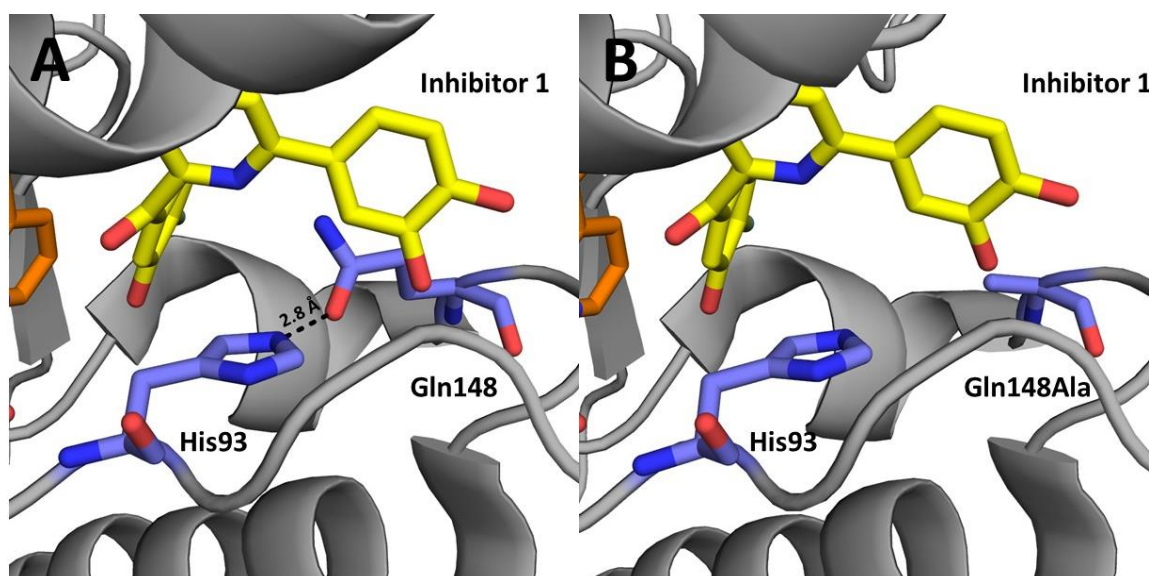
### 2.4.1. His93Ala and Gln148Ala variants

Although no double mutation was presented in a single structure in this study, in the case of His93Ala and Gln148Ala variants, they should be studied together because of their positioning (**Figure 9**) on the inner surface of the binding pocket and on the outer rim (entrance) of the binding pocket, respectively. These two residues (His93 and Gln148) are not conserved in the 17 $\beta$ -HSD family like 17 $\beta$ -HSD1 and 17 $\beta$ -HSD2, the other common trait with both residues is that they have no direct interaction with the ligand in the active site via H-bonds. Instead, they share H-bonding with each other (**Figure 10A**), thus contributing to the folding of the protein as well as forming the entrance of the binding pocket as seen in **Figure 9A**. The crystal structure of His93Ala variant (**Figure 9B**) did not indicate any significant shifts of the neighbouring amino acids or the backbone chain. However, the volume of the binding pocket did differ significantly compared to the wild type. Despite all the similarity with the wild type, even in the kinetics, the turnover for His93Ala expressed by  $k_{cat}$  in presence of E2 and 5-diol is slightly increased in comparison to the wild type, nine times higher in His93Ala with E2 than the wild type and four times with 5-diol, suggesting that His93 has to some extent a role in differentiation of velocity of the enzymatic reaction. This is intriguing, after all the similar characteristics to the wild type. The His93Ala structure was co-crystallized with inhibitor **1** like all other variants and the wild type compared to. Nevertheless, introducing the inhibitor **1** to the active site of the variant was unsuccessful, even after many trials were done to obtain a bound compound in the active site. Unfortunately, no ternary complex could be obtained. To finalise this result, there is a possibility that inhibitor **1** binds in the His93Ala variant but a crystal structure in complex with the inhibitor was not obtained in the experiments undertaken with this variant. In the case of Gln148Ala variant, the residue is positioned at the entrance rim of the binding pocket. Due to this positioning, this amino acid should regulate the exchange rate of the compounds entering the binding pocket, as seen in **Figure 9C**. The Gln148Ala variant structure (PDB ID:6FFB) is missing the dent on the entrance of the binding pocket, thus opening a large portion of the entrance that was blocked by residue Gln148 (**Figure 9A**). This might lead to a higher exchange rate of compounds through the active site and larger molecules could enter the binding pocket. The H-bonding between His93 and Gln148 in the wild type sterically hinders the access to the binding site. Although this H-bonding was dismissed in His93Ala, no structure was obtained in complex with inhibitor **1**. However, the crystal structure

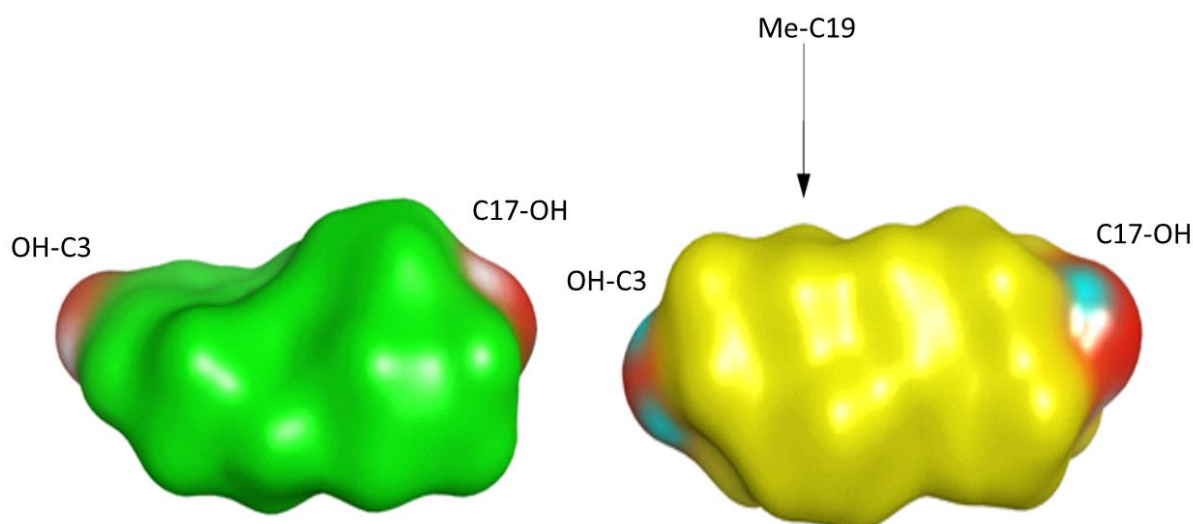
of the Gln148Ala variant has been obtained in complex with inhibitor **1**. The kinetics of Gln148Ala showed a higher turnover compared to the wild type, the variant is 22-fold more active in the presence of E2. However, this high value is less significant as with 5-diol where the turnover is only 2-fold higher than the wild type. The catalytic efficacy ( $k_{cat}/K_m$ ) is increased 30-fold for E2 against 11-fold for 5-diol, suggesting a discrimination for the variant Gln148Ala between both substrates (E2 and 5-diol). Estrogens (E2) and androgens (5-diol) differ in the presence or absence of the C19 methyl group between the steroidal A and B-ring (**Figure 11**). As discussed before, the missing H-bonded gate in Gln148Ala variant possibly facilitates the access of the substrates (E2 and 5-diol) to the binding pocket cavity. However, there is a possibility that the additional bulky C19 methyl group in 5-diol makes it harder to enter the active site, while in the case of E2, this methyl group is not present. This makes E2 slender and smaller than 5-diol, leading to greater advantage from the missing dent at the entrance of the active site after mutating Gln148 and this is one explanation of the kinetics observed for both substrates. Even more, it could be that in the wild type the rupture of the His93...Gln148 H-bonded gate defines the rate limiting step whereas in the mutated variant the diffusion into the catalytic centre becomes the limiting factor. In consequence, the catalytic turnover of the better accessible variant starts to discriminate the efficacy of substrate turnover with respect to their spatial demand.



**Figure 9.** Representation of the active site. **A.** The wild type  $17\beta$ -HSD14 in grey, highlighting His93 (red surface) and Gln148 (blue surface) PDB ID: 5ICM. **B.** The His93Ala variant in grey and the mutated residue Ala93 (red surface) PDB ID: 6HNO. **C.** The Gln148Ala variant in grey and the mutated residue Ala148 (blue surface) PDB ID: 6FFB. Both the cofactor  $NAD^+$  and the inhibitor **1** are coloured in orange and yellow, respectively.



**Figure 10.** **A.** Crystal structure of  $17\beta$ -HSD14 wild type (PDB ID: 5ICM). **B.** Crystal structure of Gln148Ala variant (PDB ID: 6FFB).

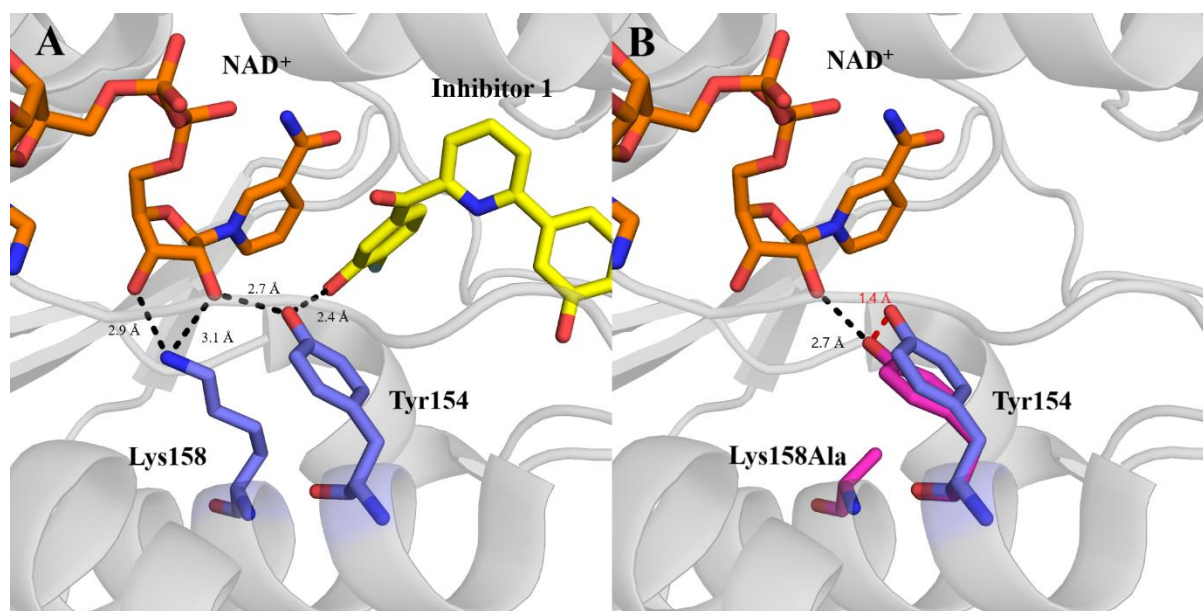


**Figure 11.** Solvent accessible surface representation to indicate the spatial requirements of E2 (left, green) and 5-diol (right, yellow). The red termini on both sides correspond to the OH group attached at C3 and C17. This figure was taken from the paper by Badran et al. [139].

#### 2.4.2. Lys158Ala variant

The catalytic triad of 17 $\beta$ -HSD14 consists of three amino acids Ser141, Tyr154 and Lys158 [140]. A study by Puranen et al. [170] suggests that the conserved catalytic triad's amino acids are essential for the catalytic activity for all members of the SDR family. Ser141 and Tyr154 of the catalytic triad interact with the OH group at C17 of the substrate which will fix the binding ligand in an optimal geometry making it possible for the hydride to transfer. Lys158 stabilizes the cofactor NAD<sup>+</sup> via H-bonding interactions with distances of 2.9Å and 3.1Å to the hydroxy group of the ribose ring of nicotinamide. There is a strong believe that the positively charged Lys158 lowers the pKa value of the adjacent catalytic triad amino acid Tyr154. This proposition is based on the a similar case from a previous study with 17 $\beta$ -HSD1 [153]. The main reason to mutate Lys158 to an alanine is to confirm that this amino acid, which is not directly involved in the substrate recognition, has a key role in 17 $\beta$ -HSD14 activity. Furthermore, when Lys158 is mutated to an alanine, the H-bonding contact to the cofactor and the interaction with Tyr154 is lost. The shift of the hydroxyl group of Tyr154 toward the vacancy is created by mutating Lys158. The OH-group moved further away from the inhibitor **1** (Figure 12B) binding position (equals 1.4 Å). Interestingly, the structure of the Lys158Ala variant could not be obtained as a ternary complex with both the cofactor and inhibitor **1**, but only with the cofactor as a holoenzyme. Apparently, the shift already mentioned had an impact on the stability of the ligand via H-bonding interactions that were present before the OH-group

of Tyr154 shifts. Regarding the cofactor NAD<sup>+</sup>, there was no shifting or displacement and it found at the exact position as it is in the wild type structure (PDB ID: 5ICM).



**Figure 12.** Representation of the active site of 17β-HSD14. **A.** structure of the wild type (PDB ID: 5ICM). **B.** Structure of Lys158Ala variant (PDB ID: 6H0M).

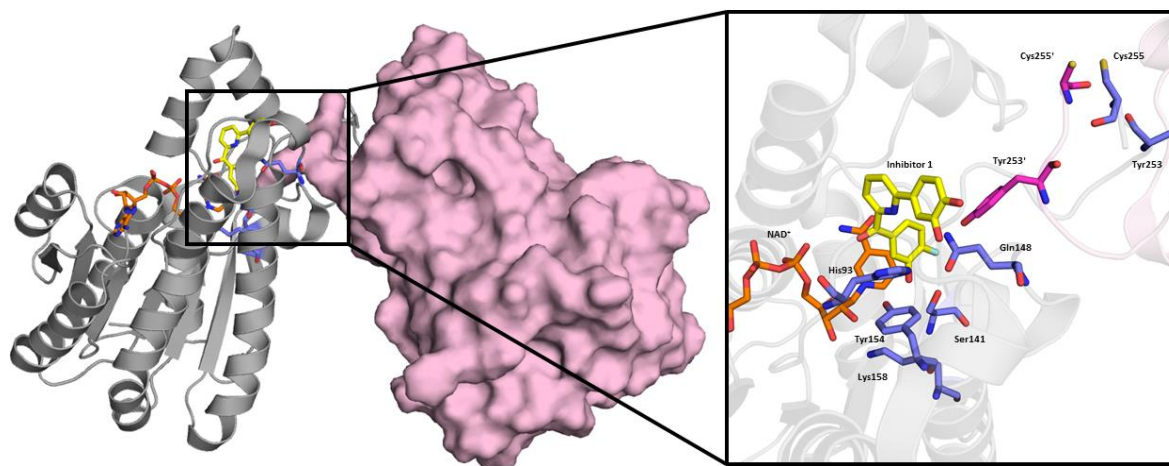
In **Figure 12B**, the amino acids of the Lys158Ala variant are coloured in pink and the Try154 in blue colour is superimposed, the distance labelled in red is the shift of the OH-group of Tyr154. The cumulative effect of the OH-group shift and the loss of the H-bonding contributes to the inactivity of the Lys158Ala variant. Furthermore, the loss of the electrostatic interactions between Lys158 and Tyr154 had most definitely a major impact on the enzymatic activity of the protein and that was already observed in the kinetic study for both the wild type and the Lys158Ala variant.

### 2.4.3. Tyr253Ala variant

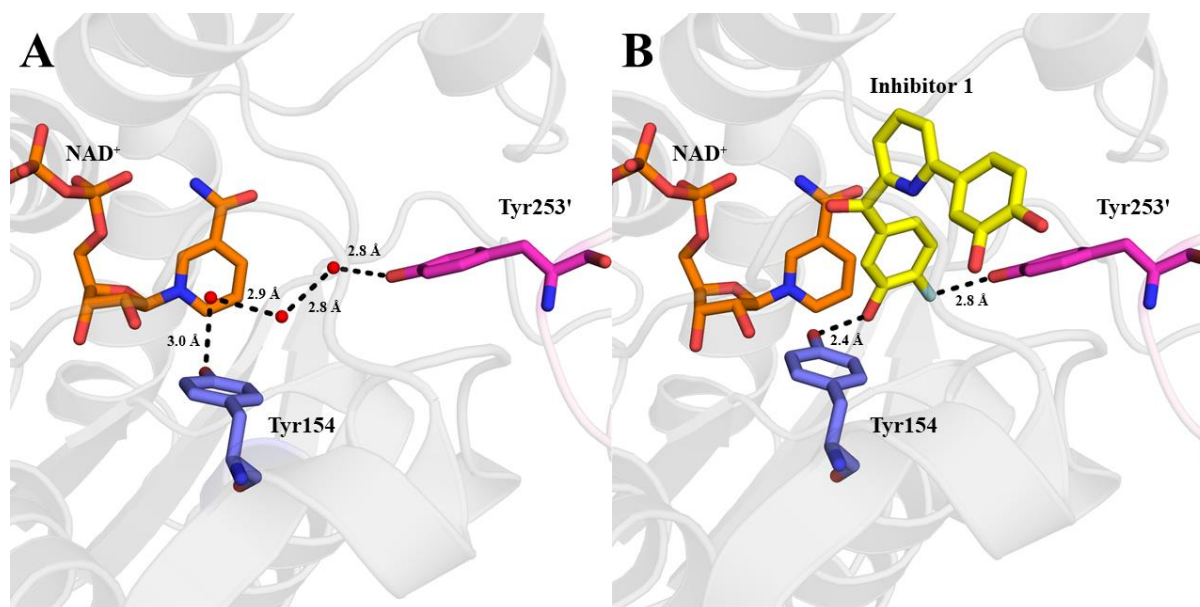
As observed in the wild type crystal structure, the residue Tyr253' from the adjacent dimer mate penetrates into the substrate binding pocket of the neighbouring monomer (**Figure 14**), which will have an effect on the catalytic properties of the enzyme. According to this observation, the Tyr253 residue was mutated to an alanine to have a better understanding of its role in the protein. After inspecting the apo structure of the wild type (PDB ID: 5ICS), a hydrogen bond interaction between Tyr253' and the catalytic Tyr154 through a chain of three conserved water molecules was found (**Figure 14B**), while in the wild type structure in



complex with inhibitor **1** (PDB ID: 5ICM), it was seen that the inhibitor **1** is able to displace these three conserved water molecules that link Tyr253' and Tyr154 (**Figure 14A**), while improving its binding pose.

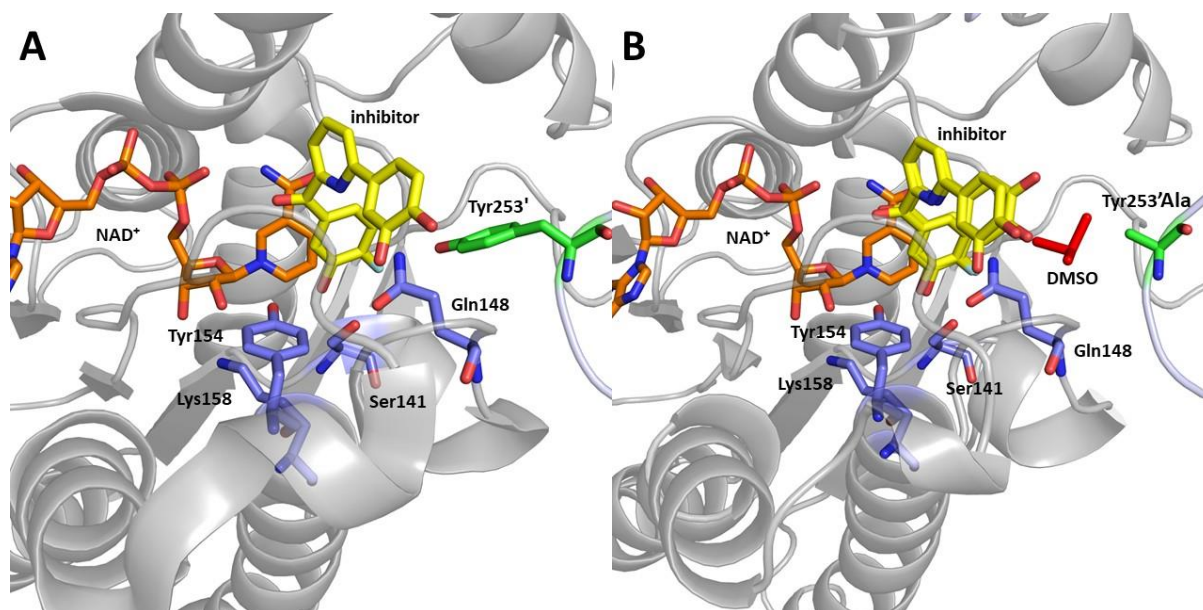


**Figure 13.** Dimer arrangement of 17β-HSD14 as found in the crystal structure and a close up view of the ligand binding site (PDB ID: 5ICM). The cofactor and inhibitor **1** are presented as sticks in orange and yellow respectively. The amino acids (catalytic triad and mutated residues) from the suggested monomer (grey) are shown in blue sticks and the amino acids from the adjacent monomer mate (light pink) are shown in pink sticks.

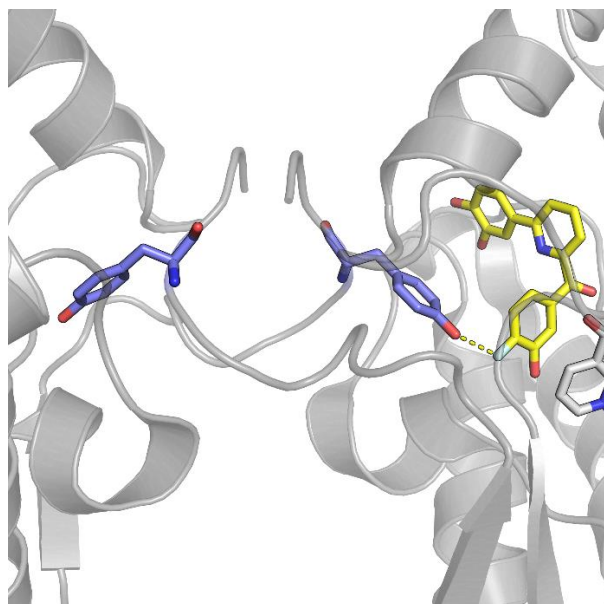


**Figure 14.** Closer view of 17β-HSD14 active site. **A.** Structure of the apo enzyme (PDB ID: 5ICS), showing the water molecules (red dots) connecting between the catalytic Tyr154 and Tyr253' from the adjacent monomer mate. **B.** Structure of ternary complex with inhibitor **1** (PDB ID: 5ICM), showing the displacement of the water molecules by the potent inhibitor.

The crystal structure of Tyr253Ala variant indicates two alternative configurations for the terminal phenyl ring in inhibitor **1** (**Figure 15B**). By taking the inhibitor complex of the wild type as a reference to compare the binding pose of the substrate. It can be argued that the steroidal substrate is stabilized through a similar H-bonding network extending from the OH-group of Tyr154 via C17-OH of Estradiol and a conserved water molecule toward the OH-group of Tyr253'. Unfortunately, after many efforts, a crystal structure for both the wild type and Tyr253Ala variant in complex with the substrate (E2) were not possible to obtain. Even with the product (estrone), no crystal structure was obtained to gain a proof of the hypothesis. Nonetheless, according to the absence of activity of the Tyr253Ala variant, it is reasonable to assume that the substrate is not positioned correctly nor sufficiently stabilized in the active site of the protein to allow the hydride transfer to occur. As observed from the crystal structure of 17 $\beta$ -HSD14 wild type, both residues Tyr253 and Tyr253' strongly support the dimer contact and stability by mutually penetrating into the active sites of the adjacent monomer mate hooking them tightly together. The interaction that is seen in **Figure 14B** between Tyr253' and fluorine moiety of inhibitor **1** is another proof how this residue is engaged in the activity of the protein as a dimer. Even after mutating Tyr253, the inhibitor is still binding in the active site, but the loss of activity in the kinetic assay is additional proof of how important this amino acid is in the activity and stability of 17 $\beta$ -HSD14.



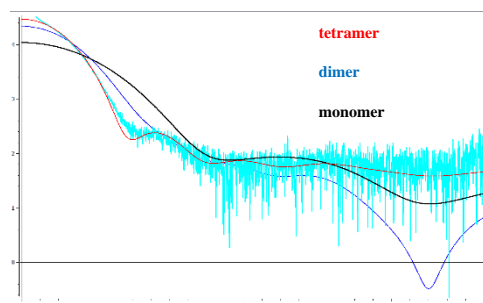
**Figure 15.** **A.** Crystal structure of 17 $\beta$ -HSD14 wild type in complex with cofactor and inhibitor (PDB ID: 5ICM). **B.** Crystal structure of Tyr253Ala variant (PDB ID: 6G4L). The protein is presented as grey cartoon. Tyr253' is depicted by green sticks while the amino acids of the neighbouring monomer are coloured in blue, cofactor and inhibitor in orange and yellow, respectively. The unoccupied space created by mutating Tyr253 to an alanine (**B**) is partially filled with a DMSO molecule.



**Figure 16.** Crystal structure of 17 $\beta$ -HSD14 (PDB ID: 6H0M) showing the positioning of both residues Tyr253 and Tyr253' (blue), each pointing in the opposite direction toward the active site of the other adjacent monomer mate.



As a last proof of the effect of this mutant on the tetrameric assembly, a SAXS (Small angle X-ray scattering) experiment at the PETRA III beamline at DESY, Hamburg was done. The experiment was undertaken on a protein sample solution of the Tyr253Ala variant and it showed a dominant tetrameric arrangement in the protein buffer (**Figure 17**).

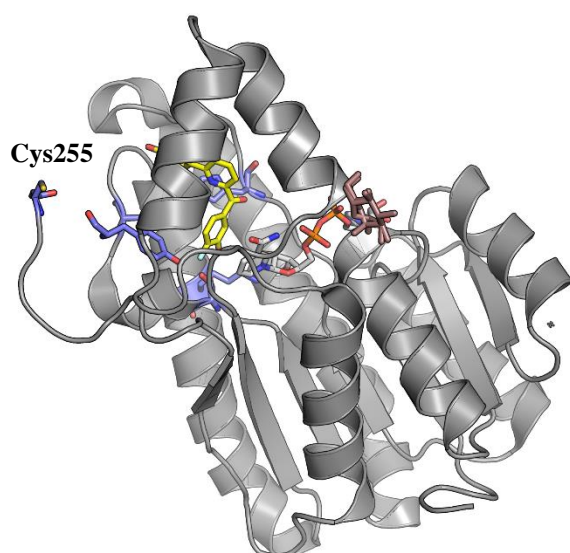


**Figure 17.** SAXS experiment. The graph shows the tetrameric signal (red curve) corresponding to the signal measured in the protein buffer.

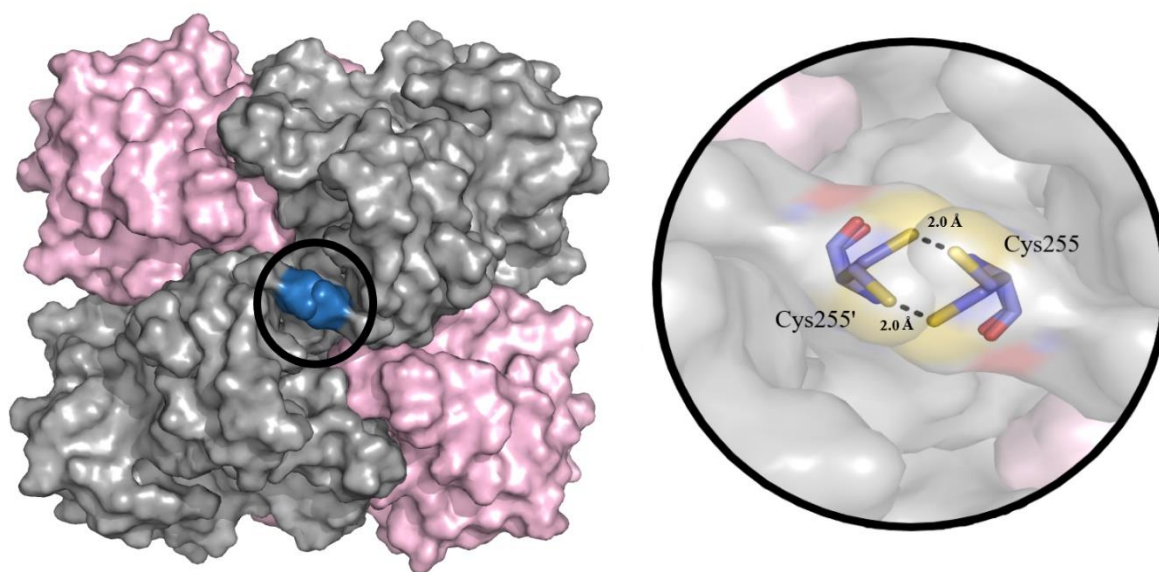
The protein concentration used for the SAXS experiment is 1 mg/mL and the sample buffer contains the same ingredients as the ones utilized in the purification step.

#### 2.4.4. Cys255Ala variant

The argument that the dimer association is the engine behind the activity of the enzyme drives the idea of mutating Cys255 to an alanine. Cys255 is located at the flexible loop of 17 $\beta$ -HSD14 (**Figure 18**). The mutation of Cys255 to an alanine means abolishing the disulfide bridge formed between Cys255 and Cys255' from the adjacent monomer mate (**Figure 19**). In the Cys255Ala variant structure, the disulfide bridge ceased to exist, which was proven by the missing electron density in the X-ray crystallography experiment and further supported by the lack of an anomalous signal alternatively observed for sulfur atoms in the electron density [160]. Also, a comparison was done between the electron density maps for both the Cys255Ala variant and the wild type.

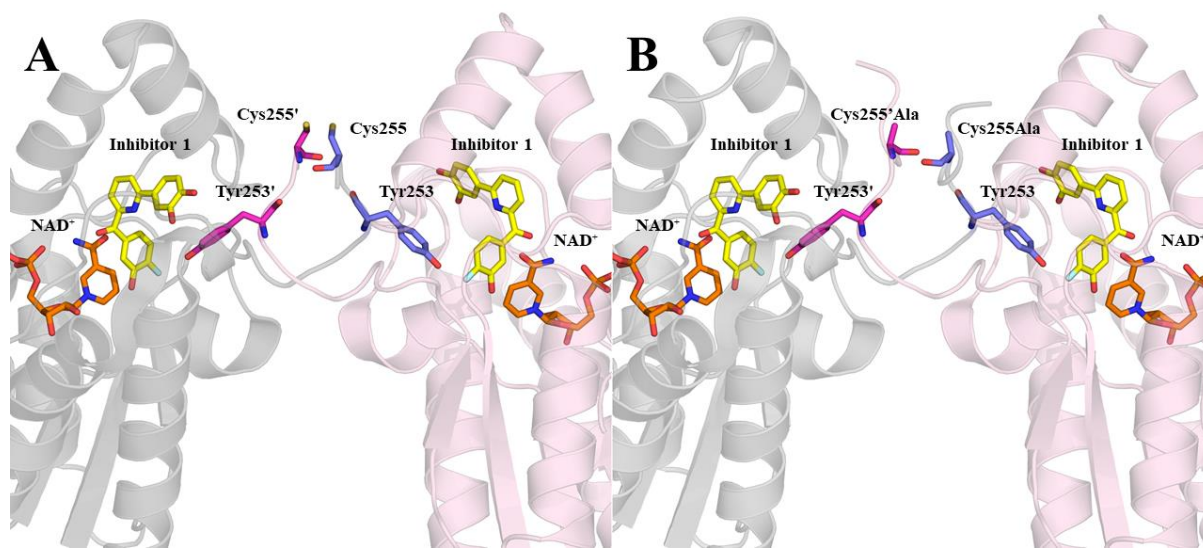


**Figure 18.** A cartoon representation of the 17β-HSD14 structure showing the residue Cys255 (blue) and the C-terminal tail. The other mutated amino acids are shown as well in blue, inhibitor 1 in yellow and glucose in light pink.



**Figure 19.** Surface representation of 17β-HSD14 Lys158Ala variant (PDB ID: 6H0M) and a closer view of the disulfide bridge formed by the two cysteine residues from the adjacent monomer mates. In this crystal structure, Cys255 adopts two alternative conformations forming an attachment (disulfide bridge) to its equivalent residue Cys255' from the adjacent monomer mate.

The role of the disulfide bridge to stabilize the dimer is not clear yet, but after breaking of the disulfide bridge, the  $k_{cat}$  values of the Cys255Ala variant with both substrates (E2 and 5-diol) are very similar to the ones of the wild type. Furthermore, inhibitor **1** was able to bind the mutated protein during the co-crystallization process and the binding pose it adopts is similar to the pose observed in the wild type complex (**Figure 20A and B**).



**Figure 20.** **A.** Crystal structure of 17β-HSD14 wild type in complex with cofactor and inhibitor **1** (PDB ID: 5ICM) are shown in the dimer conformation. The amino acids coloured in blue are from the first monomer mate (grey), the amino acids coloured in pink are from the second-adjacent monomer mate (light pink). The cofactor and inhibitor **1** are shown in both monomers in orange and yellow sticks, respectively. **B.** Crystal structure of Cys255Ala variant in complex with cofactor and inhibitor **1** (PDB ID: 6GBT). The mutated alanine residue occupies very similar space as the parent cysteine residue; however, no covalent attachment can be formed.

According to the native PAGE experiment (**Figure 6**), the Cys255Ala variant protein buffer hosts a mixture of tetramer and dimer in the gel with the dimer being the largest fraction in the protein buffer, based on the stronger dye label on the dimeric molecular weight, while all the other mutated variants show higher population at the tetrameric signal. Regarding the size exclusion chromatography, the Cys255Ala variant sample was eluted at the molecular weight of a tetramer. Therefore, it is more convincing to proclaim that both species, dimer and tetramer, exist side by side in equilibrium inside the protein sample. To further identify the assembly of the variant, a native mass spectrometry of the wild type was measured under reducing

conditions in presence of DTT (dithiothreitol) at different concentrations ranging from 5 mM to 4 M, the expectation from this experiment is the rupture of the disulfide bridges, especially the one linking Cys255 and Cys255'. The mass spectrometry measurement was performed with the addition of ammonium acetate 50 mM in the gas phase. The final result showed the protein as a tetramer, which suggest that the absence of the disulfide bridge does not influence the tetrameric assembly of the enzyme. The disulfide bridges in the protein might not be ruptured even with different incubation time with the reducing media (ranging from one hour to three days). Apparently, the detectable dimer portion is only observable in native gel for the Cys255Ala variant. These deviating results obtained through the previously mentioned experiments to determine the symmetry of the protein can be most likely explained by the significantly different conditions applied during these experiments. Unfortunately, it is not possible to reach a final argument or an evidence regarding which quaternary structure of the protein variants are populated to which extent under the applied conditions in the experiments performed. However, the assumption reached from the data gathered that the formation or rupture of the disulfide bridge is not essential for the tetramer stability and the enzymatic activity as well. However, with the Cys255 located on the flexible loop near the C-terminal tail where also Tyr253 is located, it could be important to stabilize and fix the loop at its basis because it is seen in the Tyr253Ala variant that the loss of the OH-group of Tyr253 in the protein abolishes enzymatic activity. The superimposition of all variant structures (**Figure 8**) showed a more visible portion (one extra residue) of the C-terminal tail (flexible loop). The electron density in the crystal structure was insufficient to build residue Lys256 which is rather odd, according to the hypothesis, the disulfide bridge should stabilize the flexible loop. The fact that there is an extra portion of that loop visible proves otherwise. Supposedly, the disulfide bridge has no contribution to the overall assembly and activity of 17 $\beta$ -HSD14.

## 2.5. Conclusion

This study investigated the role of several amino acids located at influential positions in 17 $\beta$ -HSD14 wild type, although these amino acids do not have any equivalents in other 17 $\beta$ -HSD family members. Gln148 was chosen because of its location at the entrance of the active site where the substrate binds. It is believed that this residue restricts the entrance of the substrate and other potent compounds to the active site. With the high activity recorded in the kinetics compared to the wild type, expectations were fulfilled regarding this amino acid and its role in the enzymatic activity of the protein and the binding pose of the potent inhibitor used in this

study, which is similar to the pose observed for the wild type structure complex with the same inhibitor. The major observation is the higher catalytic activity recorded with the estrogenic E2 whereas the androgen 5-diol has less activity. The possible explanation might be that the H-bonding between Gln148 and His93 does not exist anymore, thus, allowing a faster exchange rate and less prohibited entry to the binding pocket. In the case of androgen 5-diol, having a less slender geometry than the estrogen E2 restricts the faster exchange rate, resulting in less advantage from the additional space created. There could be another explanation for the higher activity of estrogen (E2) than the androgen (5-diol) which always brings back the question asked by multiple researchers who tried to study and further characterize the 17 $\beta$ -HSDs family: “Are E2 and 5-diol the real substrates of this enzyme?”. This still needs to be answered through more research on the 17 $\beta$ -HSDs family. In addition to Gln148, His93 is located next to it at the inner entrance of the substrate binding site, also restricting the entry to the binding pocket and the H-bond between the two residues is essential in the catalytic activity and access to the binding pocket. Another studied amino acid is Tyr253. This residue plays a key role in both stability and activity of the enzyme. The interaction formed between the OH group of Tyr253 with inhibitors in the active site of the neighbouring monomer mate has an undeniable role in geometrical orientation of the compounds binding inside the active site. On the C-terminal tail (flexible loop), Cys255 is located. The reason this residue is considered in this study is the disulfide bridge it forms to the corresponding Cys255' from the adjacent monomer mate. It is believed to strengthen the dimeric assembly of the protein and to fix the C-terminal tail at bay, thus, stabilizing the rest of the loop and in particular the Tyr253 residue that is located on the same loop and has an interaction with compounds populating the active site of the protein. In the case of Lys158, being part of the three amino acids forming the catalytic triad in 17 $\beta$ -HSD14, mutating this amino acid led to the loss of activity and the ability of potent inhibitor binding in the active site. This makes complete sense, however, studying this variant had undeniably brought extra understanding and valuable data about the protein properties. Summing things up, studying 17 $\beta$ -HSD14 variants provided a more profound understanding of the enzyme and its activity *in vitro*, hopefully, more investigations will be done in the near future to gain more reliable data about the function of the protein *in vivo* and to discover and characterize the preferred substrate of this enzyme.



# 3

## Chapter 3

---

# Serial Synchrotron X-ray Crystallography

New sample holder: The Roadrunner I chip





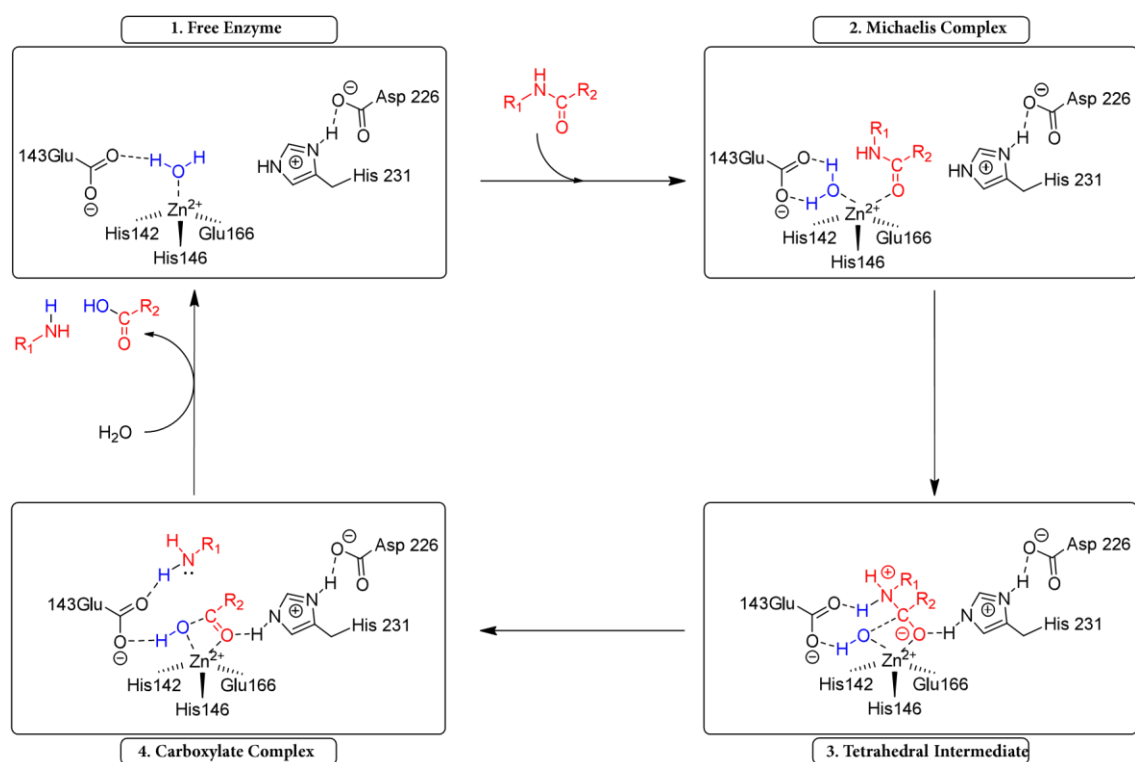
### 3.1. Introduction

For over three decades, X-ray crystallography has been the dominant method to determine protein structures. Due to the important role proteins play *in vivo*, an increasing need emerged to identify their structures and characteristics. This was the driving force to achieve higher resolution structures that allow to better define and, most importantly, to reduce the time to determine such protein structures. Serial Synchrotron X-ray crystallography (SSX) is an emerging technique to determine multiple protein structures in very short time. In addition, serial crystallography has proven to be an ideal method to obtain information about crystals of smaller size (micrometre or nanometre) [171,172]. Data collections at synchrotrons are getting increasingly faster. Most synchrotrons are trying to reduce the time and costs to be spent to measure protein crystals. Even measuring at room temperature instead of cryogenic temperature is attempted [173,174]. Synchrotrons around the world are rapidly adopting the concept of serial X-ray crystallography. One of the bottlenecks in SSX is to find an efficient and reliable system to expose the crystals to the beam which needs less protein material, reduced data collection time, and eases fast handling of the protein crystals [175]. This task virtually started a kind of competition between the various beamline scientists to design the most reliable and efficient delivery system [176–181]. The current study will make use of the mounting and delivery system developed and used at the *Deutsche Elektronen-Synchrotron* (DESY), Hamburg, Germany [176] named “Roadrunner” [48,64,182,183]. It is a micro-patterned sample holder from monocrystalline silicon (waiver technology) with micropores [64]. Like other sample holders designed at the various synchrotrons, the aim is to have a sample holder that can present hundreds to thousands of crystals to the high intensity beam without interfering with the diffraction pattern [64,182,184].

