Structure-function studies and drug discovery targeting penicillin binding protein 3 from *Escherichia coli* and *Pseudomonas aeruginosa*

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Stefan Christian Freischem
aus Düsseldorf

Düsseldorf, 25. Januar 2022

aus dem Institut für Physikalische Biologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. apl. Prof. Dr. Andrew J. Dingley

2. Prof. Dr. Dieter Willbold

Tag der mündlichen Prüfung: 25.01.2022

Eidesstattliche Erklärung

Ich versichere an Eides Statt, dass die Dissertation von mir selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf“ erstellt worden ist.

Düsseldorf, den 25.01.2022

_____

(Stefan Christian Freischem)

Abstract

The lack of antibiotics against multidrug resistant (MDR) bacteria is one of the major challenges of medicine in the 21st century. In particular, Gram-negative bacteria, such as *Pseudomonas aeruginosa* (*P. aeruginosa*) *and Escherichia coli* (*E. coli*), are responsible for thousands of deaths every year. Identifying new classes of antibiotics that are active against MDR bacteria is vital to combat against the growing number of bacteria that have developed and adapted mechanisms against current antibiotics.

In this thesis, a fragment-based drug discovery (FBDD) approach targeting the penicillin binding protein 3 (PBP3) was performed using ^{19}F nuclear magnetic resonance (NMR) as a screening method to identify small molecules, termed fragments, binding PBP3. Seventeen hit fragments were identified from the two screens conducted and may be used in subsequent chemistry approaches, e.g., linking and merging, to synthesize a drug with high (i.e., < nM) affinity toward PBP3 with inhibitory activity. To analyze the identified hit fragments and validate their interaction with PBP3, two-dimensional (2D) heteronuclear NMR experiments of PBP3 were recorded to perform a chemical shift perturbation analysis, revealing a common binding site close to the active site for four hit fragments. Competition binding using the novel β-lactam antibiotic AIC499, which covalently binds the active site, was used to gain further insights into the binding mode of the fragments. Interestingly, binding of two fragments was found to be enhanced by bound AIC499. Furthermore, in a ^{19}F NMR titration experiment, the K_d value of one of the hit fragments was estimated to be 1.23 ± 0.54 mM, which is in the expected affinity range for FBDD.

In addition to the FBDD approach targeting PBP3 from *E. coli*, AIC499 that is active in the presence of β-lactamases was co-crystallized with PBP3s from *E. coli* and *P. aeruginosa* and the binding mode was analyzed. Moreover, the apo proteins were crystallized and the structures of apo PBP3 from *P. aeruginosa* provided novel insights into the β3-β4 loop, which is a hotspot for various mutations related to antibiotic resistance. Additionally, a novel extended construct (termed *Ec*TPd*) featuring the catalytic transpeptidase domain (TPd) from *E. coli* PBP3 was produced and crystallized in the apo form and in the presence of AIC499, yielding samples with improved diffraction quality. In a parallel effort to develop a new class of antibiotics, pyrrolidine-2,3-dione derivatives were developed in high-throughput screening (HTS) and their binding to PBP3 was analyzed by surface plasmon resonance (SPR), NMR and fluorescence spectroscopy. The results from these experiments validated the interaction of these derivatives with PBP3 from *P. aeruginosa* and K_d values were determined.

In summary, fragments, found in the FBDD approach and the novel AIC499 and pyrrolidine-2,3-dione derivatives, investigated in this thesis, represent promising candidates in future development of new classes of antibiotics. Furthermore, it was possible to gain additional information on the structure of PBP3s and sequence-specifically assign approximately 40% of the peaks in the 2D ^{1}H-^{15}N transverse relaxation-optimized spectroscopy heteronuclear single quantum coherence (TROSY-HSQC) spectrum of *Ec*TPd*. These results may be crucial requirements when future antibiotic development targeting PBP3 is conducted.

Zusammenfassung

Der Mangel an Antibiotika, welche gegen multiresistente (MDR) Bakterien wirksam sind, ist eine der größten Herausforderungen der medizinischen Forschung im 21. Jahrhundert. Vor allem Gram-negative Bakterien, wie *Pseudomonas aeruginosa* (*P. aeruginosa*) und *Escherichia coli* (*E. coli*) sind verantwortlich für jährlich Tausende Tote. Aufgrund ihrer Fähigkeit Mechanismen gegen neue Antibiotika zu entwickeln und anzupassen, ist es unumgänglich neue Antibiotikaklassen gegen MDR Keime zu entwickeln.

In dieser Arbeit wurde mit Hilfe von ^{19}F-kernmagnetischer Resonanz (NMR)-Spektroskopie als Selektionsmethode eine Fragment-basierte Medikamenten-Entwicklung (FBDD) gegen das Penicillin-bindende Protein 3 (PBP3) durchgeführt, um kleine, das Protein bindende Moleküle zu identifizieren (sogenannte Fragmente). Dabei wurden in zwei Runden siebzehn Fragmente identifiziert, bei denen eine initiale Bindung gegen PBP3 nachgewiesen werden konnte. Diese können in der weiteren Entwicklung zum Vergrößern („growing“), Verknüpfen („linking“), oder Verschmelzen („merging“) genutzt werden um größerer Moleküle zu generieren, die höhere Affinitäten (< nM) gegen das Zielprotein aufweisen und dessen Aktivität inhibieren. Zur Validierung und Analyse der Interaktion von identifizierten Fragmenten mit dem Zielprotein wurden zweidimensionale (2D) NMR-Experiment für eine Analyse der Änderungen der chemischen Verschiebung aufgenommen. Dabei wurden vier Fragmente gefunden, die das PBP3 in der Nähe des aktiven Zentrums binden. In einer kompetitiven Bindungsstudie mit dem neu entwickelten β-Lactam-Antibiotikum AIC499, welches kovalent im aktiven Zentrum bindet, wurden zusätzliche Informationen über den Bindungsmodus der Fragmente erhoben. Interessanterweise wurde die Bindung zweier Fragmente durch das kovalent gebundene AIC499 verstärkt. Des Weiteren wurde mit Hilfe von ^{19}F NMR-Experimenten der K_d-Wert eines Fragments auf 1.23 ± 0.54 mM bestimmt, was im zu erwarteten Affinitätsbereich bei der Bindung von Fragmenten liegt.

Zusätzlich wurde in dieser Arbeit das neu entwickelte β-Lactam-Antibiotikum AIC499, welches sogar in Anwesenheit von β-Lactamasen antimikrobielle Aktivität zeigt, mit den PBP3s von *E. coli* und *P. aeruginosa* co-kristallisiert, um den Bindemechanismus zu analysieren. Im Zuge dessen wurden die apo-Proteine kristallisiert und neue Einblicke in die β3-β4-Schleife des *P.-aeruginosa*-PBP3 gewonnen, welche im Zusammenhang mit mehreren Mutationen von Antibiotika-resistenten PBP3s steht. Zusätzlich wurde ein neues Konstrukt der katalytischen Transpeptidasedomäne (TPd) des *E.-coli*-PBP3 produziert und kristallisiert (*Ec*TPd*), so dass *Ec*TPd*-Kristalle mit und ohne AIC499 eine verbesserte Diffraktionsqualität ergaben. In weiteren Bemühungen eine neue Antibiotikaklasse zu entwickeln, wurden Pyrrolidin-2,3-dion-Derivate in einem Hochdurchsatz-Screening (HTS) entwickelt und die Interaktion mit PBP3 analysiert. Dazu wurden die biophysikalischen Methoden der Oberflächen-Plasmon-Resonanz (SPR)-, NMR- und Fluoreszenz-Spektroskopie genutzt, die *P.-aeruginosa*-PBP3-Binding validiert und K_d-Werte ermittelt.

Letztendlich können die Fragmente, die in dem FBDD-Ansatz gefunden wurden, AIC499 und die Pyrrolidin-2,3-dion-Derivate als vielversprechende Kandidaten in der zukünftigen Antibiotika-Entwicklung betrachtet werden. Außerdem wurden neue strukturelle Daten generiert und ungefähr 40% der Peaks in einem 2D ^{1}H-^{15}N Transversalrelaxation-optimierter Spektroskopie heteronukleare Einzelquantenkohärenz (TROSY-HSQC) -Spektrum der *Ec*TPd*-Probe sequenz-spezifisch zugeordnet. Diese Ergebnisse könnten für die zukünftige Antibiotika-Entwicklung gegen PBP3 eine wichtige Voraussetzung darstellen.

Table of content

List of Abbreviations

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
4D	four-dimensional
A. baumannii	*Acinetobacter baumannii*
ADME	absorption, distribution, metabolism and excretion
BMRB	Biological Magnetic Resonance Data Bank
CD	circular dichroism
CPMG	Carr-Purcell-Meiboom-Gill
cryo-EM	cryo-electron microscopy
CSA	chemical shift anisotropy
CSAR	chemical shift-anisotropy-based affinity ranking
D-Ala	D-alanine
DESY	Deutsches Elektronen-Synchrotron
D-Glu	D-glutamate
DHP-1	dehydropeptidase-1
DHPS	dihydropteroate synthase
DMSO	dimethyl sulfoxide
DOSY	diffusion ordered spectroscopy
DSF	differential scanning fluorimetry
DSS	Na-trimethylsilylpropanesulfonate
E. coli	*Escherichia coli*
*Ec*PBP3	PBP3 from *Escherichia coli*
EDC/NHS	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide/N-hydroxysuccinimide
E. faecium	*Enterococcus faecium*
EM	electron microscopy
EMBL	European Molecular Biology Laboratory
ER	endoplasmic reticulum
ESRF	European Synchrotron Radiation Facility
FBDD	fragment-based drug discovery
FBS	fragment-based screening
GlcNAc	N-acetylglucosamine
GT	glycosyl transfer
GTase	glycosyltransferase
GTd	glycosyltransferase domain
HMM	high molecular mass
HSQC	heteronuclear single quantum coherence
HTS	high-throughput screening
ILOE	inter-ligand NOE
INPHARMA	pharmacophore mapping
ist	iterative soft threshold
ITC	isothermal titration calorimetry
K_d	equilibrium dissociation constant
L-Ala	L-alanine
LE	ligand efficiency

LMM	low molecular mass
Lpo	lipoprotein
LPS	lipopolysaccharides
LTA	lipoteichoic acids
m-DAP	meso-diaminopimelic acid
MDR	multidrug resistant
MPD	2-methyl-2,4-pentanediol
MRSA	methicillin-resistant *Staphylococcus aureus*
MST	microscale thermophoresis
MurNAc	N-acetylmuramic acid
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
n-PBd	non-penicillin binding domain
NUS	non-uniform sampling
PABA	para-aminobenzoic acid
PAINS	pan assay interference compounds
P. aeruginosa	*Pseudomonas aeruginosa*
*Pa*PBP3	PBP3 from *Pseudomonas aeruginosa*
PBP	penicillin binding protein
PEG8000	polyethylene glycol 8000
PG	peptidoglycan
PSA	polar surface area
PTFE	polytetrafluoroethylene
RMS	root-mean-square
SASA	solvent accessible surface area
S. aureus	*Staphylococcus aureus*
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SOS	structural information using Overhauser effects and selective labeling
SPAM	solubility, purity, and aggregation of the molecule
SPR	surface plasmon resonance
STD	saturation-transfer difference
TFA	Trifluoro acetic acid
TLS	translation, libration, screw
TM	transmembrane helix
T_m	midpoint of thermal unfolding
TP	transpeptidation
TPase	transpeptidase
TPd	transpeptidase domain
TROSY	transverse relaxation-optimized spectroscopy
VdW	Van der Waals
Water-LOGSY	water-ligand observed *via* gradient spectroscopy
WHO	World Health Organization
WTA	wall teichoic acids
σ	standard deviation

List of Figures

List of Tables

1. Introduction

1.1. Global threat of multidrug resistant bacteria

Prior to the development of antibiotics, many people died from bacterial infections. For example, in the Medieval Age the Black Death caused by *Yersinia pestis* infection killed up to 50% of the European population [1]. At the beginning of the 20th Century, common infectious diseases such as cholera, diphtheria, pneumonia, typhoid fever, tuberculosis, typhus and syphilis were partly responsible for the worldwide average life expectancy of ~47 years, which is about 30 years lower than the current average life expectancy [2]. During World War I, infectious diseases were responsible for approximately 10% of the German soldiers and ~20% of the British soldiers that died [3]. Fortunately, the discovery of penicillin by Sir Alexander Fleming in 1928 provided humans with a drug to treat bacterial infections, thereby reducing the morbidity rate caused by such infections dramatically [4]. However, excessive overuse of antibiotics has led to the emergence of multidrug resistant (MDR) bacterial strains that are resistant to most common antibiotics. For this reason, the term “ESKAPE” was introduced to encompass six main MDR bacteria that are responsible for most nosocomial infections: *Enterococcus faecium* (*E. faecium*), *Staphylococcus aureus* (*S. aureus*), *Klebsiella pneumoniae* (*K. pneumoniae*), *Acinetobacter baumannii* (*A. baumannii*), *Pseudomonas aeruginosa* (*P. aeruginosa*) and *Enterobacter spp. [5]*. These MDR bacteria are the major cause of nosocomial infections in hospitals and represent a massive financial burden to nations. For example, it has been estimated that national costs associated with MDR bacterial infections are at least $2.4 billion in the USA annually [6].

In 2018, the World Health Organization (WHO) assessed, prioritized and ranked the most dangerous bacteria [7]. Different criteria, such as mortality, treatability and prevalence and trend of resistance were evaluated to generate this ranking, which concluded that the most life-threatening bacteria are MDR Gram-negative bacteria *A. baumannii*, *P. aeruginosa*, *Escherichia coli* (*E. coli*), *Klebsiella spp.* and *Enterobacter spp.*, whereas the most threatening Gram-positive bacteria are *E. faecium* and the methicillin-resistant *S. aureus*. Note that all ESKAPE pathogens are included in the list of most threatening MDR bacteria. The top eleven pathogens are Gram-negative bacteria because the development of antibiotics against Gram-positive bacteria is more promising. Therefore, although responsible for high clinical and epidemiological burdens, the two Gram-positive ESKAPE bacteria are ranked lower than many Gram-negative bacteria. For example, in the last two decades, only two new classes of antibiotics have been developed against Gram-positive bacteria, whereas the last class developed against Gram-negative bacteria was the quinolones, which entered the market in 1962.

Typical nosocomial diseases caused by the Gram-negative bacterium *A. baumannii* are pneumonia and bloodstream infections, whereas *P. aeruginosa* can cause malignant external otitis, endophthalmitis, endocarditis, meningitis, septicemia and pneumonia [8, 9]. *E. coli* usually causes enteritis, urinary tract infections, septicemia and other clinical infections, such as neonatal meningitis [10]. *Enterobacter spp.* are associated with bacteremia, infections of skin and soft tissues, respiratory tract, urinary tract, bone and joints, central nervous system, gastrointestinal tract and other organs [11].

The Gram-positive *E. faecium* is responsible for infections of the urinary tract, surgery-related wounds, bloodstream, intra-abdominal space and soft tissues as well as endocarditis [12, 13]. Typical manifestations of an infection with *S. aureus* are pneumonia and surgical wound infections. Interestingly, 30% to 50% of all healthy humans are colonized by *S. aureus*, but the immune system is able to suppress the bacterium [14].

In another WHO report from 2014, it was estimated that at least 700,000 people die per year because of MDR bacterial-related infections (approximately 33,000 in Europe and approximately 2400 in Germany; [15]) and this number will climb to approximately 10 million people per year by 2050 if no further effort to develop new antibiotics is taken [16]. This mortality rate would exceed the current total number of people dying from cancer *per annum* (8.2 million). Additionally, in the context of diseases caused by MDR bacteria, it is important to note that infections have the potential to be lethal especially when the immune system is already weakened, e.g., due to organ transplantations or chemotherapy treatment of cancer.

1.2. Structure of the bacterial cell wall

Bacteria readily adapt to survive in various habitats and thus, are highly specialized to their environment, which makes classification of bacteria into specific categories challenging. In contrast to their diverse habitats, there are properties that all bacteria have in common. Firstly, all bacteria have a cell wall, consisting of a peptidoglycan (PG) layer and a cell membrane, which is composed primarily of a lipid bilayer. Secondly, unlike eukaryotic cells bacteria lack a nucleus, in which the DNA is stored. The absence of typical eukaryotic organelles such as mitochondria, chloroplasts, the endoplasmic reticulum, lysosomes and peroxisomes are additional characteristics of bacteria. Moreover, instead of the 80S ribosome found in eukaryotes, bacteria have the 70S ribosome, which is composed of the 30S and 50S subunits.

There have been numerous discussions about how to further classify bacteria and there are many different approaches that have been used previously [17]. For example, it is possible to classify bacteria based on their morphology, pathogenicity or the environment they live in. Currently, the most common methods to classify bacteria are genomic and molecular ecology studies [18]. A simple way for the classification of bacteria is the staining of the bacterial cell wall, which was invented in 1884 by Hans Christian Gram [19]. Gram staining distinguishes between Gram-negative and Gram-positive bacteria based on the different cell wall structures. The main difference in the cell walls is the thickness of the PG layer, which is five to eight times thicker for Gram-positive bacteria, allowing the dye to bind the cell wall [20]. Furthermore, Gram-negative bacteria have an additional outer membrane that prevents the dye from reaching and staining the PG.

The inner membrane of Gram-negative and Gram-positive bacteria is similar and mainly composed of phospholipids [20]. Because there are many hundred different lipids described, the exact composition of the membranes varies among different organisms and even more between bacteria and eukaryotes [21]. For example, in *E. coli* the membrane is composed of primarily zwitterionic phosphatidylethanolamine (about 75%), anionic phosphatidylglycerol (about 20%) and cardiolipin, and therefore, has an overall negative charge [22]. In comparison, the eukaryotic plasma membrane of yeast contains phosphatidylserine (about 34%), phosphatidylethanolamine (about 20%), phosphatidylinositol (about 18%) and phosphatidylcholine (about 17%) [23]. The PG layer is important for ensuring that turgor pressure is maintained and to prevent the cell from bursting [20]. In Gram-positive bacteria the

PG contains wall teichoic acids (WTA) and plasma membrane lipoteichoic acids (LTA) (Figure 1.1). WTA are directly attached to the PG by a disaccharide linker that is connected with a polymeric chain typically containing alditol phosphate. The negatively charged polymeric chain plays an important role in membrane integrity and WTA are temperature sensitive, prevents growth defects and facilitate cell elongation [24-26]. In contrast to WTA, LTA are connected to the plasma membrane *via* a glycolipid anchor and the attached polymeric chain contains glycerol phosphates. In general, LTA are considered to be essential for cell viability and are important for ensuring the correct placement of new PG layers [27].

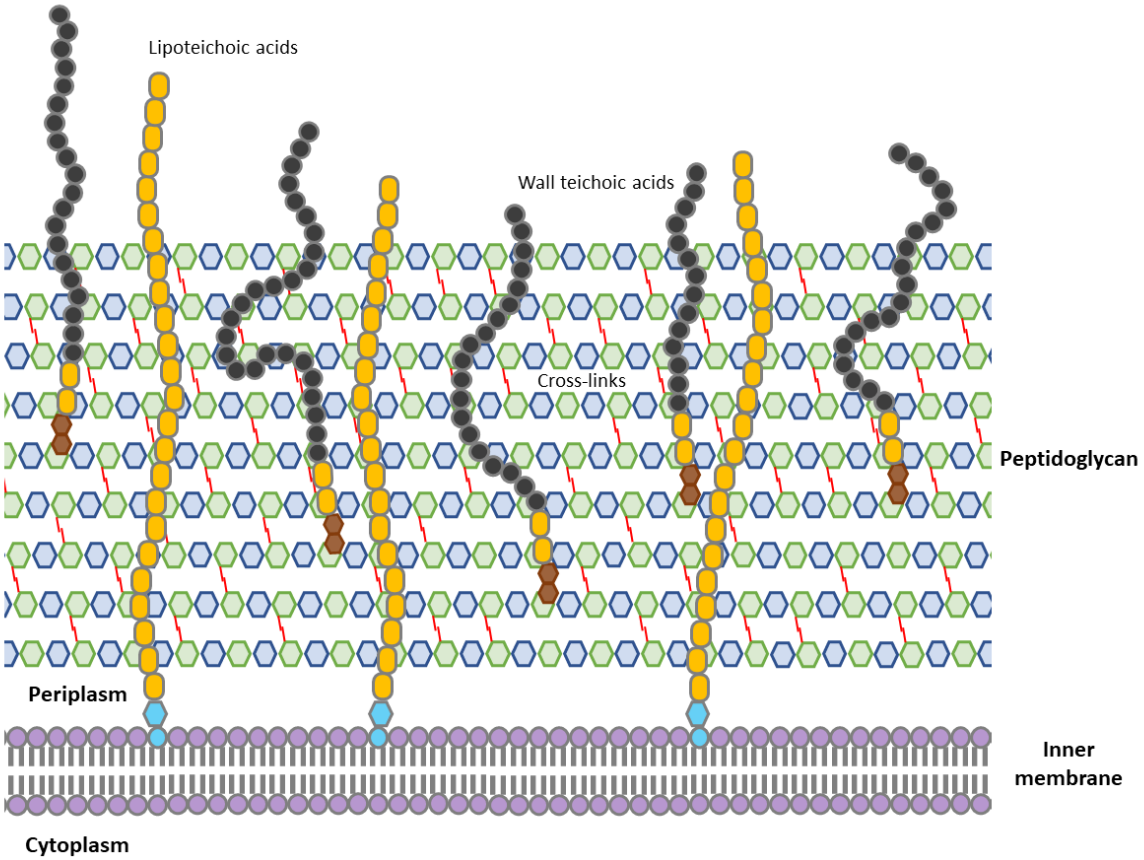


Figure 1.1: Schematic showing of the structure of the cell wall of Gram-positive bacteria.
The inner membrane consisting of phospholipids (violet and gray) separates the cytoplasm from the periplasm, whereas the peptidoglycan (PG) maintains turgor pressure. The PG is composed of β1,4-linked N-acetylglucosamine (GlcNAc, blue) and N-acetylmuramic acid (MurNAc, green) and cross-linked via peptide chains (red) that are attached to the MurNAc subunits [28]. The PG is 19–33 nm thick and contains wall teichoic acids (WTA), which are important for membrane integrity and maintenance of the PG. They are anchored to the PG by a cross-linked linkage unit (brown) and consist of glycerol phosphate (yellow) and poly(alditol phosphate) (black). With their glycolipid anchor (light blue) lipoteichoic acids (LTA) are directly attached to the inner membrane and are crucial for cell viability and correct placement of new PG layers. Figure adapted from [20].

In contrast, Gram-negative bacteria do not have LTA and WTA but lipopolysaccharides (LPS), which are anchored to the outer membrane with a lipid linker moiety (Figure 1.2). These long-chain polysaccharides function as the outer physical barrier of Gram-negative bacteria because the negative charge of LPS hampers movement of small molecules towards and through the outer membrane, resulting in a reduced sensitivity to environmental stresses [29]. Additionally, lipoproteins such as LpoA and LpoB are located in the outer membrane, undertaking an important function in cell elongation and division by forming complexes with penicillin binding proteins (PBPs; Section 1.3; [30]).

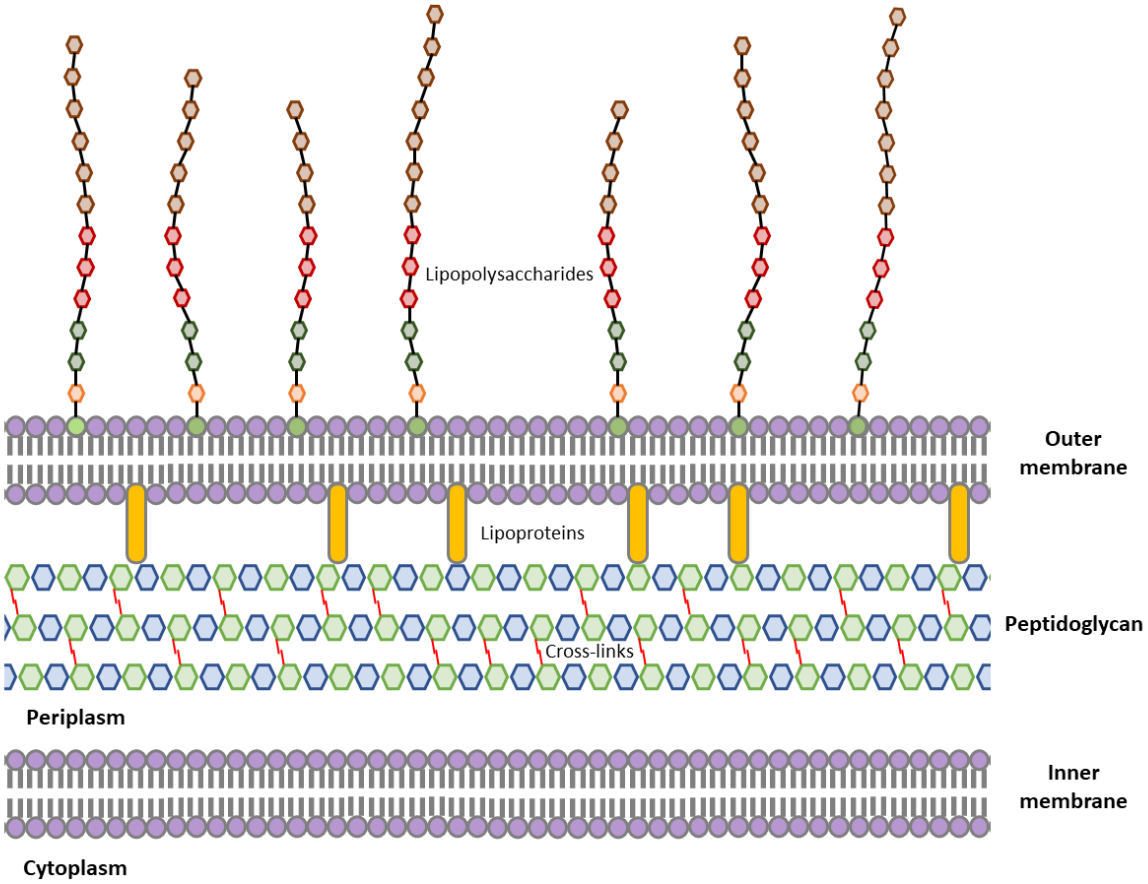


Figure 1.2: Schematic showing the structure of the cell wall of Gram-negative bacteria.
The inner membrane consists of phospholipids (violet and gray) that separate the cytoplasm from the periplasm, whereas the PG is ~4 nm in thickness and maintains turgor pressure. The PG is composed of β1,4-linked N-acetylglucosamine (GlcNAc, blue) and N-acetylmuramic acid (MurNAc, green) and cross-linked via peptide chains (red) that are attached to the MurNAc subunits [28]. The outer membrane is connected to the PG via lipoproteins (yellow), which are involved in cell elongation and division. Lipopolysaccharides (LPS) are anchored to the outer membrane (green) and hamper movement of small molecules through the membranes and entry into the cytoplasm. They consist of monosaccharides (hexagons) that are classified as inner core (orange and dark green) and outer core (red) polysaccharides. The hydrophilic O-antigens (brown) form the outer barrier of the cell wall for Gram-negative bacteria. Figure adapted from [20].