When it comes to X-ray beam exposure, there are two different types of radiation damage, the specific and global one. Specific radiation damage occurs due to the inelastic scattering of the X-ray photons with the sample on the mounting system through either photoelectric absorption or Compton scattering. This leads to the cleavage of chemical bonds or generates free radicals. Global radiation damage (tertiary damage) occurs at random and destabilizes the crystal lattice [185]. Here, the Roadrunner provides an advantage as, due to its properties as a single crystal, it significantly reduces the background scattering. It was tested at a dose up to 565.6 kGy, and a decrease in the diffracting power by half has been observed for a dose of  $D_{1/2} = 147.5 \pm 9.1$  kGy at room temperature. Nevertheless, this decrease is less under cryogenic temperature conditions [182]. The Roadrunner is compatible with room and cryogenic temperatures ( $\approx 100$

Kelvin) during data collection. The crystallization on chip can be carried out by either crystallizing immediately the protein directly on the silicon chip via the vapour diffusion method or already grown crystals are transferred onto the chip. The most popular method is to grow crystals directly on the chip, which saves a lot of time and materials. The idea is to have a system where crystallization, washing, soaking and applying cryo-protectants are all applied without removing the crystal from the delivery system. Furthermore, this will reduce the human errors in crystal handling especially of sensitive, fragile or high water-content crystals. Another reason for using the Roadrunner is the easier exposure of ligands to high-quality microcrystals [186] which is possible with the device because crystal handling by hand or typical laboratory equipment is nearly impossible to apply without damaging the crystals.

Up to now, many target proteins have been used to test the scope and efficacy of this delivery system [48,64,182,183]. In our experiment, the target protein thermolysin (TLN; EC 3.4.24.27) was used, which is a thermostable enzyme produced by gram-positive bacteria. It requires one zinc and four calcium ions for structural stability [187]. The TLN is part of the endopeptidase subfamily that is responsible of breaking the peptide bonds of non-terminal amino acids. TLN mechanism proceeds over two steps as shown in **Figure 1**.



**Figure 1.** Experimentally proposed reaction cycle for TLN (adopted from [188])

Before starting the experiment, a set of compounds must be chosen for the on-chip soaking. The compounds were picked from a 364-fragment library developed in a previous BMBF-project at the Institute of Pharmaceutical Chemistry (*Philipps-Universität Marburg*) and a 96-fragment library developed in collaboration with the *Helmholtz-Zentrum Berlin (HZB)* at BESSY II in the BMBF Frag2Xtal project. A more detailed description of the libraries can be found in **Chapter 4**. For this feasibility study, three fragments were chosen from the 364-fragment library and seven fragments from the 96-fragment library. The experimental part of this study was performed at *Deutsches Elektronen-Synchrotron (DESY)*, Hamburg at PETRA III P11 beamline in collaboration with associated laboratories at the facility.

## 3.2 Materials and methods

### 3.2.1. Preparation of the protein sample

The reason for choosing TLN for this study is its easy access in large quantities, its high tendency to crystallize in a short period of time and the various studies done with this enzyme by many researchers including our working group, even though, TLN is arguably not directly important for pathogenic processes in the human body. Freeze-dried thermolysin was purchased from Calbiochem (EMD Biosciences). TLN was suspended in pure DMSO and mixed with the same volume of crystallization buffer (all buffers used for this experiment are summarized in **Table 1**). The final protein concentration was 4 mM. The sample was centrifuged for 5 minutes at 10000 rpm to precipitate undissolved particles.

### 3.2.2. 3D printed crystal plates

Unlike the usual hanging or sitting drop plates used for other experiments done throughout this thesis, the Roadrunner has special crystal plates that are custom-made by 3D printers to suit the criteria of the method (**Figure 2**). There are various types of such plates (one-well plate, 6-well plate and adapter-like plates) [183], however for the current study, we only used the 6-well plates.



**Figure 2.** 6-well Roadrunner crystallization plate produced by a 3D printer

As seen in **Figure 2**, there is a hole at the outer wall of each well. It is covered with a magnetic seal glued to the rim of the hole to hold the sample holder.

### 3.2.3. Crystallization condition and protein buffer

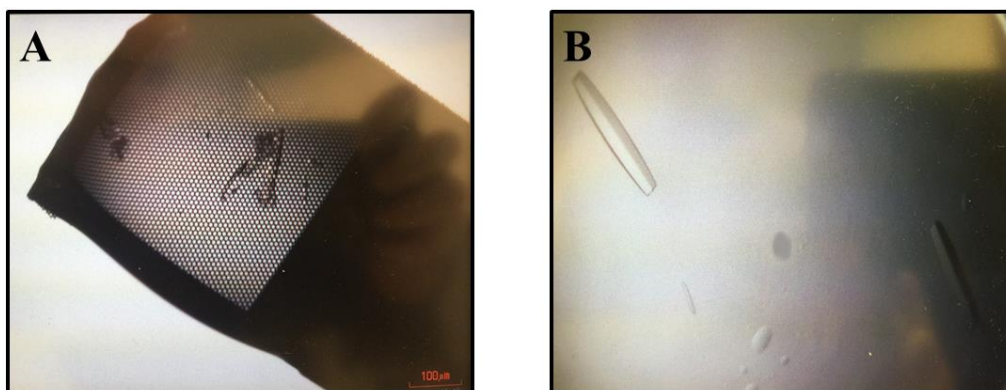
The Roadrunner chip is designed to carry out all steps of crystal handling on the chip from growing crystals until data collection. In the present study, the crystals were either grown on the chip inside the 3D printed crystal plates or additionally on usual 24-well sitting drop crystal plates (Hampton research). Here, the crystals were subsequently transferred manually to the Roadrunner chip for further application. As for the crystallization condition in the well, it consisted of demineralized water only. The crystallization condition is based on an established protocol taken from [189].

**Table 1.** *Sample, washing and soaking buffers used in the experiment*

<b>Sample buffer</b>	100 mM Tris/HCl pH 7.5 3.8 M CsCl	The sample buffer is mixed with equal volume of DMSO containing the protein
<b>Washing buffer</b>	120 mM Tris/HCl pH 7.5 2 mM Ca <sub>2</sub> Cl 20% isopropanol	Incubation duration: 4 hours
<b>Soaking buffer</b>  +  <b>cryo-protectant</b>	120 mM Tris/HCl pH 7.5 2 mM Ca <sub>2</sub> Cl 20% glycerol 30% DMSO	The fragment is dissolved in DMSO. Final fragment concentration in the buffer 100 mM  Incubation duration: 4 hours

#### 3.2.4. Setting up the plates

The smaller Roadrunner I prototype [183] was used for this experiment. All the Roadrunner I versions are manufactured from the same crystalline silicon material, but the difference lies in the chip's surface diameter (1.5 \* 1.5 mm<sup>2</sup> for small Roadrunner I), which will affect the sample drop volume. The magnetic tip holding the chip was gently removed from the vial and placed on a fixed magnetic ward with the conical shape face pointing upwards (the frame of the chip is flat from one side and has a small inner edge from the other). This prevents the sample from pouring out of the edges. 10 µL of the protein sample drop is placed on the chip, which is a sufficient volume to cover the whole surface. Each well of the 3D plate is filled with 1 mL of distilled water, then the chip is slowly and gently capped inside the well. The plates are sealed with transparent tape (SharkTape: 1.88-inch-wide Crystal Clear Tape from Hampton research) suitable for vapour diffusion method. Grown crystals appear after 8-12 hours at 18°C in both 24-well sitting drop plates and on the Roadrunner chip **Figure 3**.



**Figure 3.** Microscopic view of TLN crystals. **A.** Crystals grown on the Roadrunner chip. **B.** Crystals grown on a sitting drop plate.

It can be observed that crystals grown in a 24-well sitting drop plate (**Figure 3B**) have a more regular shape than those grown on the chip (**Figure 3A**). The reason for this are the pores on the chip, where the crystals imitate the surface they are grown on. They are thicker at the position of the pores than on the edges, but regarding the obtained diffraction pattern and resolution, no significant difference could be observed between crystals grown by either technique.

### 3.2.5. Crystal handling on the chip

#### 3.2.5.1. Hydration stream

In any normal crystallization system, the crystals are removed from the sample drop to a soaking buffer or a cryo-protectant buffer. Thus, it is obvious that moving the crystal out of solution will initiate drying out of the crystal due to water evaporation destroying the crystal. In the case of the Roadrunner chip, the sample has to be dried and the buffer needs to be exchanged. To prevent crystal dehydration of the chip, a custom-made hydro-stream nozzle is assembled and the delivery system holding the chip is placed into the stream. The hydro-stream consists of a pipe connected through a hose to multiple water bottles, an air stream is induced from the water bottles toward the pipe creating a moisture atmosphere around the tip of the pipe. Crystals can survive for quite some time (at least two hours were tested) without being soaked in a solution. Their diffraction patterns and resolution matches with that of crystals grown under the same conditions in wells other than the chip.

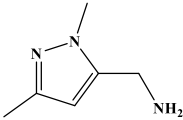
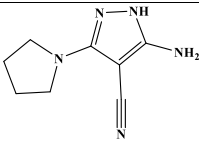
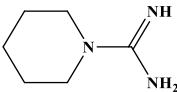
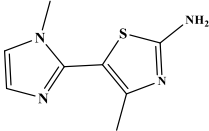
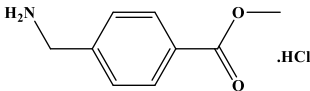
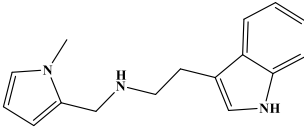
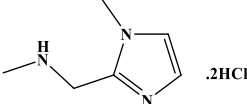
### 3.2.5.2. Incubation with isopropanol

Another challenge in handling TLN crystals is the washing step with isopropanol (IPA). Putative TLN *apo* crystals host the auto-proteolysis product Val-Lys in the catalytic site. It is possible to perform soaking without removing the dipeptide, but there is no guarantee that the fragments will replace the low micromolar Val-Lys. Likely, the fragments are weak binders and the auto-protolysis peptide has, most likely, a higher affinity to remain in the active site. IPA can be used as a small molecule to bind to the TLN active site. Applied in high concentration, it is able to wash-out the Val-Lys peptide from the active site [190]. The chip is attached to a magnetic sample holder placed in the hydro-stream. An absorption tissue is gently placed under the chip for a short period to remove all remaining sample buffer. Otherwise, the danger of crystal aggregation is given and data collection will not be successful due to overlapping crystals contributing to the total diffraction pattern which makes it impossible to process and evaluate the data. After the draining of protein buffer, a volume of 10  $\mu\text{L}$  of washing buffer was pipetted on the chip and left to incubate for 4 hours.

### 3.2.5.3. Soaking on the chip

Within the hydro-stream, the same procedure was done to drain the washing buffer with the removed Val-Lys peptide off using again a tissue paper. Several fragments from the 96-fragment library were selected (**Table 2**) for which a previous fragment screen had shown successful TLN binding (Francesca Magari, doctoral thesis [191]).

**Table 2.** Selected fragments from the 96-fragment library for the Roadrunner experiment

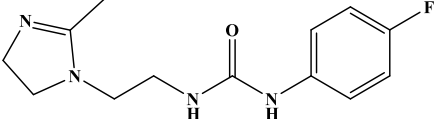
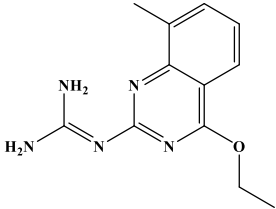
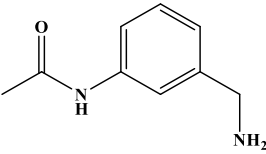
Fragment Nr.	Library ID	Structure	MW (g/mol)	Hit rate <sup>a</sup> (TLN fragment screen)
1	J13		125.1	hit PDB ID: 6SBK
2	J20		177.2	partial hit no PDB ID
3	J21		127.1	no hit
4	J26		194.1	no hit
5	J32		201.7	partial hit no PDB ID
6	J35		253.2	partial hit no PDB ID
7	J37		125.2	No hit

<sup>a</sup>The ratio of compounds that bind in the active site of TLN, hit = a fragment is found in the active site, partial hit = a portion of the fragment found in the active site (according to the electron density) and no hit = no sign of the fragment inside the active site.

The chosen fragments from the 96-fragment library were hand-picked because of their structural similarity with other fragments from the same library that were successful hits: J13, J22, J28, J62, J77, J88 and J96 (**Appendix**). The second set of fragments selected from the 364-fragment library are listed in **Table 3**.



**Table 3.** Selected fragments from the 364-fragment library for the Roadrunner experiment

Fragment Nr.	Library ID	Structure	MW (g/mol)	Hit rate (TLN fragment screen)
1	F196		264.1	n.d
2	F218		254.1	n.d
3	F285		164.1	n.d

n.d = not determined

The soaking conditions described by Magari were used in the fragment screen contain 20% MPD as a cryo-protectant [191]. Due to the appearance of ice rings, it was replaced by 20% glycerol instead, which resolved the problem. In the protocol applied by our working group, glycerol is not used as a cryo-protectant because of the appearance of ice rings during data collection. However, the ice rings did not appear during data collection while using the Roadrunner chip with glycerol as a cryo-protectant.

### 3.2.6. Data collection and processing

Data sets were collected at PETRA III beamline P11 (DESY – Hamburg) [192], at a wavelength of 1.0332 Å with a Pilatus 6M detector. The experiment was done at 100 K and the chip holding the crystals was stored in liquid nitrogen. 600 diffraction images were collected for each data sets. The chip had 30° rotation during data collection. According to *Roedig et al.* [64], the chip allows for a larger degree of rotation. Nonetheless, in the present case of TLN the large amount of TLN crystals located on most of the chips, as well as their tight mutual

placement, made interference of diffraction patterns of multiple crystals likely during data collection. As TLN crystals possess hexagonal symmetry (space group  $P6_122$ , no.177),  $30^\circ$  was in most cases sufficient to obtain data sets with acceptable completeness ( $>90\%$ ) [193]. Collected data were indexed, processed and scaled using XDSAPP [158]. Some data sets with low completeness required merging with other data sets collected from the same chip to achieve acceptable completeness using BLEND [194] from the CCP4 suite [195].

### 3.2.7. Crystal structure determination and refinement

Crystal structures were determined by molecular replacement using the program PHASER MR [159] from the CCP4 suite [195]. For data sets that did not require merging, an auto-refine pipeline [196] was used. The structure was solved using molecular replacement starting from structure PDB ID: 6SBK [191] after removing all non-protein molecules and hydrogen atoms using FCONV [161]. In the refinement, a 5% subset of the reflections was used to calculate  $R_{\text{free}}$  and they were consequently omitted from the refinement. The model was built in COOT [162] and the refinement was done using PHENIX.refine version 1.15.2-3472 [163]. First refinement steps were performed with default parameters for XYZ coordinates, occupancies and individual B-factors were altered with structural adaption in COOT. B-factors for all model atoms (except for hydrogen atoms) were refined anisotropically for resolutions better than  $1.4 \text{ \AA}$  and isotropically for resolutions worse than  $1.41 \text{ \AA}$ .

## 3.3. Results

### 3.3.1. Soaking outcome

The soaking period and fragment concentration in the soaking buffer (100 mM) is based on an established protocol used by our working group [189]. All soaking experiments done on the Roadrunner chip were successful and the crystals survived, except the crystals soaked with fragments F218 and F285, where unfortunately, the crystals dissolved in the soaking buffer after a few seconds.

### 3.3.2. Obtained data sets

About 180 data sets were collected from 20 Roadrunner chips used in the experiment. Details about the data collection are listed in **Table 4** below.

**Table 4.** List of data set collected using the Roadrunner chip.

Chip ID + Fragment ID	Data set Nr.	Resoluti on Å	Completeness	Hit rate <sup>a</sup>	comment
Chip 1 J13	001 – 014	-	-	-	Ice ring Cryo-protectant contains MPD
Chip 2 J13	001	2.4 Å	99.5%	Partial hit	Parts of the fragments appear in the electron density, but could not be added
	002	2.1 Å	99.5%	Partial hit	
	003	2.4 Å	97.2%	Partial hit	
	004	2.0 Å	95.9%	No hit	-
	005	> 4.0 Å	95.9%	-	-
Chip 3 J20	001 _ 014	-	-	-	Ice ring Cryo-protectant contains MPD
Chip 4 J20	001	1.90 Å	99.7%	No hit	-
	002	1.81 Å	92.1%	No hit	-
	003	1.95 Å	99.5%	No hit	-
	004	1.90 Å	99.3%	No hit	-
	005	1.86 Å	82 %	No hit	Data sets were merged Completeness 91%
	006	1.92 Å	77 %	No hit	
	007	1.90 Å	99.2%	No hit	-
	008	-	-	-	Chip displacement <sup>b</sup>
	009	1.80 Å	99.6%	No hit	-

	010	1.60 Å	92.1%	No hit	-
	011	-	-	-	Chip displacement <sup>b</sup>
	012	-	-	-	Chip displacement <sup>b</sup>
<b>Chip 5 J21</b>	001	2.6 Å	97.8%	No hit	-
	002	-	-	-	Crystals overlap <sup>c</sup>
	003	-	-	-	Crystals overlap <sup>c</sup>
	004	1.8 Å	99.7%	No hit	-
	005	> 3.0 Å	94.2%	-	-
	006	> 3.0 Å	96.1%	-	-
	007	1.95 Å	99.7%	No hit	-
	008	> 3.5 Å	92.3%	-	-
	009	-	-	-	Chip displacement <sup>b</sup>
	010	2.55 Å	99.7%	No hit	-
	011	2.0 Å	99.3%	No hit	Chip displacement <sup>b</sup>
	012	-	-	-	
	013	2.27 Å	75%	No hit	Data sets were merged
	014	1.91 Å	84.1%	No hit	Completeness 94.2%
<b>Chip6 J26</b>	001 _ 009	-	-	-	Ice ring Cryo-protectant contains MPD
<b>Chip 7 J26</b>	001	> 3.5 Å	92.1%	-	-
	002	1.82 Å	98.7%	No hit	-
	003	1.95 Å	95.9%	No hit	-
	004	1.91 Å	96.5%	No hit	-
	005	2.22 Å	41.2%	No hit	Data sets were merged
	006	2.71 Å	29.7%		Completeness 74.0%
	007	-	-	-	Chip displacement <sup>b</sup>
	008	-	-	-	Chip displacement <sup>b</sup>
	009	2.04 Å	99.1%	No hit	-
	010	2.02 Å	99.4%	No hit	-

	011	$> 4.0 \text{ \AA}$	-	-	-
<b>Chip 8 J26</b>	001	$> 4.0 \text{ \AA}$	-	-	-
	002	$> 3.0 \text{ \AA}$	21%	-	Chip displacement <sup>b</sup>
	003	-	-	-	Chip displacement <sup>b</sup>
	004	$1.80 \text{ \AA}$	99.8%	No hit	-
	005	$1.73 \text{ \AA}$	99.7%	No hit	-
	006	$1.82 \text{ \AA}$	98.9%	No hit	
	007	$1.87 \text{ \AA}$	99.6%	No hit	-
	008	-	-	-	Crystals overlap <sup>c</sup>
	009	-	-	-	Chip displacement <sup>b</sup>
	010	$> 3.0 \text{ \AA}$	-	-	-
	011	$1.90 \text{ \AA}$	99.6%	No hit	-
	012	-	-	-	Chip displacement <sup>b</sup>
<b>Chip 9 J32</b>	001 _ 007	-	-	-	Ice ring Cryo-protectant contains MPD
<b>Chip 10 J32</b>	001	-	-	-	Chip displacement <sup>b</sup>
	002	-	-	-	Chip displacement <sup>b</sup>
	003	$2.01 \text{ \AA}$	95.5%	No hit	-
	004	$2.62 \text{ \AA}$	99.7%	No hit	-
	005	-	-	-	Crystals overlap <sup>c</sup>
	006	-	-	-	Crystals overlap <sup>c</sup>
	007	-	-	-	Chip displacement <sup>b</sup>
	008	-	-	-	Chip displacement <sup>b</sup>
	009	-	-	-	Crystals overlap <sup>c</sup>
	010	$1.90 \text{ \AA}$	99.5%	No hit	-
	011	-	-	-	Chip displacement <sup>b</sup>
	012	-	-	-	Chip displacement <sup>b</sup>
	013	-	-	-	Crystals overlap <sup>c</sup>
	014	-	-	-	Crystals overlap <sup>c</sup>

	015	-	-	-	Crystals overlap <sup>c</sup>
	016	-	-	-	Chip displacement <sup>b</sup> + Crystals overlap <sup>c</sup>
	017	-	-	-	Chip displacement <sup>b</sup> + Crystals overlap <sup>c</sup>
	018	-	-	-	Chip displacement
<b>Chip 11 J35</b>	001 _ 005	-	-	-	Ice ring Cryo-protectant contains MPD
<b>Chip 12 J35</b>	001	-	-	-	Chip displacement <sup>b</sup>
	002	-	-	-	Chip displacement <sup>b</sup>
	003	2.03 Å	99.6%	No hit	
	004	-	-	-	Crystals overlap <sup>c</sup>
	005	-	-	-	Crystals overlap <sup>c</sup>
	006	-	-	-	Chip displacement <sup>b</sup>
	007	1.80 Å	99.7%	No hit	
	008	-	-	-	Crystals overlap <sup>c</sup>
	009	-	-	-	Crystals overlap <sup>c</sup>
	010	-	-	-	Chip displacement <sup>b</sup> + Crystals overlap <sup>c</sup>
	011	-	-	-	Chip displacement <sup>b</sup> + Crystals overlap <sup>c</sup>
<b>Chip 13 J37</b>	001 _ 007	-	-	-	Ice ring Cryo-protectant contains MPD
<b>Chip 14 J37</b>	001 _ 009	-	-	-	Chip displacement <sup>b</sup> + Crystals overlap <sup>c</sup>
<b>Chip 15 F196</b>	001 _ 008	-	-	-	Ice ring Cryo-protectant contains MPD

<b>Chip 16 F196</b>	001	1.50 Å	99.3%	-	The Val-Lys peptide is not washed from the active site
	002	1.40 Å	99.7%	-	The Val-Lys peptide is not washed from the active site
	003	> 5.0 Å	-	-	Cannot be processed
	004	1.63 Å	99.6%	No hit	
	005	-	-	-	Crystals overlap <sup>c</sup>
	006	1.88 Å	99.6%	No hit	-
	007	-	-	-	Chip displacement <sup>b</sup>
	008	-	-	-	Chip displacement <sup>b</sup>
	009	-	-	-	Chip displacement <sup>b</sup>
	010	> 5.0 Å	-	-	-
	011	-	-	-	Chip displacement <sup>b</sup>
	012	-	-	-	Chip displacement <sup>b</sup>

<sup>a</sup>Fragments that binds in the active site of TLN or at any other position. <sup>b</sup>Due to the pressure by the cryogenic stream (LN<sub>2</sub>), the chip is moving back and forward (waving), thus the crystal under measurement is displaced from the centre of the X-ray beam losing diffraction or obtaining unstable diffraction images that could not be measured. <sup>c</sup>During data collection, the crystals may overlap on top of each other or by the rotation of the chip during the process leading to extra random diffraction from two or more crystals.

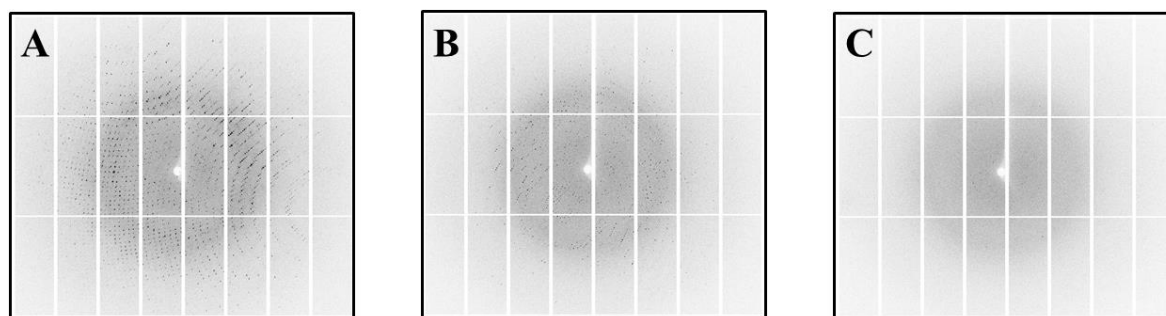
## 3.4. Discussion

### 3.4.1. Overall Roadrunner usage

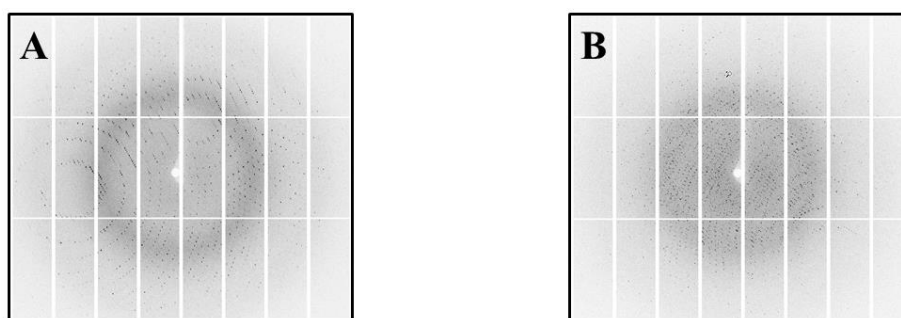
Using the Roadrunner chip provided many advantages regarding data collection and the reduction in time required to perform a crystal screening. The easy handling of micro-size crystals was possible because there was no need to handle them manually. Growing crystals on the chip was as fast similar to a normal 24-well crystallization plate but the porous surface of the chip altered the shape of the crystal and reduced their size as well. In addition, there were many micro-crystals formed on the chip during crystallization. Exchanging buffer on the chip is somehow tricky, particularly when draining off the buffer in order to exchange it with another one. Experience and manual training are essential. If the tissue paper is not in touch with the whole area under the chip there could remain some problems like the residual movement of small crystals that are not attached to the chip surface next to the point where the tissue paper

is absorbing the solution from the chip. This leads to the assembly of many crystals at a single spot piling up on top of each other. This is a major reason why many data sets have to be suspended because overlapping diffraction patterns of more than one crystal during data collection. Another problem occurred during the exchange of the buffers at the inner edge on the upper surface of the chip, even after drying out with a tissue paper. A small volume of the drained buffer remains between the inner surface and the surrounding rims. This has bad impact on the freezing process in liquid nitrogen. It was noticed during data collection under cryogenic temperature, where all the crystals that were close to the gap near the inner frame, showed ice rings. It is due to the fact that they were no longer surrounded by the cryo-protectant but instead by the remaining drained buffer. Importantly, the crystals far away from that rim (closer to the middle of the chip) showed no ice rings during data collection. Another obstacle was faced while removing the Val-Lys peptide from the catalytic centre. After checking multiple data sets that were collected on the chip, there were several ones showing some remainder of the Val-Lys peptide, still occupying the catalytic site. This was never the case applying the original protocol developed by Magari [191] after 4 hours of exposing the protein crystals to the washing buffer. The incomplete removal of Val-Lys was more frequently observed for crystals of bigger size. So it could well be that a longer exposure time to isopropanol is required for larger sized crystals. The complete removal of the dipeptide for micro-size crystals supports this hypothesis. Affinity of **J13** for TLN from the 96-fragment library was not determined, e.g. by enzyme kinetics. Fragment **J13** however, inhibits the active site of TLN and a complex with this fragment could be obtained by Magari (PDB ID: 6SBK) [191]. Unfortunately, soaking the same fragment at a concentration of 100 mM over a period of 4 hours on the Roadrunner remained unsuccessful with no residual electron density indicating the occupancy of **J13** in the active site or at any other position in the structure of TLN.





**Figure 4.** Diffraction images from data collection performed on Chip 8 – J26 (data set 003) showing the effect that occurred by the fluttering movement of the chip in the cryo-gas stream during data collection. **A.** Diffraction image #100, the diffraction pattern is well resolved. **B.** Diffraction image #250, the diffraction pattern is attenuated and less spots are visible leading to a decrease in resolution. **C.** Diffraction image #500, there are no diffraction spots visible as the measured crystal has been completely moved out of the focus of the X-ray beam.

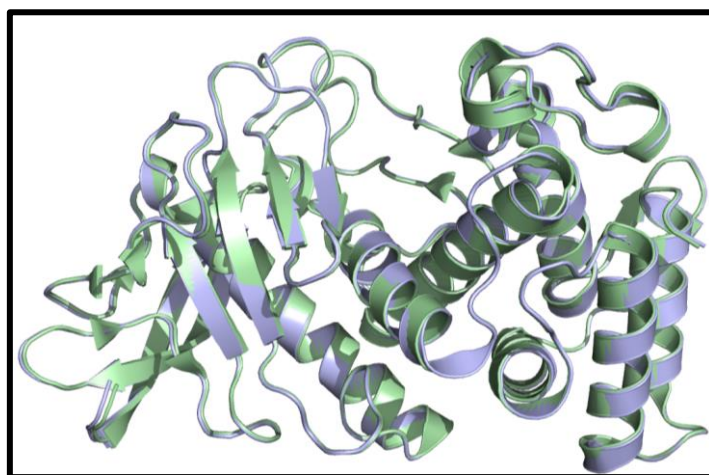


**Figure 5.** Diffraction images from data collection performed on Chip 10 – J32 (data set 005) showing the effect of crystals overlapping during data collection. **A.** Diffraction image #100 after 5° of chip rotation, the diffraction pattern is clear and single spots are well defined. **B.** Diffraction image #500 after 25° of chip rotation. There are multiple diffraction spots appearing on the detector resulting from two crystals showing overlapping diffraction patterns.

### 3.4.2. Comparison of TLN structures

Initially, we assumed that data collection using the Roadrunner instead of “classically” grown and mounted crystals would produce data sets of similar quality. TLN is a well-studied target and much information is already available about this target protein. It was therefore selected for this study as the crystal growth and handling is straightforward compared to other novel cutting-edge targets. As expected, after evaluating the diffraction data from the data collection

on the chip, no difference could be observed, although the crystals on the chip had relatively odd shape compared to the usually grown TLN crystals on a 24-well plate. They exhibit the same properties such as the space group with similar cell dimension and require similar strategies for data collection and evaluation. In **Figure 6**, a direct comparison of the crystal structure of TLN in complex with fragment **J13** from the 96-fragment library (PDB ID: 6SBK) obtained from a “classically” grown and mounted crystal is shown along with a structure determination obtained by data from the Roadrunner chip. The superimposition shows identical crystal structures.



**Figure 6.** Superimposition of TLN structure in complex with J13 (PDB ID: 6SBK) in light green and TLN structure from crystal data collected with the Roadrunner in light blue.

### 3.5. Conclusion

Indeed, there were many obstacles to overcome using the Roadrunner chip. The outcome of the experiments was not as expected. However, it is impressive to run through many crystal handling steps without the need to remove the crystals from the wells in the crystallization devices or to place them on special crystal holders. Data collection was much faster and easier to perform. The fluttering movement of the chip during data collection caused the loss of many data sets. As the chip is still a prototype, additional technical improvements are required. The silicon waiver chip offers the advantages of high transparency. A more stiff and firm material that can hold the chip at its basis will likely improve the present prototypes [183]. Another major issue is the collapsing of crystals on the X-ray beam during data collection due to agglomeration of the crystals on top of each other. An improved sample holder has to resolve these shortcomings. Nevertheless, the crystallization protocol has succeeded in producing crystals suitable for measurement. Experience shows that crystals grown on the chip, even if they remain less-well shaped or small can be handled, as no manual interference is required. SSX has a long way to go, yet, the first steps are made and will offer an exciting perspective toward faster data collection and high-throughput crystallographic fragment screening.



# 4

## Chapter 4

---

### **Solubility Study**

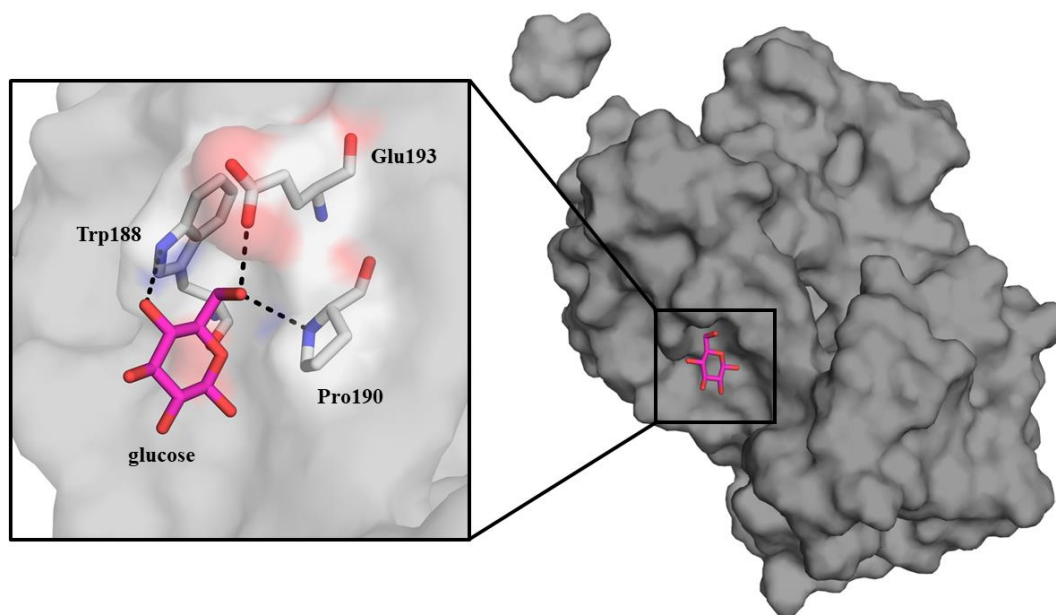
Solubility enhancement of 17 $\beta$ -HSD14 potential inhibitors using cyclodextrins



## 4.1. Introduction

A drug designer's main objective is to find a compound that binds to a given receptor in order to achieve a therapeutic effect and to pass the clinical trial and produce an end product usable as a medication. However, to reach this aim, this compound must have at least a sufficient affinity to its receptor. There are multiple factors that determine the affinity of a compound. Solubility of a compound is considered one of the most essential factors regarding affinity [197–200]. Recently, many studies focused on cyclodextrins (CDs) as a molecular solubility enhancer [79,80,84,85]. Due of their cone-shaped molecular structure (bottomless cylinder) with a hydrophobic interior and hydrophilic exterior, CDs can hide large parts of the hydrophobic functionalities of a probe compound, leaving the hydrophilic moieties pointing to the outside environment; thus, the overall solubility is increased.

The protein selected for testing ligand binding using this solubilization method is 17 $\beta$ -hydroxysteroid dehydrogenase type 14 (17 $\beta$ -HSD14). The human enzyme 17 $\beta$ -HSD14 oxidizes the hydroxyl group at position 17 of estradiol and 5-androstenediol using NAD<sup>+</sup> as cofactor (s. **Chapter 2**). However, the physiological role of this enzyme remains unclear. Through many recent studies performed by our working group [30,31,142], 17 $\beta$ -HSD14 has a good correlation with sugars. In the purification protocol, glucose is used as an additive to stabilize the protein and a glucose molecule is known to bind to the surface of the enzyme in the crystal structure. In a previous protein purification protocol used in our working group, glycerol has been used in the purification buffer instead of glucose. The protein yield tripled. However, it must be emphasized that glycerol in the presence of NAD<sup>+</sup> and the enzyme, without substrate, induces the production of a fluorescent substance which, after investigation, turned out to have the same fluorescence fingerprint as NADH. We concluded that glycerol is recognized as a substrate by h17 $\beta$ -HSD14, thereby transforming NAD<sup>+</sup> into NADH [142]. This issue has been resolved by replacing glycerol with glucose in the protein buffer. Sugars seem to have an impact on the activity and stability of 17 $\beta$ -HSD14, therefore, it has been a good subject for the solubility study with CDs.



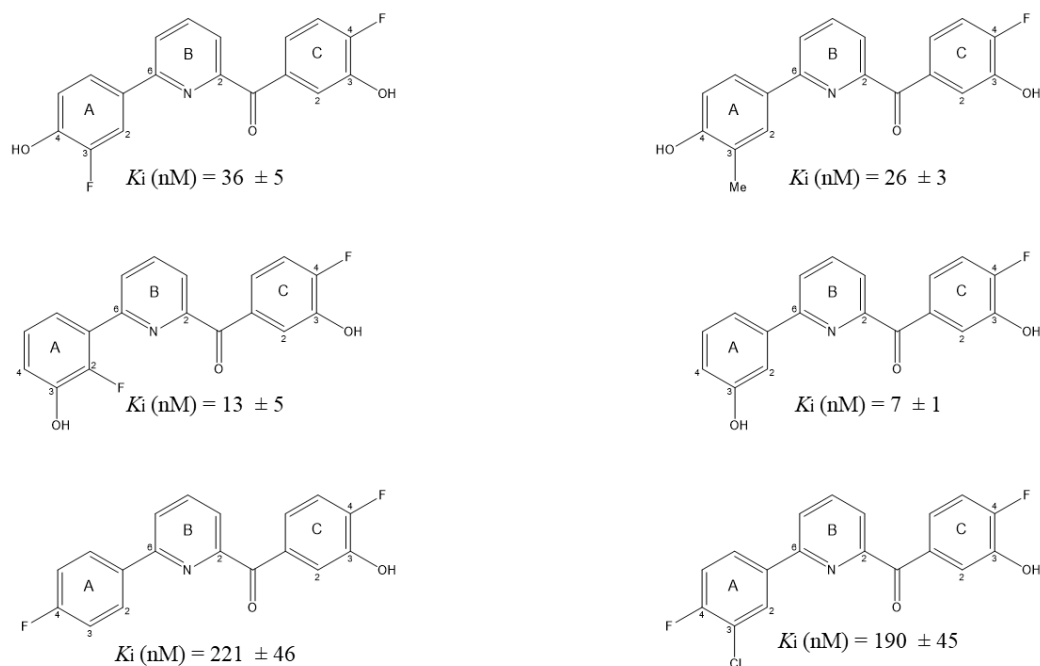
**Figure 1.** A surface representation of 17 $\beta$ -HSD14 variant (PDB ID: 6G4L) and closer view on the glucose (pink) binding location on the surface of the enzyme.

## 4.2. Materials and methods

### 4.2.1. Fluorine-compound library

According to previous studies by Braun and Bertoletti in our working group [30,31,142], aromatic compounds sharing OH-group and F-group proved to exhibit high affinity towards 17 $\beta$ -HSD14.