1.3. Synthesis of the peptidoglycan by penicillin binding proteins

Penicillin binding proteins (PBPs) are named after their ability to bind penicillin and are therefore the target of β-lactam antibiotics. PBPs are found in Gram-positive and Gram-negative bacteria and play an important role in PG synthesis by elongating the bacterial cell wall with lipid II molecules and cross-linking the peptide side chains of former lipid II molecules [31, 32]. Lipid II is synthesized inside the cytosol and consists of β1,4-linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) [28]. Additionally, MurNAc is extended by a peptide side chain, containing the amino acids L-alanine (L-Ala), D-glutamate (D-Glu), meso-diaminopimelic acid (*m*-DAP) and two D-alanines (D-Ala), and a polyisoprenoid connected *via* a phosphoester bond including two phosphates, that anchors the molecule to the inner membrane [33]. The lipid II molecule is flipped to the periplasm by flippase FtsW (Figure 1.3; [34]). The PG is extended with lipid II through two reactions catalyzed by different PBPs (Figure 1.4; [35]). In the first reaction, lipid II is connected to the PG strand. This reaction is performed by glycoslytransferases (GT) and releases two phosphate groups. The second reaction is transpeptidation (TP), which cross-links the first D-Ala and the *m*-DAP of the peptide

side chains of lipid II to increase the stability of the PG, allowing it to resist turgor pressures. In *E. coli*, PBP1b and PBP3 catalyze the GT and TP reactions, respectively [34].

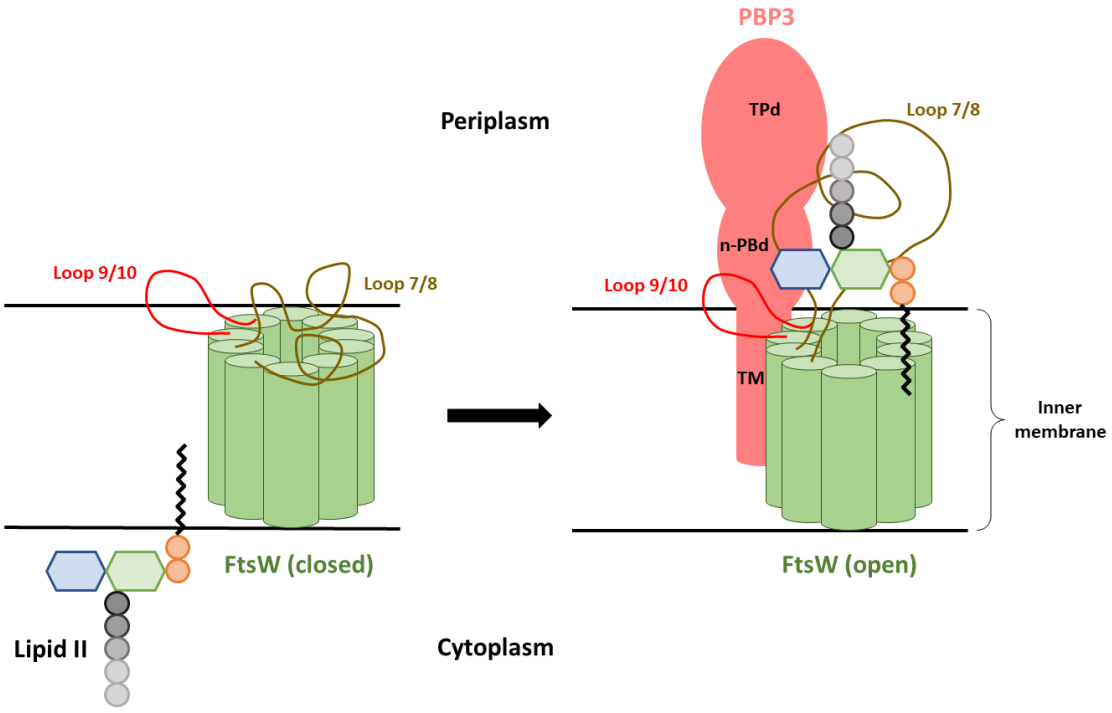


Figure 1.3: Schematic representation showing the mechanism of transport of lipid II molecules to the periplasm.

After activation by binding the non-penicillin binding domain (n-PBd) of PBP3, which is anchored in the membrane via a transmembrane helix (TM), loop 7/8 of the flippase FtsW undergoes a conformational change leading to pore gate opening and loop 9/10 facilitates positioning of PBP3 at the septum. After flipping the lipid II molecule from the cytoplasm to the periplasm, PBP3 performs the transpeptidation (TP) reaction in a further step of PG synthesis. Note that the formed protein complex called the divisome contains additional proteins, such as PBP1b, but for a clearer representation, only PBP3 is shown. Figure adapted from [34].

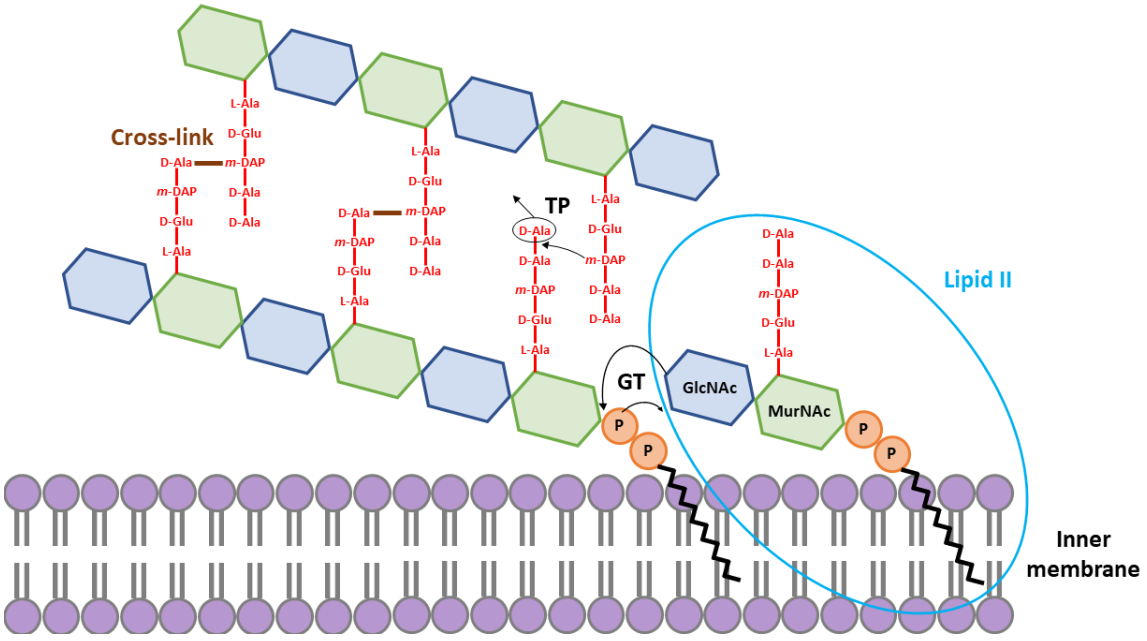


Figure 1.4: Schematic representation of the transpeptidation (TP) and glycosyltransfer (GT) reactions.

The reactions are catalyzed by penicillin binding proteins. The figure was adapted from [35].

There are various classes of PBPs that have been identified from genome sequencing, but despite differences between bacteria, the TP domain (TPd) and the GT domain (GTd) show highly conserved folds, underlining the importance of the PBPs. In the active site of the TPd, which is responsible for binding penicillins, three conserved motifs have been identified: S*xxK (contains the catalytic serine in α4), SxN in the α5-α6 loop and K(T/S)GT in β3 (for nomenclature see Figure 1.5 in Section 1.5).

Based on their molecular weights, PBPs are categorized as high molecular mass (HMM) and low molecular mass (LMM) PBPs. In general, HMM PBPs are multimodular with more than one functional domain. A subclass of HMM PBPs are the bifunctional class A PBPs, which have transpeptidase (TPase) and glycosyltransferase (GTase) activities. In contrast, class B HMM PBPs are monofunctional and do not have additional domains with catalytical functions but may have domains involved in protein-protein interactions, e.g., the first 56 amino acids of *E. coli* PBP3 are essential for positioning the protein at the division site [36, 37].

1.4. Classification of PBPs in Gram-negative bacteria

Besides sharing highly conserved motifs in the active site, PBPs from different bacteria are classified into various groups according to their molecular mass, the number of penicillin-binding domains and their biological activity. For this reason, the following section focuses on PBPs from Gram-negative bacteria and in particular on PBPs from *E. coli*.

Class A PBPs contribute primarily to the repair of cell defects while playing only a minor role in shaping the cell [38]. Based on different species of bacteria, there are different PBPs belonging to this class. Typically, they share a similar structure with a *C*-terminal TP domain and an *N*-terminal GT domain. A linker domain connecting both catalytic domains is composed of a small β-sheet structure [32]. The Gram-negative bacterium *E. coli* includes the three proteins PBP1a, 1b and 1c in class A PBPs. PBP1a and 1b interact with lipoproteins (Lpo) that are located in the outer membrane [30]. Activation of these PBPs requires specific interactions between LpoA and PBP1a and LpoB and PBP1b [39, 40]. For these essential interactions, the docking domains of PBP1a, PBP1b, LpoA and LpoB have coevolved [40]. Both PBP1a and 1b are anchored to the inner membrane with PBP1a playing a role in cell elongation and PBP1b functioning in cell division [28, 36, 40]. In this context, PBP1b was shown to interact with PBP3 and FtsW [34]. Nevertheless, PBP1b can be replaced by PBP1a, but at least one class A PBP is required for cell survival. In contrast to PBP1a and 1b, the function of PBP1c is unclear, although it is hypothesized that this protein has specialized cellular functions [28, 32]. Compared with other PBPs the affinity of PBP1c against β-lactam antibiotics is weaker [28]. However, as a class A PBP, the non-essential PBP1c contains a membrane anchor that localizes this protein to the cell membrane [28].

The monofunctional class B PBPs catalyze either the TP or GT reaction in the synthesis of the PG. *E. coli* possess two class B PBPs, PBP2 and 3, which both contain a TP domain. PBP2 is an essential elongase and important in maintaining the shape of the cell wall [32]. Furthermore, *E. coli* PBP2 ensures that the correct diameter of the new cell pole is preserved [41]. Although PBP2 localizes to the division site, this protein is not involved in cell division and is probably released from the division site just prior to cell division [32, 41]. In contrast, PBP3 is the major protein at the division site, catalyzing the cross-linking reaction in PG synthesis. During cell division, PBP3 is essential and inhibition is lethal to the bacterium, thus making this protein an ideal drug target.

LMM PBPs are involved in cell separation and PG maturation or recycling [32]. In total there are seven LMM PBPs in *E. coli*, including PBP4, 5, 6, 6b and 7, as well as AmpC and AmpH [42, 43]. The endopeptidases PBP4 and 7 cleave cross-linked PG strands, whereas PBP5, 6 and 6b are carboxypeptidases that cleave the D-alanine-D-alanine bond [32, 44]. After this reaction, the stem pentapeptide side chain is no longer available for further TP reactions. AmpH is associated with the recycling of PG and displays bifunctional DD-carboxypeptidase and DD-endopeptidase activities [45]. Furthermore, AmpH features weak β-lactamase activity [45]. In contrast, AmpC is a β-lactamase that belongs to the class of cephalosporinases, which afford resistance against cephalosporine antibiotics [46]. Although the LMM PBPs are involved in essential processes in bacteria, it was shown that they can be knocked-out with only slight morphological effects to the mutant bacterium when compared with that of wild-type *E. coli* [42].

1.5. Structure of *E. coli* PBP3

PBP3 from *E. coli* (*Ec*PBP3), which belongs to the class B3 HMM PBPs, is an essential protein for the bacterium. This PBP is active primarily during cell division by catalyzing cross-linking between peptide side chains of the PG. The structure of PBP3 is divided into the functional *C*-terminal TPd, an *N*-terminal membrane anchor and the central non-penicillin-binding domain (n-PBd). The n-PBd (also known as the *N*-terminal domain) is further divided into the anchor, head and linker subdomains (Figure 1.5; [47]). The anchor and head subdomain likely interact with other proteins at the division site, whereas the linker subdomain may function as a spacer region that enables the TPd to reach side chains of the PG. Overall, the structure of the n-PBd is highly flexible, which explains previous difficulties associated with obtaining high resolution crystal structure data. In the first apo *Ec*PBP3 structure solved by X-ray crystallography in 2010, large portions predominantly in the n-PBd were not resolved in the electron density (Figure 1.5). Additionally, the n-PBd was hypothesized to be responsible for anisotropic diffraction data and poorer electron density [48, 49]. For this reason, in 2019 a construct was produced containing only the TPd, which showed improved diffraction data with a higher resolution obtained [49]. However, this structure lacks the linker subdomain, which is hypothesized to be important in stabilizing the TPd [48].

The TPd active site is formed by residues of the α10-β3, β5-α11 and α5-α6 loops, as well as β3, α4 and the β2b-β2c-β2d region and contains the three conserved motifs S*xxK, SxN and K(T/S)GT. With their greater flexibility and proximity to the active site, the β5-α11 loop and β3-β4 loop are hotspots for various PBP3-mediated resistance mutations and play an important role in ligand binding [49].

In contrast to the *Ec*PBP3 structure, the PBP3 structure from *P. aeruginosa* (*Pa*PBP3) adopts a more stable, rigid fold, especially for the n-PBd, leading to improved diffraction quality. Therefore, three apo *Pa*PBP3 structures have been published (PDB entries 3OC2, 6HZR and 3PBN; [49-51]), with higher resolutions (2.0 Å, 1.2 Å, and 2.0 Å, respectively) when compared with the resolution of the *Ec*PBP3 structure (2.5 Å). The overall fold is similar between these structures with root-mean-square (RMS) distances of 0.397 to 0.648 Å between *Pa*PBP3 structures and 0.947 Å between PDB entries 4BJP and 6HZR. However, various secondary structure elements, such as the β5-α11 loop or β3- β4 loop (in *Pa*PBP3), are not well resolved and are assumed to only adopt stable well-defined structures upon formation of an acyl-protein complex.

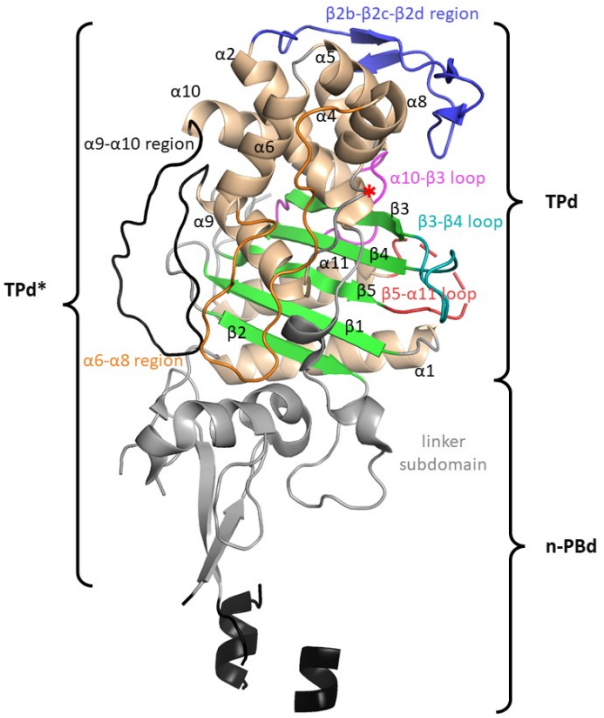


Figure 1.5: Published X-ray structure of apo PBP3 from *E. coli* lacking the transmembrane helix.
Secondary structure elements and intervening loops relevant to this thesis are labeled (PDB entry: 4BJP; [47]). The TPd*, which is introduced during this thesis, includes the n-PBd linker subdomain (gray), in addition to the TPd. The catalytic serine is marked with a red asterisk. Note that the n-PBd of the elongated molecule is incomplete because of poor electron density but was resolved in this thesis (Figure 3.37).

1.6. β-lactam antibiotics target PBPs

Antibiotics have five main bacterial targets: (i) peptidoglycan; (ii) plasma membrane; (iii) proteins involved in metabolic pathways; (iv) ribosomes; and (v) proteins that perform reactions in nucleotide synthesis (Figure 1.6). Some of these targets are found in humans, such as ribosomes, which differ in subunit size but function in a similar fashion. Thus, antibiotics targeting a bacterial protein for which a human homologue exists can have side effects, which limits their efficiency. For example, side effects arise when administrating sulfonamides such as sulfamethoxazole, which mimic the structure of the substrate para-aminobenzoic acid (PABA) and inhibit dihydropteroate synthase (DHPS) in the folic acid pathway competitively [52]. This antibiotic is usually combined with diaminopyrimidine antibiotics, such as trimethoprim, which target the final step in folic acid synthesis, to increase the effectiveness of the treatment [52]. Because the folic acid pathway also exists in humans, potential side effects are usually addressed by folic acid supplementation.

In contrast to antibiotic targets that can cause side effects when administered, the bacterial cell wall only exists in bacteria, thus reducing the risk of side effects by potential drugs that target proteins involved in cell wall synthesis. For this reason, β-lactam antibiotics, which target PBPs by covalently binding the catalytic serine and inhibiting PBP activity, are the most widely used antibiotics [53]. These antibiotics possess a β-lactam ring structure, thereby mimicking the natural substrate D-alanine-D-alanine (Figure 1.7).

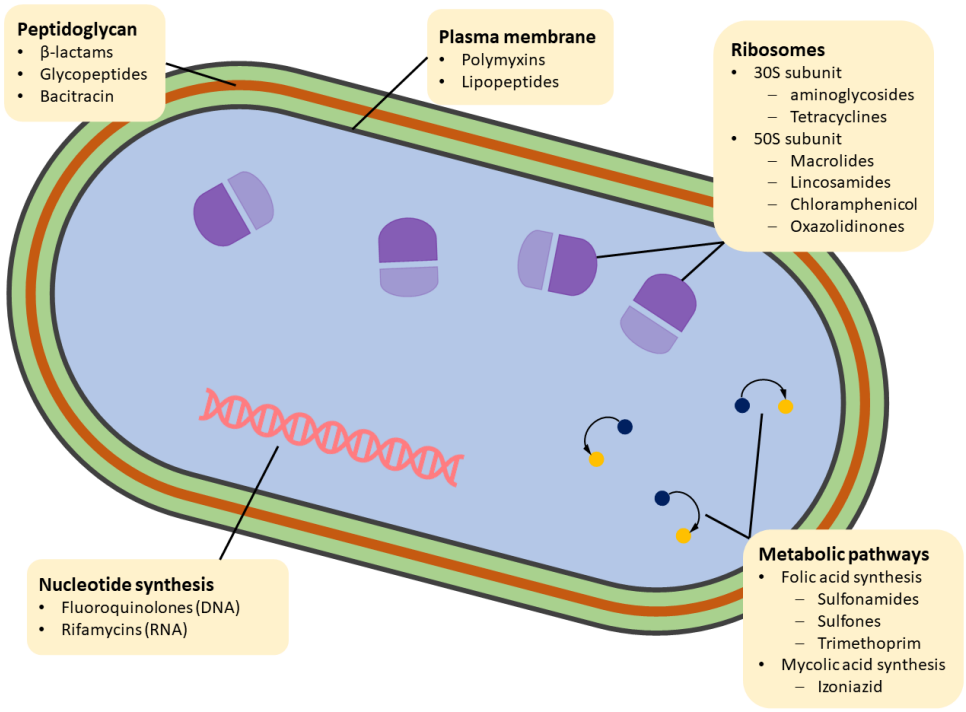


Figure 1.6: Antibiotic targets in Gram-negative bacteria.

In general, antibiotics have five main targets: (i) membrane; (ii) cell wall; (iii) nucleic acid machinery; (iv) machinery that produces proteins, e.g., ribosomes; and (v) proteins involved in metabolism.

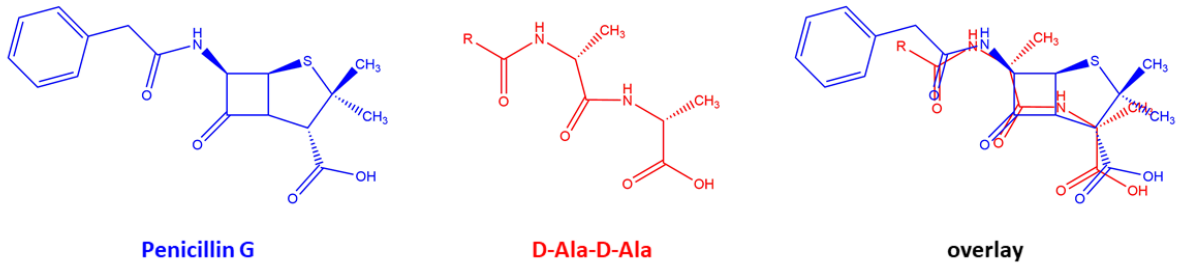


Figure 1.7: Chemical structures of penicillin G and D-alanine-D-alanine.

Overlaying their chemical structures reveals the structural similarity between both molecules. The D-alanine-D-alanine (D-Ala-D-Ala) motif is included in the structure of lipid II and is recognized by PBPs.

The first step in the inhibitory reaction is the formation of a non-covalent state, followed by nucleophilic attack by the serine hydroxyl group (E + I; Figure 1.8). After formation of a covalent product (EI), the β-lactam ring opens (EI*) and the resulting product is released (E + P). Because the final release step is slow with a turn-over constant K_3 of ~10^{-4} $M^{-1}s^{-1}$ or less, the PBP is inhibited effectively in the acyl-enzyme complex form [54]. For example, considering the time of bacterial cell division (20 to 30 min) and the half-life of penicilloyl-PBP2a (26 to 77 h) in *S. aureus*, the formation of the acyl-enzyme complex is irreversible [55].

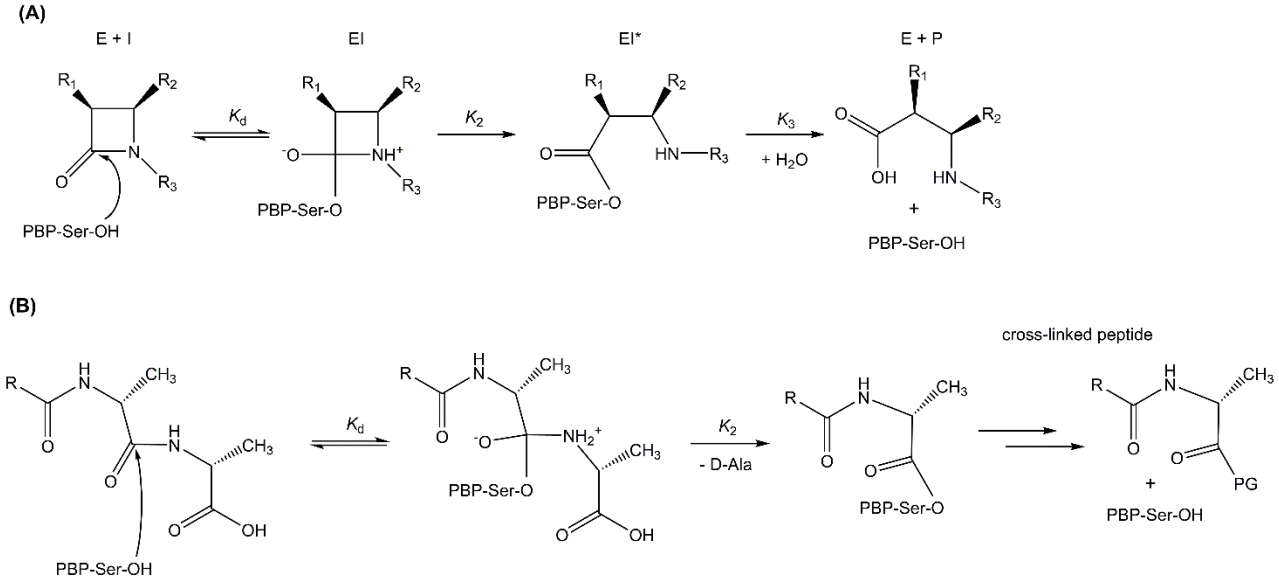


Figure 1.8: Mechanism of the reaction between PBPs and β-lactam antibiotics or D-Ala-D-Ala.

(**A**) A non-covalent bound-state between the protein and the β-lactam antibiotic (E + I) is formed to initiate the reaction, followed by covalent binding of the β-lactam (EI) through nucleophilic attack. After ring-opening of the acyl-enzyme intermediate (EI*), release of PBP and product (E + P) is negligible because of the very slow turn-over rate (K_3). Thus, the penicilloyl-PBP can be considered the final product of the reaction. (**B**) In the natural reaction mechanism, the oxygen of the catalytic serine binds the D-Ala-D-Ala motif of the Lipid II molecule through nucleophilic attack. In further reactions steps the Lipid II molecule is connected to the PG and the PBP is released.

β-lactam antibiotics are divided into five classes based on the chemical environment of the β-lactam ring and whether another ring system is attached: penicillins (penams), cephalosporins (cephems), carbapenems, monobactams and penems (Figure 1.9; [56]).

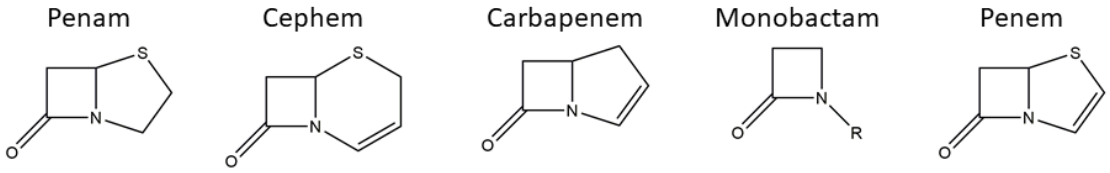


Figure 1.9: Chemical structures of β-lactam antibiotics.

Basic chemical structures of the five different classes of β-lactam antibiotics are defined as penam, cephem, carbapenem, monobactam and penem.

Historically the first class of β-lactam antibiotics was penicillins, which chemically belong to the penams (Figure 1.10). Penicillin G (benzylpenicillin) was first discovery by Sir Alexander Fleming in 1928 [4, 56, 57]. After some further investigations by Ernst Boris Chain and Howard Florey (and others), penicillin G was the first antibiotic clinically used and the three scientists received the Noble Prize in medicine in 1945. Because penicillin G has low stability in acid and alkaline buffers and is only active against Gram-positive bacteria it underwent further chemical development to generate the first semi-synthetic β-lactam antibiotic methicillin, which has been used since 1959 [58]. Still lacking reasonable stability under acidic conditions with comparatively low effectiveness against Gram-negative bacteria, methicillin was used primarily against *S. aureus* [59]. The extensive use of methicillin, including its use with farm animals, led to the emergence of methicillin-resistant *S. aureus* (MRSA), which was first reported in 1961 [58]. MRSA is also resistant to other β-lactam antibiotics and currently represents a worldwide problem in hospitals [60]. In 1961, ampicillin (α-amino-benzylpenicillin) was

produced by synthetically adding an amino group between the β-lactam and phenyl ring (Figure 1.10). Ampicillin displays sufficient stability in different buffers and is active against a broad range of bacteria, including Gram-negative species [61]. Unfortunately, although being a broad-spectrum antibiotic, ampicillin lacks activity against *Pseudomonas aeruginosa* (*P. aeruginosa*), which is a common bacterium found in hospitals. However, carbenicillin discovered in 1967 is active against *P. aeruginosa* but not *S. aureus* and is sensitive to β-lactamases [62]. There have been many types of penicillin developed over the decades but since the 1970s scientists have focused on the development of other β-lactam antibiotics.