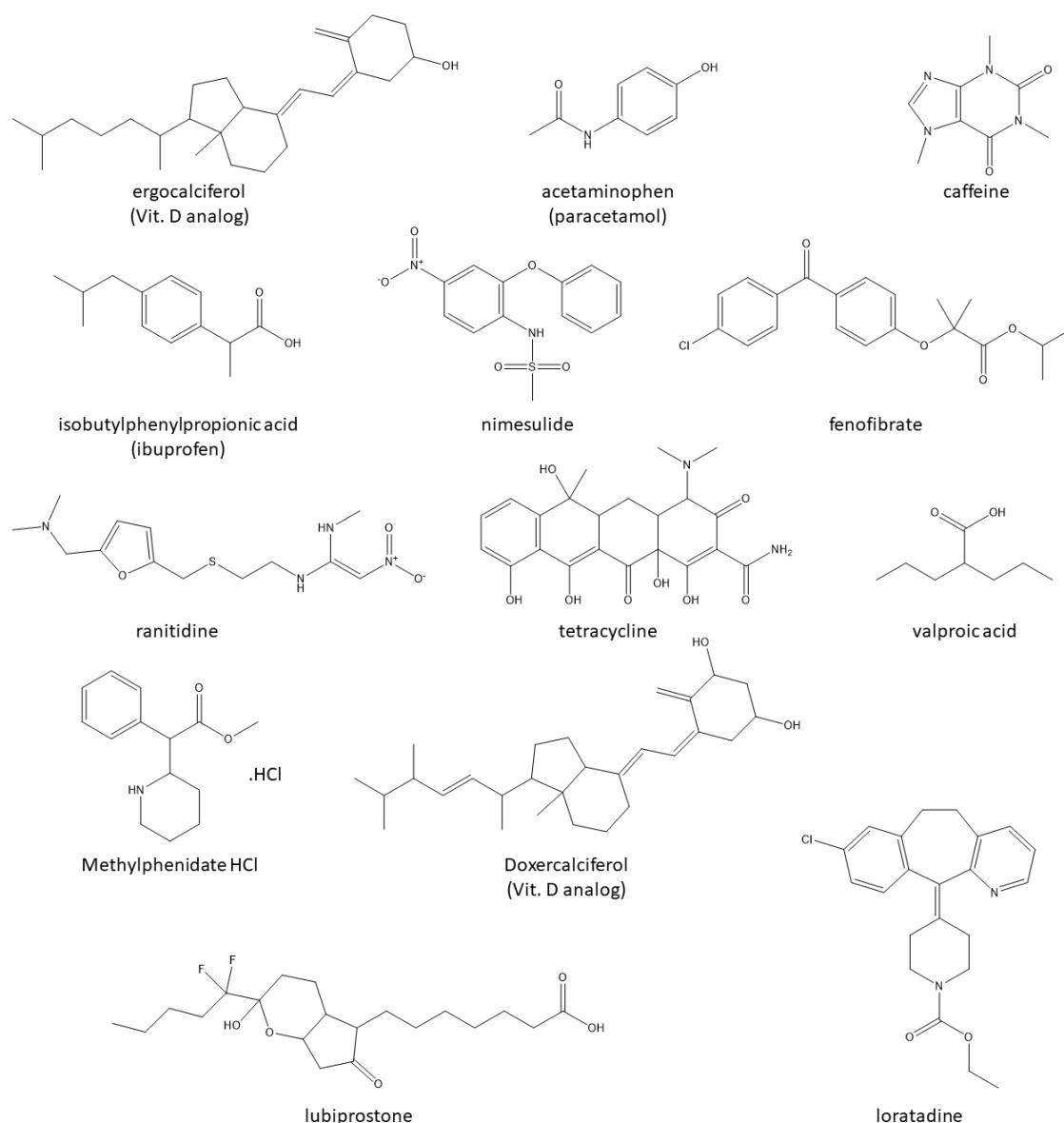
**Figure 2.** Inhibitors of 17 $\beta$ -HSD14 and inhibition constant ( $K_i$ ) with different substitution on the A-ring. The  $K_i$  values were obtained by a fluorometric assay on the 17 $\beta$ -HSD14 enzyme, substrate E2 (1.2 mM) and cofactor  $NAD^+$  (1.2 mM) at 25°C. The resulting  $K_i$  values are calculated by averaging three consecutive assay experiments done with each inhibitor under the same conditions [30].

Based on the high selectivity of 17 $\beta$ -HSD14 enzyme towards ligands comprising an OH group and F atom at an aromatic moiety, multiple compounds were purchased from different vendors or synthesized in our lab that contain these two functionalities. These gathered compounds are referred to as the fluorine-compound library, the compounds are listed in the **Appendix 4.5**.

#### 4.2.2. Hydrophobic drugs

To test the enhanced solubilization provided by the addition of CDs, many hydrophobic drugs were included in the current solubility study. The hydrophobic drugs were all selected based on the length and size of their hydrophobic part. The molecular shape of CDs containing a bottomless cone with hydrophobic functionality in the inner surface made it more convenient to test hydrophobic compounds with a long hydrophobic part while the opposing end is decorated by one or more hydrophilic functionalities. Particularly, as the interior of the CD

cone is hydrophobic, and the outer rim holds many hydrophilic functionalities, especially OH-groups, the study of amphiphilic drug-like molecules makes it notably interesting to investigate such compounds. Unlike the compounds from the fluorine-compound library, the hydrophobic part of these hydrophobic drugs does not contain any functionality other than carbon atoms that should form interactions with the interior hydrophobic surface of CDs. In theory, locking their hydrophobic part inside the hydrophobic cavity of a CD and leaving the hydrophilic moieties pointing to the surrounding aqueous environment would result in an increased solubility through drug/CD complex formation (as inclusion or host/guest complexes). Even more, the sole exposure of the hydrophilic part of the drug molecules, will provide a higher chance to achieve binding to the active site of 17 $\beta$ -HSD14. The studied hydrophobic drugs are listed in **Figure 3**.



**Figure 3.** The hydrophobic drugs used in the co-crystallization experiment with 17 $\beta$ -HSD14.

Although 17 $\beta$ -HSD14 is located in multiple tissues in the human body, the main function *in vivo* is yet to be identified. Likely, the natural substrates of the enzyme are steroid-like compounds (s. below), thus, also ligands of rather hydrophobic nature. Nevertheless, from a pragmatic point of view it appeared more convenient to select freely available highly hydrophobic drug molecules on the market that affect multiple receptors throughout the human body. As shown above in **Figure 3**, a set of well-known and widely used hydrophobic drugs has been assembled. Most of these drugs are slightly to sparingly soluble in water or even at room temperature virtually insoluble. Thus, the aim of this study was to enhance their water solubility by transferring them to the enzyme via a drug/CD complex operating as a kind of “solubilizing ferry”.

**Table 1.** *The hydrophobic product from pharmaceutical point of view*

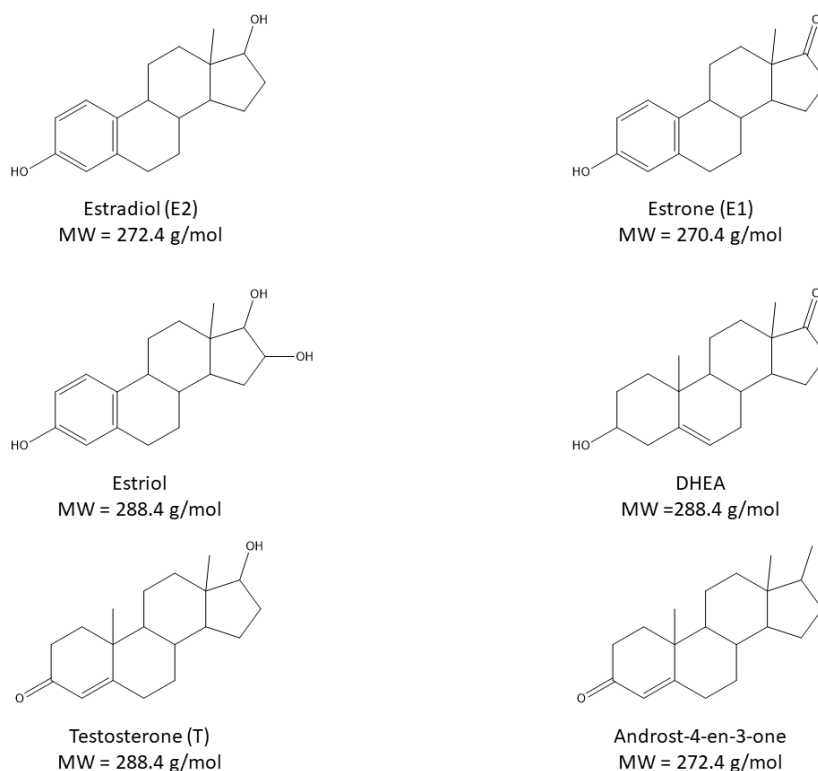
<b>Drug</b>	<b>Trade name<sup>a</sup></b>	<b>Indication<sup>b</sup></b>	<b>Dosage form<sup>c</sup></b>	<b>MW<sup>d</sup></b>	<b>Solubility<sup>e</sup></b>
Ergocalciferol	Drisdol Calcitol	Hypoparathyroidism Osteoporosis	Oral	396.65	≈ 50
Acetaminophen	Tylenol	Headache	Oral	151.16	≈ 15
	Temptra	Muscle pain	Parenteral		
		Arthritis	Rectal		
		Fever			
Caffeine	Vivarin	CNS stimulus	Oral	194.19	≈ 16
Ibuprofen	IBU	Headache	Oral	206.29	≈ 0.02
	Advil	Pain			
	Mortin	Aseptic Necrosis			
		Chronic myofascial			
Nimesulide	Nimesulide	Pain	Oral	36.83	≈ 0.01
	Mesulid	Fever			
		Osteoarthritis			
Fenofibrate	TriCor	Hypercholesterolemia	Oral	360.83	≈ 30
	Lofibra				
	Lipofen				
Ranitidine	Zantac	Gastric ulcer	Oral	314.4	≈ 500
		Gastroesophageal-reflux	IV		
			IM		
Tetracycline	Sumycin	Acne	Oral	444.43	≈ 0.2
	Actisite	Bacterial infections	Parenteral		
Valproic acid	Depakene	Mania	Oral	144.21	≈ 1
	Stavor	Epilepsy			
		Migraine prophylaxis			
Methylphenidate hydrochloride	Ritalin	ADHD	Oral	269.77	≈ 18.6
Doxercalciferol	Hectorol	Pre-dialysis	Oral	412.6	≈ 0.001
		Renal impairment	IV		
		Hepatic impairment			
Lubiprostone	Amitiza	Idiopathic constipation	Oral	390.46	≈ 0.5
		Irritable bowel-syndrome			
Loratadine	Claritin	Allergies	Oral	382.88	≈ 0.007

<sup>a</sup>the trade name of the drug manufactured by a pharmaceutical company inside EU, NA and Japan<sup>b</sup>as stated in USP<sup>c</sup>as stated in the EU, US and pharmaceutical manufacturers<sup>d</sup>molecular weight (g/mol)<sup>e</sup>solubility in distilled water (g/L) at 25°C

### 4.2.3. Sex hormones

As mentioned earlier, it is highly debated whether 17 $\beta$ HSD14's main substrates are estrogens (E2) and androgens (5-diol and testosterone) [20]. Sex hormones are well soluble in organic solvents such as DMSO, but in aqueous solution, they are only poorly soluble. For example, E2 solubility is 20 mg/mL in DMSO at 25°C, while in water, the solubility is around 3.6 mg/L at 25°C. Unfortunately, many trials have been done to obtain a crystal structure with one of the three sex hormones in the active site of 17 $\beta$ -HSD14, using co-crystallization as a method of choice. However, no detectable electron density disclosing the binding mode of such a ligand could be seen. In some cases, uninterpretable electron density was found in the active site. However, through the solubilization mechanism of host-guest complex formation with CDs, the sex hormones were introduced to the protein solution as CD/hormone complex using a lower amount of DMSO, which is described in detail in the crystallization protocol.

The sex hormones selected for the current solubility study (**Figure 4**) are the same used in previous experiments (without CDs) and other sex hormones that have similarity to E2, 5-diol and testosterone.



**Figure 4.** Sex hormones selected for the solubility study.

#### 4.2.4. Cyclodextrin derivatives

The three main cyclodextrins, considered in this study, were  $\alpha$ -,  $\beta$ - and  $\gamma$ -cyclodextrin, which consist of 6, 7, and 8 D-glucopyranosyl residues, respectively. The derivatives of CDs are developed by adding more hydrophilic functionalities to the original three CDs to increase their water solubility. Many CD derivatives are considered safe for human consumption by the FDA and EMA. In this study, only the human consumable derivatives that are applicable in daily life have been used. The only exception is  $\beta$ -cyclodextrin, which, due to its low solubility in aqueous solutions (5 mg/mL), was not applied. A list of the CDs and some other derivatives that have been used to obtain crystal structures with 17 $\beta$ -HSD14 are listed in **Table 2**.

**Table 2.** List of Cyclodextrins used in the solubility study and their solubility in water

Cyclodextrin's derivative	MW (Dalton)	Solubility in water (g/L) at 25°C
$\alpha$ -cyclodextrin	972	45
methyl- $\beta$ -cyclodextrin	1312	350
2-hydroxypropyl- $\beta$ -cyclodextrin	1400	300
carboxymethyl- $\beta$ -cyclodextrin	1541	50
$\gamma$ -cyclodextrin	1312	350
2-hydroxypropyl- $\gamma$ -cyclodextrin	1576	400

Depending on the solubility rate of cyclodextrins shown in **Table 2** above, all CDs were dissolved in water at room temperature to accomplish a concentration of 200 mM ( $\approx$  260 mg/mL on average for all CDs). However, in the case of  $\alpha$ -cyclodextrin and carboxymethyl- $\beta$ -cyclodextrin, both could only be resolved by heating the mixture up to 60°C in a water bath for 24 h to achieve the required concentration.

The reason for preparing a CD solution with a concentration of 200 mM, was that the CD solution mixture was subsequently mixed with the same volume of the probe compound dissolved or in some cases suspended in DMSO with the same concentration (200 mM) to reach a final concentration of 100 mM in the CD/compound mixture. After incubation, a

volume of 5% v/v of the CD/compound mixture was added to the crystallization drop; thus, the required 5 mM concentration of the compound is present in the drop.

#### 4.2.5. Expression and purification of 17 $\beta$ -HSD14

For this study the wild-type variant S205 17 $\beta$ -HSD14 was overexpressed using a pET-based vector, p11-Toronto1 (SGC), containing the coding sequences of the human gene HSD17B14 variant S205 with a *N*-terminal 6His-tag and a TEV (tobacco etch virus) protease cleavage site in *E. coli* BL21 (DE3) pLysS cells. Purification followed an established procedure [142]. The purity of the protein was verified by SDS-Page (SDS-gel not shown) after size exclusion chromatography (SEC200 HiLoad 26/200 Superdex, eluent buffer containing 25 mM Tris HCl, 500 mM NaCl, 250 mM Glucose and 0.5 mM TCEP at a pH = 8, volume of 200 mL at flow rate 2 mL/min). The protein concentration was quantified routinely using the UV-vis spectrometer Nanodrop™ at 280 nm and taking the buffer as blank. This method was previously validated by comparison with the Bradford assay, which afforded a similar protein concentration as observed spectrophotometrically at 280 nm. The yield of the purified protein varied between 10-14 mg of protein per liter of bacterial culture.

#### 4.2.6. Crystallization protocol

The 17 $\beta$ -HSD14 complex crystal structures were obtained by co-crystallization. A solution of the ligand dissolved in DMSO (or suspended if the ligand was too hydrophobic) and cyclodextrin (CD) dissolved in distilled water, each with a concentration of 200 mM, were used (s. above). Equal volumes from each stock solution were mixed together making the final concentration of each (ligand and CD) 100 mM in the mixture.

To ensure that all particles of the ligand were completely dissolved, the mixture was heated at 60°C in a water bath for 6 hours and gently shaken from time to time during the process. The mixture was kept at room temperature for 24 hours. Samples that have been incubated for longer time at room temperature showed results that were more promising (2-3 weeks). Due to the pronounced hydrophobicity of some ligands, there could be some parts of the material that remained not fully dissolved in the mixture. In such cases, more heating or extended incubation time were applied. If even then, some ligand material remained incompletely dissolved,

centrifugation for 15 min at 10000 rpm was performed and the solid material was discarded (any undissolved material present in the crystallization drop will hamper the nucleation process to form crystals).

A volume ratio of 10 % of the ligand/CD mixture was added to the protein solution (10 mg/mL) making the final concentration of each ligand and CD in the protein solution to 5 mM. An incubation time of 30 min at room temperature was necessary before centrifugation for 10 min at 10000 rpm and the supernatant was used.

The concentration of the crystallization buffers in the crystallization well (reservoir) were used as follows: 100 mM HEPES, pH 7, 2.5% DMSO, 2.5% cyclodextrin (the same derivative of cyclodextrin used in the drop) and 15-30% PEG6000 (depending on the derivative used, the crystals appear in different PEG6000 range). The final volume of the reservoir was 800  $\mu$ L. A 3  $\mu$ L drop of the protein was mixed with an equal amount of the reservoir solution in a sitting drop plate and the plates were sealed with tape. Crystals with quality sufficient for data collection were grown at 18°C between 1-3 weeks. The crystals were dipped for a few seconds into cryo-buffer solution composed by the crystallization buffer with an addition of glucose (20% v/v), 5 mM CD and 5 mM ligand.

#### 4.2.7. Data collection and processing

For the current study, all data were collected at Helmholtz-Zentrum Berlin (Berlin-Adlershof, Germany) [157] at BESSY II using beamlines MX14.1 and MX14.2. The data collection was done at cryogenic temperature of 100 K (-173°C) and wavelength 0.9184Å using a silicon Pilatus 6M pixel detector at atmospheric pressure. Data sets collected were indexed, processed and scaled using XDSAPP [158].

#### 4.2.8. Crystal structure determination and refinement

Crystal structures were determined by molecular replacement using the program PHASER MR [159] from the CCP4 suite [195]. The structure of the wild type 17 $\beta$ -HSD14 S205 variant was used as a search model for molecular replacement (PDB ID: 5ICM [142]) after removing all non-protein molecules and hydrogen atoms using FCONV [161]. In the refinement, a 5% subset of the reflections was used for the  $R_{\text{free}}$  calculations and they were



consequently, omitted from the refinement. The model was built in COOT [162] and the refinement was done using PHENIX.refine version 1.18.2-3874 [163]. The restraints assigned to the ligands were generated using the Grade Web Server [164] or eLBOW from PHENIX [163]. They were also used for energy minimization and restraint generation, SMILES codes were obtained from Molinspiration [165]. First refinement step was performed with default parameters for XYZ coordinates, occupancies and individual B-factors were alternated with structural adaption in COOT [162].

#### 4.2.9. Inhibition assay of 17 $\beta$ -HSD14

The inhibition of 17 $\beta$ -HSD14 by the various test compounds was evaluated with a fluorimetric assay following an established protocol [30], using the purified, recombinantly expressed enzyme, E2 as substrate and NAD<sup>+</sup> as cofactor. A high enzyme (between 3.4  $\mu$ M) and substrate concentration (E2 = 32  $\mu$ M) had to be applied because of the low sensitivity of the assay. To a mixture of NAD<sup>+</sup> (1.2 mM) and E2 (32  $\mu$ M) in 100 mM phosphate buffer pH 8, the compound was added in DMSO (final DMSO concentration in assay: 1%) and the enzymatic reaction was started by adding the purified enzyme (1 mg/mL). The production of the fluorescent NADH by the enzymatic reaction was measured continuously for 15 min on a Tecan Sapphire 2.

### 4.3. Results and discussion

#### 4.3.1. Compounds solubility with cyclodextrins

##### 4.3.1.1. The fluorine-compound library solubility

The compounds from the fluorine-compound library (**Appendix**) were purchased from different vendors or synthesized by our working group. They are soluble in organic solvents (such as DMSO), except for most of the compounds which were synthesized by our working group (**FB20**, **FB133**, **ES19**, **ES24** and **AM4.6**).

The first attempts were to solubilize the insoluble library compounds in distilled water by the addition of CDs. However, even when in principle a CD/compound complex should be formed

in aqueous solution enhancing overall hydrophilicity, no improved solubility in water was observed.

Then, all compounds were first solubilized or suspended in DMSO and then mixed with the same volume of CD solution. There was a large portion of the compound that precipitated in the mixture. The mixtures have been centrifuged at 10000 rpm for 10 min and the supernatant was filtered (to obtain a clear mixture), to be later used in the crystallization experiment.

#### **4.3.1.2. Hydrophobic drugs solubility**

Solubilizing the hydrophobic drugs in water was not possible without heating. A water bath at 80°C was needed for 24 – 72 h to solubilize the hydrophobic drugs. All drugs showed an enhancement in solubility when heated in water in presence of an equal amount of CD. The protocol was successful except for the Non-Steroidal Anti-Inflammatory Drugs (NSAIDs, Ibuprofen and nimesulide) and vitamin D analogs (ergocalciferol and doxercalciferol), which were practically insoluble in water.

Since all of the hydrophobic drugs were soluble in organic solvents, they were first solubilized in DMSO. Then, by mixing with an equal amount of aqueous CD solution, the organic drug solution was successfully prepared. The NSAIDs solutions precipitated after mixing with the CD solution, while solution of the Vitamin D analogs turned into a kind of emulsion (assessed by visual inspection).

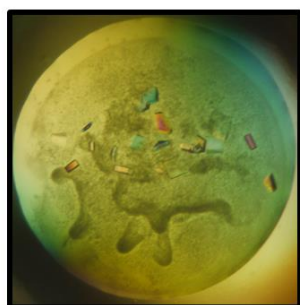
#### **4.3.1.3. Sex hormones solubility**

Sex hormones (estrogens and androgens), are badly soluble and water repellent, thus dissolving them directly in water is not an option. All of the sex hormones used in this study are well soluble in DMSO. Thus, mixing the sex hormones dissolved in DMSO with an equal volume of aqueous CD solution led to a clear solution.

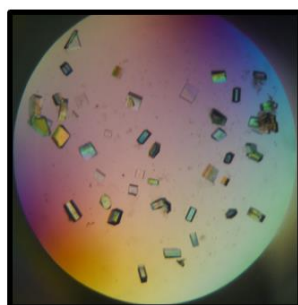
### 4.3.2. Crystal quality and stability with cyclodextrins

In the crystallization protocol used for 17 $\beta$ -HSD14 without CDs (**Chapter 1**), measurable protein crystals were obtained after 7-21 days. Their resolution varied from 1.4 - 2.2 Å.

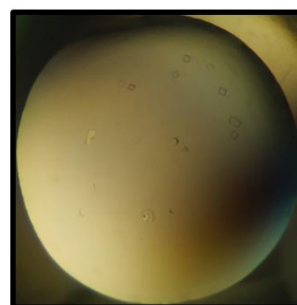
The addition of the CD solution to the crystallization condition resulted in an increased size and number of 17 $\beta$ -HSD14 crystals in the crystallization drop. However, poorly shaped or no crystal growth was observed with compounds which were insoluble or poorly soluble in the CD/compound mixture. All of the crystals obtained are co-crystallized with compounds which were previously dissolved in DMSO before mixing with the CD solution. No crystals were observed for the crystallization condition with only water as a solvent.



Compound: bad17  
CD:  $\alpha$ -cyclodextrin



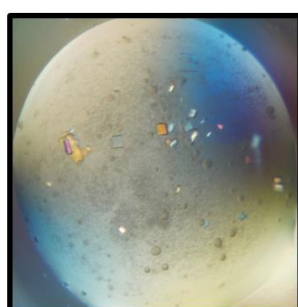
Compound: FB20  
CD: 2-hydroxypropyl- $\beta$ -cyclodextrin



Compound: Estradiol (E2)  
CD: methyl- $\beta$ -cyclodextrin



Compound: acetaminophen  
CD: methyl- $\beta$ -cyclodextrin



Compound: bad27  
CD: 2-hydroxypropyl- $\beta$ -cyclodextrin

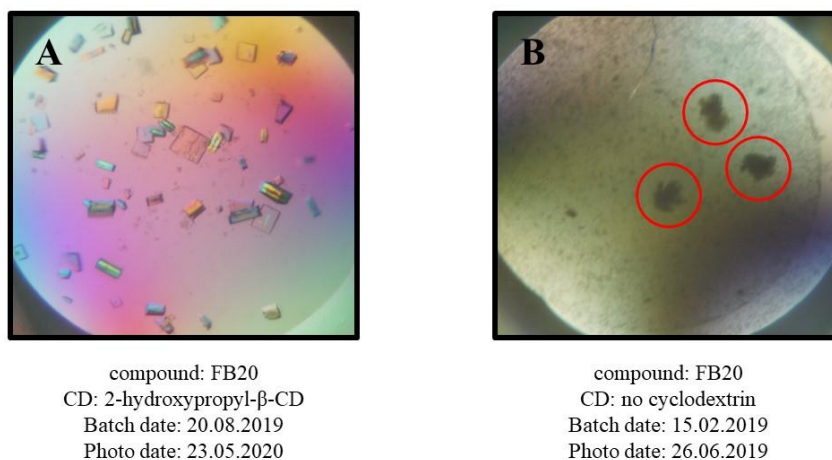


Compound: bad20  
CD: 2-hydroxypropyl- $\beta$ -cyclodextrin

**Figure 5.** Crystals of 17 $\beta$ -HSD14 co-crystallized with hydrophobic probe compounds (marked below each image) in presences of different cyclodextrin derivatives.

The majority of the crystals are obtained under crystallization conditions in which  $\beta$ -CD derivatives are present, especially, 2-hydroxypropyl- $\beta$ -CD.

Unfortunately, even with the increased size and number of 17 $\beta$ -HSD14 crystals, there was only a small improvement in the diffraction resolution. 17 $\beta$ -HSD14 crystals do not have a long duration in the crystallization drop (2 weeks up to 3 months maximum). Longer than that, the crystals dissolve in the crystallization drop or lose their diffraction power.



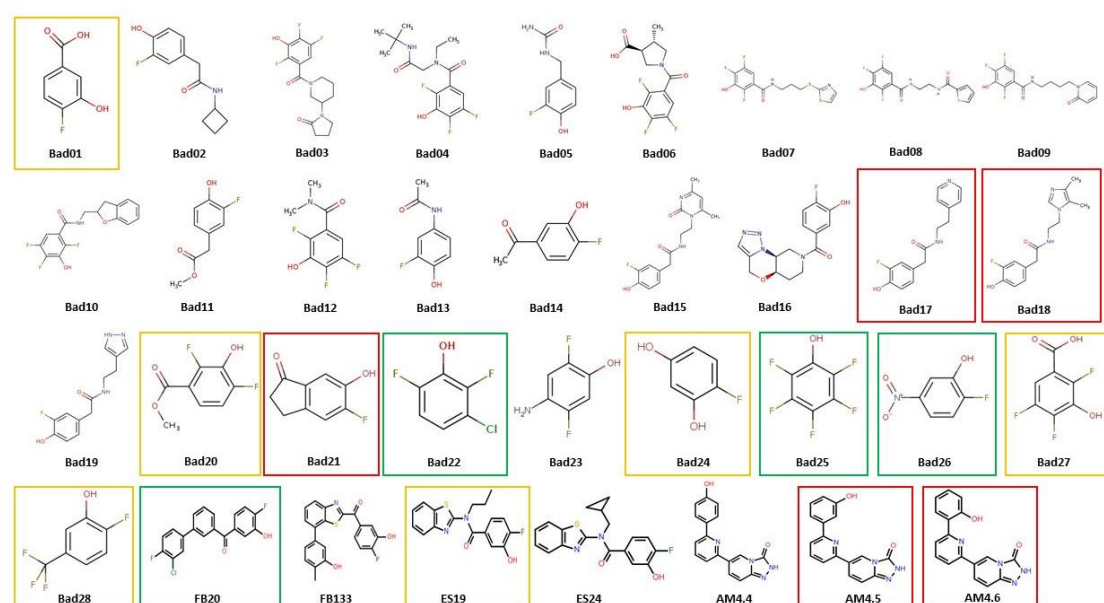
**Figure 6.** Comparison between 17 $\beta$ -HSD14 crystals co-crystallized with ligand FB20. **A.** crystals grown with cyclodextrin. **B.** crystals grown without cyclodextrin.

As seen in **Figure 6A**, the crystals grown in presence of CD have a sustained durability in the crystallization drop. After 9 months from preparing the initial crystal batch, the crystals were still in good shape. The resolution of the data set collected after 9 months did not decrease compared to those collected after a few weeks from starting the preparation of crystal plates without adding the CDs. However, the crystals that were grown in higher concentration of PEG6000 (> 25% w/v) showed more durability than crystals grown in lower PEG6000 concentrations (< 20% w/v).

### 4.3.3. The fluorine-compound library

#### 4.3.3.1. Hits from the Fluorine-compound library

As mentioned earlier, all the compounds from the fluorine-compound library were screened with 17 $\beta$ -HSD14 without the addition of CDs. No successful hits could be recorded. Instead, all the hits obtained from screening the library were revealed in the presence of CDs in the crystallization buffer.



**Figure 7.** The compounds from the fluorine-compound library. All compounds surrounded by a box did crystallize with 17 $\beta$ -HSD14 (red box = compound is not binding, yellow box = compound is partially binding and green box = compound is binding).

Throughout this study, the only successful method used to obtain 17 $\beta$ -HSD14 crystals was to co-crystallize the test compound with the enzyme. However, the enzyme had to be crystallized at 18°C, due to a thermostability issue of the enzyme. When leaving the enzyme for 1 hour at room temperature, the protein concentration decreases by 40% (concentration measured by using spectrometer NanoDrop™ from Thermo Fisher SCIENTIFIC) and crystallizing at the conditions of the cold room ( $\approx$  4°C) does not result in crystal growth. The growth period of 17 $\beta$ -HSD14 crystals ranged from 7 - 21 days. In previous inhibition assays done in our working group by Florian Braun or in the present study, mostly, any compound that has no affinity to

the enzyme did not result in the production of protein crystals when co-crystallized with the enzyme. 17 $\beta$ -HSD14 needs to be in the complex with a potent inhibitor to be sufficiently stable or in conformationally homogeneous form in order to crystallize. Nevertheless, even when a compound had sufficient affinity to the enzyme, it does not necessarily mean that it will crystallize (especially for compounds with rather poor solubility). All the compounds rendered by boxes in **Figure 7** did produce protein crystals when subjected to co-crystallization with the enzyme. Nevertheless, small compounds (fragments) usually have very weak affinity to bind to a biological target macromolecule. In the case of compounds **FB20**, **FB133**, **ES19**, **ES24**, **AM4.4**, **AM4.5** and **AM4.6**, they were all synthesized as putatively potent 17 $\beta$ -HSD14 inhibitors. They all share an aromatic ring with hydroxyl and fluorine functionalities which make favourable interactions in the active site. However, some of these compounds did not produce crystals while others did, but without a visibly bound hit indicated by an appropriate difference electron density in the active site. The issue with these compounds is not the affinity, but possibly their minor solubility. The presence of cyclodextrins made it possible to obtain protein crystals with compounds, such as **FB20**, **ES19**, **AM4.5** and **AM4.6**, while the first attempts using the original crystallization protocol (without CD) was not successful in obtaining any crystals.

#### 4.3.3.1.1. Compound **FB20**

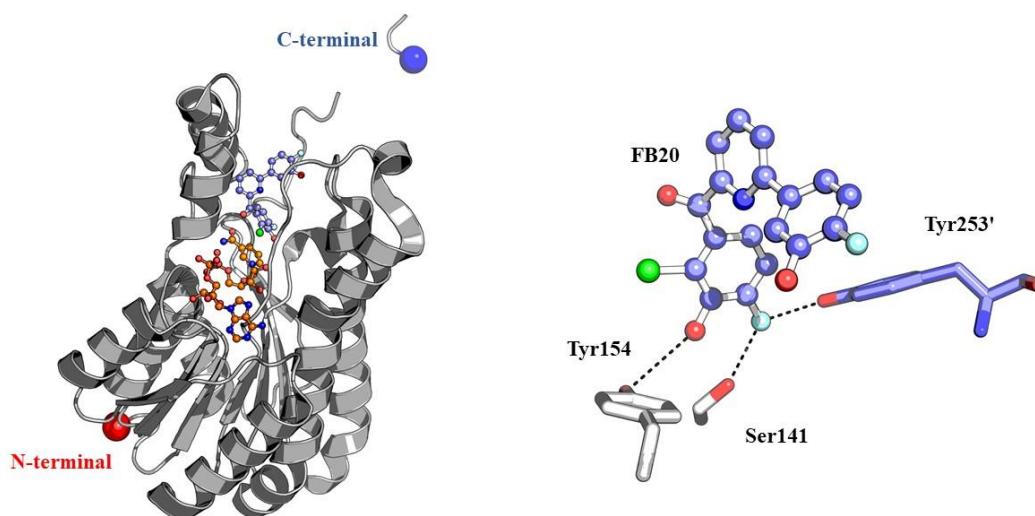
Ligand **FB20** was synthesized in our working group by Florian Braun. The inhibition constant  $K_i$  of **FB20** was determined by the described fluorometric assay using the 17 $\beta$ -HSD14 enzyme ( $K_i = 190 \pm 45$  nM). The poor water solubility made it difficult to obtain protein crystals in complex with this ligand. Fortunately, when the protein is co-crystallized in presence of the ligand/CD complex, protein crystals were obtained.

Before co-crystallizing the protein with **FB20**, the ligand was dissolved (or suspended due to its low solubility) in DMSO. This solution was then mixed with an equal volume of different CD derivatives dissolved in water (each at a time). The molarities of both, **FB20** and CD, were adjusted equal so that stoichiometric complex formation was possible. Protein crystals were obtained from conditions that contained  $\alpha$ -CD, methyl- $\beta$ -CD, 2-hydroxypropyl- $\beta$ -CD,  $\gamma$ -CD or 2-hydroxypropyl- $\gamma$ -CD. The ligand is visible in the electron density when 2-hydroxypropyl- $\beta$ -CD and 2-hydroxypropyl- $\gamma$ -CD were applied.

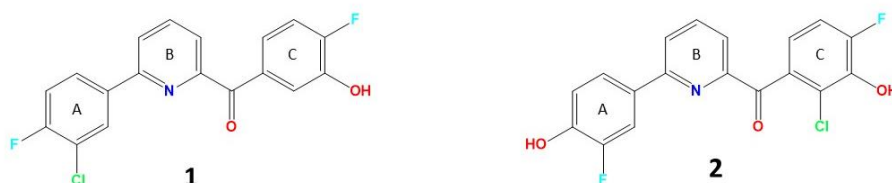
However, after resolving the crystal structure, a structural difference to the ligand topology predicted from synthesis was found. **FB20** contains a chlorine atom which is easily detected in the difference electron density due to its massive signal in the mFo-DFc map. Even more, an anomalous map has been created to check the presence of the chlorine atom in the active site (**Figure 10 A**), not to mention the bond length difference which is obvious in the electron density (**Figure 10 B**). There are three functional groups attached to the aromatic ring C of **FB20**, a chlorine, fluorine and hydroxyl group. A Cl...C bond length is usually 1.73 - 1.75 Å, while a F...C and HO...C bond length are both about 1.34 - 1.36 Å in length, respectively. Thus, it is easy to distinguish the Cl attachment at the aromatic ring from the other bond lengths in crystal structure with an acceptable quality and resolution (1.5 Å in our case).

In X-ray crystallography, the difference electron density is the compass to determine whether a compound is binding or not and how a compound is oriented in the binding site. The electron density in the binding pocket shows that ligand FB20 fits (**Figure 10 B**), but the chemical structure seems to deviate from the synthesis proposal. Even more, the average *B*-factor for the ligand after refinement is 28.0 Å<sup>2</sup> at an occupancy of 0.75 with all atoms fitting in the electron density (except for a carbon atom at the pyridine ring which is located slightly outside the selected electron density contour level).

Both the electron density and the anomalous map showed that the ligand **FB20** is clearly bound to the active site, but a different chemical structure has to be assigned compared to that originally proposed by synthesis.



**Figure 8.** A cartoon representation of 17 $\beta$ -HSD14 (left) with cofactor NAD<sup>+</sup> and ligand FB20 presented as balls and sticks in orange and blue respectively and a closer view of ligand FB20 (right) inside the binding pocket. Residue Tyr253' (blue) is from the adjacent monomer mate.



**Figure 9.1.** The proposed chemical structure of the synthesized ligand FB20. **2.** The chemical structure of ligand FB20 as found in the crystal structure.

17 $\beta$ -HSD14 crystals in complex with ligand FB20 were obtained with five different CD derivatives. Only crystal structures that were crystallized with 2-hydroxypropyl- $\beta$ -CD and 2-hydroxypropyl- $\gamma$ -CD, showed the ligand's electron density in the binding pocket. The crystal structure shown in **Figure 8** has been crystallized with 2-hydroxypropyl- $\beta$ -CD.

The protein did not yield any proper crystals with bound ligand **FB20** without exposing the protein to the ligand/CD complex. However, the bound ligand in the active site showed a deviating chemical structure in comparison to the originally anticipated chemical structure as obtained from synthesis. In **Figure 9 (A)** the chemical structure of the originally synthesized ligand showing on the aromatic ring (A) both a fluorine and a chlorine atom at *para* and *meta* positions respectively, while in **Figure 9 (B)** the chlorine is found at the aromatic ring (C) at *ortho* position while at ring (A) a fluorine and hydroxyl group are detected with mutually exchanged positions.

The protein has been crystallized multiple times in complex with **FB20**/CD mixture over a period of time spanning about 18 months. During the first attempts, many methods to analyse the chemical structure of **FB20** were performed. Mass spectrometry was performed and a mass of 345 Da was found. Nevertheless, as both chemical structures in **Figure 9** show the same chemical composition they share the same mass. Furthermore, Nuclear Magnetic Resonance (NMR) was performed suggesting correctness of the topology of ligand **FB20** as displayed in **Figure 9 A**, but the interpretation of the aromatic substitution pattern is quite delicate in the



present case. The first crystal structure was collected six months after the NMR was performed, and the sample was stored at -20°C in an air tight and inert dark glass container, which makes structural changes to the ligand during storage unlikely. Also, structural changes of **FB20** once exposed to 2-hydroxypropyl- $\beta$ -CD appear unlikely as stable C-halogen bonds would be necessary to break and the ligand forms only an inclusion complex with the CDs. Thus, no covalent bonds are formed or broken during the complex formation. The fact that **FB20** only forms with 2-hydroxypropyl-CDs a detectable complex is likely explained by the deviating physicochemical property which render these CDs prominent (**Table 3**).

**Table 3.** Cavity size and some physiochemical properties natural cyclodextrin and some derivatives.

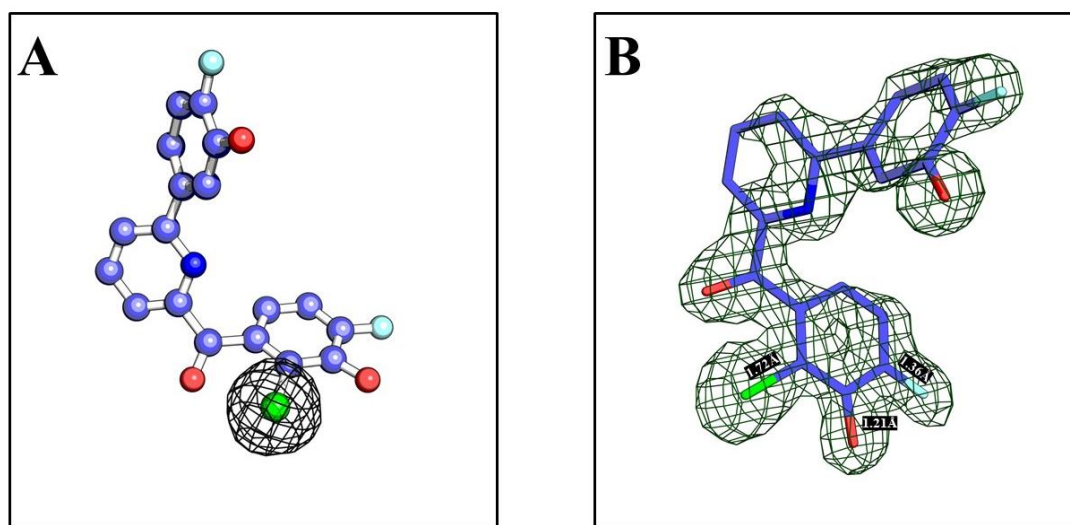
CD type	Substituent	Inner cavity diameter(Å)	Hydrogen donor	Hydrogen acceptor	Log $P_{o/w}$ <sup>a</sup>
$\alpha$ -CD	H	4.7_5.3	18	30	-13
$\beta$ -CD	H	6.0_6.5	21	35	-14
$\gamma$ -CD	H	7.5_8.3	24	40	-17
methyl- $\beta$ -CD	-CH <sub>3</sub>	5.8_6.5	7	35	-
2-hydroxypropyl- $\beta$ -CD	-CH <sub>2</sub> -CHOH-CH <sub>3</sub>	6.0	25	39	-11
2-hydroxypropyl- $\gamma$ -CD	-CH <sub>2</sub> -CHOH-CH <sub>3</sub>	8.0	24	45	-13

<sup>a</sup>The partition coefficient, abbreviated P, is defined as a particular ratio of the concentrations of a solute between the two solvents (octanol/water).