Initially found in *Cephalosporium acremonium*, the group of cephalosporins belongs chemically to cephems because the β-lactam ring is fused with a dihydrothiazine ring (Figure 1.10; [56]). Interestingly, the acetate group at the C3 position is eliminated during nucleophilic attack of the catalytic serine and the opening of the β-lactam ring (marked with a red cycle in Figure 1.10). Cephalosporin C, the first cephalosporin identified, displays low antibacterial activity and was not used as a drug but as a starting scaffold for developing other cephalosporins. In general, there are five generations of cephalosporins. The first generation, including cephalexin, was introduced in the 1960s. These first generation cephalosporins were effective against Gram-positive bacteria and particular Gram-negative species. Because some stability issues existed under acidic conditions, the second generation (such as cefuroxime) was designed to be more polar and therefore more hydrophilic, increasing the chemical stability and facilitating passage of the antibiotic through the outer membrane of Gram-negative bacteria. This was mainly achieved by changing the chemical group at the C3 position, which is released when binding the catalytic serine of PBPs. However, both the first and second generations were ineffective against *P. aeruginosa*. This problem was addressed in the third generation, where hydrophilicity was increased further to facilitate passive passage across the LPS and outer membrane by porin channels (ceftizoxime in Figure 1.10). In the fourth generation of cephalosporins (e.g., cefpinome), effectiveness against Gram-negative bacteria was not improved but optimization against β-lactamases from Gram-positive bacteria, such as *S. aureus*, was achieved. The fifth generation (represented by ceftaroline in Figure 1.10) was developed to be active against MRSA bacteria, but unfortunately, bacteria developed resistance. Currently cephalosporins are administered in combination with β-lactamase inhibitors (e.g., avibactam (see below)) to ensure drug efficacy.

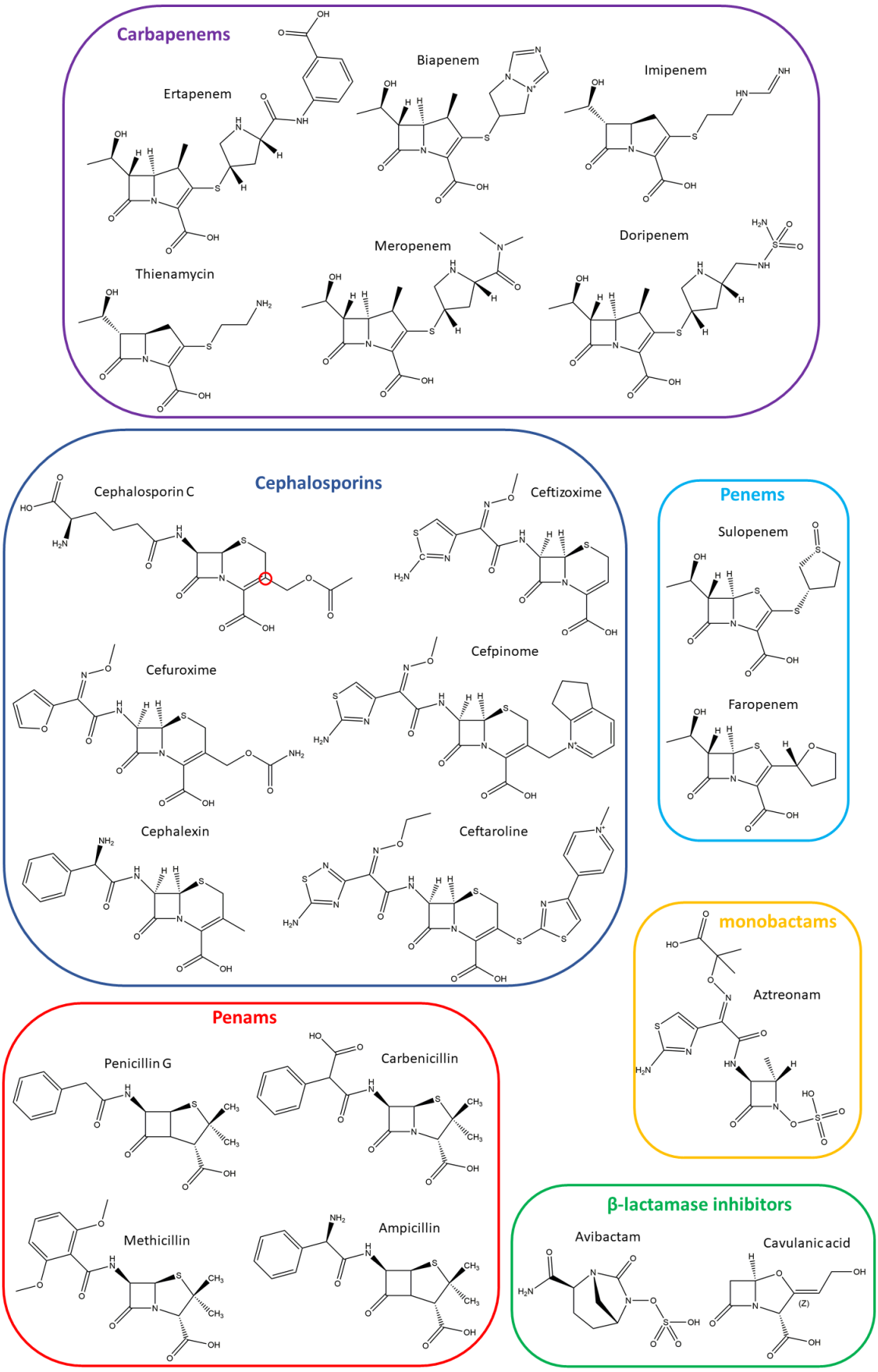


Figure 1.10: Chemical structures of β-lactam antibiotics grouped according to their class or activity.
Structures of two β-lactamase inhibitors are also shown in the bottom right corner. The C3 position of cephalosporin C is marked with a red circle.

As another class of β-lactam antibiotics, carbapenems are less vulnerable to β-lactamases and target Gram-positive and Gram-negative bacteria as broad-spectrum antibiotics. The structure of imipenem has a 2,3-dihydro-1H-pyrrole ring attached to the β-lactam ring and was the first carbapenem used in medicine in 1985 [63-65]. Imipenem is a structural analogue of thienamycin, which was found in *Streptomyces cattleya*, with the amide group exchanged by a formamidine group, leading to five to ten times greater stability when compared to thienamycin [66]. Imipenem is not stable in the human body and is degraded by kidney dehydropeptidase-1 (DHP-1). Thus, imipenem is typically combined with the DHP-1 inhibitor cilastatin in a 1:1 ratio [66]. To avoid potential side effects in humans, the DHP-1 resistant meropenem was developed and approved in 1996 as an alternative to imipenem. Meropenem is a broad-spectrum antibiotic with inhibitory activity against Gram-positive and Gram-negative bacteria. Based on the structure of meropenem, doripenem, ertapenem and biapenem have been developed since the turn of the 21st century to increase stability in the human body. Although carbapenems are considered to be less vulnerable to β-lactamases, there are cases of mainly Gram-negative bacteria acquiring resistance. Carbapenemases, members of the β-lactamase family, are the major reason for this resistance [67].

The simplest basic structure of β-lactam antibiotics is the monobactam, containing only the β-lactam ring with additional atoms attached to the nitrogen in the β-lactam ring. Currently, aztreonam is the only monobactam antibiotic available on the market. A precursor of aztreonam was originally found in nature and was chemically modified to increase its antimicrobial activity [68, 69]. Aztreonam displays high resistance to β-lactamases and has the highest affinity against PBPs from Gram-negative bacteria [69]. Although able to inhibit the growth of most clinically relevant Gram-negative strains, aztreonam does not bind to PBPs from Gram-positive or anaerobic bacteria, such as *S. aureus* [70]. Currently, LYS228 (originally discovered at Novartis) and MC-1 (a siderophore conjugate developed by Pfizer) are examples of monobactam drug candidates tackling Gram-negative bacteria such as *Enterobacteriaceae* or *P. aeruginosa*, with LYS228 currently in clinical phase II trials [51, 71-74].

The final class of β-lactam antibiotics is the penems, which are completely synthetic. The penems are also named thiopenems because the pyrroline ring of carbapenems is exchanged by a thiazoline ring. The first developed penem was faropenem. The trans conformation of the C6 side chain makes faropenem stable against many β-lactamases; however, *P. aeruginosa* is resistant to faropenem because of poor outer membrane permeability [75]. FDA approval of faropenem was applied for in 2005 but was rejected because the clinical trial was not performed correctly [75]. Currently, faropenem is available in many Asian countries but not in Europe or the USA [76, 77]. A second penem that was developed and approved for the market is sulopenem, which has properties similar to those of faropenem. Sulopenem is effective against MDR bacteria but is not active against *P. aeruginosa*. Application for FDA approval was submitted in early 2021 by Iterum Therapeutics [78].

Inhibitors against β-lactamases have been developed to combat MDR bacteria with β-lactam resistance. Most of these inhibitors share a scaffold that is similar to penicillins, acting as suicide inhibitors, such as clavulanic acid [56, 79]. Thus, they share the same mechanism featuring slow turn-over rates K_3 (Figure 1.8). The first non-β-lactam β-lactamase inhibitor with a broad inhibition against different classes of β-lactamases was avibactam (Figure 1.10; [80]). Based on a non-β-lactam structure, the inhibition mechanism is unusual. The usual ring-opening event occurs when binding to the target protein, but unlike previous β-lactamase inhibitors this step is reversible [81, 82]. In summary, β-lactamase inhibitors are an adequate approach to treat MDR bacterial infections when administered with regular β-lactam antibiotics.

1.7. Novel antibacterial molecules targeting PBP3

1.7.1. AIC499

In 2013, AiCuris patented the development of a novel monobactam (AIC499), which binds covalently in the active site and shows remarkable inhibitory activity against PBPs from Gram-negative bacteria, including clinical isolates harboring several β-lactamase classes [83, 84]. During optimization towards AIC499, five important groups were explored to improve the antibacterial potency of the lead structure against *Enterobacteriaceae* and the non-fermenter *P. aeruginosa* (Figure 1.11). Target optimization was mainly driven by investigating the inhibition of *Pa*PBP3 (unpublished data).

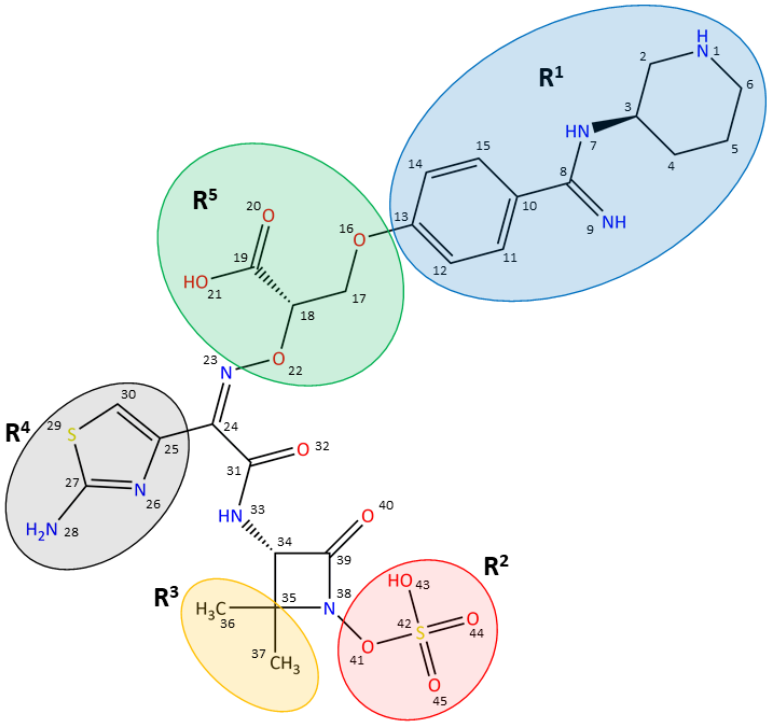


Figure 1.11: Chemical structure of the monobactam AIC499.
The relevant functional groups are shaded in different colors: benzamidine-based head group (R^1), blue; β-lactam N-1 position (R^2), red; β-lactam C-4 position (R^3), orange; amino-thiazole (R^4), gray; linker (R^5), green.

The amino-thiazole (R^4) was found to be crucial for potent antimicrobial activity, whereas variation of the chemical moieties at the R^3 and R^2 positions revealed that two methyl groups on the β-carbon (C35) and a sulfate group attached to the nitrogen of the β-lactam ring, respectively, were optimal. The optimal linker (R^5) length was found to be ethylene and introducing the carboxylate function improved the overall antibacterial and absorption, distribution, metabolism and excretion (ADME) profile. Finally, substituting the head group (R^1) with a piperidine moiety increased antibacterial activity.

1.7.2. Discovery of pyrrolidine-2,3-diones as inhibitors of *P. aeruginosa* PBP3

Pyrrolidine-2,3-diones were recently identified as novel inhibitors of *Pa*PBP3 [85]. In high-throughput screening (HTS) using an optimized fluorescence-based approach, hit compounds were clustered based on the similarity of their chemical structures. In the following step, the

most antibacterial active cluster of compounds was further optimized by exchanging moieties of the molecule that showed a high variance among various structures within the cluster (referred as Cluster J; Figure 1.12). Because the compounds were identified by performing HTS, only data about the inhibition of the target protein was available. The IC_{50} value represents the compound concentration at which the *Pa*PBP3 loses 50% of its activity, thus lower IC_{50} values indicate increased potency. The binding of the pyrrolidine-2,3-diones to *Pa*PBP3 should be validated by orthogonal methods, such as nuclear magnetic resonance (NMR) heteronuclear single quantum coherence (HSQC) perturbation shift analysis or surface plasmon resonance (SPR).

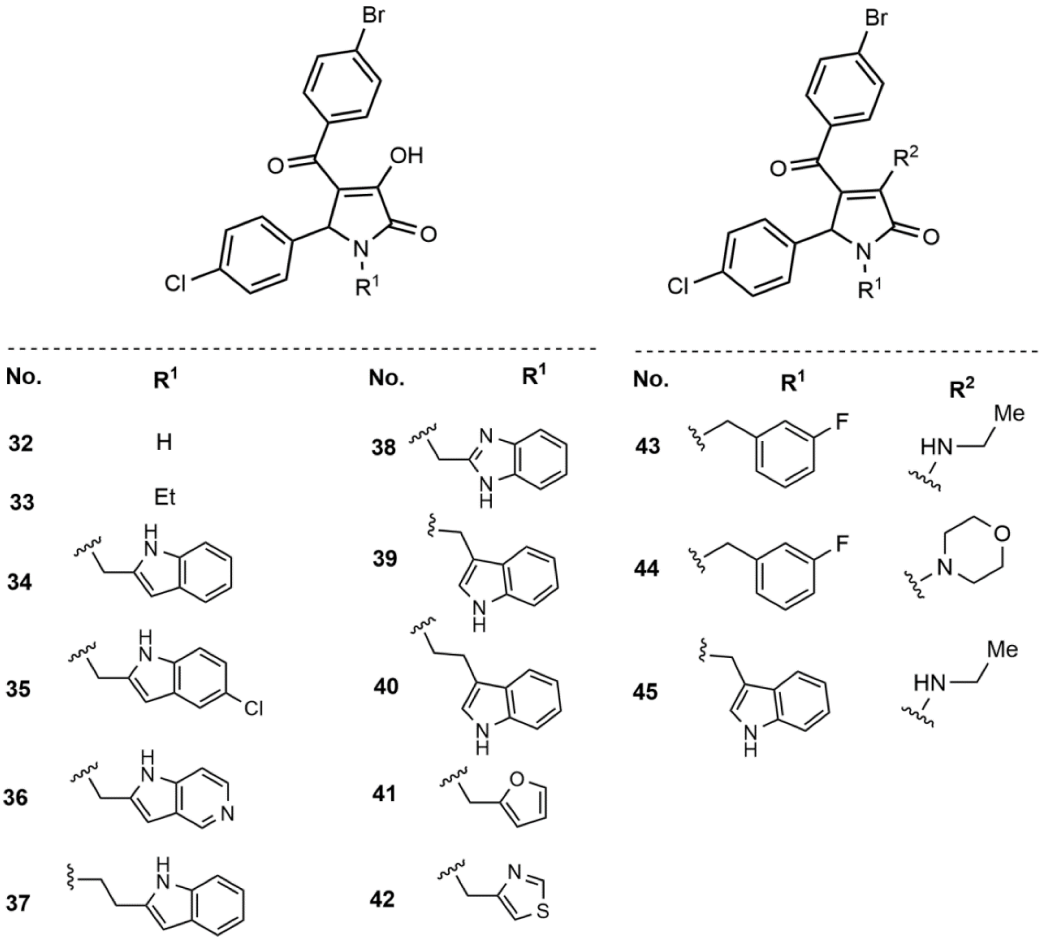


Figure 1.12: Structures of the pyrrolidine-2,3-diones derivatives.
They were initially identified in an HTS approach and chemically optimized. Figure taken from [85].

Based on docking studies it was hypothesized that the pyrrolidine-2,3-dione compounds bind the *Pa*PBP3 non-covalently in the active site (Figure 1.13). However, as shown for compounds 34 and 35, despite similar structures, molecular docking results suggest that the compounds are positioned in different orientations. Therefore, for further structure-based optimization, it is essential to collect additional data that describes the binding mode and affinity of these compounds toward *Pa*PBP3, including, X-ray crystallography data.

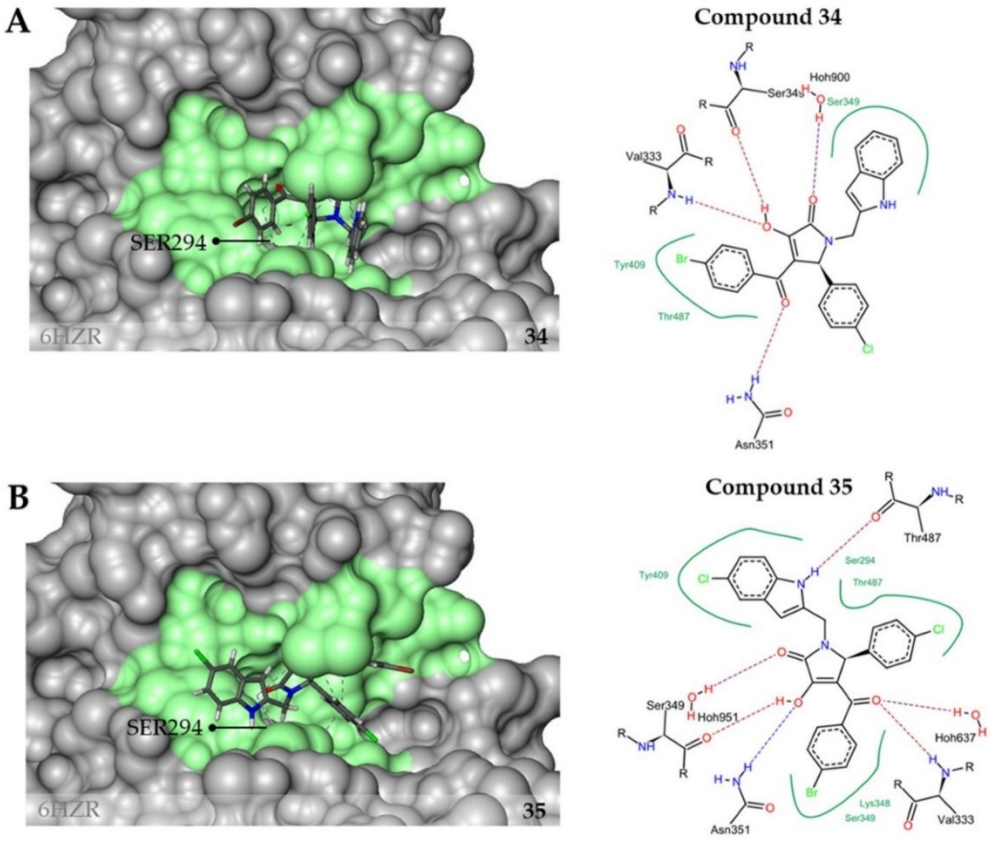


Figure 1.13: Docking models of pyrrolidine-2,3-dione compounds and *Pa*PBP3.

Best-fit position of compound 34 and compound 35 in PDB entry 6HZR are shown in (**A**) and (**B**), respectively [49]. Dotted lines indicate hydrogen-bonds, whereas green lines indicate potential hydrophobic interactions. The proposed active site, including the catalytic serine 294, is shown in green in the space-filled model, whereas the remaining part of *Pa*PBP3 is colored gray. Figure taken from [85].

1.8. Drug discovery pipeline

In drug discovery, an approach that is historically very successful starts with an organism (for example, plants and fungi) that is used to treat a disease. To develop a drug with improved properties, the active molecule has to be extracted from the organism, isolated and physicochemical properties determined. If the chemical synthesis of this lead molecule is available, modifications such as addition or exchange of chemical moieties might further improve affinity and bioavailability (see Section 1.6 as examples). Unfortunately, isolation and characterization of natural products with suitable therapeutic properties from organisms is challenging and time-consuming. Thus, this natural product discovery process is no longer the mainstay in drug discovery efforts.

Many approaches in drug development start with the identification of a target, which can be a protein, nucleic acid or cell membrane. For example, typical targets for developing antibiotics are depicted in Figure 1.6. The next step is the identification or selection of molecules that bind the target. The primary approaches available for identifying molecules that bind targets include phage display, HTS and fragment-based drug discovery (FBDD). Currently, HTS is the major method for lead molecule discovery. Libraries with up to a million different molecules are used with different screening techniques to identify hits that bind the target with affinities in the nM range [86]. For the generation of a molecular library to perform HTS, the “rule of five” is the favored approach for selecting potential molecules [87]. The “rule of five” was established,

because after the comparison of market approved drugs, it was found that more than 90% share similar properties. Therefore, the “rule of five” is applied to select compounds to screen because this approach increases the likelihood that selected compounds reach the market. Basically, this rule includes molecules with less than five H-bond donors, less than 10 H-bond acceptors, a logP value (i.e., logarithm of the n-octanol-water partition ratio with negative values indicating the molecule is hydrophilic and positive values indicating the molecule is hydrophobic) smaller than 5 and compounds with molecular weights lower than 500 Da [87]. Molecules that are soluble in aqueous conditions are usually obtained by following the “rule of five” and for example, as antibacterial lead molecules, they should also be able to penetrate the bacterial cell wall. Taking the molecular weight of the screening molecules (~400 Da; [86]) into account, a library of compounds that cover a broad range of chemical space are used. Identified lead molecules usually have high affinities in the nM range, making these lead molecules promising candidates for further development [86].

The chosen screening approach for the identification of lead molecules that bind the target must be developed and adjusted according to the target. Lead molecules identified from initial screening results usually have weak-to-modest affinities that range from mM (FBDD) to nM (HTS) based on the approach used. Consequently, these lead molecules are not suitable drugs to enter clinical trials. Nonetheless, these lead compounds are used during an optimization process to improve the affinity of the compound toward the target molecule, for example by exchanging or adding chemical moieties. In this process, biochemical assays are used to determine equilibrium dissociation constant (K_d) or IC_{50} values. Once compounds are chemically optimized to give high affinities in the nM range or higher, there are other criteria to consider, such as cell wall or cell membrane permeability [88, 89]. Additionally, if the drug is administered to humans, its pharmacokinetics and pharmacodynamics have to be examined. For example, molecules with logP values below 0.5 may not be able to diffuse through membranes, whereas logP values above 3.0 may dissolve the lead molecule in membranes [90]. Therefore, it is favorable to work with lead molecules featuring logP values between 0.5 and 3.0. However, there are examples of drug-like molecules featuring low absorption rates (azithromycin), or high logP values (dihydropyridines) that are still effective drugs [91, 92]. In summary, all “rule of five” values should be considered in parallel, without rejecting a lead molecule based on only one criterion. Nevertheless, compounds that fail in cytotoxicity assays are rejected. Finally, the administered concentrations, the potential for side effects and the effectiveness in humans are tested in clinical phases I to III before the drug candidate is approved for the market by the FDA (in the US) or the EMA (in Europe). During the process of drug development, many compounds turn out to have adverse properties and their development is discontinued. For example, 85% of all lead molecules entering the first clinical phase fail prior to market approval [93]. For this reason, the number of investigated molecules is reduced dramatically at each development stage, and it usually takes between 10 to 15 years to obtain market approval of a new drug (Figure 1.14; [94]).

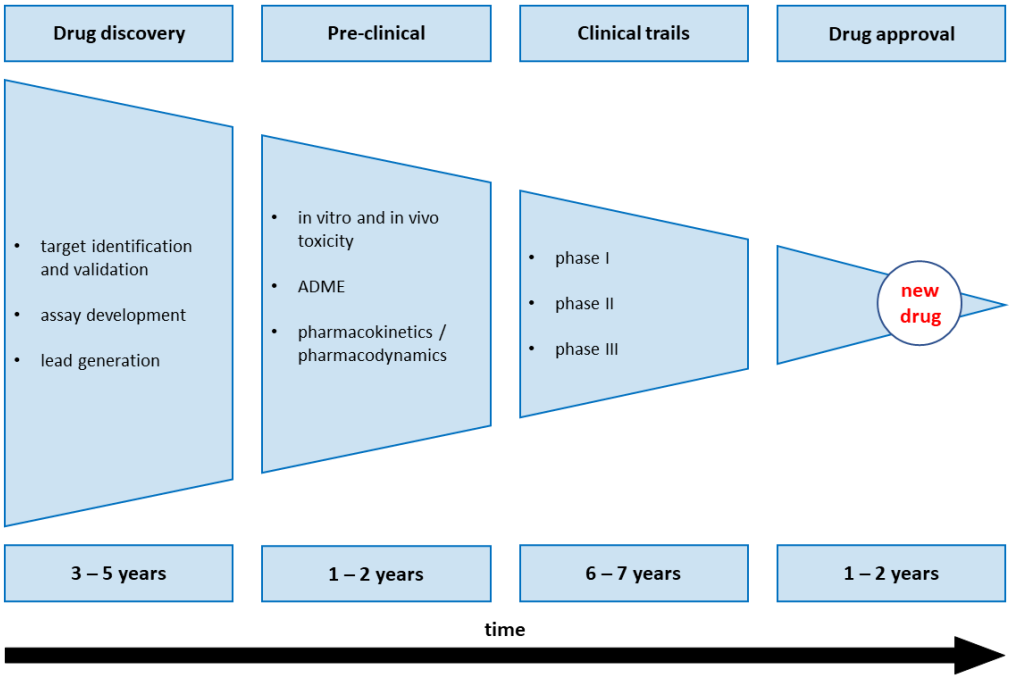


Figure 1.14: Typical drug discovery pipeline.
Tapering indicates a decrease in the number of molecules in the development pipeline. The number of starting molecules varies between different approaches. Usually, the timeframe for market approval of a new drug is 10 to 15 years. Figure adapted from [94].