Data is taken from [201–205]

**Table 3** shows the different physicochemical properties for each CD derivative used to produce a complex with ligand **FB20**. As mentioned earlier, the interior cavity of CD is lipophilic and the exterior is hydrophilic, which means that the OH groups are located on the outer surface of the CD. Especially, 2-hydroxypropyl- $\beta$ -CD has the highest number of hydrogen-bond donors while having relatively the smallest cavity size compared to other CDs. The chances of compound binding might be for these reasons higher along with the better aqueous solubility

of the more polar CD. Nevertheless, this does not explain the observed differences in the chemical structure anticipated from synthesis and found by crystal structure analysis. Exposure to X-ray can break chemical bond, especially to electron-rich atoms. However, we believe that also this modification does not explain the quantitative transformation of a chlorine atom from ring A to C along with the topology change of hydroxyl and fluorine substitution in ring A. Definitely, a more detailed analysis of the synthesis protocol will be required. The chlorine atom has been introduced to aromatic ring by electrophilic halogenation, following the use of  $\text{Cl}_2$  in the presence of an appropriate Lewis acid catalyst (e.g.  $\text{FeCl}_3$ ). Possibly in this step, the other ring was substituted leading to the isomeric product. Mass spectrometry and NMR can hardly distinguish the two isomers. However, to clarify this aspect, a more detailed analysis will be required, which is beyond the scope of this thesis. For the question studied in the chapter, the issue is of minor importance as it could be shown that the use of a more polar CD definitely allowed to obtain a crystal structure with a badly water-soluble ligand, which could not be cocrystallized otherwise.

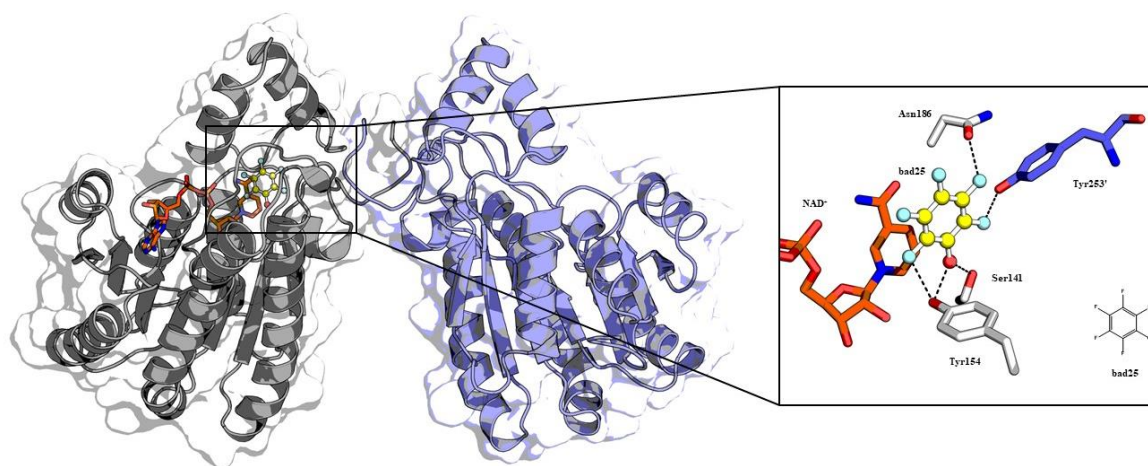


**Figure 10.** *A. Ligand FB20 represented as ball and stick in the binding pocket, where the black mesh around the chlorine atom (green) is the anomalous peak. B. Ligand FB20 represented as stick model in the binding pocket, where the black mesh surrounding it is the  $m\text{Fo\_DFc}$  map at  $3.0\text{\AA}$   $\sigma$  level.*

#### 4.3.3.1.2. Compound bad25

The fluorine-compound library contains many organo-fluorine compounds, but bad25 (pentafluorophenol, MW = 184 g/mol) which is a colourless solid that melts just above room temperature (melting point 32.8°C) with a  $pK_a$  value of 5.5 is the most acidic phenol from the library. Pentafluorophenol has one hydrogen-bond donor (OH group) and five fluorine atoms which make favourable interactions in the binding pocket with the catalytic triad and the OH group of Tyr253' from the adjacent monomer mate. The compound's small size makes it easier to travel through the water channels to the active site. However, no protein crystals could be obtained in complex with **bad25** when crystallized using the original protocol without CDs. It seems that the high acidity of the compound has a detrimental effect on the crystallization process which is performed under neutral pH conditions.

**Bad25** has no solubility issue and it is well soluble in water and organic solvents like most small compounds (fragments). Just like other small compounds, **bad25** tends to have lower affinity than larger compounds. Unfortunately, the inhibition constant ( $K_i$ ) could not be determined for **bad25**, as the assay is not sensitive enough to record weakly binding ligands.



**Figure 11.** A cartoon representation of 17 $\beta$ -HSD14 as a dimer (PDB ID: 6ZDE) and a closer view of the active site where compound bad25 (balls and sticks) is binding.

Compound **bad25** is hydrophilic with five fluorine functionalities on the aromatic ring, which means that it will likely bind on the hydrophilic outer surface of the CD. However, the electron

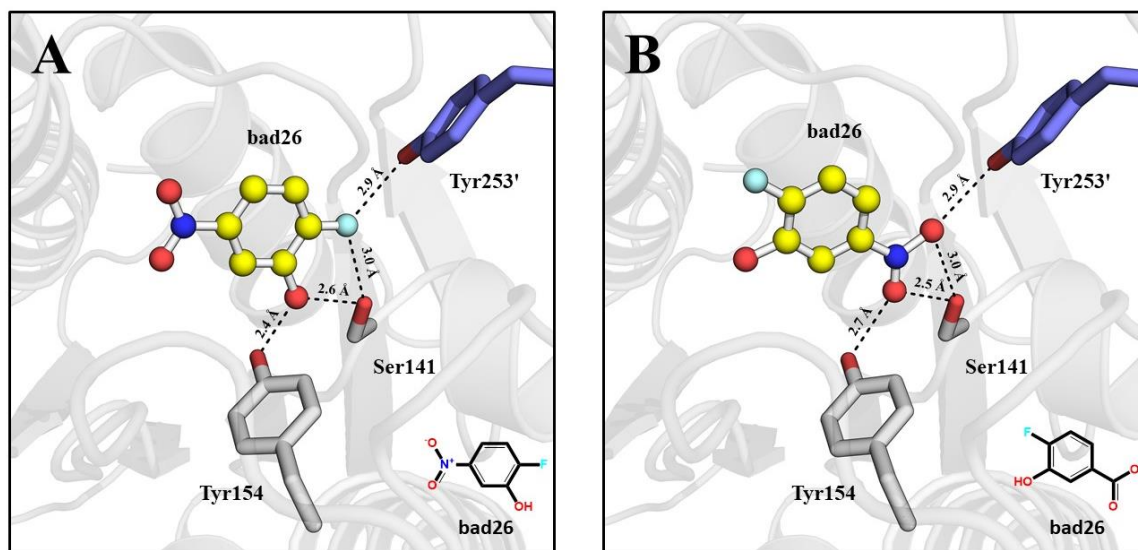
density does not show traces of cyclodextrin or any glucose subunits, beside the conserved glucose molecule on the surface of the enzyme.

17 $\beta$ -HSD14 in complex with **bad25** was crystallized with  $\gamma$ -CD. The interaction of compounds with cyclodextrins leads to apparent changes in their chemical properties, such as shifts in their absorption spectra and acid-base equilibrium. Formation of inclusion complexes between acidic and/or basic forms of a compound with CDs can change the amount of the conditional acidity constant [206]. Apparently, the effect of the **bad25**/ $\gamma$ -CD inclusion body made it possible to obtain protein crystals by masking the high acidity of **bad25** in the crystallization drop.

The interactions of the compound in the active site (**Figure 11**) with the enzyme show a typical pattern similar to that observed for potent inhibitors synthesized by our working group. The OH group interacts with residues Ser141 and Tyr154, while one of the fluorine atoms is interacting with residue Tyr253' from an adjacent monomer mate. The interaction between fluorine and residue Asn186 is not seen for other potent compounds. However, the presence of multiple fluorine atoms on one ring in **bad25** makes the difference compared to synthesized potent inhibitors, which contain only one fluorine atom on the aromatic ring.

#### 4.3.3.1.3. Compound bad26

**Bad26** (2-fluoro-5-nitrophenol, MW = 157.1 g/mol) is harmful if swallowed (toxic), inhaled (targeting organ is the lung) or contacted the skin (irritant), so all experiments done with this compound were performed under a chemical fume hood. Unfortunately, due to its toxic and slightly volatile chemical properties, no experiments apart from crystallization was performed (e.g. inhibition assay). 17 $\beta$ -HSD14 crystals were obtained in complex with compound **bad25**, the co-crystallization of the protein was done by introducing a **bad25**/CD mixture to the protein sample in the crystallization drop. Protein crystals were obtained from conditions that contained  $\alpha$ -,  $\gamma$ -, methyl- $\beta$ -, 2-hydroxypropyl- $\beta$ - and 2-hydroxypropyl- $\gamma$ -cyclodextrin. Data sets were collected for each crystal with a different CD.



**Figure 12.** A close view of the active site of 17 $\beta$ -HSD14 in complex with compound **bad26** (balls and sticks). The catalytic triad residues Ser141 and Tyr154 (grey sticks) and residue Tyr253' from the adjacent monomer mate (blue sticks). **A.** the OH group and F atom forming interactions to Ser141 and Tyr154 (PDB ID: 6ZDI). **B.** The nitro group is forming very similar interactions as the OH and F substituent (PDB ID: 6ZRA).

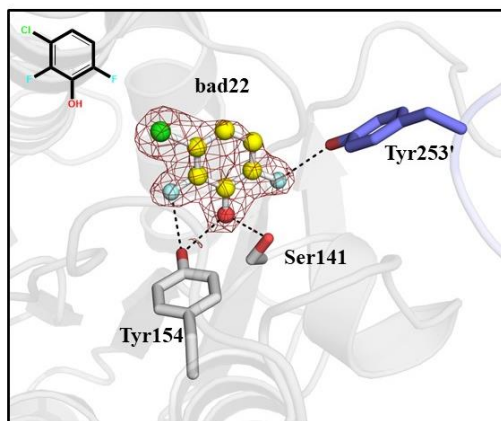
**Figure 12** shows two crystal structures of 17 $\beta$ -HSD14 in complex with compound **bad26**, but the established interactions found in each structure are different (the compound orientation changed). Both structures were crystallized using 2-hydroxypropyl- $\beta$ -CD, which throughout this study, is the CD that gave the better results (crystal quality, crystal growth and resolution). **Figure 12A** shows the OH group and F atom are forming interactions with residues Ser141 and Tyr154 from the catalytic triad and residue Tyr253' from the adjacent monomer mate, which is the pattern seen with other potent inhibitors containing these two functionalities, such as inhibitor **1** (**Chapter 1**). However, in **Figure 12B** the compound changes orientation and the nitro group is making interactions instead of the OH group and F atom at the aromatic ring.

The first data set from 17 $\beta$ -HSD14 in complex with **bad26** is the one shown in **Figure 12A** where the OH group and F atom are forming the interactions. Another data set was collected from different protein batch aiming for a better resolution. The newly collected data set did not

have a better resolution. However, it has been noticed that the interaction pattern has changed. To validate these results, data sets for 17 $\beta$ -HSD14 in complex with **bad26**/2-hydroxypropyl- $\beta$ -CD complex were collected multiple times (4 times) to assess which orientation of the compound in the active site is the predominant one. Crystal plates have been prepared, crystals from different plates were tested and collected, they all show the interaction pattern seen in **Figure 12A**, where the OH group and F atom at the aromatic ring are forming interactions with the protein. Thus, the structure where the nitro group was interacting with the enzyme's amino acids was assumed to be a rare case grown under slightly different conditions which are difficult to reproduce. Likely, both orientations found for the fragment are very close in energy.

#### 4.3.3.1.4. Compound bad22

Compound **bad22** (3-chloro-2,6-difluorophenol, MW = 164.5 g/mol) is a chlorophenol. The compound is corrosive, irritant, environmental hazardous with fair vitality. The handling of the compound and the preparation of crystals plate were all done under a chemical fume hood.

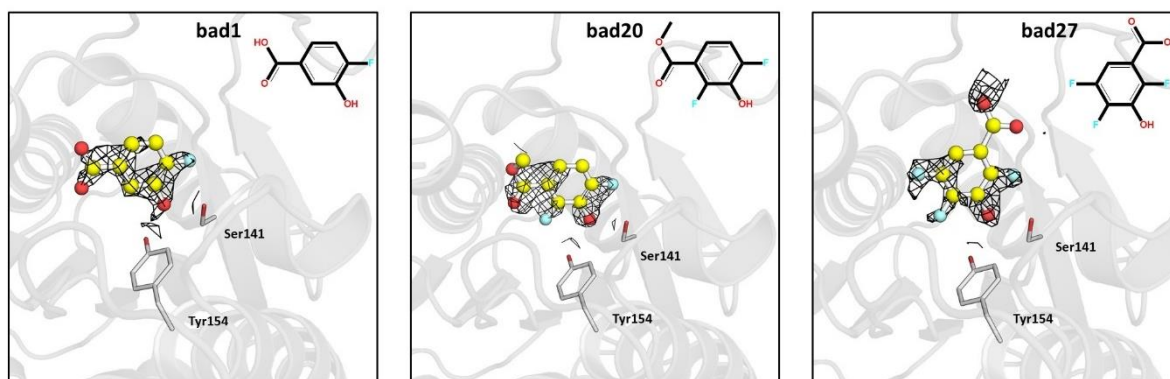


**Figure 13.** A close view of the active site of 17 $\beta$ -HSD14 in complex with compound **bad26** (PDB ID: 6ZT2). The catalytic triad residues Ser141 and Tyr154 (grey sticks) and residue Tyr253' from the adjacent monomer mate (blue sticks). The electron density is shown at 3.0 Å  $\sigma$  level.

Compound **bad22** shares similarity to compound **FB20** in terms of attached functionalities which are forming interactions with the enzyme. Even with an extra F atom, the positioning of the Cl atom of **bad22** in the active site is similar to that of **FB20** that binds to the active site of the enzyme (**Figure 8**). The orientation seems to be favoured when a compound with F, HO, and Cl substituents are present in the active site. The chlorine atom is easily localized due to its strong anomalous signal, but it does not form any interactions with active-site residues.

#### 4.3.3.2. The fluorine-compound library partial hits

Many compounds from the fluorine-compound library shown in **Figure 7** were labelled as partial hits. In this study, a compound is considered a “partial hit” (no full population of the bound fragment) when there is a change in the mFo-DFc map of the collected data set. A “change” is defined as newly visible difference electron density which is not present in 17 $\beta$ -HSD14 *apo* crystal structure, independent whether the difference is in the active site or other areas throughout the structure.



**Figure 14.** Examples of the compounds considered as “partial hits” (balls and sticks) displayed together with mFo-DFc map at 3.0  $\sigma$  (black mesh) inside 17 $\beta$ -HSD14 active site. The electron density shown, was enhanced using a polder map [207].

All compounds selected for the fluorine-compound library share the presence of an OH group and an adjacent F substituent at an aromatic ring in common. However, only a few of these compounds were found to bind to the active site, while some others were classified as “partial hits”. Apparently, the presence of a favoured pattern of functional groups in a compound is not



sufficient to observe binding to a given macromolecule target. The compounds of bigger size do not bind to the enzyme and some even do not crystallize with the protein. The volume of the binding pocket of 17 $\beta$ -HSD14 (2619Å<sup>3</sup>) [139] is big enough to host most of the compounds of the library, but the spatial fixation of the flexible loops contributed by two adjacent monomer mates on top of the entrance of the binding pocket makes it more difficult for larger compounds to enter.

#### 4.3.4. The hydrophobic drugs

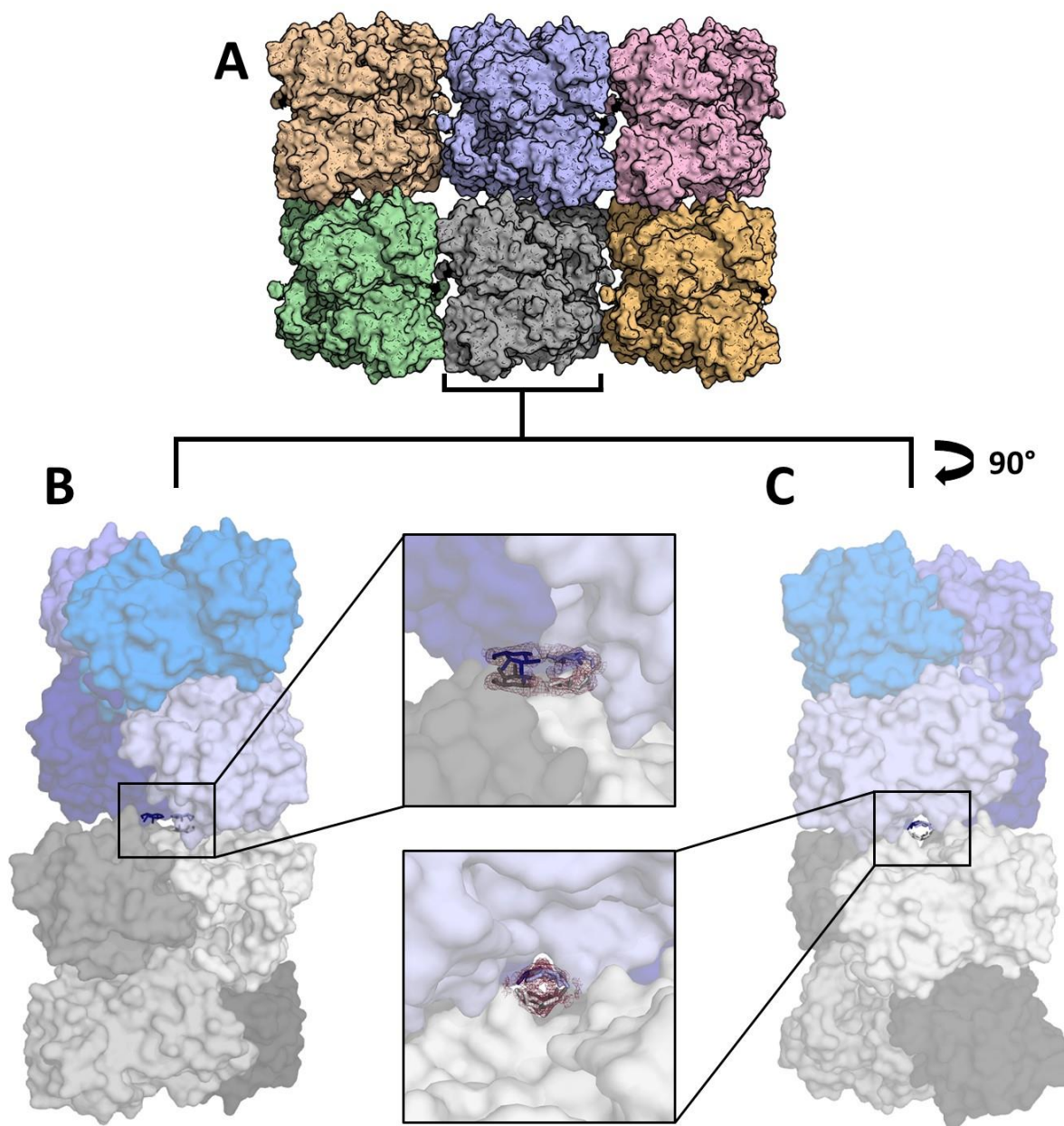
Even with the use of CDs to make inclusion complexes with the hydrophobic drugs, it was not possible to completely solubilize them in an aqueous solution. Still, the crystallization experiment was conducted after centrifugation and filtration of the drug/CD complex to remove any suspended particles which were not dissolved. The only protein crystals that formed were those with acetaminophen (paracetamol) that was incubated with 2-hydroxypropyl- $\gamma$ -CD. Unfortunately, no visible electron density for acetaminophen could be traced within the structure.

#### 4.3.5. Sex hormones

A crystal structures containing one of the sex hormones such as estrogens (E2) and androgens (5-diol), arguably being 17 $\beta$ -HSD14 substrate, would be useful to study their interaction pattern in the active site of the enzyme. Protein crystals formed when co-crystallized with estrone (E1), estradiol (E2) and Androst-4-en-3-one. Sadly, none of the collected data sets contained a bound hormone molecule in the active site or any location elsewhere in the crystal packing. Even after enhancing the electron density, no signal of a bound compound could be observed. To enhance the chances to obtain hormone-enzyme structures, CDs were added to the crystallization conditions. The crystals that grew from co-crystallization with the sex hormones/CD complexes had inadequate properties in terms of crystal shape, diffraction pattern and resolution, thus also this protocol did not result in a novel crystal structure.



## 4.3.6. Water channels and cyclodextrin



**Figure15.** *A. Surface representation of 17 $\beta$ -HSD14 tetramer as arranged in the crystal lattice (each tetramer indicated with a different colour). B. Two adjacent tetramer mates represented as surface (each monomer is labelled with a different colour, but the same colour category for each tetramer) with a closer view on the water channel between the two tetramers where glucose molecules are found. C. The same as B, but from 90° vertical view. The red mesh around the glucose units is the mFo-DFc map at 3.0  $\sigma$  level.*

It has been noticed in the current investigation that multiple crystal structures, crystallized in the presence of CD, contain massive electron density (at 3.0 Å  $\sigma$  level) located in the water channels of 17 $\beta$ -HSD14 between each two adjacent tetramer mates (**Figure 15**). The electron density shape is oval and connected to one another, thus it might be a CD molecule. However, the size and diameter of the water channels at the position where the electron density was found is not sufficiently large to host a CD molecule. Even the smallest CD, used for this study ( $\alpha$ -CD), could not be fitted in the water channel without clashing with neighbouring amino acids. This difference in the electron density between crystals grown in presence or absence of CD gave strong evidence of a substantial involvement of CDs in compound delivery to the active site. Even if the molecule located throughout the water channels of the enzyme is not a CD molecule, and observing that water molecules, which occupied the water channels are displaced by this other compound are an indication of the definite CD effect on the delivery system of the enzyme (water channels).

Many attempts were made to find a suitable molecule to fit in the unexplained electron density found in the water channels. After failing to fit a CD molecule inside, the ligands that were crystallized in each sample set that showed this electron density were tried to fit. Unfortunately, the compounds were not able to explain the residual electron density as well. Finally, a glucose molecule proved to be a good candidate to explain the electron density. In the close-up view of the water channel in **Figure 15A** and **B**, four neighbouring glucose molecules were placed, each molecule is from a different monomer that are forming the water channel (two monomers from each neighbouring tetramer in the asymmetric unit). During refinement, the glucose occupancy was set to 100% because each molecule is part of a separated monomer. The glucose molecule proved to be a good fit to explain the residual electron density. However, the high fluctuation of the refined *B*-factors between the different glucose atoms was an issue. All glucose atoms forming interactions with other atoms from the neighbouring glucose molecules had high *B*-factor values, so the molecules were discarded. When the occupancy is set to 25% for each glucose molecule, the *B*-factor average for all atoms converted to an acceptable value ( $\approx 35$  Å<sup>2</sup>), but the electron density was not explained properly after refinement. Surprisingly, all crystals were grown with glucose present in the crystallization buffer independent of crystallization with or without presence of CDs. Thus, the observation of a newly hosted glucose molecules in the water channels only accommodated once a CD was added to the crystallization trials is not fully comprehensible. Possibly, the CD molecules

dissociated (glycosidic linkage broke) and multiple glucose molecules roam freely through the water channels.

#### 4.3.7. Crystallographic table

Details of the data collection, processing, diffraction data, refinement and other parameters related to 17 $\beta$ -HSD14 crystal structures are listed in **Table 5**.

**Table 5.** Data collection and refinement statistics for all crystal structures

PDB code	6ZDE	6ZDI	6ZR6	6ZT2	6ZRA
<u>Data collection and processing</u>					
Space group	I 4 2 2	I 4 2 2	I 4 2 2	I 4 2 2	I 4 2 2
Unit cell	91.2 91.2 133.0	91.2 91.2 133.0	91.1 91.1 132.1	91.2 91.2 133.5	91.1 91.1 132.5
Matthews coefficient <sup>a</sup>	2.34	2.43	2.4	2.9	2.4
Solvent content (%)	49.4	49.4	48.9	58.3	49.1
<u>Diffraction data</u> <sup>b</sup>					
Resolution range	50-1.87 (1.98-1.87)	50-2.13 (2.26-2.13)	50-1.50 (1.59-1.50)	50-1.95 (2.06-1.95)	50-1.73 (1.83-1.173)
Unique reflections	23587(2322)	16003(2517)	44723(7128)	20921(3271)	29510(4670)
R(I) <sub>sym</sub> (%) <sup>c</sup>	8.4(52.3)	11.7(48.5)	4.8(52.4)	10.2(57.7)	6.7(56.4)
Wilson B factor	22.5	21.6	18.0	24.1	23.4
completeness	99.8(99.7)	99.3(98.4)	99.9(99.7)	99.7(98.4)	99.9(99.8)
redundancy	13.1(13.6)	8.5(8.1)	8.7(8.9)	8.5(8.8)	13.2(13.5)
<I/σ(I)>	24.2(5.22)	14.8(4.2)	25.9(3.9)	15.1(3.5)	24.2(4.6)
<u>Refinement</u>					
Resolution range	29.65-1.87	46.34-2.13	45.12-1.50	46.37-1.95	46.2-1.73
Reflections used in refinement (work/free)	23603 (22425/1178)	16003 (15200/803)	44723 (42496/2236)	20921 (19875/1046)	28510 (27083/1426)
Final R value for all reflections (work <sup>d</sup> /free <sup>e</sup> ) (%)	16.4/19.3	20.8/21.4	15.1/17.9	17.5/21.5	17.7/16.7
Protein residues	252	252	257	252	256
Water molecules	149	52	165	96	132
rmsd from ideality: bond length (Å)	0.007	0.007	0.005	0.007	0.006
rmsd from ideality: bond angle (°)	0.81	0.88	0.77	0.83	0.81
Ramachandran most favored(%) <sup>f</sup>	98.4	98.5	98.4	98.4	98.4
Ramachandran additionally allowed (%) <sup>f</sup>	1.6	1.5	1.6	1.6	1.6
Mean B factor protein (Å <sup>2</sup> ) <sup>g</sup>	23.84	29.6	22.2	24.2	25.2
Mean B factor ligand (cofactor/ligand) (Å <sup>2</sup> ) <sup>g</sup>	20.8/27.9	26.0/30.7	17.0/31.5	22.4/25.6	23.5/26.6
Mean B factor water molecules (Å <sup>2</sup> ) <sup>g</sup>	29.9	28.0	27.2	27.3	30.8

<sup>a</sup> Calculated with Matthews\_coef program from CCP4 suite version 6.4.0. <sup>b</sup> Values in parenthesis describe the highest resolution shell. <sup>c</sup>  $R(I)_{\text{sym}} = [\sum_h \sum_i |I_i(h) - \langle I(h) \rangle| / \sum_h \sum_i I_i(h)] \times 100$ , in which  $I(h)$  is the mean of the  $I(h)$  observation of reflection  $h$ . <sup>d</sup>  $R_{\text{work}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$ . <sup>e</sup>  $R_{\text{free}}$  was calculated as shown for  $R_{\text{work}}$  but on refinement-excluded 5% of data. <sup>f</sup> Calculated with PROCHECK [167]. <sup>g</sup> Mean B factors were calculated with MOLEMAN [168].

## 4.4. Conclusion

Using cyclodextrins to form an inclusion complex with a compound to be transferred into the 17 $\beta$ -HSD14 enzyme in order to grow subsequently a protein crystal proved to be successful throughout this study. Without the usage of CDs, most of the hydrophobic compounds did not bind successfully to the enzyme. Some even did not lead to any successful crystal formation when co-crystallized with the enzyme. The fluorine-compound library contains hydrophobic and hydrophilic compounds. No crystals could be obtained without the usage of CDs when dealing with the hydrophilic compounds from the library, mainly, because of their chemical properties (acidity) was sufficiently masked by the CD, e.g. in the case of compound **bad25**. Some hydrophobic compounds, which showed promising results (e.g. **FB20** and **ES19**), showed an improvement due to increased solubility of the compounds. Multiple data sets could be collected from protein crystals formed as complexes with these compounds. Even more, through the enhancement of electron density maps (Polder maps), multiple compounds from the library could be identified as potential binders to the enzyme. Unfortunately, no successful hits could be obtained in the case of the hydrophobic drugs and sex hormones when the protein was exposed to CD complexes with these probe molecules. However, protein crystals have been obtained with these compounds, which is an advantage over the original crystallization protocol used in absence of CDs.

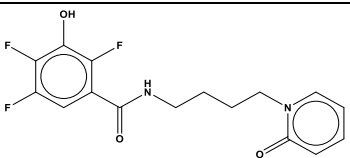
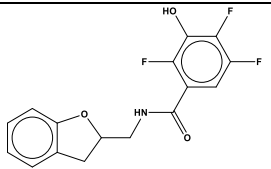
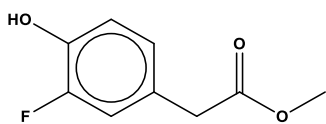
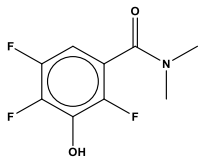
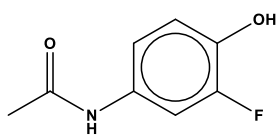
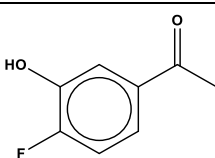
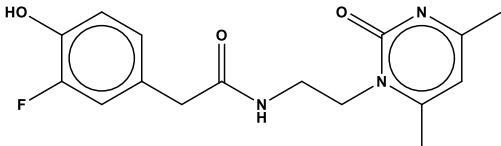
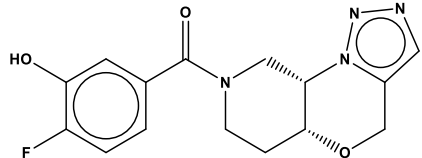
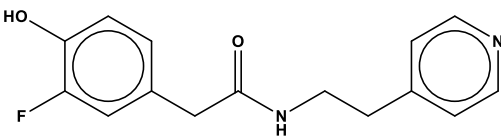
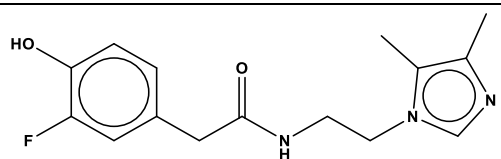
Glucose is seen in almost every 17 $\beta$ -HSD14 crystal structure deposited by our working group (PDB ID: 6G4L, 6H0M, 6ZRA) as bound on the surface of the enzyme. The glucose was present in the crystallization conditions as it was used in the purification protocol of the enzyme. Surprisingly, in the crystal structures obtained with the presence of CDs, the glucose molecules are supposedly in the water channels. Obviously, 17 $\beta$ -HSD14 is compatible with sugars and this was another reason to use CDs in the crystallization condition of the protein. The stability and quality of the protein crystals showed an observable enhancement. The crystals could survive for longer times in the crystallization plates without losing the diffraction power.

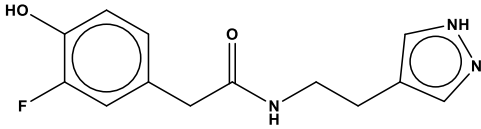
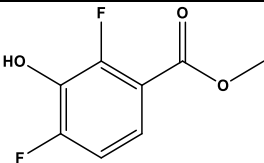
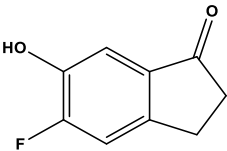
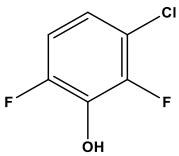
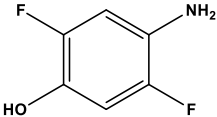
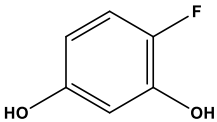
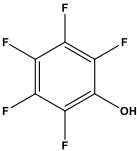
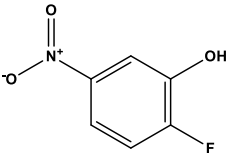
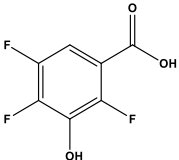
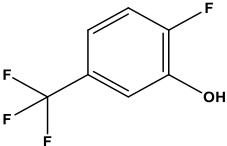
Among all CDs, 2-hydroxypropyl- $\beta$ -CD is the derivative that gave the best results when compared to other CD derivatives used in this study in terms of crystal formation, hit generation or data set quality and resolution. Still, further research is required to identify the effect of CDs on enhanced compound binding to 17 $\beta$ -HSD14 and the mechanism how the solubility improvement and compound transfer operates.

## 4.5. Appendix

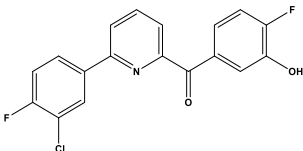
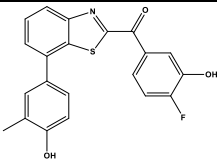
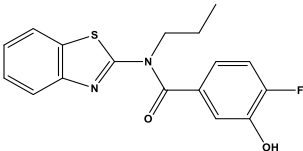
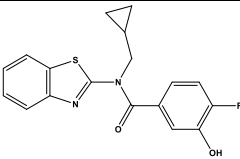
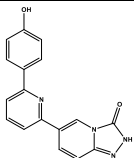
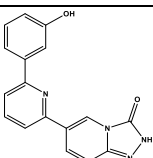
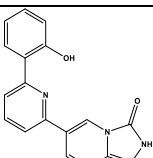
## The fluorine-compound library

	ID	Structure	MW (g/mol)
1	bad1		156.1
2	bad2		223.2
3	bad3		342.3
4	bad4		332.3
5	bad5		184.2
6	bad6		303.2
7	bad7		348.4
8	bad8		344.3

9	bad9		340.3
10	bad10		323.3
11	bad11		184.2
12	bad12		219.2
13	bad13		169.2
14	bad14		154.1
15	bad15		319.336
16	bad16		318.3
17	bad17		274.3
18	bad18		291.2

19	bad19		263.3
20	bad20		188.1
21	bad21		166.2
22	bad22		164.5
23	bad23		145.1
24	bad24		128.1
25	bad25		184.1
26	bad26		157.1
27	bad27		192.1
28	bad28		180.1



29	FB20		345.7
30	FB133		379.4
31	ES19		330.4
32	ES24		342.4
33	AM4.4		304.3
34	AM4.5		304.3
35	AM4.6		304.3



# 5

## Chapter 5

---

# Fragment-based drug discovery

Fragment screening against trypsin in a trigonal and orthorhombic crystal form



## 5.1. Introduction

Fragment-based lead discovery (FBLD) or in a broader sense named fragment-based drug discovery (FBDD) is an established approach and widely used method applied in drug discovery to find lead compounds. FBDD is increasingly used in pharmaceutical industry to reduce attrition and obtain leads for intractable or new biological targets.

Fragments are identified as low-molecular weight ligands which “usually” follow “The Rule of Three”, which is having a maximum of three H-bond donors/acceptors,  $\text{clogP} \leq 3$ , less than 20 non-hydrogen atoms ( $\text{MW} \leq 300 \text{ Da}$ ) not more than three rotatable bonds and Polar Surface Area ( $\text{PSA} \leq 60 \text{ \AA}^2$ ). However, fragments have weak affinity which requires a highly sensitive biophysical technique to check their binding mode with a designated target, such as nuclear magnetic resonance (NMR) or X-ray crystallography which is the method of choice in this study [97,98,105,111,113,114,208].

The high success rate in generating chemical series that possess lead-like properties is one of the advantages of FBDD. Fragment libraries contain a number of compounds to screen against a targeted receptor. However, small-scale screening compounds from a library can be customized or specifically picked depending on the target. In small-scale screening against a biological macromolecule, many criteria should be considered, such as the volume of the active site, the unrestricted movement through the water channels if soaking techniques are applied and the residues inside the corresponding active site.

Even with low affinity and small size fragments, hits obtained from a fragment screening would lead to a subsequent expansion of hits after optimizing a fragment candidate, through the addition of better matching functionality and making it more compatible to its corresponding receptor. When observing hits from a fragment screen, it is noticeable that the number of interactions is relatively low because of the low number of atoms. Unfavourable or unexpected interactions could form in regions other than the targeted ones [111].

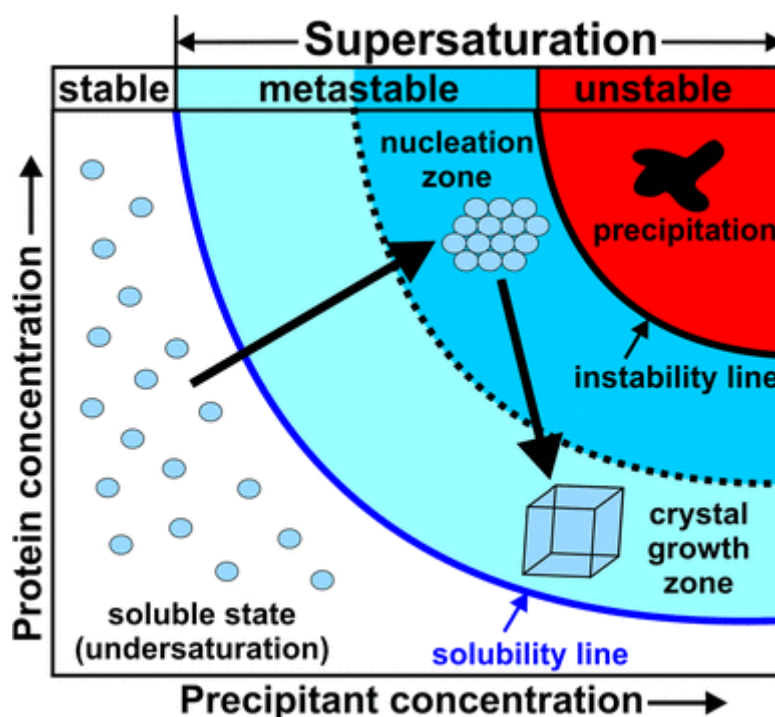
Weak intermolecular interactions such as hydrogen bonding and hydrophobic interactions are key players in stabilizing energetically-favoured ligands. The low water solubility of many drug-like compounds makes it harder to obtain a hit. Fragments of small size and low molecular weight offer higher hydrophilic hits that can boost affinity. Occasionally, adding a hydrophobic group can increase affinity. Starting from a hydrophilic fragment and adding a hydrophobic group would ease the process and compromise between hydrophilicity and affinity [209–213].

In this study, X-ray crystallography is the method of choice for FBDD to screen against cationic trypsin (EC: 3.4.21.4). There are other methods used to identify fragment hits, such as SPR, NMR or TSA, but X-ray crystallography is the method used in this study. The availability of third generation synchrotron sources and the modern instrumentation used in automated crystal mounting systems, data collection and processing enable FBDD to be done with more ease and speed. The automated refinement pipeline developed in the working group of Prof. Gerhard Klebe (AG. Klebe) contributed enormously to reduce the time required for manual structural refinement [196].

Trypsin (EC: 3.4.21.4) was first described in the late 1800s as a proteolytic activity present in pancreatic secretions [128]. Subsequent studies revealed that this enzyme specifically hydrolyzes peptide bonds C-terminal to the amino acid residues lysine (Lys) and arginine (Arg)  $10^9$  times faster than hydrolysis by hydroxide ion. Since its initial discovery, trypsin has been identified in all animals, including insects, fish, and mammals. The natural substrate for the enzyme is generally any peptide that contains Lys or Arg. The specificity of trypsin allows it to serve both digestive and regulatory functions. As a digestive agent, it degrades large polypeptides into smaller fragments. As a regulatory protease, it activates other proteins through proteolysis at specific Lys or Arg bonds.

Trypsin crystallization conditions can reveal different crystal packing. In this study trypsin crystals obtained are either trigonal or orthorhombic. The fragment screening was done on each crystal form of trypsin crystals using the soaking method. However, for orthorhombic trypsin crystals, both soaking and cocrystallization methods were applied.

For the description and structural characterization of protein-ligand interactions using X-ray structure analysis, protein crystals are grown in solution. For this purpose, the saturation point of the protein in the solution needs to be exceeded, so that the protein can crystallize. It is important that this process takes place slowly in order to form crystallization nuclei that can grow to larger crystals.