1.9. Fragment-based drug discovery

An approach for the development of lead compounds and new drugs that is growing in popularity is fragment-based drug discovery (FBDD). The FBDD pipeline includes four main steps: fulfilling initial requirements, screening, validation of hit fragments and optimization of the fragments (Figure 1.15).

To initiate a FBDD study, a fragment library and a target protein are required. Additionally, a structure of the target protein is required during the optimization stage of FBDD. In contrast to HTS, FBDD uses different libraries that contain significantly lower numbers of molecules. To gain similar efficiency when compared with that of HTS, molecules screened against in FBDD have molecular weights of ~150 Da (typically less than 300 Da) and cover the same broad chemical space used in HTS [86, 95]. Basically, the “rule of five” is adjusted to the “rule of three” [96]. Following this rule, fragments in the library should not have more than 3 H-bond donors, nor acceptors, the molecular weight should be below 300 Da and the logP value should be below 3. Additionally, the number of rotational bonds should be less than 3 and a polar surface area (PSA) value >60 Å^2 is considered ideal [97]. After the selection of a library, the fragments are screened using an available method. These methods include standard NMR approaches such as ^{19}F NMR, saturation-transfer difference (STD) NMR, water-ligand observed *via* gradient spectroscopy (Water-LOGSY), ^{1}H-^{15}N HSQC experiments or relaxation-based NMR experiments, whereas popular non-NMR methods are differential scanning fluorimetry (DSF), virtual docking or SPR [94]. Binding to the target protein has to be validated once hit fragments are identified by the screening approach [98]. For this, NMR, X-ray crystallography, SPR, isothermal titration calorimetry (ITC) or biochemical assays can be used to determine K_d values and gather structural information that defines the binding mode [94]. Finally, validated hit fragments are used to build lead compounds using fragment growing, merging and linking.

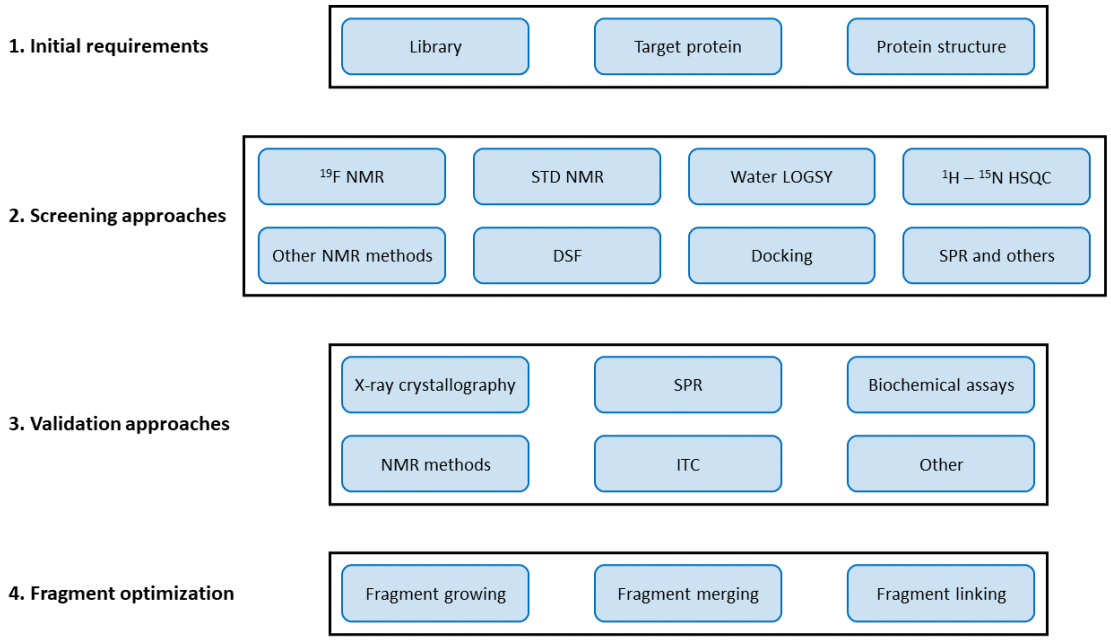


Figure 1.15: Pipeline of the FBDD approach.
A fragment library and a target protein are required for carrying out fragment screening. Additionally, a protein structure is beneficial at the fragment optimization step, thus it is listed as an initial requirement. Next, in the screening approach, various methods can be used to identify hit fragments, which must be validated by an orthogonal method (STD NMR: saturation-transfer difference NMR; DSF: differential scanning fluorimetry; SPR: surface plasmon resonance spectroscopy; ITC: isothermal titration calorimetry). In the final step of FBDD, fragment growing, merging and linking are used to generate a lead compound. Figure adapted from [99].

Fragment growing is the most widely used strategy in FBDD (Figure 1.16A) and is used when there is only one fragment found in a binding pocket that has been proven to show inhibitory activity [100]. Using this approach, a fragment hit that binds the target molecule with moderate affinity (μM-mM) is extended by covalently adding chemical groups to increase the number of non-covalent interactions in the binding site. The aim is to generate second generation molecules with increased affinities. In addition to K_d and IC_{50} values, ligand efficiency (LE) is another indicator used in drug development. This value is calculated by taking the quotient between the logarithm of the IC_{50} value and the number n of heavy atoms in the molecule ($\log(IC_{50})/n$]) and therefore represents the efficiency per heavy atom. Ideally, the LE does not change or even increases when the number of heavy atoms is expanded. The development of inhibitors against the bacterial DNA gyrase and methyltransferases from the erythromycin-resistance methylase family are examples where fragment growing approaches were used successfully [101, 102].

Different fragments from screening that target the same binding pocket of the protein can be merged to generate a lead compound (Figure 1.16B). In this case, an overlapping part of the fragments is used together with additional chemical groups of both molecules to build a larger molecule with improved affinity and LE. The merging approach has been used successfully for developing inhibitors against the Jun *N*-terminal kinase 3, thymidylate synthase and the 3-phosphoinositide-dependent protein kinase [88, 89, 103].

Two fragments that are identified to bind in neighboring regions but do not share similar structures and do not have a common merging component can be connected by a linker (Figure 1.16C). Here, the linker is added to enlarge the molecule with beneficial binding properties. The linking approach has been used to develop inhibitors against FK506 binding protein, matrix metalloproteinase-3 and B-cell lymphoma – extra-large [104-106].

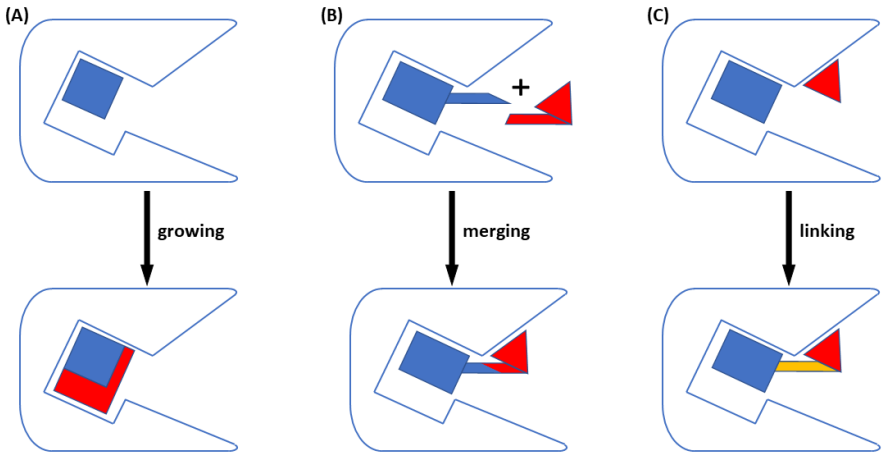


Figure 1.16: Schematic representation of growing, merging and linking approaches used in fragment-based drug discovery (FBDD).

(**A**) Shows the principle of growing in which a starting molecule (blue) is found and used to build a larger compound that fully occupies the binding pocket and has a higher affinity than the original pharmacophore scaffold. The optimized molecule is represented in blue and red. (**B**) When the merging approach is used, overlapping chemical groups of two molecules (red and blue) are merged together to build one larger molecule (red / blue). (**C**) Two molecules (red and blue) can be connected with a linker (yellow), when found to bind pockets on the protein surface that are close proximity to generate a larger molecule. Figure was adapted from [100] and [107].

A major problem in FBDD are the properties of the fragment molecules. Many of the fragments are hydrophobic because they contain aromatic rings and thus show low solubility in water or aqueous buffers. This can be problematic in screening and hit validation of fragments that are initially identified as lead fragments. Measuring the affinities of identified lead fragments can also be challenging because the K_d values are usually between 100 µM and 10 mM, which is often beyond the solubility limit of the fragments in aqueous solutions [86]. Moreover, soaking or co-crystallizing experiments with hit fragments can fail because of their low affinities and solubility. This inability to obtain structural data hampers further optimization of the hit fragment. Alternatively, *in silico* methods such as docking simulations can be used to gain structural insights. In situations where no structure is available, homology modeling may provide a suitable model.

A general problem in drug development is the occurrence of pan assay interference compounds (PAINS; [108, 109]). These are false-positive hits that cannot be distinguished easily and are often not noticed immediately. These compounds do not bind in a specific drug-like fashion to a binding pocket but inhibit the protein after a chemical reaction [108]. In addition, false-positive results may be caused by colloidal aggregation, non-specific binding or interference with the assay detection method [110]. In summary, the interaction between the target protein and PAINS are not specific and cannot be optimized for the target protein. If PAINS are present in the library during the screening approach, it is possible that they are identified as hits. Since they often show beneficial proprieties in the initial screening phase, “real” hit molecules can be excluded in early stages of FBDD, because they show weaker interactions with the target protein when compared with those of PAINS, and therefore, are not further investigated. Although there are common molecules known as PAINS and many commercially available libraries apply filters to exclude them, “harmless” molecules that are not classified as PAINS may bypass the filter but behave as PAINS when, for example, used in a different buffer system [109, 111]. To avoid PAINS in further development steps, it is indispensable to gather structural information about the binding mechanism to ensure selection of non-specific binders is avoided.

Another widespread problem is the emergence of impurities during the synthesis of the compounds. Usually there are several quality criteria, such as 1D NMR or HPLC experiments, but it is still possible to have a small number of impurities. Similar to PAINS, to avoid false-positive hits because of impurities, structural information is necessary. Furthermore, a solubility, purity and aggregation of the molecule (SPAM) NMR filter may be used prior to screening [112]. This filter identifies impurities, degraded fragment components or aggregated fragments prior the screening with the target protein. To perform a SPAM filter two ^{1}H NMR spectra (1D referencing spectrum and WaterLOGSY) and two ^{19}F NMR spectra (1D and T_2 filter spectra) are required for each fragment. Thus, performing a SPAM filter is time- and resource-consuming but may be valuable to avoid false-positive hits and facilitates the analysis.

1.9.1. Using NMR for FBDD

Compared to HTS, the expected affinities of hit fragments in FBDD are much lower (i.e., 100 µM to 10 mM; [86]). Thus, the use of higher concentrations of fragments in the mM range is ideal in FBDD; however, the typically low aqueous solubility of fragments prohibits the use of high fragment concentrations during the fragment-based screening (FBS) process, resulting in low fragment concentrations being used and consequent weaker responses in the performed screening method. Thus, highly sensitive methods have to be used for the FBS in FBDD. Most FBS studies use NMR because the readout from NMR is exquisitely sensitive to any changes in the sample, for example, a small population (e.g., a few percent) of a fragment in the protein-bound state. There are numerous NMR approaches to screen fragments. Below are brief descriptions of some standard, commonly used NMR methods to screen fragment interactions with a target protein.

Recording two-dimensional (2D) ^{1}H-^{15}N HSQC spectra of the protein before and after the addition of a fragment mixture can provide valuable information [99]. In the chemical shift perturbation analysis, changes in the chemical shifts or intensity of peaks indicate identification of a hit fragment. The biggest advantage of this method is the potential to identify immediately the binding pocket if the peaks in the 2D ^{1}H-^{15}N HSQC are sequence-specifically assigned. Even without assignment information it is possible to identify fragments affecting the same peaks and thus, binding in the same binding site. Disadvantages include the requirement of a considerable amount (mg quantities) of ^{15}N-labeled protein material and the acquisition time of the experiment is often much longer than other NMR-based methods, which reduces the rate of fragment screening.

Acquisition of standard one-dimensional (1D) ^{1}H experiments of fragments can also be used to detect interactions between fragments and targets. Peaks in the spectrum representing fragments that bind a target typically show line-broadening and changes in chemical shifts. Line-broadening arises from the bound fragment adopting the correlation time (τ_{cor}) of the protein and chemical exchange between the bound and free ligand states on the chemical shift timescale. Analysis of spectra of fragments in the presence and absence of the protein identifies interacting hit fragments. Automated sample loading and short acquisition times enable high-throughput screening of fragments. However, every fragment gives rise to several peaks in the 1D ^{1}H spectrum and thus, typically no more than five fragments are pooled together to reduce spectral overlap and facilitate analysis [95, 113].

In the STD experiment, which effectively identifies fragments with µM to mM affinities toward the target protein, a cocktail of up to ten fragments is screened against a target protein [114]. A long selective pulse is applied to saturate a small subset of protons of the target protein (e.g., upfield methyl signals) without irradiating nuclei arising from the fragments. This saturation is transferred by dipole-dipole interactions to a protein-bound ligand. In 1D ^{1}H spectra of a fragment cocktail, the intensity of peaks arising from a bound ligand decrease because of saturation transfer and can be readily identified by subtracting a reference spectrum acquired with off-resonance saturation using the same sample. Saturation transfer is most efficient for ligand protons in closest proximity to the protein. Thus, ligand binding epitope maps that contain information about the orientation of the ligand can be obtained. Furthermore, K_d determination is possible by performing a ligand titration. For the STD NMR approach, low protein concentrations between 10 and 100 µM can be used, but a high fragment concentration of 500 µM to 5 mM is required, which may not be possible if the ligand has low aqueous solubility. Ligands that are not in fast exchange and feature high affinities with K_d values below the µM range cannot be examined. Because of the high ligand excess, the protein is fully bound with the ligands and because of the slow exchange the lifetime of the bound-state exceeds the time of saturation. Therefore, only a few percent of the population of the ligand are saturated and differences to the referencing spectrum can be overlooked.