**Figure 1.** Phase diagram for protein crystallization taken from Bijelic [214].

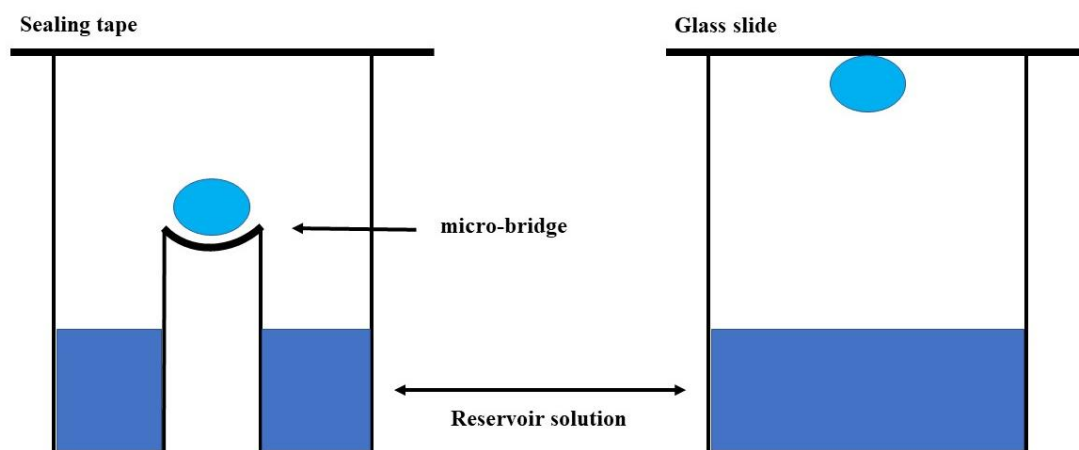
As seen in **Figure 1**, a distinction is made between an unsaturated and saturated range. At the boundary of both states, the maximum solubility of the protein is present at a defined concentration.

In addition, the saturated area is divided into a metastable and an unstable phase. In the metastable phase (**Figure 1**), the first crystallization nuclei are obtained from which no protein crystals can yet be formed. This process only takes place in equilibrium with the metastable phase, so that the crystals can grow slowly. If the saturation point is exceeded, an imbalance is created and a state is reached in which the protein can crystallize.

In order to achieve such supersaturation, the temperature can be lowered so that the solubility of the protein is reduced. This can also be additionally influenced by adding other precipitants such as PEG. Since proteins have, among other properties, titratable functional groups which can change protonation state, the crystallization can be supported by changing the pH value and the ionic strength of the solutions used in the protein sample buffer of the crystallization buffer.

The "salting out" process removes the solvent from the protein, thus achieving increased saturation and facilitating crystallization. Usually the hanging drop or the sitting drop method

is used in practice. If crystallization is carried out by hanging drop, the protein and reservoir solution are mixed together in the same ratio on a glass plate and then the reservoir is covered with it according to **Figure 2**. Depending on the difference in vapour pressure, the aqueous phase slowly passes into or out of the reservoir, resulting in a shift in concentration in the drop by vapour diffusion.



**Figure 2.** The well is covered by a glass plate in the hanging drop method (right), while in the sitting drop method (left) the well is sealed with a sealing tape.

Both the protein concentration and the precipitant concentration are increased and consequently a supersaturated solution is obtained. Vapour diffusion works in a similar way via the sitting drop method in which the drop is placed on a micro-bridge. As soon as the crystals formed have a sufficient size ( $> 0.05$  mm), they can be subjected to diffraction experiments. Here, X-rays are used which, with their short wavelength of approx.  $1 \text{ \AA}$ , should exert elasticity (scattering without energy loss) on the sample.

When X-rays hit the electrons of an atom, they are scattered in different directions. The atoms in the crystal are periodically arranged and through them, series of parallel lattice planes can be laid out, which differ in their orientations, their relative distance and their density of atoms. A detector can then collect the reflected rays and obtain a diffraction image from which the intensity and amplitude of these reflected rays can be determined. From this, differential electron density maps are calculated, which define the positioning of the respective atoms of the protein in the elementary cell.



## 5.2. Materials and methods

### 5.2.1. The 96-fragment library

The 96-fragment library used in this study is an inhouse library developed in collaboration with the Helmholtz Zentrum Berlin (HZB) MX-group at BESSY II (AG Weiss). The library is used for fragment screening by X-ray crystallography. The fragments are selected with respect to many criteria, such as, size, molecular weight, functional groups and number of hydrogen bond acceptor/donor properties.

The assembled fragments are selected from existing protein-ligand complexes, carbohydrates, nucleotides, amino acids, peptide like fragments and drug-like organic compounds. 52 entries of the 96-fragment library are extracted from a 361-fragment library developed by the working group of Prof. Gerhard Klebe (AG Klebe) at the Institute of Pharmaceutical Chemistry (Philipps University Marburg), 34 compounds from a HZB screening set, which was part of the PDBeChem database.

Most of the fragments from the 96-fragment library follow “The Rule of Three”. However, all of the fragments have a molecular weight less than 300 Da. The smallest fragment from the library is J95 (MW= 115 Da) and the biggest one is J24 (MW= 288 Da). The average molecular weight of all the compounds is around 190 Da.

### 5.2.2. Protein Sample

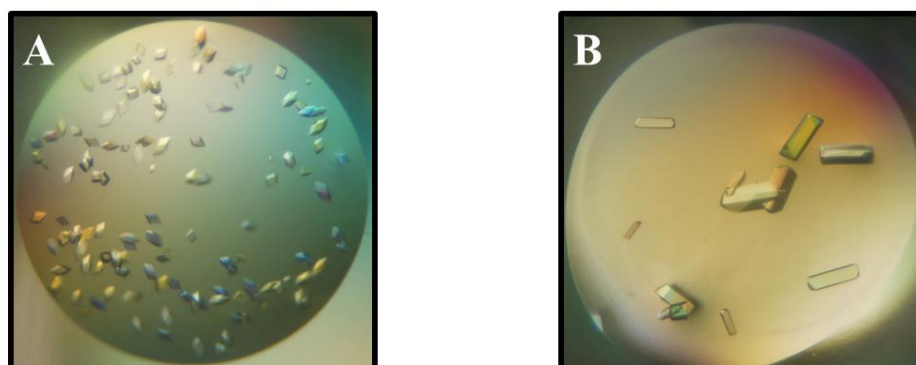
Trypsin from bovine pancreas (EC: 3.4.21.4) is purchased from Sigma Aldrich. The sample comes as essentially salt-free lyophilized powder. Trypsin consists of a single chain polypeptide of 223 amino acid residues (MW= 23.8 kDa), produced by the removal of the N-terminal hexapeptide from trypsinogen which is cleaved at the Lys - Ile peptide bond. The sequence of amino acids is cross-linked by 6 disulfide bridges. Trypsin is a member of the serine protease family. Solid trypsin samples were stored at -20°C.

### 5.2.3. Competitive fragment screening on two crystal forms

This study focuses on two crystal forms with **trigonal** and **orthorhombic** lattice geometry. Both crystal forms were obtained from the same protein sample, but the difference is in the crystallization conditions used to obtain crystals. The fragment screening was done against each crystal form by using the crystallization method of **soaking**.

The trigonal crystal form can be obtained as ligand-free crystals by direct crystallization conditions (see **2.4.1**). These crystals were subjected to a soaking protocol exposing them to a fragment concentration of 100 mM (see **5.2.4.1**). The orthorhombic crystal form is more difficult to obtain, particularly if uncomplexed crystals are desired. Trypsin tends to suffer from autoproteolysis. This can be countered by adding an inhibitor such as benzamidine and crystals of the orthorhombic form are readily formed. However, since benzamidine is a fragment that blocks the active site of trypsin, the fragment has either to be washed out of the formed crystals to apply a soaking protocol. An alternative option will be the displacement of benzamidine by the fragment of interest. This however, requires that, under equilibrium conditions, the probe fragment achieves higher potency. Since benzamidine is a two-digit micromolar inhibitor of trypsin, the displacement affords rather potent ligands exposed at very high concentration. This is difficult to achieve. In a preliminary trial (see **5.2.4.3**), the 96-fragment library was added to the crystallization buffer to obtain the orthorhombic crystal form of trypsin. Only with some fragment, crystals were obtained. Remarkably with fragment **J24** a co-crystal structure was produced, where the active site showed no significant residual electron density indicating a bound ligand. Only in a remote position a bound copy of fragment **J24** could be detected. Since this protocol produced crystals of the orthorhombic form with unoccupied active site, these crystals were subjected to soaking experiments of the 96-fragment library at a fragment concentration of 100 mM.

In addition to these two soaking scenarios, the successful cocrystallization experiments mentioned above with some of the of 96-fragments will be considered in the following. These cocrystallization trials were performed at a fragment concentration of 5 mM, much lower than those used for soaking. However, we have to keep in mind that under cocrystallization conditions the fragment complexes are formed in solution and the complexes then crystallized. Usually, such conditions require adjustment of different concentration than soaking. The latter usually works at higher fragment concentrations.



**Figure 3.** *A. Trypsin crystals with trigonal space group  $P 3_1 1 2$  (space group #152). B. Trypsin crystals with orthorhombic space group  $P 2_1 2_1 2_1$  (space group #19)*

#### 5.2.4. Crystallization protocol

##### 5.2.4.1. Trigonal crystal form (soaking)

Trypsin powder is dissolved in distilled water containing 10 mM  $\text{CaCl}_2$ . The final trypsin concentration is 40 mg/mL in the sample buffer. Sitting and hanging drop plates (Hampton research) are used for growing the crystals. The crystallization buffer (reservoir) contains 200 mM  $(\text{NH}_4)_2\text{SO}_4$ , PEG8000 15% w/v, ethylene glycol 25% w/v and 0.1 M HEPES pH = 7.5. Each well is filled up to 500  $\mu\text{L}$  with the crystallization buffer. A 2  $\mu\text{L}$  drop from the protein sample is placed on the micro-bridge of the sitting drop plate and mixed with an equal volume (2  $\mu\text{L}$ ) of the crystallization buffer. The crystallization buffer drop added should not be mixed with the protein and should be placed directly on top of the protein sample drop (if both drops are not pipetted directly on top of each other, it will lead to a higher crystal population in the contact area and less (or none) crystals in the unmixed area). The higher population of crystals in a certain area of the crystallization drop would let crystals merge together during their growth and require more effort to fish the crystals out of the drop. In 24-48 h crystals with sufficient quality for data collection should be seen in the drop.

In the case of the hanging drop plate, the same procedure as the sitting drop applies. The difference in the hanging drop plate is the sealing of the well, which is sealed by a glass slide.

The glass slide is fixed on top of the well with the crystallization drop facing the reservoir. A medium viscosity hydrophobic grease is applied to the outer rim of the well to fix the glass slide at its base. The crystals were grown at 18°C. In 24-48 h crystals with sufficient quality for data collection should be seen in the drop.

Regarding PEG8000 concentration, using a ratio of 15% w/v would yield medium size crystals. Decreasing the PEG8000 concentration down to 11% w/v would give less crystals in the drop, but of a bigger size. However, it has been noticed that smaller trigonal trypsin crystals showed better results with soaking experiment (the crystals durability in the soaking buffer is better for smaller crystals than for larger ones). Higher crystal count with a smaller size would be possible by increasing the PEG8000 concentration in the crystallization buffer to 18% w/v. This protocol is taken from [215] with minor modifications.

The crystals were fished out of the drop and transferred into a soaking buffer containing the same crystallization condition as the crystals were grown in with an extra 20% glycerol and a concentration of 100 mM of the fragment dissolved in DMSO (obtained from a 1 M stock solution of fragment in DMSO by adding 10% v/v to the soaking solution). Smaller soaking buffer drops make it easier to fish the crystals (drop size = 1-2  $\mu$ L). The crystals are soaked for 24 h in a sitting drop well which should be sealed with tape so no drying occurs. After soaking the crystals were dipped in liquid nitrogen and stored for later data collection.

#### **5.2.4.2. Orthorhombic crystal form (Soaking/co-crystallization)**

Orthorhombic crystals are obtained via co-crystallization first with benzamidine added to the crystallization conditions. Then soaking was carried out on the co-crystallized crystals. The compound in the soaking buffer should displace the benzamidine present in the active site that was used in co-crystallization. Unfortunately, fragments have relatively low affinity to their biological receptor, so a displacement of benzamidine in the active site of trypsin by a fragment in the soaking buffer would require high concentrations of the fragment, as likely, benzamidine used to co-crystallize the orthorhombic trypsin crystals has a higher affinity to the protein.

Trypsin powder is dissolved in a solution containing 10 mM CaCl<sub>2</sub> and 1 mM benzamidine HCl (dissolved in DMSO, final DMSO concentration in the protein sample buffer is 5% v/v). The final trypsin concentration is 40 mg/mL in the protein sample buffer. The protein sample

buffer was left to incubate at room temperature for 20-30 min. Benzamidine is used as an igniter to start crystal growth. The crystallization buffer contains 1.7 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 50 mM MES pH = 6 in distilled water. The crystal plate well is filled with 500 µL of the crystallization buffer. A 2 µL drop of the protein sample buffer is placed on the micro-bridge (Sitting drop plate) or on a glass slide (hanging drop plate) and an additional 2 µL of the crystallization buffer is added to the drop. The sitting drop plates are sealed with transparent tape (Hampton Research - SharkTape). The hanging drop well rims are coated with medium viscosity hydrophobic grease and the glass slide is placed gently on top of it, the drop is facing the interior of the well. Crystals are grown at 18°C, after 24-72 h crystals with sufficient quality for data collection should be seen in the drop.

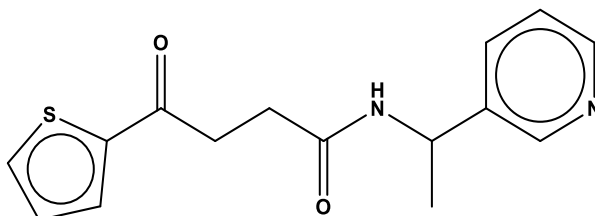
For soaking experiment, the crystals are moved to a soaking buffer containing the same condition as the crystallization buffer with an additional 20% glycerol v/v and a concentration of 100 mM of the fragment dissolved in DMSO, final DMSO concentration in the soaking buffer is 10% v/v.

For co-crystallization experiment using all 96 fragments from the screening sample, a concentration of 5 mM of the fragment was added to the protein sample prior to crystallization (obtained from a 100 mM stock solution of fragment in DMSO by adding 5% v/v to the sample protein buffer). Instead of plunging the crystals in the soaking buffer, they were dipped into the cryo-protectant buffer containing the same crystallization buffer as the crystals are grown in with an additional 20% glycerol v/v and 5 mM of the fragment used in the protein sample buffer dissolved in DMSO. This crystallization protocol was adapted from the one developed by Rauh et al. [216] with minor modifications.

#### **5.2.4.3. Benzamidine displacement**

The affinity of benzamidine to trypsin is fairly high for a small compound of this size ( $K_i = 10^{-7}$  M) [217]. This has led to an issue with the soaking experiment. Many collected data sets from the soaking experiment with different fragments from the 96-fragment library showed the unchanged presence of benzamidine in the S1 binding pocket of trypsin. Fragments with an affinity to trypsin will not bind to the enzyme unless they have higher affinity than benzamidine to replace it in the active site.

As a consequence, the preliminary soaking experiments performed on the orthorhombic crystal form of trypsin, with the idea to displace benzamidine with a fragment from the 96-fragment library was not successful, except fragment **J21** which was indeed able to displace benzamidine from the active site.



**Figure 4.** Fragment **J24** from the 96-fragment library

**J24** is the biggest fragment in the 96-fragment library (MW = 288). It did not bind in the S1 binding pocket but on the surface of the protein making interactions with residues Tyr20 and Lys159. From the electron density seen inside the S1 binding pocket, some conserved waters were absent and a trace of fragment **J24** could be seen, but the electron density was not sufficient to add **J24**. Fortunately, many orthorhombic trypsin crystals could be grown when co-crystallized with fragment **J24** instead of benzamidine, which made it possible to do soaking experiments with the entire fragment library. In due course of this study, a second crystal structure was obtained by co-crystallization with fragment **J24**, and there the fragment was found to bind in the surface depression and in the S1 binding pocket as well (PDB ID: 6ZFK). In this second, the electron density in the active site was sufficient to add **J24** as shown in **Table 1**. (see 5.3.1).

#### 5.2.5. Data collection and processing

For this study, all data were collecting at Helmholtz-Zentrum Berlin (Berlin-Adlershof, Germany) [157] at BESSY II at beam line MX14.1 and MX14.2. The data collection was done under cryogenic temperature at 100 K (-173°C) and a wavelength of 0.9184 Å on a silicon Pilatus 6M pixel detector at atmospheric pressure. Data sets collected were indexed, processed and scaled using XDSAPP [158]. Details of the data collection and processing are given in **Table 2-4** (see 5.3.2).

### 5.2.6. Crystal structure determination and refinement

Crystal structures were determined by molecular replacement using the program PHASER MR [159] from the CCP4 suite [160]. The structure of trypsin was used as a search model for molecular replacement (PDB ID: 5MNK) after removing all non-protein molecules and hydrogen atoms using FCONV [161]. During refinement, a 5% subset of the reflections was used for the  $R_{\text{free}}$  calculation and they were consequently omitted from the refinement. The model was built in COOT [162] and the refinement was done using PHENIX.refine version 1.16-3549 [163]. The restraints of fragments were generated using the Grade Web Server [164] or by eLBOW from PHENIX [163] which were also used for energetically minimization and restraint generation. A first refinement step was performed with default parameters, XYZ coordinates, occupancies and individual B-factors were alternated with structural adaption performed in COOT [162]. A series of repeated refinement steps were applied until the electron density was well explained by the built model. All crystal structures were anisotropically refined using PHENIX.refine [163]. Water molecules in all data sets were refined (isotropically refined for resolution  $\geq 1.31$  Å and anisotropically refined for resolution  $< 1.30$  Å) and hydrogen atoms were included in the final model at the calculated positions. Refinement statistics are included in **Tables 2-4** (see **5.3.2**).

### 5.2.7. Volume calculation of the binding pocket

The volume of the binding pocket was calculated using FCONV [161], a program devolved by Gert Neudert. The first step in this program is to identify the residues surrounding the S1 binding pocket. In the case of trypsin, the protein is studied as a monomer. Fragment **J2** is bound in every structure of trypsin obtained by the three different protocols (trigonal/soaking, orthorhombic/soaking, orthorhombic/co-crystallization). These three structures have been used as input for the binding pocket volume calculation. All amino acids clustered for this study should have an interaction with fragment **J2** in the binding pocket or located in the cavity of the active site. Before starting the calculation, water molecules and hydrogen atoms were removed from the PDB files. All unbound molecules (not amino acids from trypsin) except fragment **J2** were deleted (sulfate, DMSO, ethylene glycol and calcium atoms). The first run was focused on the area surrounding **J2** (diameter 10 Å) and fragment **J2** deleted. The amino

acids that are used in this study to calculate the volume of the active site cavity are clustered into three continuous chains and few other adjacent residues: Val17, His57, Leu99, Asp102, Gly142 – Asn143, Gly188 – Ser195, Gly211 – Tyr228.

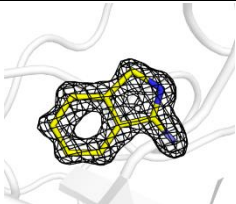
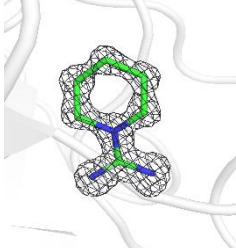
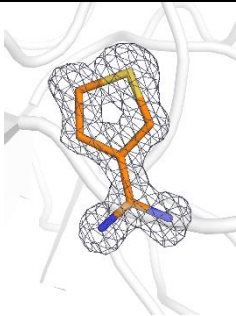
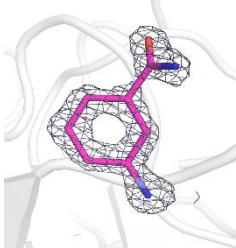


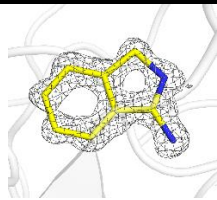
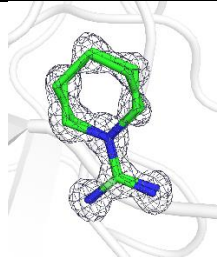

## 5.3. Results and discussion

### 5.3.1. The 96-fragment library hits

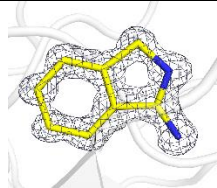
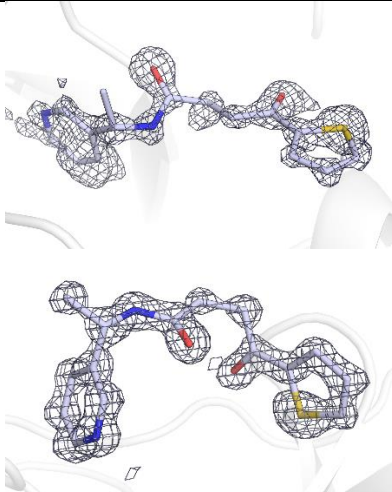
In this work, crystal structures of the bovine trypsin (pancreatic trypsin) in complex with fragments from the 96-fragment library were determined. The structures were identified using X-ray crystallography. The data ranged from 0.98-1.6 Å in resolution.

**Table 1.** Hits obtained from the 96-fragment library screening against trypsin

Trigonal / Soaking		
Fragment J2	PDB ID: 6YIY	
Fragment J21	PDB ID: 6YIS	
Fragment J29	PDB ID: 6YIT	
Fragment J85	PDB ID: 6YIU	
Orthorhombic / Soaking		

J2	PDB ID: 6YIW	
J21	PDB ID: 6YIV	
J29	PDB ID: 6YIX	

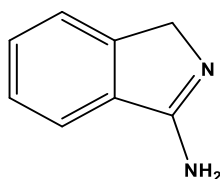
### Orthorhombic / Co-crystallization

Fragment J2	PDB ID: 6ZFJ	
Fragment J24	PDB ID: 6ZFK	

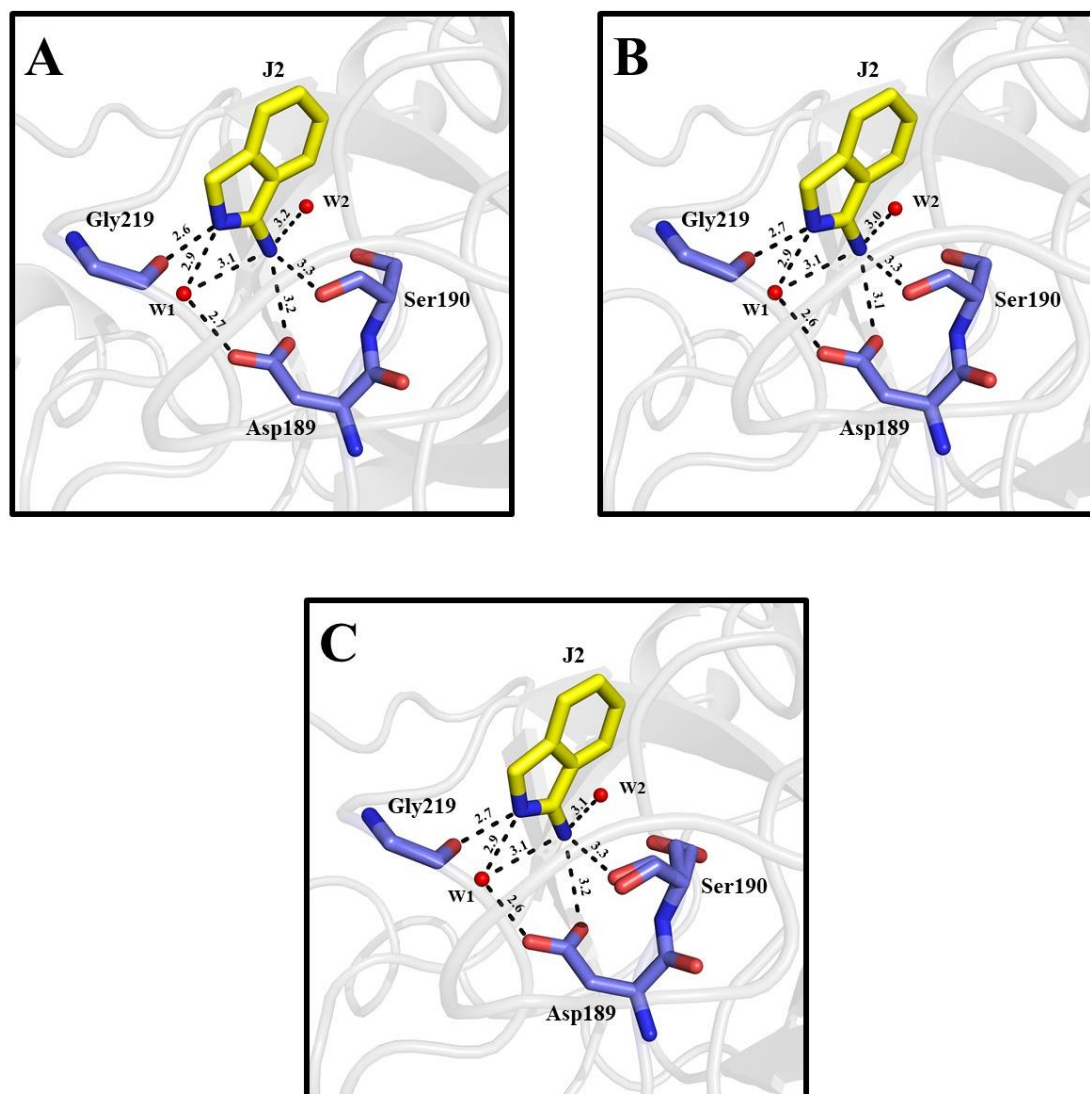
The electron density around each fragment represented in grey mesh is the mFo<sub>-</sub>DFc map at 3.0  $\sigma$  level.

The hits obtained from the 96-fragment library with trypsin are summarized in **Table 1**. Each hit will be discussed in detail in the following.

## 5.3.1.1. Fragment J2



**Figure 5.** Fragment J2 from the 96-fragment library

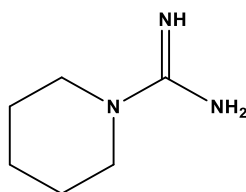


**Figure 6.** *A.* Crystal structure of cationic trypsin (trigonal space group) in complex with fragment J2 (soaking) (PDB ID: 6YIY). *B.* Crystal structure of cationic trypsin (orthorhombic space group) in complex with fragment J2 (soaking) (PDB ID: 6YIW). *C.* Crystal structure of cationic trypsin (orthorhombic space group) in complex with fragment J2 (co-crystallization) (PDB ID: 6ZFJ). All distances are given in Å.

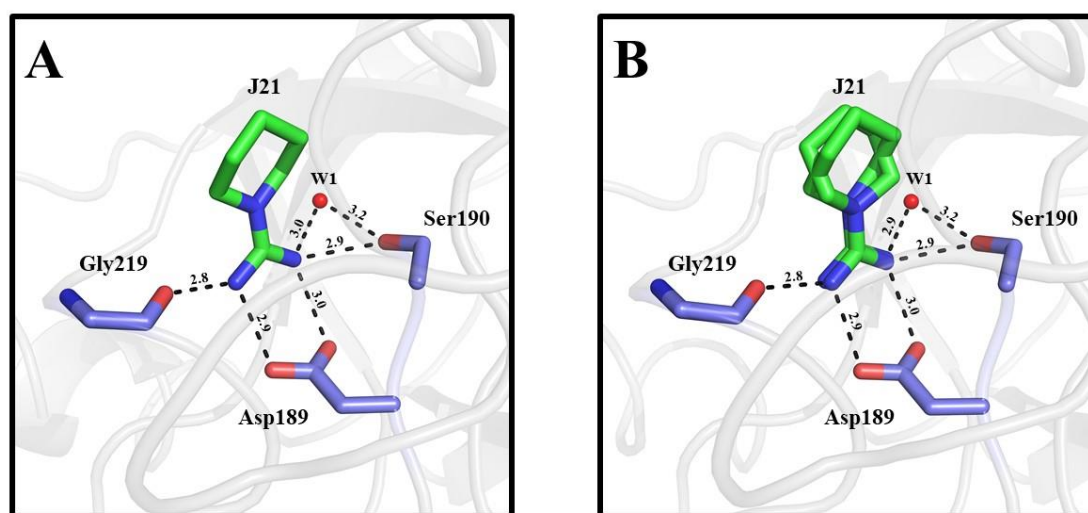
Fragment **J2** (3H-isoindol-1-yl amine, MW = 132 Da) binds to trypsin in both crystal forms (trigonal and orthorhombic) and is obtained following both crystallization protocols (soaking and co-crystallization) in the orthorhombic crystal form. In all trypsin crystal structures in complex with **J2**, the fragment binds in the S1 specificity pocket (**Figure 6**). With respect to for the binding mode, there are no differences in the three structures that could be linked to the interactions formed between trypsin's amino acids in the S1 pocket and fragment **J2**. The interactions with **J2** are conserved for all structures (**Figure 6A, B, and C**), in which the amino group is interacting with two water molecules (W1 and W2), Ser190 and Asp189. In native trypsin, residue Asp189 is located at the bottom of the primary substrate binding pocket and interacts with the arginine and lysine side chain of the substrate peptide chain. Therefore, also in case of fragments, a basic functional group is favoured.

Soaking of small fragments is different from soaking a highly potent ligand. The fragment concentration in the soaking buffer is 100 mM. Occasionally, soaking for long times at such high concentrations (depending on the fragment influence on the crystals in the soaking buffer) can have a damaging effect on the crystal quality and integrity. The crystals which were soaked in buffers containing 100 mM of **J2** could survive for at least 72 h, which is the longest duration a trypsin crystal has survival in the soaking buffer considering the present screening of the 96-fragment library.

### 5.3.1.2. Fragment J21



**Figure 7.** Fragment J21 from the 96-fragment library



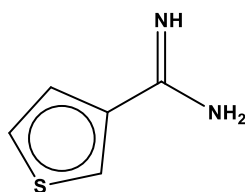
**Figure 8.** **A.** Crystal structure of cationic trypsin (trigonal space group) in complex with fragment J21 (soaking) (PDB ID: 6YIS). **B.** Crystal structure of cationic trypsin (orthorhombic space group) in complex with fragment J21 (soaking) (PDB ID: 6YIV). All distances are given in Å.

Fragment **J21** (N-amidinopiperidine, MW = 127 Da) binds in the S1 pocket of trypsin in both crystal forms via soaking protocol (**Figure 8**). Unfortunately, no crystal structure of trypsin in complex with fragment **J21** could be obtained by co-crystallization in orthorhombic trypsin crystal form. The protein tends to precipitate in the crystallization drop under the applied conditions.

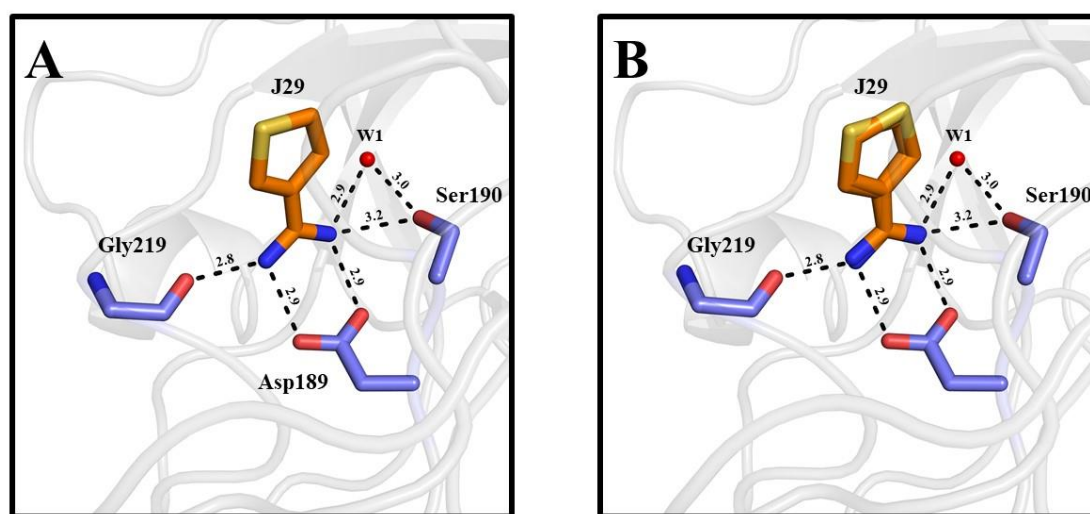
As discussed earlier in the protein crystallization procedure, benzamidine was first used to obtain orthorhombic trypsin crystals for soaking which was later exchanged by fragment **J24** (see **5.2.4.3**). Fragment **J21** was the only fragment in the soaking experiment with

orthorhombic trypsin crystals that was able to replace benzamidine. In a previous study by Schiebel et al [215] the affinities of benzamidine and fragment **J21** (N- amidinopiperidine) to trypsin were determined and in solution benzamidine is about a factor of 10 more potent. However, it does not mean that the solution affinities must correspond exactly to the affinity against a crystal form. However, the initial concentration of benzamidine used to obtain orthorhombic trypsin crystal is 1 mM using cocrystallization protocol, while **J21** was introduced to the orthorhombic crystal using soaking protocol with concentration of 100 mM. This can compensate for a possibly slightly lower binding affinity of **J21**.

### 5.3.1.3. Fragment J29



**Figure 9.** Fragment J29 from the 96-fragment library



**Figure 10.** **A.** Crystal structure of cationic trypsin (trigonal space group) in complex with fragment J29 (soaking) (PDB ID: 6YIT). **B.** Crystal structure of cationic trypsin (orthorhombic space group) in complex with fragment J29 (soaking) (PDB ID: 6YIX). All distances are given in Å.

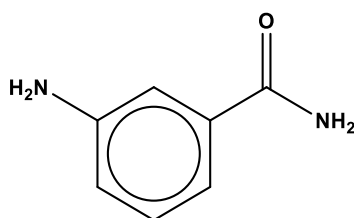
Fragment **J29** (thiophene-3-amidine, MW = 126 Da) binds into the S1 pocket of trypsin in the trigonal and orthorhombic trypsin crystal form (**Figure 10**). However, no crystal structure in complex with fragment **J29** could be obtained with the orthorhombic trypsin crystals using the co-crystallization protocol.

An anomalous map was generated to validate the presence of the ligand in the S1 pocket (**Figure 18**). An anomalous signal for the sulfur was visible at 3.0  $\sigma$  contour level in the orthorhombic crystal structure, while in the trigonal crystal structure, it was a weaker signal at 3.0  $\sigma$  contour level. From the refinement steps that were done for the trypsin structures in complex with **J29**, a second alternative conformation of **J29** with lower occupancy (<11%) is

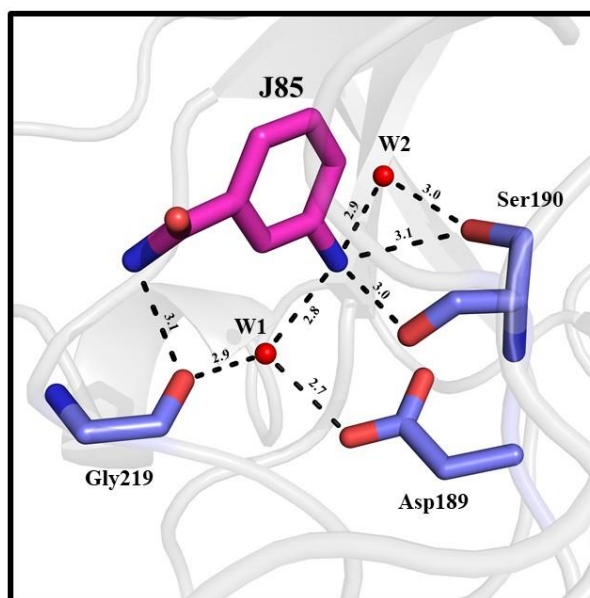
indicated in the trigonal crystal form, while in the orthorhombic crystal form the occupancy of the alternative conformation is 25% and the electron density was well explained (PDB ID: 6YIX).

The interactions and the positioning of **J29** in the S1 pocket is similar to the one of **J21** and benzamidine as well. However, fragment **J29** was able to replace **J24** in the S1 pocket, but it was not able to displace benzamidine out of the pocket, which supposedly indicates that the affinity of **J29** is lower than the other two compounds. The alternative conformations show the thiophen sulfur either oriented to Gly219 (75% occupancy) or to Ser190 (25% occupancy).

#### 5.3.1.4. Fragment J85



**Figure 11.** Fragment J85 from the 96-fragment library

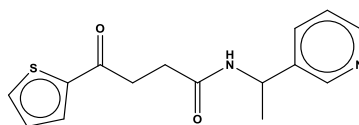


**Figure 12.** Crystal structure of cationic trypsin (trigonal space group) in complex with fragment J85 (PDB ID: 6YIU). All distances are given in Å.

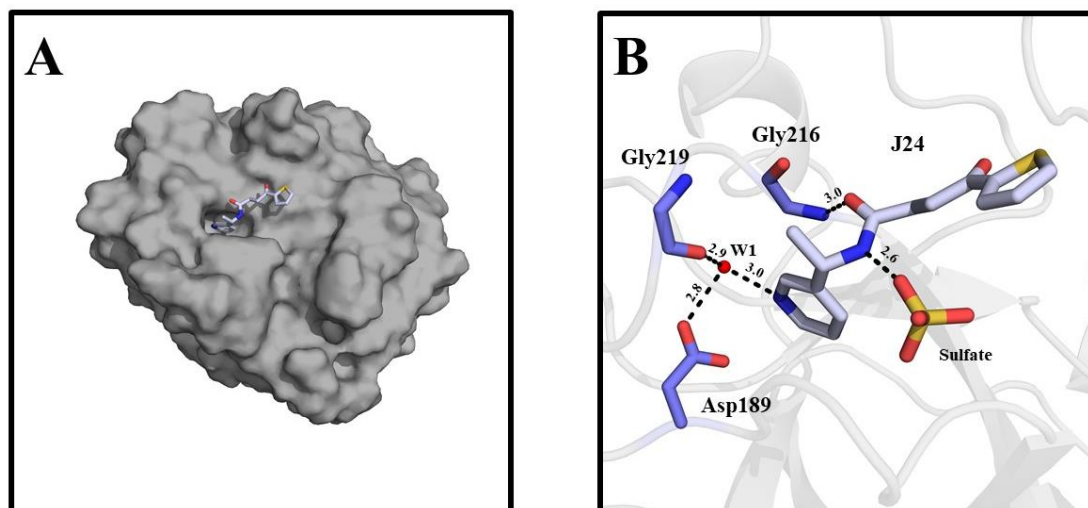


Fragment **J85** (3-aminobenzamide, MW = 136 Da) binds into the S1 pocket of trypsin in trigonal trypsin crystal form only (**Figure 12**). Unfortunately, no crystal structure in complex with fragment **J85** could be obtained with the orthorhombic trypsin crystal form. **J85** does not have a direct interaction with Asp189 at the bottom of the S1 specificity pocket, instead, a water molecule (W1) interacts with both **J85** and Asp189 (**Figure 12**). In the case of the structures in complex with fragments **J21** and **J29**, the amino group has a direct interaction with Asp189 (**Figure 8** and **10**) at the bottom of the S1 pocket. In the **J85** case, the amino group is interacting with Gly219; thus, fixing **J85** away from Asp189. This is also observed in the crystal structures in complex with fragment **J2** (**Figure 6**), where the indole is interacting with Gly219 and the amino group of the carboxamide is interacting with residue Asp189. Even with similar functionality, the structural differences of a fragment (the position of the functionalities at the ring in meta, ortho or para position) plays a key role in the positioning of the fragments in the active site, which could be an argument for not having a direct interaction between the amino group of fragment **J85** and Asp189 at the bottom of the S1 pocket.

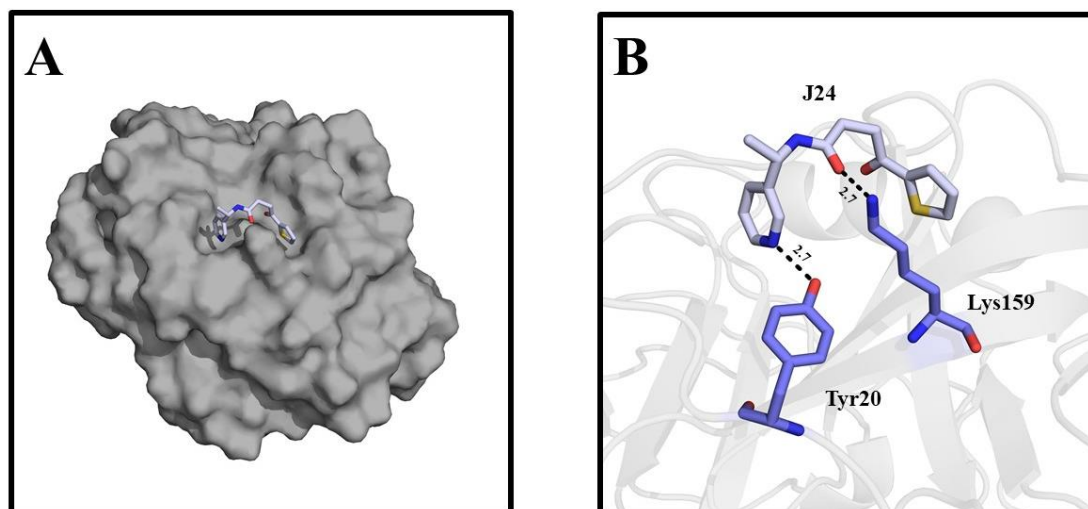
## 5.3.1.5. Fragment J24



**Figure 13.** Fragment J24 from the 96-fragment library



**Figure 14.** Crystal structure of cationic trypsin (orthorhombic space group) in complex with fragment J24 (co-crystallization) (PDB ID: 6ZFK). **A.** Surface representation of the enzyme showing fragment J24 located in the S1 binding pocket. **B.** Closer view at fragment J24 interactions with the enzyme. All distances are given in Å.



**Figure 15.** Crystal structure of cationic trypsin (orthorhombic space group) in complex with fragment J24 (co-crystallization) (PDB ID: 6ZFK). **A.** Surface representation of the enzyme showing fragment J24 binding on the surface at the opposite side of the S1 binding pocket. **B.** Closer view at fragment J24 interactions on the surface of the enzyme. All distances are given in Å.

Fragment **J24** (MW = 288 Da) is the fragment from the 96-fragment library with the highest molecular weight. The first crystal structure obtained by co-crystallization of the orthorhombic trypsin crystals proved to be useful for soaking experiments using the orthorhombic crystal form (instead of benzamidine). Although **J24** binds in the S1 binding pocket (**Figure 14**), in the first obtained co-crystallized structure of the protein with **J24**, it did not bind to the S1 pocket. Surprisingly, in that structure, fragment **J24** was binding only on the surface of the protein at a site opposite of the S1 pocket (**Figure 15**). There was some electron density indicated in the binding pocket, but not sufficient to assign **J24**. Obviously, **J24** has some weak affinity to bind to the S1 pocket of trypsin. After preparing a new batch of crystals, the binding of **J24** in the S1 pocket could be confirmed (PDB ID: 6ZFK). Overall, there were only three hits with S1 binders obtained from the fragment screening of the orthorhombic trypsin crystal form by soaking into crystals obtained from cocrystallization. Nevertheless, all of the data sets collected from the soaked crystals obtained by that protocol did not show any trace of **J24** binding on the surface of the protein (as evaluated from the electron density). Although **J24** has some affinity to bind trypsin, it cannot be very strong as the fragment must have been washed out by simply transferring the crystals into the soaking buffer not containing **J24** apart from the additives.

There is no direct interaction in the S1 pocket between Asp189 and the nitrogen of the pyridine ring of fragment **J24**, but both interact with a water molecule (W1) (**Figure 14B**). The sulfate ion seen in **Figure 14B** is present near the entrance of the S1 pocket in almost every structure of cationic trypsin, but it establishes no interaction to any fragment from the library, except to **J24**.

Regarding the hit on the surface of the protein (**Figure 15**), there is a depression on the surface where Tyr20 and Lys159 are present. In other crystal structures, only water molecules are present in this area and it is one of the narrowings along the water channels in this trypsin crystal form. The distance between the amino group of Lys159 and the carbonyl groups of Ser96 and Asn97 of the adjacent crystal mate by symmetry is 3.0 Å and 2.9 Å, respectively.

Water channels in the crystal are the bottlenecks that the compound has to travel through during soaking. Due to the narrow channels in this area, **J24** accumulated there and was easily detected at that position, even in the cocrystallized structure which was washed by the transfer to the cryo buffer. Possibly the exchange is so strongly slowed down that the electron density of the

ligand is still visible in the crystals at that position whereas the wash-out from the S1 pocket is more easily achieved.

### **5.3.2. Crystallographic tables**

**Table 2.** Data collection and refinement statistics for trigonal trypsin crystal structures obtained through soaking

PDB code	6YIY (J2)	6YIS (J21)	6YIT (J29)	6YIU (J85)
<b>Data collection and processing</b>				
Space group	P 31 2 1	P 31 2 1	P 31 2 1	P 31 2 1
Unit cell	54.6 54.6 107.5	54.6 54.6 107.5	54.6 54.6 107.5	54.6 54.6 107.5
Matthews coefficient a	1.8	1.8	1.8	1.8
Solvent content (%)	32.3	31.4	31.5	31.7
<b>Diffraction data<sup>b</sup></b>				
Resolution range	50-1.11 (1.18-1.11)	50-1.19 (1.26-1.19)	50-1.25 (1.33-1.25)	50-1.36 (1.44-1.36)
Unique reflections	73703 (11641)	58894 (9206)	52150 (8335)	40740 (6477)
R(I)sym (%) <sup>c</sup>	5.8(58.6)	5.5(58.0)	6.0(59.9)	4.7(57.9)
Wilson B factor	12.0	12.1	13.0	14.8
completeness	99.3(98.2)	97.7(95.8)	99.8(99.7)	99.7(99.3)
redundancy	6.4(6.1)	6.6(6.4)	6.5(6.5)	6.4(6.0)
<I/σ(I)>	15.9(2.9)	17.4(2.9)	16.6(3.0)	21.4(2.9)
<b>Refinement</b>				
Resolution range	47.29-1.11	47.25-1.19	43.29- 1.25	43.34-1.36
Reflections used in refinement (work/free)	73703 (70017/3686)	58894 (55949/2945)	52150 (49542/2608)	40740 (38703/2037)
Final R value for all reflections (work d /free e) (%)	14.0/15.7	15.0/17.9	14.0/16.3	14.3/16.9
Protein residues	223	223	223	223
Water molecules	193	156	171	148
rmsd from ideality: bond length (Å)	0.006	0.006	0.006	0.006
rmsd from ideality: bond angle (°)	0.93	0.92	0.95	0.90
Ramachandran most favored(%) <sup>f</sup>	99.1	98.2	99.1	98.2
Ramachandran additionally allowed (%) <sup>f</sup>	0.9	1.8	0.9	1.8
Mean B factor protein (Å <sup>2</sup> ) <sup>g</sup>	15.1	16.0	16.2	17.8
Mean B factor (ligand) (Å <sup>2</sup> ) <sup>g</sup>	17.1	17.2	16.1	21.7
Mean B factor water molecules (Å <sup>2</sup> ) <sup>g</sup>	25.1	24.6	25.7	27.0

<sup>a</sup> Calculated with Matthews\_coef program from CCP4 suite version 6.4.0. <sup>b</sup> Values in parenthesis describe the highest resolution shell. <sup>c</sup>  $R(I)_{\text{sym}} = [\sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)] \times 100$ , in which  $I(h)$  is the mean of the  $I(h)$  observation of reflection  $h$ . <sup>d</sup>  $R_{\text{work}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$ . <sup>e</sup>  $R_{\text{free}}$  was calculated as shown for  $R_{\text{work}}$  but on refinement-excluded 5% of data. <sup>f</sup> Calculated with PROCHECK [167]. <sup>g</sup> Mean B factors were calculated with MOLEMAN [168].

**Table 3.** Data collection and refinement statistics for orthorhombic trypsin crystal structures obtained through soaking

PDB code	6YIW (J2)	6YIV (J21)	6YIX (J29)
<b>Data collection and processing</b>			
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	54.2 58.1 66.8	54.2 58.1 66.8	54.2 58.1 66.8
Matthews coefficient <sup>a</sup>	2.0	2.0	2.0
Solvent content (%)	40.1	39.9	39.8
<b>Diffraction data</b> <sup>b</sup>			
Resolution range	50-0.97 (1.03-0.97)	50-0.95 (1.01-0.95)	50-1.31 (1.39-1.31)
Unique reflections	125036 (19736)	130966 (19891)	50997 (8098)
R(I) <sub>sym</sub> (%) <sup>c</sup>	6.7(48.0)	2.5(22.5)	4.7(49.4)
Wilson B factor	7.5	7.5	7.5
completeness	99.5(98.1)	97.7(92.6)	98.9(98.3)
redundancy	6.4(4.8)	4.6(3.6)	4.6(4.7)
<I/σ(I)>	12.0(2.2)	29.4(4.7)	19.0(2.3)
<b>Refinement</b>			
Resolution range	42.13-0.97	29.06-0.95	43.29- 1.31
Reflections used in refinement (work/free)	125036 (118784/6252)	130966 (124417/6549)	50997 (48447/2550)
Final R value for all reflections (work <sup>d</sup> /free <sup>e</sup> ) (%)	13.0/14.2	12.4/13.4	17.6/19.3
Protein residues	223	223	223
Water molecules	287	303	215
rmsd from ideality: bond length (Å)	0.006	0.006	0.006
rmsd from ideality: bond angle (°)	0.97	0.97	0.96
Ramachandran most favored(%) <sup>f</sup>	99.1	98.6	99.1
Ramachandran additionally allowed (%) <sup>f</sup>	0.9	1.4	0.9
Mean B factor protein (Å <sup>2</sup> ) <sup>g</sup>	10.2	9.0	13.5
Mean B factor (ligand) (Å <sup>2</sup> ) <sup>g</sup>	10.7	7.0	13.7
Mean B factor water molecules (Å <sup>2</sup> ) <sup>g</sup>	20.7	19.4	22.9

<sup>a</sup> Calculated with Matthews\_coef program from CCP4 suite version 6.4.0. <sup>b</sup> Values in parenthesis describe the highest resolution shell. <sup>c</sup>  $R(I)_{\text{sym}} = [\sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)] \times 100$ , in which  $I(h)$  is the mean of the  $I(h)$  observation of reflection  $h$ . <sup>d</sup>  $R_{\text{work}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$ . <sup>e</sup>  $R_{\text{free}}$  was calculated as shown for  $R_{\text{work}}$  but on refinement-excluded 5% of data. <sup>f</sup> Calculated with PROCHECK [167]. <sup>g</sup> Mean B factors were calculated with MOLEMAN [168].

**Table 4.** Data collection and refinement statistics for orthorhombic trypsin crystal structures obtained through co-crystallization

PDB code	6ZFI (J2)	6ZFK (J24)
<b><u>Data collection and processing</u></b>		
Space group	P 21 21 21	P 21 21 21
Unit cell	54.3 58.2 66.8	54.7 58.2 67.0
Matthews coefficient <sup>a</sup>	2.3	2.2
Solvent content (%)	45.3	45.6
<b><u>Diffraction data</u></b> <sup>b</sup>		
Resolution range	50-1.00 (1.06-1.00)	50-1.10 (1.17-1.10)
Unique reflections	113783 (18318)	84897 (12117)
R(I) <sub>sym</sub> (%) <sup>c</sup>	6.1(40.0)	6.7(47.2)
Wilson B factor	6.1	8.6
completeness	99.1(99.5)	97.5(86.9)
redundancy	5.3(5.0)	5.1(3.4)
<I/σ(I)>	13.7(8.3)	14.6(2.6)
<b><u>Refinement</u></b>		
Resolution range	43.90-1.00	43.94-1.10
Reflections used in refinement (work/free)	113783 (108093/5690)	84897 (80652/4245)
Final R value for all reflections (work <sup>d</sup> /free <sup>e</sup> ) (%)	11.3/12.6	12.5/13.8
Protein residues	223	223
Water molecules	295	244
rmsd from ideality: bond length (Å)	0.006	0.006
rmsd from ideality: bond angle (°)	1.02	0.94
Ramachandran most favored(%) <sup>f</sup>	98.2	98.2
Ramachandran additionally allowed (%) <sup>f</sup>	1.8	1.8
Mean B factor protein (Å <sup>2</sup> ) <sup>g</sup>	6.6	9.3
Mean B factor (ligand) (Å <sup>2</sup> ) <sup>g</sup>	8.2	16.4
Mean B factor water molecules (Å <sup>2</sup> ) <sup>g</sup>	20.3	24.2

<sup>a</sup> Calculated with Matthews\_coef program from CCP4 suite version 6.4.0. <sup>b</sup> Values in parenthesis describe the highest resolution shell. <sup>c</sup>  $R(I)_{\text{sym}} = [\sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)] \times 100$ , in which  $I(h)$  is the mean of the  $I(h)$  observation of reflection  $h$ . <sup>d</sup>  $R_{\text{work}} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$ . <sup>e</sup>  $R_{\text{free}}$  was calculated as shown for  $R_{\text{work}}$  but on refinement-excluded 5% of data. <sup>f</sup> Calculated with PROCHECK [167]. <sup>g</sup> Mean B factors were calculated with MOLEMAN [168].

### 5.3.3. Volume of the binding pocket

A crystal structure was obtained in complex with fragment **J2** with both space groups (trigonal and orthorhombic) and with both crystallization methods in orthorhombic trypsin crystal structures (soaking and co-crystallization). The volume of the binding pocket was calculated for each structure in complex with **J2** using FCONV program [161].

**Table 5.** *Volume of the binding pocket in trypsin in complex with fragment J2*

PDB ID	Space group	Method of fragment delivery	Volume of the binding pocket (Å <sup>3</sup> ) <sup>a</sup>
6YIY	Trigonal	Soaking	123
6YIW	Orthorhombic	Soaking	138
6ZFJ	Orthorhombic	Co-crystallization	144

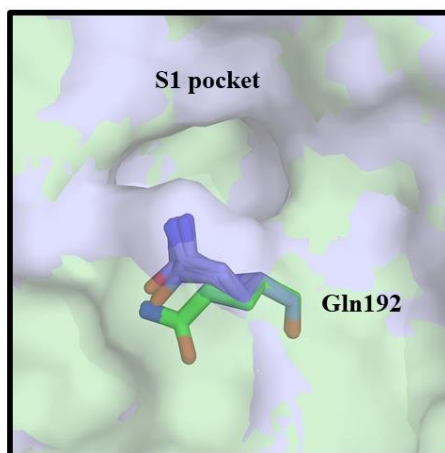
<sup>a</sup> The volume of the binding pocket is measured using FCONV [161].

The results from the binding pocket volume calculations listed in **Table 5** show a slight difference in the volume of the binding pocket between trypsin structures in the different crystal forms. As the volume difference is low, this unlikely serves as an explanation for the differentiation in the number of the hits obtained from the fragment screens with trigonal and orthorhombic trypsin crystals (one hit less from the screening performed on orthorhombic crystal form using soaking protocol compared to the trigonal crystal form). The residues of the S1 pocket where fragment **J2** binds exhibit no difference in their positioning, except Gln192 at the entrance of the S1 pocket, when superimposing the corresponding residues (**Figure 16**).

From each crystal form, data sets have been refined and the positioning of Gln192 is fixed according to the electron density. As shown in **Figure 16**, in the trigonal crystal form, about 80% of the data sets collected from the 96-fragment library screening show that Gln192 is pointing toward the entrance of the binding pocket. However, about 70% of the collected data sets from the 96-fragment library on the orthorhombic crystal form show that Gln192 is pointing away from the entrance of the binding pocket. That was the case in crystal structures in complex with fragment **J2**. Residue Gln192 is involved in the binding pocket calculation



because it is part of the amino acids forming the S1 cavity and it has interactions to residues that are interacting with fragment **J2**.



**Figure 16.** Superimposition of trigonal crystal structures (blue) and orthorhombic crystal structures (green). Residue Gln192 at the entrance of the S1 pocket is shown as sticks.

In total 40 superimposed structures are shown in **Figure 16**, 20 structures from each crystal form (trigonal and orthorhombic). The dent covering a small portion from the entrance of the binding pocket in the trigonal trypsin crystal structure, which is made by residue Gln192 shifting toward the entrance of the S1 pocket that may restrict fragments entering the active site. This slight shift of residue Gln192 has created a small difference in the value of the binding pocket volume. However, for other crystal structures that do not show the same positioning of Gln192 in both trigonal and orthorhombic structures, Gln192 is either seen in alternative conformation, insufficient electron density for Gln192 to be assigned or Gln192 is pointing in the other direction unlike the majority of the data sets collected from the screening for each crystal form.

According to the data obtained from calculating the binding pocket volume and the superimposition of multiple crystal structures from both crystal forms, orthorhombic crystal structures should have better tendency to bind fragments due to the slightly bigger size of the S1 cavity. However, the fragment screening against trigonal trypsin crystals gave more hits (four hits) compared to the orthorhombic crystals (three hits). Since the volume of the S1 binding pocket is slightly bigger for the orthorhombic crystal structures the outcome was not

expected, especially for the soaking method, in which the fragment concentration used is the same (**100 mM**).

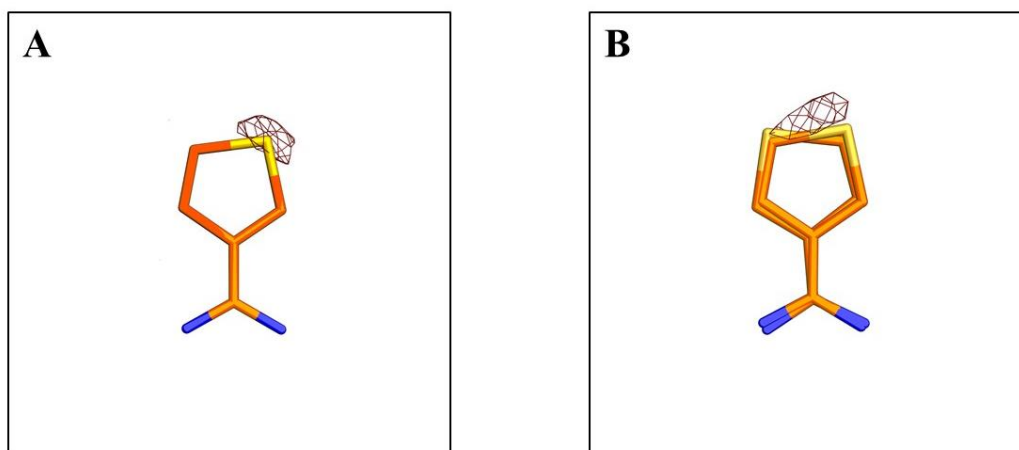
By observing different trypsin crystal structures found in the Protein Data Bank, the positioning of residue Gln192 closer or further from the entrance of the active site is not a characteristic of a space group (trigonal or orthorhombic). It was simply the case in the collected data sets presented in this study.

#### 5.3.4. Anomalous map

Anomalous map calculation is a method to identify anomalous scattering elements in a structure by electron density maps. An anomalous map is supportive to improve the model and to investigate the presence of any anomalous scattering in the model. In this study, the anomalous maps are created to investigate the presence of sulfur.

A python script used by our working group was developed by Dr. Alexander Metz. The script requires a pdb file (fully refined structure), mtz file and XDS\_ASCII.HKL file from processing with XDSAPP [158] in which the Friedel's law is set to false. The output file contains an anomalous map and anomalous map peaks.

The reason for using an anomalous map is the different conformation noticed for fragment **J29** in the orthorhombic space group crystal structure and the absence or at least lower population of the alternative conformation in the trigonal space group crystal structure. An anomalous map was created for all structures in complex with fragment **J29** to check the positioning of the sulfur (**Figure 17**).



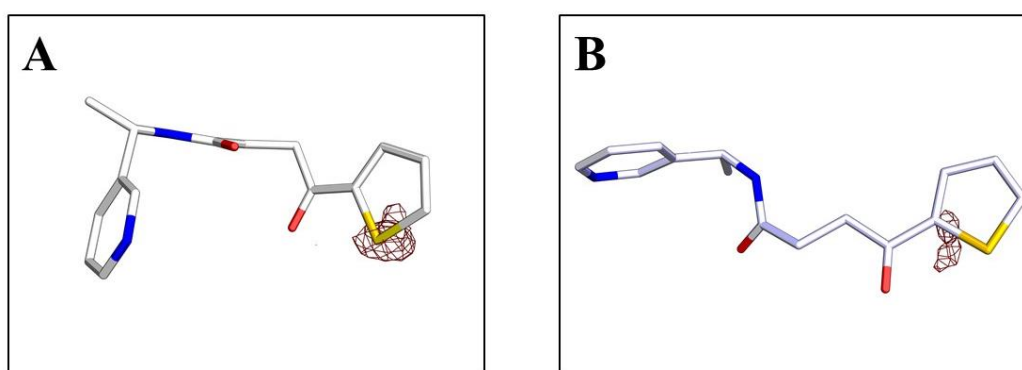
**Figure 17.** A stick representation of fragment J29 (orange) and the anomalous map peak (crimson red mesh). **A.** Anomalous peak of the trigonal trypsin crystal structure (soaking) in complex with J29 (PDB ID: 6YZA). **B.** Anomalous peak of the orthorhombic trypsin crystal structure (soaking) in complex with J29 (PDB ID: 6YIX). All the peaks from the anomalous map are displayed at  $3.0\sigma$ .

The anomalous signal for sulfur is observed for both crystal structures in complex with fragment **J29** (**Figure 17**). A difference is seen in the orthorhombic crystal structure obtained through soaking (**Figure 17B**). The signal is extended compared to the trigonal crystal structure. During refinement, fragment **J29** is added after explaining all possible electron density surrounding the S1 pocket and all around the structure. The occupancy for each fragment added is fixed at 50% before starting the refinement. Then, the occupancy is refined using refine.Phenix [163]. The occupancy of **J29** refined to 80% and the m2Fo-DFc map does not show any negative signal after the occupancy refinement, but in the orthorhombic crystal structure the occupancy was refined to 60%. After refinement with a negative signal surrounding the sulfur. After adding the alternative conformation, the occupancy was refined to 75% and 25% for each conformation of fragment **J29**.

The output obtained by creating an anomalous map is highly dependent on the quality of the data set, the resolution and the wavelength used for collecting the data set. All of the trypsin structures in complex with **J29** are collected at the same beamline under identical conditions. Furthermore, the trigonal crystal structure obtained by soaking and the orthorhombic crystal structure obtained by soaking are in the same resolution range 1.25 and 1.31Å, respectively. Although, the alternative conformation is present in the orthorhombic crystal structure, the

sulfur atom in the alternative conformation is not forming any interactions with the enzyme amino acids or the solvent molecules surrounding it.

Another reason for investigating the peaks in the anomalous map is the co-crystallized orthorhombic crystal structure in complex with fragment **J24** (PDB ID: 6ZFK). The first try to obtain a crystal structure in complex with fragment **J24** was successful. However, fragment **J24** was not binding in the S1 binding pocket but on the surface of the enzyme. The absence and rearrangement of some conserved water molecules (compared to the apo structure) in the S1 pocket indicates that **J24** could bind in the S1 pocket. The trypsin crystals were co-crystallized again with **J24** and this time it could be seen on the surface and in the S1 pocket as well. Fortunately, **J24** contains a sulfur atom to compare both binding fragments (surface and S1 pocket) in the crystal structure by creating an anomalous map.



**Figure 18.** A stick representation of fragment **J24** and the anomalous signal as crimson red mesh verifying the presence of sulfur in yellow (PDB ID: 6ZFK). **A.** Fragment **J24** binding at the surface of the binding pocket. **B.** Fragment **J24** binding in the S1 pocket. Peaks from the anomalous map are displayed at 3.0  $\sigma$  level.

Fragment **J24** is seen binding on the surface of the enzyme of every trypsin crystal co-crystallized with **J24**, but it is easily washed away by soaking the crystal in buffer not containing any fragment. Washed orthorhombic crystals showed water molecules in the position of the fragment binding on the surface in their structures, which makes these crystals the starting point to do soaking experiments. However, not every crystal structure co-crystallized with fragment **J24** showed the binding of the fragment in the S1 pocket of the structure. Even when trypsin crystals in complex with fragment **J24** were obtained under

identical conditions, some structure showed electron density of **J24** in the active site while others did not and the reason for that is unclear. In the orthorhombic crystal form cocrystallized with **J24** (PDB ID: 6ZFk), fragment **J24** binding on the surface has an occupancy of 100% while the one binding in the S1 pocket has an occupancy of 78%. That would explain the weaker anomalous signal seen for the sulfur atom for the fragment binding in the S1 pocket (**Figure 18B**) compared to the fragment binding on the surface (**Figure 18A**).

### 5.3.5. Water channels

Map-Channels [218] is a computational tool that facilitates visualization and characterization of solvent channels or pores within macromolecular crystals. Mapping the shortest distance to protein surfaces is calculated on a grid embedding the unit cell. The grid values are written to a map file. Multiple map files were selected from both packing of trypsin crystal forms (trigonal and orthorhombic) to visualize and calculate multiple positions in the water channels of the crystal structures. The results showed a very similar fit of all the generated water channel maps with no significant differences to be observed. The area around fragment **J24** that binds on the surface was most intriguing. However, no significant changes or modifications in the diameter of the water channels in that area could be observed.

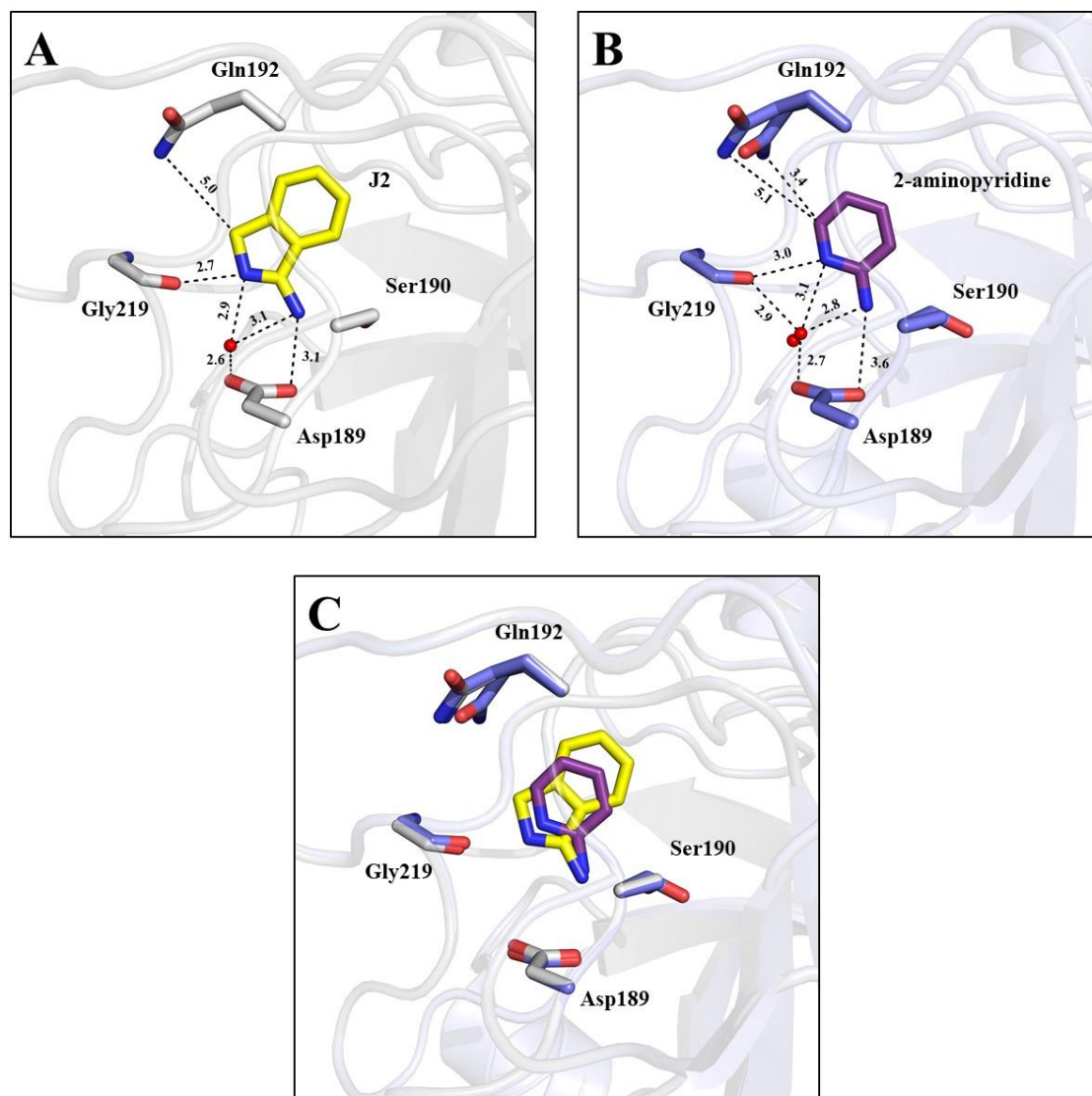
Using Map-Channels, the area in the water channels with the minimum diameter (bottleneck) could be identified and calculated. The minimum distance of each structure in complex with fragment **J2** are as follow: 3.6 Å (trigonal/soaking – PDB ID: 6YIY), 3.4 Å (orthorhombic/soaking – PDB ID: 6YIW) and 3.6 Å (orthorhombic/co-crystallization – PDB ID: 6ZFJ).

It seems that water channels in both trypsin space groups do not affect the hits obtained from the 96-fragment library. Even more, with the small fragment size (<300 Da), the narrowest positions of the water channels of trypsin would not create a barrier to the movement of the fragments through the water channels.

### 5.3.6. Comparison of crystal structures

Some fragments from the 96-fragment library which were hits from the fragment screening against trypsin share structural similarities to compounds that were already obtained in complex with trypsin crystals. The main difference between the crystal structures presented in this study and other crystals structures are the crystallization condition used to obtain trypsin crystals. As mentioned in the crystallization protocol section, different crystallization conditions used to obtain trypsin crystals led to different crystal forms (trigonal and orthorhombic).

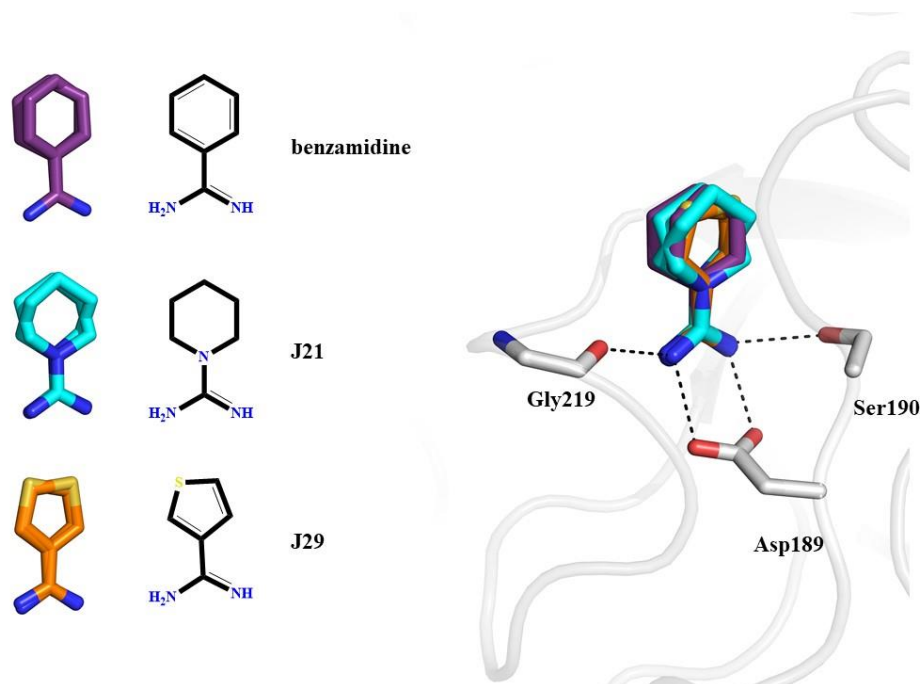
One example of structural similarities between the fragments from the 96-fragment library which were hits and other compounds that were obtained in complex with trypsin is fragment **J2** (1H-isoindol-3-amine) and 2-aminopyridine. Another example are the two fragments **J21** and **J29** sharing structural similarity to benzamidine.



**Figure 19.** Overview of the S1 binding pocket in cationic trypsin. **A.** Crystal structure of cationic trypsin in complex with fragment J2 (PDB ID: 6YIW). **B.** Crystal structure of cationic trypsin in complex with 2-aminopyridine (PDB ID: 5MN1) [215]. **C.** Superimposition of both structures. All distances are given in Å.

The amino group of fragment **J2** has a direct interaction with residue Asp189 (**Figure 19A**) and that is the case for all other trypsin structures in complex with fragment **J2**. However, the amino group of 2-aminopyridine does not have a direct interaction with residue Asp189 (**Figure 19B**). When both trypsin crystal structures in complex with **J2** and 2-aminopyridine were superimposed (**Figure 19C**), the shift of the amino group of the crystal structure in complex with 2-aminopyridine (PDB ID: 5MN1) is observed. The amino group of 2-aminopyridine shows a long interaction of 3.6 Å with residue Asp189 compared to 3.1 Å with

fragment **J2**. As seen in **Figure 19C**, beside the more remote positioning of the amino group of 2-aminopyridine away (2.7 Å/3.0 Å) from residue Asp189 at the bottom of the S1 pocket, the nitrogen on the ring has shifted away from residue Gly219. The binding pocket volume calculation showed a slightly bigger cavity site for trypsin structures in the orthorhombic crystal form. The positioning of residue Gln192 was the reason for the different cavity volume. The majority of trypsin crystal structures in an orthorhombic crystal form showed that residue Gln192 is positioned away from the entrance of the S1 cavity (**Figure 16**), while in a few cases residue Gln192 is positioned closer to the entrance of the S1 cavity or adopts both alternative conformations. The positioning of 2-aminopyridine in the S1 cavity is restricted by residue Gln192 which is positioned closer to the entrance of the binding pocket shifting it away from residue Gly219. Unlike fragment **J2**, where residue Gln192 is positioned away from the binding pocket making no restrictions for **J2** in the S1 cavity to move closer to residue Gly219 and interact with residues Asp189



**Figure 20.** Superimposition of crystal structures in complex with benzamidine (PDB ID: 5MNG) [215], fragment J21 (PDB ID: 6YIS) and fragment J29 (PDB ID: 6YIT).



Since benzamidine was mentioned plentifully during this study, comparing the positioning of this fragment with other similar binding fragments in the S1 pocket was executed. **Figure 20** shows three different trypsin crystal structures, one in complex with benzamidine (PDB ID: 5MNG), another in complex with fragment **J21** (PDB ID: 6YIS) and the last one in complex with fragment **J29** (PDB ID: 6YIT). The amidino nitrogens of all fragments share identical interactions with three residues inside the S1 specificity pocket (Asp189, Ser190 and Gly219). The difference in the distances between the amino group of each compound and residue Asp189 at the back of the S1 specificity pocket is  $< 0.2 \text{ \AA}$ . The orientation and the interactions of these compounds in trypsin active site are similar.

## 5.4. Conclusion

Fragment screening's main objective in drug discovery is finding lead compounds that will prove to be good starting point for the development of drug candidates. Throughout this study, the biological macromolecule receptor is the same (trypsin) for every screen performed. However, when the fragment screen was performed on two different trypsin crystal forms (trigonal and orthorhombic) using soaking technique, the outcome showed one additional hit compared to that performed on the trigonal crystal form. When the fragment screening was performed on the same crystal form (orthorhombic), but with a different crystallization technique (soaking and co-crystallization), an extra hit was obtained in favour of soaking technique.

If trypsin crystals were crystallized under different conditions (buffers, pH or temperature), the hits obtained could vary from the ones obtained in this study. However, all the hits obtained with trypsin in this study represent the smallest fragments from the 96-library with a basic moiety (except fragment **J24**), which has been excepted due to trypsin favouring basic compounds.

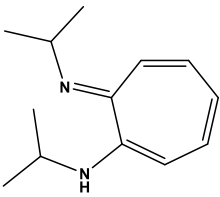
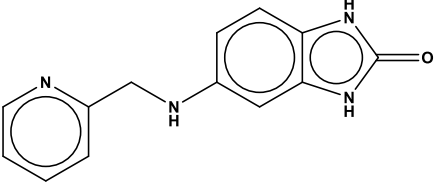
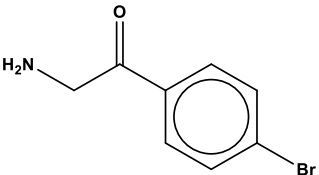
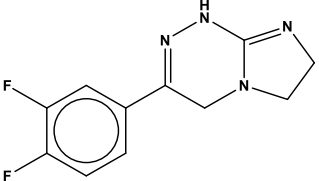
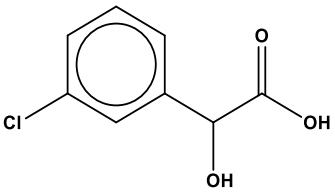
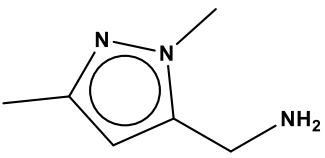
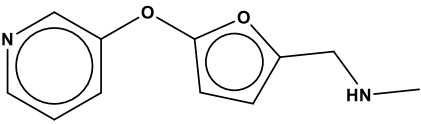
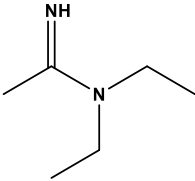
When performing a fragment screen with trypsin, soaking proved to be the method of choice. The ease of crystallization without depending on which fragment is used in co-crystallization, the higher crystal count in the crystallization drop and less material consumption (protein, fragments and buffers) makes it more favourable. All the fragments from the library were screened against trypsin through the soaking protocol and data sets were collected for each soaked fragment. While using the method of co-crystallization for the orthorhombic trypsin crystals.

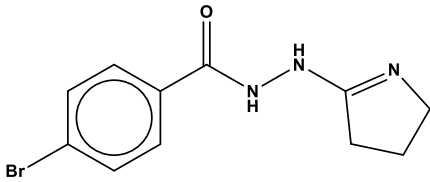
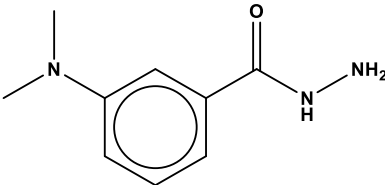
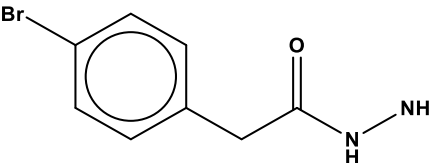
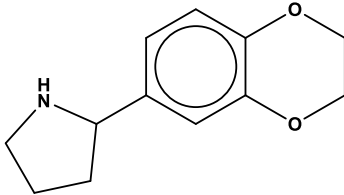
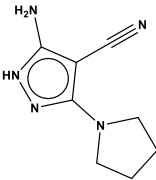
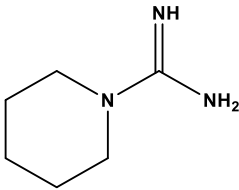
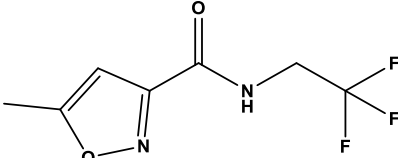
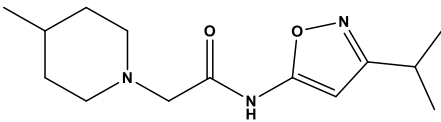
Trypsin is a well-studied protein throughout the last century. However, the aim of this project was observing the outcome of a fragment screening against the same protein, but with different crystal forms.

## 5.5. Appendix

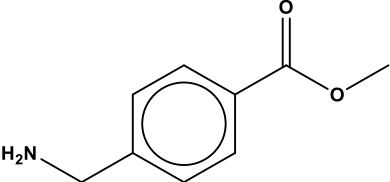
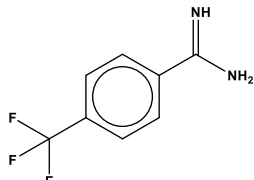
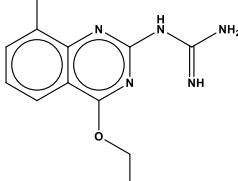
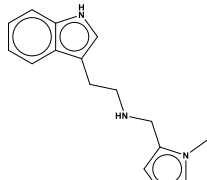
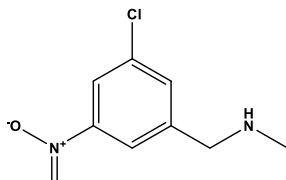
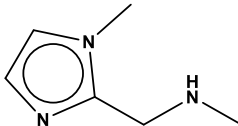
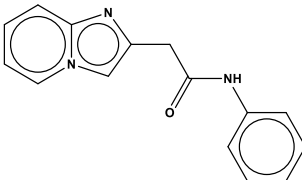
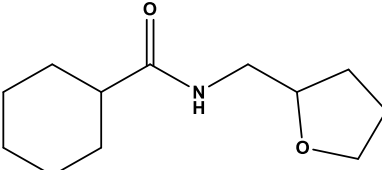
Table 6. The 96-fragment library

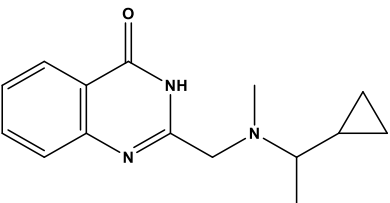
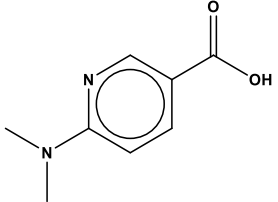
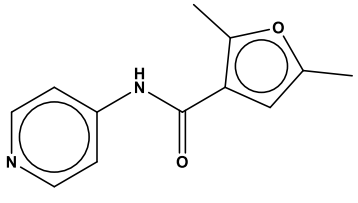
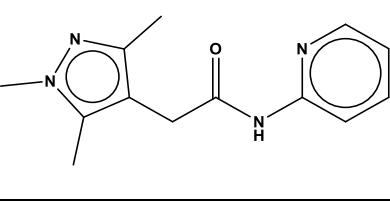
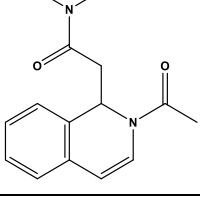
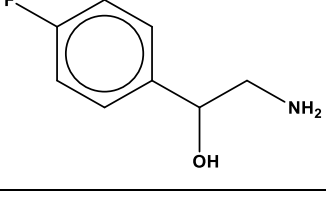
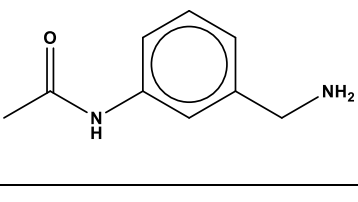
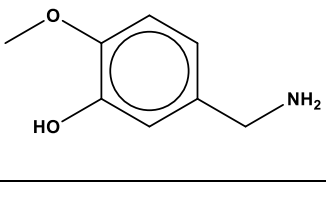
ID	Structure	MW (g/mol)	SMILES code
J1		150.1	<chem>CC1=CC(=CC=C1)C(=O)NN</chem>
J2		132.1	<chem>C1C2=CC=CC=C2C(=N1)N</chem>
J3		277.9	<chem>CNC(=S)NC1=C(C=C(C=C1)Br)Cl</chem>
J4		191.1	<chem>CC1=NC=CC(=N1)N2CCCCC2</chem>
J5		172.2	<chem>CCC(C)(CN)N1CCOCC1</chem>
J6		219.1	<chem>C1CCC(C1)NCC2=CC3=C(C=C2)OCOC3</chem>
J7		262.1	<chem>O=C(CN1CCCCC1)Nc1ccc2OCOC2c1</chem>

J8		204.2	<chem>CC(C)NC1=CC=CC=CC1=NC(C)C</chem>
J9		240.1	<chem>C1=CC=NC(=C1)CNC2=CC3=C(C=C2)NC(=O)N3</chem>
10		212.9	<chem>C1=CC(=CC=C1C(=O)CN)Br</chem>
J11		236.1	<chem>C1CN2CC(=NNC2=N1)C3=CC(=C(C=C3)F)F</chem>
J12		186.0	<chem>C1=CC(=CC(=C1)C1C(C(=O)O)O)Cl</chem>
J13		125.1	<chem>CC1=NN(C(=C1)CN)C</chem>
J14		204.1	<chem>CNCC1=CC=C(O1)OC2=CN=CC=C2</chem>
J15		252.2	<chem>CCN(CC)C(=N)C</chem>

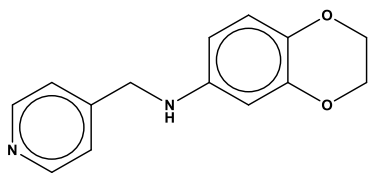
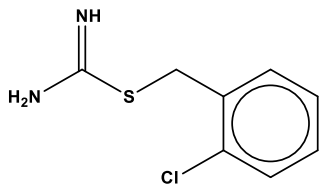
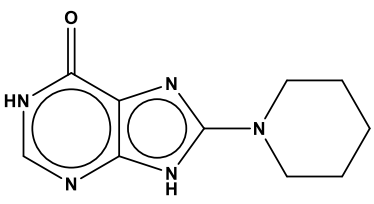
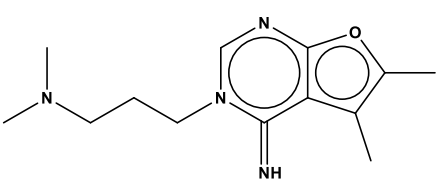
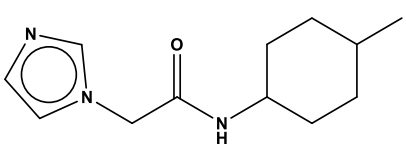
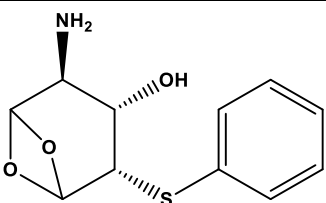
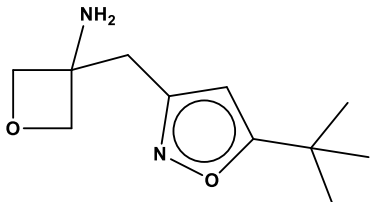
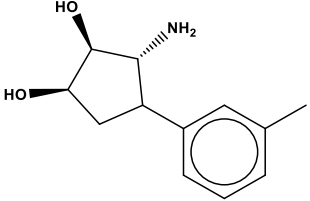
J16		218.1	<chem>C1CC(=NC1)NNC(=O)C2=CC=C(C=C2)Br</chem>
J17		179.1	<chem>CN(C)C1=CC=CC(=C1)C(=O)NN</chem>
J18		168.1	<chem>C1=CC(=CC=C1CC(=O)NN)Br</chem>
J19		205.1	<chem>C1CC(NC1)C2=CC3=C(C=C2)OCCO3</chem>
J20		177.1	<chem>C1CCN(C1)C2=NNC(=C2C#N)N</chem>
J21		127.1	<chem>NC(=N)N1CCCCC1</chem>
J22		208.0	<chem>CC1ON=C(C(NCC(F)(F)F)=O)C=1</chem>
J23		265.2	<chem>CC(C1=NOC(NC(CN2CCC(C)CC2)=O)=C1)C</chem>

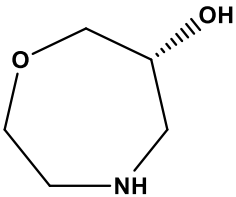
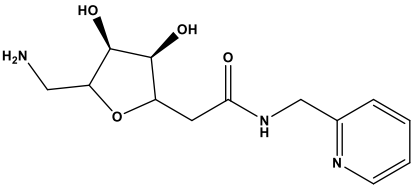
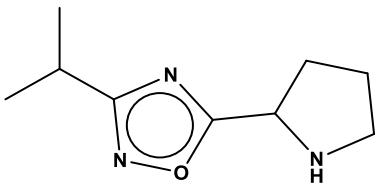
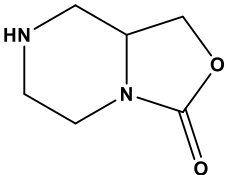
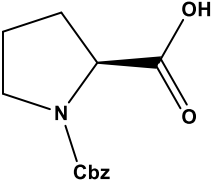
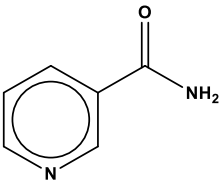
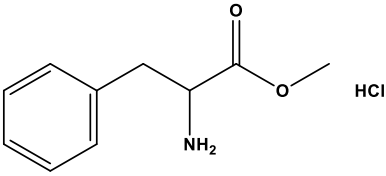
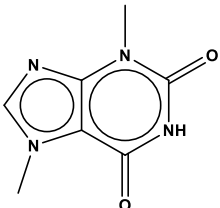
J24		288.1	<chem>CC(C1=CN=CC=C1)NC(=O)CCC(=O)C2=CC=CS2</chem>
J25		274.2	<chem>CC1=CC(=NC(=C1C#N)N)CCCCN2CCOCC2)C</chem>
J26		194.1	<chem>CC1=C(SC(=N1)N)C2=NC=CN2C</chem>
J27		264.2	<chem>CC1CC(CN(C1)CC2=NC(=NC(=N2)N(C)C)N)C</chem>
J28		122.1	<chem>C1(=CC=CN=C1)CCN</chem>
J29		126.0	<chem>C1=CSC=C1C(=N)N</chem>
J30		165.0	<chem>NCC1OC(C(F)(F)F)CC1</chem>
J31		196.2	<chem>O=C(C1CCCN1)N1CCCCC1</chem>

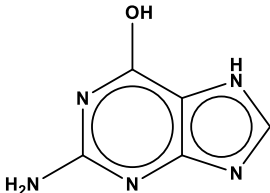
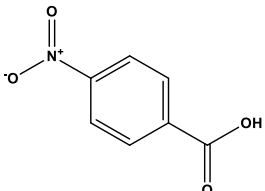
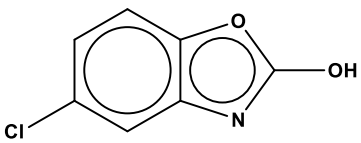
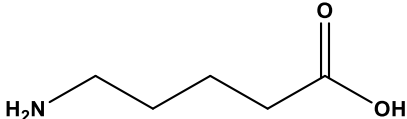
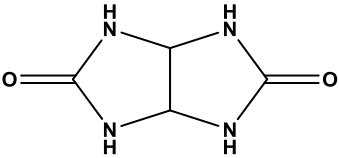
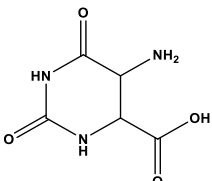
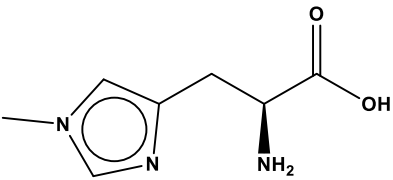
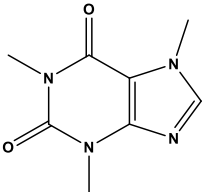
J32		165.1	<chem>COC(=O)C1=CC=C(C=C1)CN</chem>
J33		188.1	<chem>C1=CC(=CC=C1C(=N)N)C(F)(F)F</chem>
J34		270.2	<chem>CCOC1=NC(=NC2=C(C=CC=C21)C)N=C(N)N</chem>
J35		253.2	<chem>CN1C=CC=C1CNCCC2=CNC3=CC=CC=C32</chem>
J36		200.0	<chem>CNCC1=CC(=CC(=C1)Cl)[N+](=O)[O-]</chem>
J37		125.1	<chem>CNCC1=NC=CN1C</chem>
J38		251.1	<chem>C1=CC=C(C=C1)NC(=O)CC2=CN3C=CC=CC3=N2</chem>
J39		211.2	<chem>C1CCC(CC1)C(=O)NCC2CCCCO2</chem>

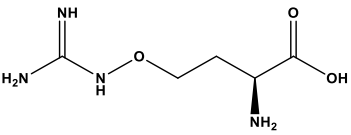
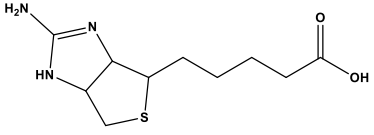
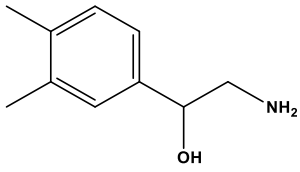
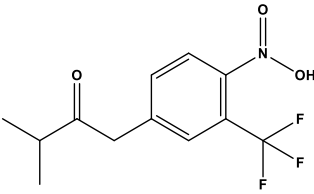
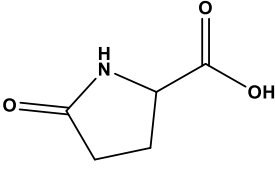
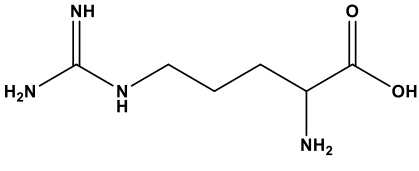
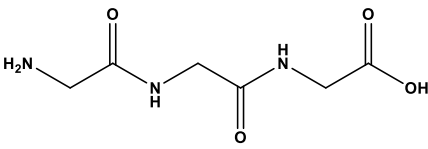
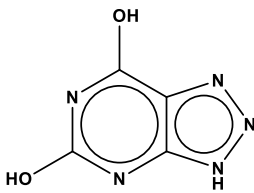
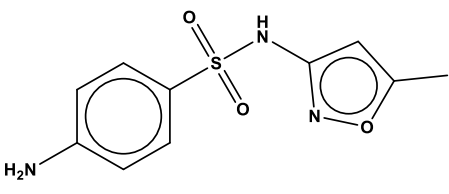
J40		257.2	<chem>CC(N(CC1NC(=O)C2=C(C=CC=C2)N=1)C)C1CC1</chem>
J41		166.1	<chem>CN(C)C1=NC=C(C=C1)C(=O)O</chem>
J42		216.1	<chem>CC1=CC(=C(O1)C)C(=O)NC2=CC=NC=C2</chem>
J43		244.1	<chem>CC1=C(C(=NN1C)C)CC(=O)NC2=CC=CC=N2</chem>
J44		258.1	<chem>CN(C)(CC1C2=C(C=CC=C2)C=CN1C(C)=O)O)C</chem>
J45		155.1	<chem>C1=CC(=CC=C1C(N)O)F</chem>
J46		164.1	<chem>CC(=O)NC1=CC=CC(=C1)CN</chem>
J47		153.1	<chem>COC1=C(C=C(C=C1)CN)O</chem>



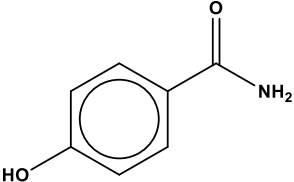
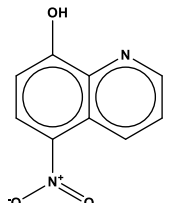
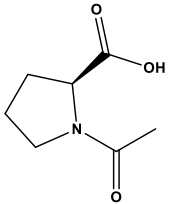
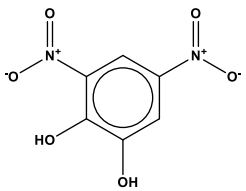
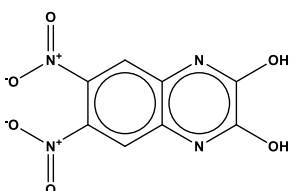
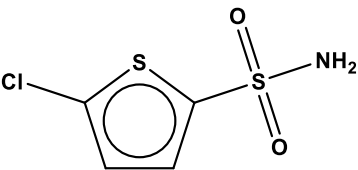
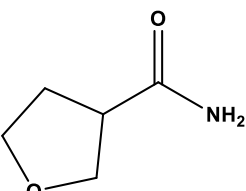
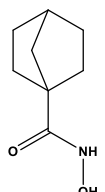
J48		242.1	<chem>C1COC2=C(O1)C=CC(=C2)NCC3=CC=NC=C3</chem>
J49		200.0	<chem>C1=CC=C(C(=C1)CSC(=N)N)Cl</chem>
J50		219.1	<chem>C1CCN(CC1)C2=NC3=C(N2)C(=O)NC=N3</chem>
J51		248.2	<chem>CC1=C(OC2=C1C(=N)N(C=N2)CCCN(C)C)C</chem>
J52		221.2	<chem>CC1CCC(CC1)NC(=O)CN2C=CN=C2</chem>
J53		253.1	<chem>C12[C@H]([C@@H]([C@@H](C(O1)O2)SC3=CC=C(C=C3)O)N</chem>
J54		210.1	<chem>CC(C)(C)C1=CC(=NO1)CC2(COC2)N</chem>
J55		207.1	<chem>CC1=CC(=CC=C1)C2C[C@H]([C@H]([C@@H]2N)O)O</chem>

J56		117.1	<chem>O[C@@H]1CNCCOC1</chem>
J57		281.1	<chem>C1(=NC=CC=C1)CNC(CC2[C@@H]([C@@H](C(CN)O2)O)O)=O</chem>
J58		181.1	<chem>CC(C)C1=NOC(=N1)C2CCCN2</chem>
J59		143.1	<chem>C2C1COC(=O)N1CCN2</chem>
J60		249.1	<chem>C1CC(N(C1)C(=O)OCC2=CC=CC=C2)C(=O)O</chem>
J61		122.1	<chem>C1=CC(=CN=C1)C(=O)N</chem>
J62		179.1	<chem>COC(=O)C(CC1=CC=CC=C1)N.Cl</chem>
J63		180.1	<chem>CN1C=NC2=C1C(=O)NC(=O)N2C</chem>

J64		151.1	<chem>C1=NC2=C(N1)C(=O)NC(=N2)N</chem>
J65		167.0	<chem>C1=CC(=CC=C1C(=O)O)[N+](=O)[O-]</chem>
J66		168.9	<chem>C1=CC2=C(C=C1Cl)NC(=O)O2</chem>
J67		117.1	<chem>C(CCN)CC(=O)O</chem>
J68		142.1	<chem>O=C1NC2NC(=O)NC2N1</chem>
J69		171.0	<chem>C1(C(O)=O)NC(=O)NC(=O)C1N</chem>
J70		169.1	<chem>CN1C=C(N=C1)CC(C(=O)O)N</chem>
J71		194.1	<chem>C12N=CN(C)C=1C(N(C)C(=O)N2C)=O</chem>

J72		176.1	<chem>[C@H](N)(CCONC(=N)N)C(=O)O</chem>
J73		243.1	<chem>C1C2C(C(S1)CCCC(=O)O)N=C(N2)N</chem>
J74		169.1	<chem>C1=CC(=C(C=C1C(CN)O)O)O</chem>
J75		276.1	<chem>C1=CC(=C(C=C1CC(C(C)C)=O)C(F)(F)F)[N](=O)O</chem>
J76		129.0	<chem>OC(=O)C1CCC(=O)N1</chem>
J77		174.1	<chem>C(CC(C(=O)O)N)CN=C(N)N</chem>
J78		189.1	<chem>NCC(=O)NCC(=O)NCC(O)=O</chem>
J79		153.0	<chem>C12=NNN=C1NC(=O)NC2=O</chem>
J80		253.1	<chem>CC1=CC(=NO1)NS(=O)(=O)C2=CC=C(C=C2)N</chem>

J81		161.1	<chem>C(CC(C(=O)[O-])N)CN=C(N)N</chem>
J82		155.0	<chem>C1=CC(=C(C=C1[N+](=O)[O-])O)O</chem>
J83		160.1	<chem>C1=CC(=CC(=C1)N)C2=CN=CO2</chem>
J84		133.1	<chem>C1=CC2=C(C=CN2)C=C1O</chem>
J85		136.1	<chem>C1=CC(=CC(=C1)N)C(=O)N</chem>
J86		157.1	<chem>C1CC(CCC1CN)C(=O)O</chem>
J87		192.1	<chem>CSCCC(NC(N)=O)C(=O)O</chem>
J88		160.1	<chem>CN(C)NC(=O)CCC(O)=O</chem>

J89		137.1	<chem>C1=CC(=CC=C1C(=O)N)O</chem>
J90		190.0	<chem>C1=CC2=C(C=CC(=C2N=C1)O)[N+](=O)[O-]</chem>
J91		157.1	<chem>CC(=O)N1CCC[C@H]1C(=O)O</chem>
J92		200.0	<chem>C1=C(C=C(C(=C1[N+](=O)[O-])O)O)[N+](=O)[O-]</chem>
J93		252.0	<chem>C1=C2C(=CC(=C1[N+](=O)[O-])[N+](=O)[O-])NC(=O)C(=O)N2</chem>
J94		196.9	<chem>C1=C(SC(=C1)Cl)S(=O)(=O)N</chem>
J95		115.1	<chem>NC(=O)C1CCOC1</chem>
J96		155.1	<chem>ONC(=O)C12CCC(CC1)C2</chem>

## References

- [1] Morris DJ, Brem AS, Ge R, Jellinck PH, Sakai RR, Hardy MP. The functional roles of 11 $\beta$ -HSD1: vascular tissue, testis and brain. *Molecular and Cellular Endocrinology* 2003;203(1-2):1–12.
- [2] Nobel S, Abrahmsen L, Oppermann U. Metabolic conversion as a pre-receptor control mechanism for lipophilic hormones. *European Journal of Biochemistry* 2001;268(15):4113–25.
- [3] Lukacik P, Kavanagh KL, Oppermann U. Structure and function of human 17 $\beta$ -hydroxysteroid dehydrogenases. *Molecular and Cellular Endocrinology* 2006;248(1-2):61–71.
- [4] Le Bail J-C, Champavier Y, Chulia A-J, Habrioux G. Effects of phytoestrogens on aromatase, 3 $\beta$  and 17 $\beta$ -hydroxysteroid dehydrogenase activities and human breast cancer cells. *Life Sciences* 2000;66(14):1281–91.
- [5] Marc Meier, Gabriele Moeller, Jerzy Adamski. Perspectives in understanding the role of human 17 $\beta$ -hydroxysteroid dehydrogenases in health and disease. *Annals of The New York Academy of Sciences* 2009;1155(1):15–24.
- [6] Marie Berube, Yannick Laplante and Donald Poirier. Design, Synthesis and In Vitro Evaluation of 4-Androstene-3,17-dione/Adenosine Hybrid Compounds as Bisubstrate Inhibitors of Type 3 17 $\beta$ -Hydroxysteroid Dehydrogenase. *Medicinal Chemistry* 2006;2(4):329–47.
- [7] Marioli DJ, Saltamavros AD, Vervita V, Koika V, Adonakis G, Decavalas G, Markou KB, Georgopoulos NA. Association of the 17-hydroxysteroid dehydrogenase type 5 gene polymorphism (-71A/G HSD17B5 SNP) with hyperandrogenemia in polycystic ovary syndrome (PCOS). *Fertility and Sterility* 2009;92(2):648–52.
- [8] Moeller G, Adamski J. Multifunctionality of human 17 $\beta$ -hydroxysteroid dehydrogenases. *Molecular and Cellular Endocrinology* 2006;248(1-2):47–55.
- [9] Moeller G, Adamski J. Integrated view on 17 $\beta$ -hydroxysteroid dehydrogenases. *Molecular and Cellular Endocrinology* 2009;301(1-2):7–19.
- [10] Negri M, Recanatini M, Hartmann RW. Insights in 17 $\beta$ -HSD1 enzyme kinetics and ligand binding by dynamic motion investigation. *PLoS One* 2010;5(8):e12026.
- [11] Peltoketo H, Luu-The V, Simard J, Adamski J. 17 $\beta$ -hydroxysteroid dehydrogenase (HSD)/17-ketosteroid reductase (KSR) family; nomenclature and main characteristics of the 17HSD/KSR enzymes. *J Mol Endocrinol* 1999;23(1):1–11.
- [12] Penning TM, Burczynski ME, Jez JM, Hung CF, Lin HK, Ma H, Moore M, Palackal N, Ratnam K. Human 3 $\alpha$ -hydroxysteroid dehydrogenase isoforms (AKR1C1-AKR1C4) of the aldoketo reductase superfamily: functional plasticity and tissue distribution reveals roles in the inactivation and formation of male and female sex hormones. *Biochem J* 2000;351(Pt 1):67–77.
- [13] Penning TM. Human hydroxysteroid dehydrogenases and pre-receptor regulation: insights into inhibitor design and evaluation. *J Steroid Biochem Mol Biol* 2011;125(1-2):46–56.

- [14] Persson B, Kallberg Y. Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chem Biol Interact* 2013;202(1-3):111–5.
- [15] Puranen TJ, Poutanen MH, Peltoketo HE, Vihko PT, Vihko RK. Site-directed mutagenesis of the putative active site of human 17 beta-hydroxysteroid dehydrogenase type 1. *Biochem J* 1994;304(Pt 1):289–93.
- [16] van Luu-The. Analysis and characteristics of multiple types of human 17 $\beta$ -hydroxysteroid dehydrogenase. *J Steroid Biochem Mol Biol* 2001;76(1-5):143–51.
- [17] Wisniewski AB, Mazur T. 46, XY DSD with Female or Ambiguous External Genitalia at Birth due to Androgen Insensitivity Syndrome, 5 $\alpha$ -Reductase-2 Deficiency, or 17 $\beta$ -Hydroxysteroid Dehydrogenase Deficiency: A Review of Quality of Life Outcomes. *Int J Pediatr Endocrinol* 2009;2009(1):567430.
- [18] Zeitoun K, Takayama K, Sasano H, Suzuki T, Moghrabi N, Andersson S, Johns A, Meng L, Putman M, Carr B, Bulun SE. Deficient 17 $\beta$ -hydroxysteroid dehydrogenase type 2 expression in endometriosis: failure to metabolize 17 $\beta$ -estradiol. *J Clin Endocrinol Metab* 1998;83(12):4474–80.
- [19] Zhongyi S, Rantakari P, Lamminen T, Toppari J, Poutanen M. Transgenic male mice expressing human hydroxysteroid dehydrogenase 2 indicate a role for the enzyme independent of its action on sex steroids. *Endocrinology* 2007;148(8):3827–36.
- [20] Labrie F, Luu-The V, Lin SX, Simard J, Labrie C, El-Alfy M, Pelletier G, Bélanger A. Intracrinology: role of the family of 17 beta-hydroxysteroid dehydrogenases in human physiology and disease. *J Mol Endocrinol* 2000;25(1):1–16.
- [21] Imperato-McGinley J, Peterson RE, Stoller R, Goodwin WE. Male pseudohermaphroditism secondary to 17 beta-hydroxysteroid dehydrogenase deficiency: gender role change with puberty. *J Clin Endocrinol Metab* 1979;49(3):391–5.
- [22] Hikima T, Maibach HI. Gender differences of enzymatic activity and distribution of 17 $\beta$ -hydroxysteroid dehydrogenase in human skin in vitro. *SPP* 2007;20(4):168–74.
- [23] Hedlund J, Jörnvall H, Persson B. Subdivision of the MDR superfamily of medium-chain dehydrogenases/reductases through iterative hidden Markov model refinement. *BMC Bioinformatics* 2010;11(1):534.
- [24] Haller F, Moman E, Hartmann RW, Adamski J, Mindnich R. Molecular framework of steroid/retinoid discrimination in 17 $\beta$ -hydroxysteroid dehydrogenase type 1 and photoreceptor-associated retinol dehydrogenase. *J Mol Biol* 2010;399(2):255–67.
- [25] Ghosh D, Pletnev VZ, Zhu D-W, Wawrzak Z, Duax WL, Pangborn W, Labrie F, Lin S-X. Structure of human estrogenic 17 $\beta$ -hydroxysteroid dehydrogenase at 2.20 Å resolution. *Structure* 1995;3(5):503–13.
- [26] Geissler WM, Davis DL, Wu L, Bradshaw KD, Patel S, Mendonca BB, Elliston KO, Wilson JD, Russell DW, Andersson S. Male pseudohermaphroditism caused by mutations of testicular 17 beta-hydroxysteroid dehydrogenase 3. *Nat Genet* 1994;7(1):34–9.
- [27] Gargano EM, Allegretta G, Perspicace E, Carotti A, van Koppen C, Frotscher M, Marchais-Oberwinkler S, Hartmann RW. 17 $\beta$ -Hydroxysteroid Dehydrogenase Type 2 Inhibition: Discovery of Selective and Metabolically Stable Compounds Inhibiting Both the Human Enzyme and Its Murine Ortholog. *PLOS ONE* 2015;10(7):e0134754.



- [28] Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem* 2002;277(28):25677–84.
- [29] Dufort I, Rheault P, Huang XF, Soucy P, Luu-The V. Characteristics of a highly labile human type 5 17 $\beta$ -hydroxysteroid dehydrogenase. *Endocrinology* 1999;140(2):568–74.
- [30] Braun F, Bertoletti N, Möller G, Adamski J, Steinmetzer T, Salah M, Abdelsamie AS, van Koppen CJ, Heine A, Klebe G, Marchais-Oberwinkler S. First Structure-Activity Relationship of 17 $\beta$ -Hydroxysteroid Dehydrogenase Type 14 Nonsteroidal Inhibitors and Crystal Structures in Complex with the Enzyme. *J Med Chem* 2016;59(23):10719–37.
- [31] Braun F, Bertoletti N, Möller G, Adamski J, Frotscher M, Guragossian N, Madeira Gírio PA, Le Borgne M, Ettouati L, Falson P, Müller S, Vollmer G, Heine A, Klebe G, Marchais-Oberwinkler S. Structure-based design and profiling of novel 17 $\beta$ -HSD14 inhibitors. *Eur J Med Chem* 2018;155:61–76.
- [32] Jörnvall H, Persson B, Krook M, Atrian S, González-Duarte R, Jeffery J, Ghosh D. Short-chain dehydrogenases/reductases (SDR). *Biochemistry* 1995;34(18):6003–13.
- [33] Oppermann U, Filling C, Hult M, Shafqat N, Wu X, Lindh M, Shafqat J, Nordling E, Kallberg Y, Persson B, Jörnvall H. Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem Biol Interact* 2003;143-144:247–53.
- [34] Kallberg Y, Oppermann U, Jörnvall H, Persson B. Short-chain dehydrogenases/reductases (SDRs). *European Journal of Biochemistry* 2002;269(18):4409–17.
- [35] The PyMOL Molecular Graphics System, Version 1.7.x Schrödinger, LLC.
- [36] Bertoletti N. Structural Characterization of 17 $\beta$ -Hydroxysteroid Dehydrogenase Type 14 and Inhibitor Optimization Using Crystallography and Computational Techniques; 2017. <https://archiv.ub.uni-marburg.de/diss/z2017/0671/pdf/dnb.pdf>
- [37] Wolff AM, Young ID, Sierra RG, Brewster AS, Martynowycz MW, Nango E, Sugahara M, Nakane T, Ito K, Aquila A, Bhowmick A, Biel JT, Carbajo S, Cohen AE, Cortez S, Gonzalez A, Hino T, Im D, Koralek JD, Kubo M, Lazarou TS, Nomura T, Owada S, Samelson AJ, Tanaka T, Tanaka R, Thompson EM, van den Bedem H, Woldeyes RA, Yumoto F, Zhao W, Tono K, Boutet S, Iwata S, Gonen T, Sauter NK, Fraser JS, Thompson MC. Comparing serial X-ray crystallography and microcrystal electron diffraction (MicroED) as methods for routine structure determination from small macromolecular crystals. *IUCrJ* 2020;7(Pt 2):306–23.
- [38] Zander U, Bourenkov G, Popov AN, Sanctis D de, Svensson O, McCarthy AA, Round E, Gordeliy V, Mueller-Dieckmann C, Leonard GA. MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines. *Acta Crystallogr D Biol Crystallogr* 2015;71(Pt 11):2328–43.
- [39] Sugahara M, Motomura K, Suzuki M, Masuda T, Joti Y, Numata K, Tono K, Yabashi M, Ishikawa T. Viscosity-adjustable grease matrices for serial nanocrystallography. *Sci Rep* 2020;10(1):1371.
- [40] Beale JH, Bolton R, Marshall SA, Beale EV, Carr SB, Ebrahim A, Moreno-Chicano T, Hough MA, Worrall JAR, Tews I, Owen RL. Successful sample preparation for serial crystallography experiments. *J Appl Crystallogr* 2019;52(Pt 6):1385–96.

- [41] Andersson R, Safari C, Båth P, Bosman R, Shilova A, Dahl P, Ghosh S, Dunge A, Kjeldsen-Jensen R, Nan J, Shoeman RL, Kloos M, Doak RB, Mueller U, Neutze R, Brändén G. Well-based crystallization of lipidic cubic phase microcrystals for serial X-ray crystallography experiments. *Acta Crystallogr D Struct Biol* 2019;75(Pt 10):937–46.
- [42] Wang B, Zou X, Smeets S. Automated serial rotation electron diffraction combined with cluster analysis: an efficient multi-crystal workflow for structure determination. *IUCrJ* 2019;6(Pt 5):854–67.
- [43] Davy B, Axford D, Beale JH, Butryn A, Docker P, Ebrahim A, Leen G, Orville AM, Owen RL, Aller P. Reducing sample consumption for serial crystallography using acoustic drop ejection. *J Synchrotron Radiat* 2019;26(Pt 5):1820–5.
- [44] Wijn R de, Hennig O, Roche J, Engilberge S, Rollet K, Fernandez-Millan P, Brillet K, Betat H, Mörl M, Roussel A, Girard E, Mueller-Dieckmann C, Fox GC, Olieric V, Gavira JA, Lorber B, Sauter C. A simple and versatile microfluidic device for efficient biomacromolecule crystallization and structural analysis by serial crystallography. *IUCrJ* 2019;6(Pt 3):454–64.
- [45] Ebrahim A, Appleby MV, Axford D, Beale J, Moreno-Chicano T, Sherrell DA, Strange RW, Hough MA, Owen RL. Resolving polymorphs and radiation-driven effects in microcrystals using fixed-target serial synchrotron crystallography. *Acta Crystallogr D Struct Biol* 2019;75(Pt 2):151–9.
- [46] Smeets S, Zou X, Wan W. Serial electron crystallography for structure determination and phase analysis of nanocrystalline materials. *J Appl Crystallogr* 2018;51(Pt 5):1262–73.
- [47] Lan T-Y, Wierman JL, Tate MW, Philipp HT, Martin-Garcia JM, Zhu L, Kissick D, Fromme P, Fischetti RF, Liu W, Elser V, Gruner SM. Solving protein structure from sparse serial microcrystal diffraction data at a storage-ring synchrotron source. *IUCrJ* 2018;5(Pt 5):548–58.
- [48] Roedig P, Ginn HM, Pakendorf T, Sutton G, Harlos K, Walter TS, Meyer J, Fischer P, Duman R, Vartiainen I, Reime B, Warmer M, Brewster AS, Young ID, Michels-Clark T, Sauter NK, Kotecha A, Kelly J, Rowlands DJ, Sikorsky M, Nelson S, Damiani DS, Alonso-Mori R, Ren J, Fry EE, David C, Stuart DI, Wagner A, Meents A. High-speed fixed-target serial virus crystallography. *Nat Methods* 2017;14(8):805–10.
- [49] Wierman JL, Paré-Labrosse O, Sarracini A, Besaw JE, Cook MJ, Oghbaey S, Daoud H, Mehrabi P, Kriksunov I, Kuo A, Schuller DJ, Smith S, Ernst OP, Szebenyi DME, Gruner SM, Miller RJD, Finke AD. Fixed-target serial oscillation crystallography at room temperature. *IUCrJ* 2019;6(Pt 2):305–16.
- [50] Fry EE, Grimes J, Stuart DI. Virus Crystallography. *Mol Biotechnol* 1999;12(1):13–24.
- [51] Olmos JL, Pandey S, Martin-Garcia JM, Calvey G, Katz A, Knoska J, Kupitz C, Hunter MS, Liang M, Oberthuer D, Yefanov O, Wiedorn M, Heyman M, Holl M, Pande K, Barty A, Miller MD, Stern S, Roy-Chowdhury S, Coe J, Nagaratnam N, Zook J, Verburgt J, Norwood T, Poudyal I, Xu D, Koglin J, Seaberg MH, Zhao Y, Bajt S, Grant T, Mariani V, Nelson G, Subramanian G, Bae E, Fromme R, Fung R, Schwander P, Frank M, White TA, Weierstall U, Zatsepin N, Spence J, Fromme P, Chapman HN, Pollack L, Tremblay L, Ourmazd A, Phillips GN, Schmidt M. Enzyme intermediates captured "on the fly" by mix-and-inject serial crystallography. *BMC Biol* 2018;16(1):59.
- [52] Coquelle N, Sliwa M, Woodhouse J, Schirò G, Adam V, Aquila A, Barends TRM, Boutet S, Byrdin M, Carbajo S, La Mora E de, Doak RB, Feliks M, Fieschi F, Foucar L, Guillon V, Hilpert M, Hunter MS, Jakobs S, Koglin JE, Kovacsova G, Lane TJ, Lévy B, Liang M, Nass

- K, Ridard J, Robinson JS, Roome CM, Ruckebusch C, Seaberg M, Thepaut M, Cammarata M, Demachy I, Field M, Shoeman RL, Bourgeois D, Colletier J-P, Schlichting I, Weik M. Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat Chem* 2018;10(1):31–7.
- [53] Garman EF. Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallogr D Biol Crystallogr* 2010;66(Pt 4):339–51.
- [54] Sanishvili R, Yoder DW, Pothineni SB, Rosenbaum G, Xu S, Vogt S, Stepanov S, Makarov OA, Corcoran S, Benn R, Nagarajan V, Smith JL, Fischetti RF. Radiation damage in protein crystals is reduced with a micron-sized X-ray beam. *PNAS* 2011;108(15):6127–32.
- [55] Sliz P, Harrison SC, Rosenbaum G. How does Radiation Damage in Protein Crystals Depend on X-Ray Dose? *Structure* 2003;11(1):13–9.
- [56] Taberman H. Radiation Damage in Macromolecular Crystallography—An Experimentalist's View. *Crystals* 2018;8(4):157.
- [57] Garman EF, Weik M. Radiation Damage in Macromolecular Crystallography. *Methods Mol Biol* 2017;1607:467–89.
- [58] Holton JM. A beginner's guide to radiation damage. *J Synchrotron Rad* 2009;16(Pt 2):133–42.
- [59] Burkhardt A, Wagner A, Warmer M, Reimer R, Hohenberg H, Ren J, Fry EE, Stuart DI, Meents A. Structure determination from a single high-pressure-frozen virus crystal. *Acta Crystallogr D Biol Crystallogr* 2013;69(Pt 2):308–12.
- [60] Garman EF, Owen RL. Cryocooling and radiation damage in macromolecular crystallography. *Acta Cryst D* 2006;62(Pt 1):32–47.
- [61] Southworth-Davies RJ, Medina MA, Carmichael I, Garman EF. Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography. *Structure* 2007;15(12):1531–41.
- [62] Feiler CG, Wallacher D, Weiss MS. An All-in-one Sample Holder for Macromolecular X-ray Crystallography with Minimal Background Scattering. *JoVE (Journal of Visualized Experiments)* 2019;(149):e59722.
- [63] Owen RL, Axford D, Sherrell DA, Kuo A, Ernst OP, Schulz EC, Miller RJD, Mueller-Werkmeister HM. Low-dose fixed-target serial synchrotron crystallography. *Acta Crystallogr D Struct Biol* 2017;73(Pt 4):373–8.
- [64] Roedig P, Vartiainen I, Duman R, Panneerselvam S, Stübe N, Lorbeer O, Warmer M, Sutton G, Stuart DI, Weckert E, David C, Wagner A, Meents A. A micro-patterned silicon chip as sample holder for macromolecular crystallography experiments with minimal background scattering. *Sci Rep* 2015;5:10451.
- [65] Huang CY, Olieric V, Ma P, Panepucci E, Diederichs K, Wang M, Caffrey M. In meso in situ serial X-ray crystallography of soluble and membrane proteins. *Acta Crystallogr D Biol Crystallogr* 2015;71(Pt 6):1238–56.
- [66] Gati C, Bourenkov G, Klinge M, Rehders D, Stellato F, Oberthür D, Yefanov O, Sommer BP, Mogk S, Duzenko M, Betzel C, Schneider TR, Chapman HN, Redecke L. Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ* 2014;1(Pt 2):87–94.

- [67] Coquelle N, Brewster AS, Kapp U, Shilova A, Weinhausen B, Burghammer M, Colletier JP. Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. *Acta Crystallogr D Biol Crystallogr* 2015;71(Pt 5):1184–96.
- [68] Diederichs K, Wang M. Serial Synchrotron X-Ray Crystallography (SSX). In: Wlodawer A, Dauter Z, Jaskolski M (editors). *Protein crystallography: Methods and protocols* / edited by Alexander Wlodawer, Zbigniew Dauter, Mariusz Jaskolski. New York: Humana Press; 2017. p. 239–72.
- [69] Saenger W, Jacob J, Gessler K, Steiner T, Hoffmann D, Sanbe H, Koizumi K, Smith SM, Takaha T. Structures of the Common Cyclodextrins and Their Larger Analogues-Beyond the Doughnut. *Chem Rev* 1998;98(5):1787–802.
- [70] Stick RV, Williams SJ (editors). *Carbohydrates: The essential molecules of life* / Robert V. Stick, Spencer J. Williams. 2nd ed. Amsterdam, London: Elsevier; 2009.
- [71] Schneider H-J, Hacket F, Rüdiger V, Ikeda H. NMR Studies of Cyclodextrins and Cyclodextrin Complexes. *Chem Rev* 1998;98(5):1755–86.
- [72] Saenger W, Steiner T. Cyclodextrin Inclusion Complexes: Host–Guest Interactions and Hydrogen-Bonding Networks. *Acta Cryst A* 1998;54(6):798–805.
- [73] Pinjari RV, Joshi KA, Gejji SP. Theoretical studies on hydrogen bonding, NMR chemical shifts and electron density topography in alpha, beta and gamma-cyclodextrin conformers. *The Journal of Physical Chemistry A* 2007;111(51):13583–9.
- [74] Pinjari RV, Joshi KA, Gejji SP. Molecular electrostatic potentials and hydrogen bonding in alpha-, beta-, and gamma-cyclodextrins. *The Journal of Physical Chemistry A* 2006;110(48):13073–80.
- [75] Onda M, Yamamoto Y, Inoue Y, Chûjô R. <sup>1</sup>H NMR Study of Intramolecular Hydrogen-Bonding Interaction in Cyclodextrins and Their Di- O -methylated Derivatives. *BCSJ* 1988;61(11):4015–21.
- [76] Niether D, Kawaguchi T, Hovancová J, Eguchi K, Dhont JKG, Kita R, Wiegand S. Role of Hydrogen Bonding of Cyclodextrin-Drug Complexes Probed by Thermodiffusion. *Langmuir* 2017;33(34):8483–92.
- [77] Deshmukh MM, Bartolotti LJ, Gadre SR. Intramolecular hydrogen bond energy and cooperative interactions in  $\alpha$ -,  $\beta$ -, and  $\gamma$ -cyclodextrin conformers. *Journal of Computational Chemistry* 2011;32(14):2996–3004.
- [78] Attoui-Yahia O, Khatmi D, Krim K, Ferkous F. Hydrogen bonding investigation in Pyridoxine/ $\beta$ -cyclodextrin complex based on QTAIM and NBO approaches. *Journal of the Taiwan Institute of Chemical Engineers* 2015;47:91–8.
- [79] Yoshida K-i, Shimomura T, Ito K, Hayakawa R. Inclusion Complex Formation of Cyclodextrin and Polyaniline. *Langmuir* 1999;15(4):910–3.
- [80] Saenger W, Noltemeyer M, Manor PC, Hingerty B, Klar B. “Induced-fit”-type complex formation of the model enzyme  $\alpha$ -cyclodextrin. *Bioorganic Chemistry* 1976;5(2):187–95.
- [81] Harada A, Kamachi M. Complex formation between cyclodextrin and poly(propylene glycol). *J. Chem. Soc., Chem. Commun.* 1990;0(19):1322.

- [82] Harada A, Kawaguchi Y, Nishiyama T, Kamachi M. Complex formation of poly( $\epsilon$ -caprolactone) with cyclodextrin. *Macromol. Rapid Commun.* 1997;18(7):535–9.
- [83] Harada A, Li J, Suzuki S, Kamachi M. Complex formation between polyisobutylene and cyclodextrins: inversion of chain-length selectivity between  $\beta$ -cyclodextrin and  $\gamma$ -cyclodextrin. *Macromolecules* 1993;26(19):5267–8.
- [84] Connors KA. Application of a Stoichiometric Model of Cyclodextrin Complex Formation. In: Szejtli J (editor). *Proceedings of the First International Symposium on Cyclodextrins: Budapest, Hungary, 30 September-2 October 1981*. Dordrecht: Springer Netherlands; 2013. p. 163–72.
- [85] Charoenchaitrakool M, Dehghani F, Foster NR. Utilization of supercritical carbon dioxide for complex formation of ibuprofen and methyl- $\beta$ -cyclodextrin. *International Journal of Pharmaceutics* 2002;239(1-2):103–12.
- [86] Archontaki HA, Vertzoni MV, Athanassiou-Malaki MH. Study on the inclusion complexes of bromazepam with  $\beta$ - and  $\beta$ -hydroxypropyl-cyclodextrins. *Journal of Pharmaceutical and Biomedical Analysis* 2002;28(3-4):761–9.
- [87] Yadav VR, Suresh S, Devi K, Yadav S. Effect of cyclodextrin complexation of curcumin on its solubility and antiangiogenic and anti-inflammatory activity in rat colitis model. *AAPS PharmSciTech* 2009;10(3):752–62.
- [88] Backensfeld T, Müller BW, Wiese M, Seydel JK. Effect of cyclodextrin derivatives on indomethacin stability in aqueous solution. *Pharm Res* 1990;7(5):484–90.
- [89] Kfoury M, Landy D, Auezova L, Greige-Gerges H, Fourmentin S. Effect of cyclodextrin complexation on phenylpropanoids' solubility and antioxidant activity. *Beilstein J. Org. Chem.* 2014;10(1):2322–31.
- [90] Hadžiabdić J, Elezović A, Rahić O, Mujezin I. Effect of Cyclodextrin Complexation on the Aqueous Solubility of Diazepam and Nitrazepam: Phase-Solubility Analysis, Thermodynamic Properties. *AJAC* 2012;03(12):811–9.
- [91] Cirri M, Maestrelli F, Corti G, Furlanetto S, Mura P. Simultaneous effect of cyclodextrin complexation, pH, and hydrophilic polymers on naproxen solubilization. *Journal of Pharmaceutical and Biomedical Analysis* 2006;42(1):126–31.
- [92] Hoshino T, Uekama K, Pitha J. Increase in temperature enhances solubility of drugs in aqueous solutions of hydroxypropylcyclodextrins. *International Journal of Pharmaceutics* 1993;98(1-3):239–42.
- [93] Loukas YL, Vraka V, Gregoriadis G. Novel non-acidic formulations of haloperidol complexed with  $\beta$ -cyclodextrin derivatives. *Journal of Pharmaceutical and Biomedical Analysis* 1997;16(2):263–8.
- [94] Tommasini S, Raneri D, Ficarra R, Calabrò ML, Stancanelli R, Ficarra P. Improvement in solubility and dissolution rate of flavonoids by complexation with  $\beta$ -cyclodextrin. *Journal of Pharmaceutical and Biomedical Analysis* 2004;35(2):379–87.
- [95] Zarzycki PK, Lamparczyk H. The equilibrium constant of  $\beta$ -cyclodextrin–phenolphthalein complex; influence of temperature and tetrahydrofuran addition. *Journal of Pharmaceutical and Biomedical Analysis* 1998;18(1-2):165–70.
- [96] Ho B, Baryshnikova A, Brown GW. Unification of Protein Abundance Datasets Yields a Quantitative *Saccharomyces cerevisiae* Proteome. *Cell Systems* 2018;6(2):192-205.e3.

- [97] Schiebel J, Radeva N, Köster H, Metz A, Krotzky T, Kuhnert M, Diederich WE, Heine A, Neumann L, Atmanene C, Roecklin D, Vivat-Hannah V, Renaud J-P, Meinecke R, Schlinck N, Sitte A, Popp F, Zeeb M, Klebe G. One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* 2015;10(9):1511–21.
- [98] Giordanetto F, Jin C, Willmore L, Feher M, Shaw DE. Fragment Hits: What do They Look Like and How do They Bind? *J Med Chem* 2019;62(7):3381–94.
- [99] Bauman JD, Patel D, Baker SF, Vijayan RSK, Xiang A, Parhi AK, Martínez-Sobrido L, LaVoie EJ, Das K, Arnold E. Crystallographic fragment screening and structure-based optimization yields a new class of influenza endonuclease inhibitors. *ACS Chemical Biology* 2013;8(11):2501–8.
- [100] Blomberg N, Cosgrove DA, Kenny PW, Kolmodin K. Design of compound libraries for fragment screening. *J Comput Aided Mol Des* 2009;23(8):513–25.
- [101] Cheng Y, Judd TC, Bartberger MD, Brown J, Chen K, Freneau RT, Hickman D, Hitchcock SA, Jordan B, Li V, Lopez P, Louie SW, Luo Y, Michelsen K, Nixey T, Powers TS, Rattan C, Sickmier EA, St Jean DJ, Wahl RC, Wen PH, Wood S. From fragment screening to in vivo efficacy: optimization of a series of 2-aminoquinolines as potent inhibitors of beta-site amyloid precursor protein cleaving enzyme 1 (BACE1). *J Med Chem* 2011;54(16):5836–57.
- [102] Congreve M, Aharony D, Albert J, Callaghan O, Campbell J, Carr RAE, Chessari G, Cowan S, Edwards PD, Frederickson M, McMenamin R, Murray CW, Patel S, Wallis N. Application of fragment screening by X-ray crystallography to the discovery of aminopyridines as inhibitors of beta-secretase. *J Med Chem* 2007;50(6):1124–32.
- [103] Davis BJ, Erlanson DA. Learning from our mistakes: the 'unknown knowns' in fragment screening. *Bioorganic & Medicinal Chemistry Letters* 2013;23(10):2844–52.
- [104] Duong-Thi M-D, Meiby E, Bergström M, Fex T, Isaksson R, Ohlson S. Weak affinity chromatography as a new approach for fragment screening in drug discovery. *Analytical Biochemistry* 2011;414(1):138–46.
- [105] Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, Jhoti H. Fragment-based lead discovery using X-ray crystallography. *J Med Chem* 2005;48(2):403–13.
- [106] Howard N, Abell C, Blakemore W, Chessari G, Congreve M, Howard S, Jhoti H, Murray CW, Seavers LCA, van Montfort RLM. Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *J Med Chem* 2006;49(4):1346–55.
- [107] Leach AR, Hann MM, Burrows JN, Griffen EJ. Fragment screening: an introduction. *Molecular BioSystems* 2006;2(9):430–46.
- [108] Liu G, Xin Z, Pei Z, Hajduk PJ, Abad-Zapatero C, Hutchins CW, Zhao H, Lubben TH, Ballaron SJ, Haasch DL, Kaszubska W, Rondinone CM, Trevillyan JM, Jirousek MR. Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor. *J Med Chem* 2003;46(20):4232–5.
- [109] Murray CW, Callaghan O, Chessari G, Cleasby A, Congreve M, Frederickson M, Hartshorn MJ, McMenamin R, Patel S, Wallis N. Application of fragment screening by X-ray crystallography to beta-secretase. *J Med Chem* 2007;50(6):1116–23.

- [110] Navratilova I, Hopkins AL. Fragment screening by surface plasmon resonance. *ACS Medicinal Chemistry Letters* 2010;1(1):44–8.
- [111] Scott DE, Coyne AG, Hudson SA, Abell C. Fragment-based approaches in drug discovery and chemical biology. *Biochemistry* 2012;51(25):4990–5003.
- [112] Siegal G, AB, Eiso, Schultz J. Integration of fragment screening and library design. *Drug Discovery Today* 2007;12(23-24):1032–9.
- [113] Zhang C, Ibrahim PN, Zhang J, Burton EA, Habets G, Zhang Y, Powell B, West BL, Matusow B, Tsang G, Shellooe R, Carias H, Nguyen H, Marimuthu A, Zhang KYJ, Oh A, Bremer R, Hurt CR, Artis DR, Wu G, Nespi M, Spevak W, Lin P, Nolop K, Hirth P, Tesch GH, Bollag G. Design and pharmacology of a highly specific dual FMS and KIT kinase inhibitor. *Proc Natl Acad Sci U S A* 2013;110(14):5689–94.
- [114] Köster H, Craan T, Brass S, Herhaus C, Zentgraf M, Neumann L, Heine A, Klebe G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J Med Chem* 2011;54(22):7784–96.
- [115] Murray CW, Verdonk ML, Rees DC. Experiences in fragment-based drug discovery. *Trends Pharmacol Sci* 2012;33(5):224–32.
- [116] Kühne W. Ueber die Verdauung der Eiweisstoffe durch den Pankreassaft. *Archiv f. pathol. Anat.* 1867;39(1):130–74.
- [117] Northsop JH, Kunitz M. Isolation of protein crystals possessing tryptic activity. *Science (Washington)* 1931;73:262–3.
- [118] Adjémian J, Anne A, Cauet G, Demaille C. Cleavage-sensing redox peptide monolayers for the rapid measurement of the proteolytic activity of trypsin and alpha-thrombin enzymes. *Langmuir* 2010;26(12):10347–56.
- [119] Franco OL, dos Santos RC, Batista JAN, Mendes ACM, Araújo MAM de, Monnerat RG, Grossi-de-Sá MF, Freitas SM de. Effects of black-eyed pea trypsin/chymotrypsin inhibitor on proteolytic activity and on development of *Anthonomus grandis*. *Phytochemistry* 2003;63(3):343–9.
- [120] Tanka-Salamon A, Bóta A, Wacha A, Mihály J, Lovas M, Kolev K. Structure and Function of Trypsin-Loaded Fibrinolytic Liposomes. *BioMed Research International* 2017;2017:5130495.
- [121] Wang H, Li S, Wang J, Chen S, Sun X-l, Wu Q. N-glycosylation in the protease domain of trypsin-like serine proteases mediates calnexin-assisted protein folding. *eLife Sciences Publications, Ltd*, 6 November 2018; Available from: <https://elifesciences.org/articles/35672> (20 May 2020).
- [122] Simon LM, Kotormán M, Garab G, Laczkó I. Structure and activity of alpha-chymotrypsin and trypsin in aqueous organic media. *Biochemical and Biophysical Research Communications* 2001;280(5):1367–71.
- [123] Schmidt A, Jelsch C, Ostergaard P, Rypniewski W, Lamzin VS. Trypsin revisited: crystallography AT (SUB) atomic resolution and quantum chemistry revealing details of catalysis. *J Biol Chem* 2003;278(44):43357–62.
- [124] Olsen JV, Ong S-E, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* 2004;3(6):608–14.

- [125] Miike S, McWilliam AS, Kita H. Trypsin induces activation and inflammatory mediator release from human eosinophils through protease-activated receptor-2. *J Immunol* 2001;167(11):6615–22.
- [126] Kotzia GA, Lappa K, Labrou NE. Tailoring structure-function properties of L-asparaginase: engineering resistance to trypsin cleavage. *Biochem J* 2007;404(2):337–43.
- [127] Koshikawa N, Hasegawa S, Nagashima Y, Mitsuhashi K, Tsubota Y, Miyata S, Miyagi Y, Yasumitsu H, Miyazaki K. Expression of trypsin by epithelial cells of various tissues, leukocytes, and neurons in human and mouse. *Am J Pathol* 1998;153(3):937–44.
- [128] Hirota M, Ohmuraya M, Baba H. The role of trypsin, trypsin inhibitor, and trypsin receptor in the onset and aggravation of pancreatitis. *J Gastroenterol* 2006;41(9):832–6.
- [129] Gui F, Zhang Y, Wan J, Zhan X, Yao Y, Li Y, Haddock AN, Shi J, Guo J, Chen J, Zhu X, Edenfield BH, Zhuang L, Hu C, Wang Y, Mukhopadhyay D, Radisky ES, Zhang L, Lugea A, Pandol SJ, Bi Y, Ji B. Trypsin activity governs increased susceptibility to pancreatitis in mice expressing human PRSS1R122H. *J Clin Invest* 2020;130(1):189–202.
- [130] Christiansen S. The Health Benefits of Trypsin. Verywell Health, 11/30/2019; Available from: <https://www.verywellhealth.com/trypsin-4777708> (20 May 2020).
- [131] Bajaj-Elliott M. Trypsin and host defence: a new role for an old enzyme. *Gut* 2003;52(2):166–7.
- [132] Blankenship E, Vukoti K, Miyagi M, Lodowski DT. Conformational flexibility in the catalytic triad revealed by the high-resolution crystal structure of *Streptomyces erythraeus* trypsin in an unliganded state. *Acta Crystallogr D Biol Crystallogr* 2014;70(Pt 3):833–40.
- [133] Carter P, Wells JA. Dissecting the catalytic triad of a serine protease. *Nature* 1988;332(6164):564–8.
- [134] Ekici OD, Paetzel M, Dalbey RE. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci* 2008;17(12):2023–37.
- [135] Ma W, Tang C, Lai L. Specificity of trypsin and chymotrypsin: loop-motion-controlled dynamic correlation as a determinant. *Biophys J* 2005;89(2):1183–93.
- [136] Sichler K, Hopfner K-P, Kopetzki E, Huber R, Bode W, Brandstetter H. The influence of residue 190 in the S1 site of trypsin-like serine proteases on substrate selectivity is universally conserved. *FEBS Letters* 2002;530(1-3):220–4.
- [137] Perona JJ, Hedstrom L, Rutter WJ, Fletterick RJ. Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* 1995;34(5):1489–99.
- [138] Czapinska H, Otlewski J. Structural and energetic determinants of the S1-site specificity in serine proteases. *European Journal of Biochemistry* 1999;260(3):571–95.
- [139] Badran MJ, Bertolotti N, Keils A, Heine A, Klebe G, Marchais-Oberwinkler S. Mutational and structural studies uncover crucial amino acids determining activity and stability of 17 $\beta$ -HSD14. *J Steroid Biochem Mol Biol* 2019;189:135–44.
- [140] Lukacik P, Keller B, Bunkoczi G, Kavanagh KL, Kavanagh K, Lee WH, Hwa Lee W, Adamski J, Oppermann U. Structural and biochemical characterization of human orphan DHRS10



- p reveals a novel cytosolic enzyme with steroid dehydrogenase activity.
- Biochem J*
- 2007;402(3):419–27.
- [141] Marchais-Oberwinkler S, Henn C, Möller G, Klein T, Negri M, Oster A, Spadaro A, Werth R, Wetzel M, Xu K, Frotscher M, Hartmann RW, Adamski J. 17 $\beta$ -Hydroxysteroid dehydrogenases (17 $\beta$ -HSDs) as therapeutic targets: protein structures, functions, and recent progress in inhibitor development. *J Steroid Biochem Mol Biol* 2011;125(1-2):66–82.
  - [142] Bertoletti N, Braun F, Lepage M, Möller G, Adamski J, Heine A, Klebe G, Marchais-Oberwinkler S. New Insights into Human 17 $\beta$ -Hydroxysteroid Dehydrogenase Type 14: First Crystal Structures in Complex with a Steroidal Ligand and with a Potent Nonsteroidal Inhibitor. *J Med Chem* 2016;59(14):6961–7.
  - [143] Poirier D. Inhibitors of 17 beta-hydroxysteroid dehydrogenases. *Current medicinal chemistry* 2003;10(6):453–77.
  - [144] Brozic P, Lanisnik Risner T, Gobec S. Inhibitors of 17beta-hydroxysteroid dehydrogenase type 1. *Current medicinal chemistry* 2008;15(2):137–50.
  - [145] Frotscher M, Ziegler E, Marchais-Oberwinkler S, Kruchten P, Neugebauer A, Fetzer L, Scherer C, Müller-Vieira U, Messinger J, Thole H, Hartmann RW. Design, synthesis, and biological evaluation of (hydroxyphenyl)naphthalene and -quinoline derivatives: potent and selective nonsteroidal inhibitors of 17beta-hydroxysteroid dehydrogenase type 1 (17beta-HSD1) for the treatment of estrogen-dependent diseases. *J Med Chem* 2008;51(7):2158–69.
  - [146] Bey E, Marchais-Oberwinkler S, Negri M, Kruchten P, Oster A, Klein T, Spadaro A, Werth R, Frotscher M, Birk B, Hartmann RW. New insights into the SAR and binding modes of bis(hydroxyphenyl)thiophenes and -benzenes: influence of additional substituents on 17beta-hydroxysteroid dehydrogenase type 1 (17beta-HSD1) inhibitory activity and selectivity. *J Med Chem* 2009;52(21):6724–43.
  - [147] Schuster D, Nashev LG, Kirchmair J, Laggner C, Wolber G, Langer T, Odermatt A. Discovery of nonsteroidal 17beta-hydroxysteroid dehydrogenase 1 inhibitors by pharmacophore-based screening of virtual compound libraries. *J Med Chem* 2008;51(14):4188–99.
  - [148] Wetzel M, Marchais-Oberwinkler S, Hartmann RW. 17 $\beta$ -HSD2 inhibitors for the treatment of osteoporosis: Identification of a promising scaffold. *Bioorganic & medicinal chemistry* 2011;19(2):807–15.
  - [149] Haeseleer F, Palczewski K. [24] Short-chain dehydrogenases/reductases in retina. In: *Methods in Enzymology Proteolytic Enzymes: Serine and Cysteine Peptidases*, Vol. 316. Academic Press; 1994. p. 372–83.
  - [150] Sivik T, Gunnarsson C, Fornander T, Nordenskjöld B, Skoog L, Stål O, Jansson A. 17 $\beta$ -Hydroxysteroid dehydrogenase type 14 is a predictive marker for tamoxifen response in oestrogen receptor positive breast cancer. *PLOS ONE* 2012;7(7):e40568.
  - [151] Sivik T, Vikingsson S, Gréen H, Jansson A. Expression patterns of 17 $\beta$ -hydroxysteroid dehydrogenase 14 in human tissues. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et métabolisme* 2012;44(13):949–56.
  - [152] Breton R, Housset D, Mazza C, Fontecilla-Camps JC. The structure of a complex of human 17 $\beta$ -hydroxysteroid dehydrogenase with estradiol and NADP<sup>+</sup> identifies two principal targets for the design of inhibitors. *Structure* 1996;4(8):905–15.

- [153] Azzi A, Rehse PH, Zhu DW, Campbell RL, Labrie F, Lin SX. Crystal structure of human estrogenic 17 beta-hydroxysteroid dehydrogenase complexed with 17 beta-estradiol. *Nat Struct Mol Biol* 1996;3(8):665–8.
- [154] Agilent Technologies. Agilent Technologies; Available from: <https://www.chem-agilent.com/pdf/strata/200133.pdf>.
- [155] GraphPadPrism 7.0.
- [156] Native Gel Electrophoresis, <http://www.proteinchemist.com/tutorial/natpage.html>; Available from: <http://www.proteinchemist.com/tutorial/natpage.html>.
- [157] Mueller U, Darowski N, Fuchs MR, Förster R, Hellmig M, Paithankar KS, Pühringer S, Steffien M, Zocher G, Weiss MS. Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J Synchrotron Radiat* 2012;19(Pt 3):442–9.
- [158] Krug M, Weiss MS, Heinemann U, Mueller U. XDSAPP a graphical user interface for the convenient processing of diffraction data using XDS. *J Appl Cryst* 2012;45(3):568–72.
- [159] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Cryst* 2007;40(Pt 4):658–74.
- [160] Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Cryst D* 1994;50(Pt 5):760–3.
- [161] Neudert G, Klebe G. fconv: Format conversion, manipulation and feature computation of molecular data. *Bioinformatics* (Oxford, England) 2011;27(7):1021–2.
- [162] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Cryst D* 2004;60(Pt 12 Pt 1):2126–32.
- [163] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010;66(Pt 2):213–21.
- [164] O. S. Smart, T. O. Womack, A. Sharff, C. Flensburg. Grade.
- [165] Molinspiration Cheminformatics. Molinspiration. Molinspiration Cheminformatics.
- [166] [http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/).
- [167] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26(2):283–91.
- [168] Kleywegt GJ, Zou J-Y, Kjeldgaard M, Jones TA. Around O. In: Rossmann MG, Arnold E (editors). *International Tables for Crystallography*. 1st ed. New York: Springer; 2006. p. 353–6.
- [169] 43. [www.bioinf.org.uk](http://www.bioinf.org.uk) Dr. Andrew C.R. Martin's Group at UCL, (n.d.). <http://www.bioinf.org.uk/profit/> (January 2019). 43. [www.bioinf.org.uk](http://www.bioinf.org.uk) Dr. Andrew C.R. Martin's Group at UCL, (n.d.). <http://www.bioinf.org.uk/profit/> (January 2019). 43. [www.bioinf.org.uk](http://www.bioinf.org.uk) Dr. Andrew C.R. Martin's Group at UCL, (n.d.). <http://www.bioinf.org.uk/profit/> (January 2019).

- [170] Puranen T, Poutanen M, Ghosh D, Vihko P, Vihko R. Characterization of structural and functional properties of human 17 beta-hydroxysteroid dehydrogenase type 1 using recombinant enzymes and site-directed mutagenesis. *Mol Endocrinol* 1997;11(1):77–86.
- [171] Mehrabi P, Müller-Werkmeister HM, Leimkohl JP, Schikora H, Ninkovic J, Krivokuca S, Andriček L, Epp SW, Sherrell D, Owen RL, Pearson AR, Tellkamp F, Schulz EC, Miller RJD. The HARE chip for efficient time-resolved serial synchrotron crystallography. *J Synchrotron Radiat* 2020;27(Pt 2):360–70.
- [172] Wolff AM, Young ID, Sierra RG, Brewster AS, Martynowycz MW, Nango E, Sugahara M, Nakane T, Ito K, Aquila A, Bhowmick A, Biel JT, Carbajo S, Cohen AE, Cortez S, Gonzalez A, Hino T, Im D, Koralek JD, Kubo M, Lazarou TS, Nomura T, Owada S, Samelson AJ, Tanaka T, Tanaka R, Thompson EM, van den Bedem H, Woldeyes RA, Yumoto F, Zhao W, Tono K, Boutet S, Iwata S, Gonen T, Sauter NK, Fraser JS, Thompson MC. Comparing serial X-ray crystallography and microcrystal electron diffraction (MicroED) as methods for routine structure determination from small macromolecular crystals. *IUCrJ* 2020;7(Pt 2):306–23.
- [173] Nam KH. Shortening injection matrix for serial crystallography. *Sci Rep* 2020;10(1):107.
- [174] La Mora E de, Coquelle N, Bury CS, Rosenthal M, Holton JM, Carmichael I, Garman EF, Burghammer M, Colletier J-P, Weik M. Radiation damage and dose limits in serial synchrotron crystallography at cryo- and room temperatures. *Proc Natl Acad Sci U S A* 2020;117(8):4142–51.
- [175] Chapman HN. X-Ray Free-Electron Lasers for the Structure and Dynamics of Macromolecules. *Annu Rev Biochem* 2019;88:35–58.
- [176] Deutsches Elektronen-Synchrotron DESY. Deutschlands größtes Beschleunigerzentrum - Deutsches Elektronen-Synchrotron DESY; Available from: <https://www.desy.de/> (31 March 2020).
- [177] Swiss Light Source - SLS; Available from: <https://www.psi.ch/en/sls> (31 March 2020).
- [178] MAX IV – We make the invisible visible; Available from: <https://www.maxiv.lu.se/> (31 March 2020).
- [179] European Synchrotron Radiation Facility (ESRF); Available from: <https://www.esrf.eu/> (31 March 2020).
- [180] Materialien und Energie, Helmholtz-Zentrum Berlin für. BESSY II Electron storage ring; Available from: [https://www.helmholtz-berlin.de/forschung/quellen/bessy/index\\_en.html](https://www.helmholtz-berlin.de/forschung/quellen/bessy/index_en.html) (31 March 2020).
- [181] Source DL. Diamond Light Source; Available from: <https://www.diamond.ac.uk/Home.html?jsessionid=992C01462B17998AB9DA3EB07BB1ABD1> (31 March 2020).
- [182] Roedig P, Duman R, Sanchez-Weatherby J, Vartiainen I, Burkhardt A, Warmer M, David C, Wagner A, Meents A. Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *J Appl Crystallogr* 2016;49(Pt 3):968–75.
- [183] Lieske J, Cerv M, Kreida S, Komadina D, Fischer J, Barthelmess M, Fischer P, Pakendorf T, Yefanov O, Mariani V, Seine T, Ross BH, Crosas E, Lorbeer O, Burkhardt A, Lane TJ, Guenther S, Bergtholdt J, Schoen S, Törnroth-Horsefield S, Chapman HN, Meents A. On-chip

- p>crystallization for serial crystallography experiments and on-chip ligand-binding studies.
- IUCrJ*
- 2019;6(Pt 4):714–28.
- [184] Zarrine-Afsar A, Barends TRM, Müller C, Fuchs MR, Lomb L, Schlichting I, Miller RJD. Crystallography on a chip. *Acta crystallographica. Section D, Biological crystallography* 2012;68(Pt 3):321–3.
- [185] Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction. *Proc. R. Soc. Lond. B* 1990;241(1300):6–8.
- [186] Kupitz C, Grotjohann I, Conrad CE, Roy-Chowdhury S, Fromme R, Fromme P. Microcrystallization techniques for serial femtosecond crystallography using photosystem II from *Thermosynechococcus elongatus* as a model system. *Philos Trans R Soc Lond B Biol Sci* 2014;369(1647):20130316.
- [187] Tajima M, Urabe I, Yutani K, Okada H. Role of calcium ions in the thermostability of thermolysin and *Bacillus subtilis* var. *amylosacchariticus* neutral protease. *European Journal of Biochemistry* 1976;64(1):243–7.
- [188] Pelmeshnikov V, Blomberg MRA, Siegbahn PEM. A theoretical study of the mechanism for peptide hydrolysis by thermolysin. *J Biol Inorg Chem* 2002;7(3):284–98.
- [189] Krimmer SG, Cramer J, Schiebel J, Heine A, Klebe G. How Nothing Boosts Affinity: Hydrophobic Ligand Binding to the Virtually Vacated S1' Pocket of Thermolysin. *Journal of the American Chemical Society* 2017;139(30):10419–31.
- [190] English AC, Done SH, Caves LSD, Groom CR, Hubbard RE. Locating interaction sites on proteins: The crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins: Structure, Function, and Bioinformatics* 1999;37(4):628–40.
- [191] Francesca Magari. The power of fragments: fblD aprouch to investigate protein struture; 2020.
- [192] Alke Meents, Bernd Reime, Nicolas Stuebe, Pontus Fischer, Martin Warmer, Dennis Goeries, Jan Roever, Jan Meyer, Janine Fischer, Anja Burkhardt, Ismo Vartiainen, Petri Karvinen, Christian David. Development of an in-vacuum x-ray microscope with cryogenic sample cooling for beamline P11 at PETRA III. In: *International Society for Optics and Photonics*; 2013. 88510K.
- [193] Dauter Z. [21] Data collection strategy. In: *Methods in Enzymology Macromolecular Crystallography Part A*, Vol. 276. Academic Press; 1997. p. 326–44.
- [194] Foadi J, Aller P, Alguel Y, Cameron A, Axford D, Owen RL, Armour W, Waterman DG, Iwata S, Evans G. Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* 2013;69(Pt 8):1617–32.
- [195] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* 2011;67(Pt 4):235–42.
- [196] Schiebel J, Krimmer SG, Röwer K, Knörlein A, Wang X, Park AY, Stieler M, Ehrmann FR, Fu K, Radeva N, Krug M, Huschmann FU, Glöckner S, Weiss MS, Mueller U, Klebe G, Heine A. High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits. *Structure* 2016;24(8):1398–409.

- [197] Carr R, Jhoti H. Structure-based screening of low-affinity compounds. *Drug Discovery Today* 2002;7(9):522–7.
- [198] Konkel MJ, Packiarajan M, Chen H, Topiwala UP, Jimenez H, Talisman IJ, Coate H, Walker MW. Amino substituted analogs of 1-phenyl-3-phenylimino-2-indolones with potent galanin Gal3 receptor binding affinity and improved solubility. *Bioorganic & Medicinal Chemistry Letters* 2006;16(15):3950–4.
- [199] Luong JH, Male KB, Nguyen AL. Synthesis and characterization of a water-soluble affinity polymer for trypsin purification. *Biotechnology and Bioengineering* 1988;31(5):439–46.
- [200] Cogan U, Kopelman M, Mokady S, Shinitzky M. Binding affinities of retinol and related compounds to retinol binding proteins. *European Journal of Biochemistry* 1976;65(1):71–8.
- [201] Jansook P, Ogawa N, Loftsson T. Cyclodextrins: structure, physicochemical properties and pharmaceutical applications. *International Journal of Pharmaceutics* 2018;535(1-2):272–84.
- [202] Kalepu S, Nekkanti V. Insoluble drug delivery strategies: review of recent advances and business prospects. *Acta Pharmaceutica Sinica B* 2015;5(5):442–53.
- [203] Legrand F-X, Sauthier M, Flahaut C, Hachani J, Elfakir C, Fourmentin S, Tilloy S, Monflier E. Aqueous hydroformylation reaction mediated by randomly methylated  $\beta$ -cyclodextrin: How substitution degree influences catalytic activity and selectivity. *Journal of Molecular Catalysis A: Chemical* 2009;303(1-2):72–7.
- [204] Loftsson T, Jarho P, Másson M, Järvinen T. Cyclodextrins in drug delivery. *Expert opinion on drug delivery* 2005;2(2):335–51.
- [205] Szente L, Fenyvesi É. Cyclodextrin-Lipid Complexes: Cavity Size Matters. *Struct Chem* 2017;28(2):479–92.
- [206] Kalyani VS, Malkhede DD, Mohanty J. Cyclodextrin-assisted modulation of the photophysical properties and acidity constant of pyrene-armed calix4arene. *Physical chemistry chemical physics PCCP* 2017;19(32):21382–9.
- [207] Liebschner D, Afonine PV, Moriarty NW, Poon BK, Sobolev OV, Terwilliger TC, Adams PD. Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Cryst D* 2017;73(Pt 2):148–57.
- [208] Congreve M, Carr R, Murray C, Jhoti H. A ‘Rule of Three’ for fragment-based lead discovery? *Drug Discovery Today* 2003;8(19):876–7.
- [209] Erlanson DA, McDowell RS, O'Brien T. Fragment-based drug discovery. *J Med Chem* 2004;47(14):3463–82.
- [210] Patil R, Das S, Stanley A, Yadav L, Sudhakar A, Varma AK. Optimized hydrophobic interactions and hydrogen bonding at the target-ligand interface leads the pathways of drug-designing. *PLoS One* 2010;5(8):e12029.
- [211] Haglund ÅC, Marsden NVB. Hydrophobic and polar contributions to solute affinity for a highly crosslinked water-swollen (sephadex) gel. *J. Polym. Sci. B Polym. Lett. Ed.* 1980;18(4):271–9.
- [212] Lee S, Barron MG. Structure-Based Understanding of Binding Affinity and Mode of Estrogen Receptor  $\alpha$  Agonists and Antagonists. *PLoS One* 2017;12(1):e0169607.

- [213] Klebe G. Protein–Ligand Interactions as the Basis for Drug Action. In: Klebe G (editor). Drug design. New York: Springer; 2013. p. 61–88.
- [214] Bijelic A, Rompel A. Polyoxometalates: more than a phasing tool in protein crystallography. *ChemTexts* 2018;4(3):10.
- [215] Schiebel J, Gaspari R, Wulsdorf T, Ngo K, Sohn C, Schrader TE, Cavalli A, Ostermann A, Heine A, Klebe G. Intriguing role of water in protein-ligand binding studied by neutron crystallography on trypsin complexes. *Nat Commun* 2018;9(1):3559.
- [216] Rauh D, Reyda S, Klebe G, Stubbs MT. Trypsin mutants for structure-based drug design: expression, refolding and crystallisation. *Biological Chemistry* 2002;383(7-8):1309–14.
- [217] Renatus M, Bode W, Huber R, Stürzebecher J, Stubbs MT. Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: implications for the inhibition of factor Xa, tPA, and urokinase. *J Med Chem* 1998;41(27):5445–56.
- [218] Juers DH, Ruffin J. MAP\_CHANNELS: a computation tool to aid in the visualization and characterization of solvent channels in macromolecular crystals. *J Appl Cryst* 2014;47(Pt 6):2105–8.

## Acknowledgments

Better to light a candle than to curse the darkness. I would like to express my sincere gratitude and appreciation to my mentor and supervisor **Prof. Gerhard Klebe**. Who was the beacon of light and hope that guided me through my Ph.D study. Without his patience, knowledge and support both scientifically and personally, I would have not been able to proceed with this work. It has been an honour and privilege to know him. A great scientist and an even greater human being.

I am very thankful to my second supervisor **Prof. Andreas Heine** for introducing me to macromolecular X-ray crystallography, for sharing his vast knowledge with me and for the great amount of time he invested in me. Thanks for the many insightful discussions and for always having an open door for questions and valuable suggestion.

I want to thank **Dr. Sandrine Marchais-Oberwinkler** who believed in me. It would not have been possible to proceed with my Ph.D study without her support. I can not express enough how her enthusiasm, experience, knowledge and liveliness have been essential during my journey.

I want to thank a dear friend and college **Dr. Stefan Merkl** for all the wonderful time we had together. His care, wisdom, friendly and “adventurous” nature defined him for me. Knowing him during my study was a pleasure and a great advantage. Signature move “remove glasses and points towards when he thinks that he got the right argument”.

A thanks from the heart goes to **Dr. Felix Terwesten** who was a great friend and college to me during our study together and even after it. His friendship and support were most valuable to me throughout my journey. His hard work and effort ensured our network and internal system to run flawlessly. Signature move “head shaking horizontally the vertically with a questioning face”.

A big thanks goes to **Dr. Alexander Metz** for his help and support during my study. His friendly and smooth approach were a huge help to me to complete my work, He was able to draw a smile on my face during the hardest times. We experienced a lot together as part of the Foodies club.

I want to thank **Dr. Khang Ngo** and **Ahmed Merabet** for their friendship and help. The memories we made together during our study (scientifically and personally) will be cherished forever.

A big thanks goes to my longest serving office mate **Dr. Francesca Magari** (maybe) for all the lovely memories and for standing my bad taste in humour. I want to thank all of my office mates that I was fortunate to be in their company throughout my study, **Dr. Marko Kljajic**, **Dr. Engi Hassaan**.

A big fat thanks goes to **Magdalena Scharf** (Biceps), a dear friend whose intelligence is matched only by her kindness and liveliness.

To my neighbours in office 03A16, thank for all the great time we had together. **Maurice Sebastiani** (Schnorrise) for being an amazing friend and colleague and **Stefanie Dörr** for her support in the lab and her kindness. I enjoyed the fresh air with both of them. I want to thank **Matthias Oebbeke** (Captain), **Lea Hubert** and **Marina Ley**.

I would like to thank **William Scanlan** for all the wonderful memories we had in the past.

I would like to express my thanks and gratitude to the HSD team **Dr. Nicole Bertoletti** and **Dr. Florian Braun**. For their help and support in the beginning of my Ph.D work

I want to thank my friends and colleagues from the **AG Klebe**. In particular, **Dr. Christoph Sager**, **Dzung Nguyen**, **Andreas Nguyen**, **Anna Sandner**, **Steffen Glöckner**.

I am grateful to **Lydia Hartleben** for her administrative support, patience and kindness, to know her was an honour and a pleasure. **Christian Sohn** for his technical help and great support in several occasions. Thanks to **Hans-Dieter Gerber** for his kindness, support and for sharing his knowledge and saving my work during the water crisis.

I am very grateful to the MX-team at BESSY II (Helmholtz-Zentrum Berlin, Germany) for their enthusiasm, advice during data collection and sharing their valuable knowledge with me. I would particularly like to acknowledge the help and support of **Dr. Manfred Weiss**, **Dr. Christian Feiler**, **Dr. Martin Gerlach** and **Dr. Jan Wollenhaupt**. Every time I went to BESSY, I always learned something new from each of them.

I want to thank **AG Kolb** for contributing to the nice working environment. In particular, I would like to **Lim Jun Yu Victor**, **Johanna Senst**, **Janik Hedderich** and **Torben Gutermuth** for the lovely memories and conversations.



Finally, a heartfelt thanks to my family for their constant encouragement and support during my studies. Everything that is good in me I learned from them. They always push me to be a better version of myself. I hope I can be the person who they aspire me to be and make them proud.



# Curriculum Vitae

## Personal Details

---

**Name** Mohammed Badran  
**Date of birth** June 6<sup>th</sup>, 1991  
**Place of birth** Damascus, Syria  
**Nationality** Syrian

## Education

---

10/2020 **PhD (Dr. rer. nat) in Life Science**  
(expected) Institute of Pharmaceutical Chemistry, Philipps University of Marburg, Germany  
03/2015 **National State Examination, Board License as a Pharmacist in Syria**  
2009-2014 **Study of Pharmacy**  
AIU university (Damascus)

## Research Experience

---

Since 01/2016 **Dissertation**  
Drug design group of Prof. Dr. Gerhard Klebe, Institute of Pharmaceutical Chemistry, Philipps University of Marburg, Germany  
08/2016-11/2016 **Student Internship**  
Group of Prof. Dr. Gerhard Klebe, Institute of Pharmaceutical Chemistry, Philipps University of Marburg, Germany  
01/2013-06/2013 **Graduation research project (fifth year)**  
Pharmacology and Toxicology department, AIU university (Damascus)

## Professional Experience

---

2017/2018 **Teaching Assistant**, Philipps University of Marburg, Germany  
09/2013-06/2014 **Pharmacist Intern**, Central Pharmacy in Damascus City-Centre  
04/2015-06/2016 **Pharmacist**, Ashrafiah Pharmacy, Damascus  
09/2018-09/2020 **Community Servant**  
*Bürgerbeteiligungsrat/Citizen Participation Council, University city of Marburg*  
Since 05/2020 **Corona Hilfsgruppe/Corona help team**, The magistrate of Marburg

## Journal Publications

---

- Badran M.J., Bertoletti N, Keils A, Heine A, Klebe G, Marchais-Oberwinkler S. “Mutational and structural studies uncover crucial amino acids determining activity and stability of 17 $\beta$ -HSD14”. *Journal of Steroid Biochemistry and Molecular Biology* **2019**. 189: 135-144
- Magari F, Hassan E, Abazi N, Badran M.J., Bertoletti N, Glöckner S, Huschmann F, Metz A, Siefker C, Scanlan J.H.W, Heine A, Weiss M, Klebe G. “Exhaustive X-ray Crystallographic Screening of a Hit-Enriched 96 fragment library. Manuscript in preparation”.

## Poster presentation

---

### **DPhG Jahrestagung 2018 / Annual meeting, October 3–5, 2018 in Munich, Germany**

Poster on “Mutated variants of 17 $\beta$ -Hydroxysteroid dehydrogenase type 14 Studying the functional role of residues next to the catalytic center of the enzyme”



# Erklärung

Ich versichere, dass ich meine Dissertation

*“Mutational Studies on 17 $\beta$ -HSD14, Serial Synchrotron X-ray Crystallography, Solubility  
Enhancement using Cyclodextrins and Fragment-Based Drug Discovery  
Multiple Blocks to Pave the Road of Drug Design “*

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen sind Zitate als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den .....

.....

(Mohammed Badran)