In Water-LOGSY NMR experiments the bulk water is selectively magnetized before the magnetization is transferred *via* the protein-ligand complex to free ligands in a selective manner [115, 116]. Water molecules that interact with the ligand in the binding site usually have low B factors. Therefore, in the nuclear Overhauser effect (NOE) mixing period, the magnetization can be transferred to bound ligands, whereas water molecules diffuse prior to transferring the magnetization towards free ligands. Consequently, peaks arising from fragments that do not interact with the target protein appear with opposite sign and tend to be weaker when compared with those signals arising from fragments that interact with the protein. This approach effectively identifies fragments with µM affinities against the target protein and only a limited amount of protein is required (µM range). Similar to other NMR methods using ^{1}H, the pool size is typically small, thus screening is relatively slow.

In recent years, the advent of probe heads with higher sensitivity and performance combined with improved methods and other hardware have led to a dramatic increase in the use of ^{19}F NMR for FBS, which will be discussed in Section 1.9.2 [99]. Other NMR methods in addition to the abovementioned approaches include T_2-filters, transfer NOE, diffusion ordered spectroscopy (DOSY), inter-ligand NOE (ILOE), pharmacophore mapping (INPHARMA) and SOS (structural information using Overhauser effects and selective labeling) NMR experiments [117].

1.9.2. Using ^{19}F NMR for FBS

Approximately 25% of all drugs contain at least one ^{19}F atom (e.g., fluoroquinolones that target DNA synthesis of bacteria), underlining the importance of fluorine atoms in drugs [118]. ^{19}F is the only stable fluorine isotope and has a spin ½ nucleus and thus detectable by NMR. Additionally, because the gyromagnetic ratio of ^{19}F is 40.05 MHz/T (~94% compared to ^{1}H) and a sensitivity of 83% compared to ^{1}H nuclei, it is possible to record ^{19}F NMR spectra in a timeframe similar to that of ^{1}H NMR experiments [98, 119]. In contrast to 1D ^{1}H NMR spectra of non-fluorinated fragments where the complexity of the NMR spectra limits the size of each pool to usually 2–3 compounds, molecules with one fluorinated group give rise to only a single

intense peak (or complex multiplets if *J*-couplings are active, see below) in a 1D ^{19}F NMR experiment [113]. Moreover, the presence of protonated chemicals (e.g., buffer components) and solvent are not observed in ^{19}F NMR datasets. Additionally, the spectral width of ^{19}F NMR is considerably larger when compared with that of ^{1}H NMR. In proton NMR, most of the signals derived from small organic molecules are observed within a range of 15 ppm, whereas the chemical shift of ^{19}F NMR ranges from approximately +80 ppm (axial fluorine of SF_5) to –272 ppm (CH_3F) [98, 120, 121]. This considerably larger chemical shift range reduces potential spectral overlap. These favorable features enable screening of 20–50 fragments per pool. Thus, the number of experiments required to screen a fragment library is reduced by a factor 15 to 20 when compared with ^{1}H NMR screening methods [98].

Peaks in 1D ^{19}F spectra arising from tri-fluorinated fragments appear as singlets and are not split by active *J*-couplings because (i) no active $^{2}J_{FH}$ couplings are present and (ii) because of the free rotation of the tri-fluoromethyl group, $^{n}J_{FH}$ couplings, where $n > 2$, are too small to be resolved [122, 123]. In comparison, active $^{2}J_{FH}$ couplings for mono- and di-fluorinated organic compounds range between 46 to 79 Hz and $^{3}J_{FH}$ couplings range between 14 and 27 Hz [119]. In aromatic systems, $^{3}J_{FH}$ and $^{4}J_{FH}$ couplings have been determined to be between 8.6–11.2 Hz and 4.0–5.7 Hz, respectively [124, 125]. Consequently, signals in the ^{19}F NMR spectrum are invariably split into complex multiplets by these active J_{FH} couplings, which drastically reduce the intensities of signals arising from di- and mono-fluorinated compounds. Proton decoupling schemes included in the pulse program combined with specific NMR hardware ensure that signals appear as singlet resonances. In some cases, split signals are observed when two conformational states are present and in slow exchange, or the chemical environment is slightly different for ^{19}F nuclei in diastereoisomers [112].

Similar to ^{1}H NMR experiments, in ^{19}F NMR FBS, changes in chemical shift are observed when a population of the fragment is bound to the protein target because of a difference in the chemical environment between bound and free states (i.e., $\Delta\delta_{obs} = p_b(\delta_{free} - \delta_{bound})$, where p_b is the fraction of bound fragment and δ_{free} and δ_{bound} are the free- and bound-state ^{19}F chemical shifts of the fragment, respectively). In ^{19}F NMR, the difference in chemical shift between fully bound and free fragments states ($\Delta\delta$) can be several ppm, which on high field NMR spectrometers corresponds to a few thousand Hz [126, 127]. Although dependent on $\Delta\delta$, $\Delta\delta_{obs}$ can possibly be used in ^{19}F NMR FBS to identify hits.

The observed ^{19}F transverse relaxation rate ($R_{2,obs}$) is a parameter often used to monitor binding of fragments to proteins because the large chemical shift anisotropy (CSA) of fluorine and dipolar contributions make the difference in linewidth for a fragment in the free and bound state very large (Figure 1.17), especially at high magnetic fields, i.e., > 500 MHz ^{1}H frequency. $R_{2,obs}$ is a population-weighted average of R_2 of the fragment in the free and bound states, $R_{2,f}$ and $R_{2,b}$, respectively. Peak intensities arising from fragments that bind a protein broaden essentially because the bound fragment adopts the much longer τ_{cor} of the protein, and larger proteins cause greater line broadening. Additionally, the kinetics of exchange between the free and bound states are typically in the intermediate-to-fast exchange regime, i.e., $k_{ex} \geq 1000\ s^{-1}$, where $k_{ex} = k_{on}[F]_f + k_{off}$, where $[F]_f$ is the fragment concentration of fragments in the free state. Conversion between the two states with different chemical shifts gives rise to an exchange term (R_{ex}), which accounts for line broadening due to chemical exchange. Thus, $R_{2,obs}$ can be defined as:

$$R_{2,\mathrm{obs}} = p_\mathrm{f} R_{2,\mathrm{f}} + p_\mathrm{b} R_{2,\mathrm{b}} + R_\mathrm{ex} \quad (1),$$

where p_b is the fraction of bound fragment (i.e., [PF]/[F]$_{total}$) and p_f is the fraction of free fragment (i.e., [1–PF]/[F]$_{total}$). Importantly, the reduction in the peak intensity from screening experiments is not proportional to the affinity of the ligand simply because $R_{2,obs}$ contains contributions from various effects and only $R_{2,f}$ can be readily measured. In contrast, the chemical shift can be used to measure fragment binding affinity through a titration experiment; however, the affinity is usually weak, and saturation of the protein with fragment is challenging and therefore the endpoint of the titration (i.e., δ_{bound}) is poorly defined or cannot be reached because fragment and/or protein solubility issues. Nonetheless, a number of approaches have been developed, including the use of spy molecules, CSA-based affinity ranking and lineshape analysis, to measure K_d values of ligand-protein interactions using ^{19}F NMR data [128-130].

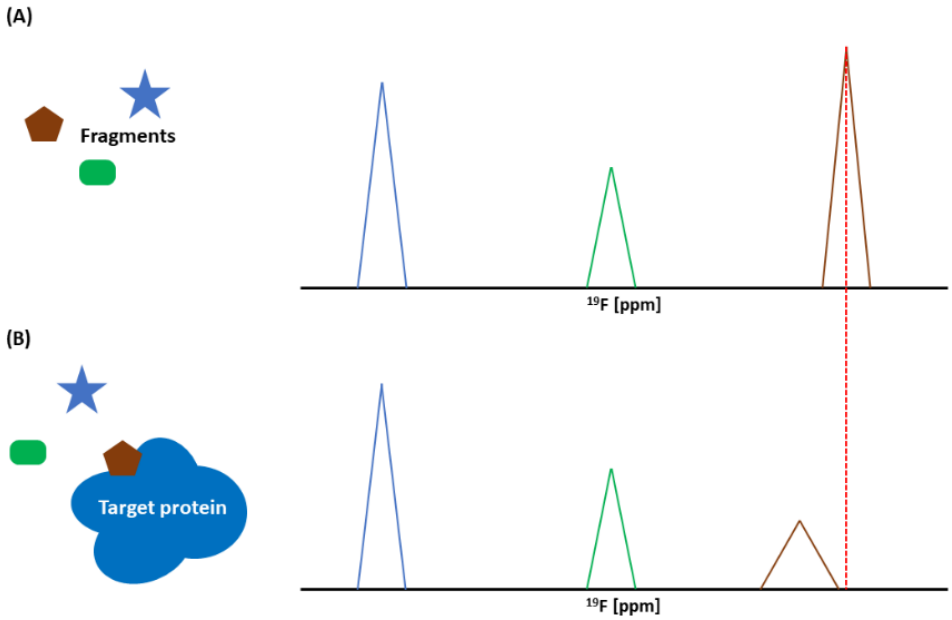


Figure 1.17: Schematic representation of the results expected from 1D ^{19}F NMR fragment screening.
(**A**) Fluorinated fragments (indicated by different shapes on the left) in solution are observed as unique peaks in the corresponding 1D ^{19}F NMR spectrum. (**B**) Once the target protein is added to the sample, fragments are able to bind (brown shape) and a change in intensity (i.e., linewidth at half height increases) and/or chemical shift is observed in the spectrum. Peaks representing non-binding fragments (blue and green shapes) do not show chemical shift or intensity changes after the addition of the protein. Figure adapted from [116].

1.10. Sequence-specific backbone assignments for proteins

NMR is a powerful technique to address many questions in structural biology and can be used to determine a protein structure and to probe protein dynamics, protein-protein interactions or protein-ligand interactions at the atomic level [131, 132]. In drug discovery, obtaining sequence-specific backbone assignments of the peaks in the 2D ^{1}H-^{15}N HSQC has a number of advantages that can advance the drug discovery process. Peaks in the 2D ^{1}H-^{15}N HSQC arise from amide group correlations (i.e., ^{1}H–^{15}N correlations). Thus, backbone amide groups of every amino acid (except proline) in a protein give rise to a peak in the 2D ^{1}H-^{15}N HSQC spectrum. Additionally, ^{1}H–^{15}N correlations from amino acids such as asparagine or glutamine with side chain ^{1}H–^{15}N moieties also give rise to signals in the 2D ^{1}H-^{15}N HSQC spectrum. The ^{1}H and ^{15}N chemical shifts of these correlations differ from one another because the chemical environment of each backbone amide group in a folded protein is typically unique. Therefore, each spectrum of a protein is unique and the 2D ^{1}H-^{15}N HSQC is often referred to as the fingerprint spectrum of a protein. Sequence-specific assignment of the 2D ^{1}H-^{15}N HSQC of a protein provides valuable information that can be used to probe protein function and structure

in different environments. For example, a sequence-specific assigned 2D ^{1}H-^{15}N HSQC can be used in FBDD for fast identification of the binding pocket of a hit fragment, and fragments binding in the active site can be specifically selected for further development to increase the possibility of inhibiting the protein. Nonetheless, obtaining these assignments can be challenging, especially for large proteins that have molecular weights greater than approximately 25–30 kDa [133-135], because the increased τ_{cor} causes the linewidth to broaden and therefore, reduce peak intensities and resolution, which can lead to spectral overlap.

Multi-dimensional (e.g., three-dimensional (3D), four-dimensional (4D)) heteronuclear NMR experiments are required to obtain sequence-specific backbone assignment information of proteins larger than approximately 10 kDa [133]. With these experiments, sequence-specific $^{1}H_N$, ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$ and $^{13}C'$ chemical shift information can be obtained. Moreover, each amino acid has a specific set of ^{13}C chemical shifts that facilitate the identification of the amino acid type. Typically, the $^{1}H_N$ and ^{15}N nuclei of the backbone amide groups are correlated with ^{13}C nuclei in 3D heteronuclear NMR experiments, referred as spin systems. Thus, the 2D ^{1}H-^{15}N HSQC acts as an anchor experiment with the third dimension providing carbon chemical shift information. The sequence-specific assignment process begins with identifying such spin systems obtained from 2D, 3D and 4D heteronuclear *J*-coupled NMR experiments. In particular, 3D heteronuclear NMR experiments that record the C_α and C_β chemical shifts are essential because these nuclei have characteristic chemical shift ranges for each amino acid (see Biological Magnetic Resonance Data Bank; BMRB; [136]), with some chemical shift ranges being unique, e.g., C_α chemical shift of glycine. Next, sequence-specific assignment of resonances can be achieved by using combinations of complementary multidimensional heteronuclear NMR experiments to link spin systems. Once several spin systems are sequentially linked, and some or all amino acid types have been identified for this stretch of linked spin systems, comparison with the primary protein sequence enables unambiguous assignment of the linked resonances.

Typically, the combinations of 3D HNCA/HN(CO)CA and 3D HN(CA)CO/HNCO are used together to assign backbone resonances for large proteins. The more powerful 3D HNCACB/HN(CO)CACB combination, which provides both C_α and C_β chemical shifts, can also be used, but often the signal-to-noise ratio of the 3D HNCACB experiment is low for large proteins, which prohibits the acquisition of good quality datasets. As an example, the 3D HN(CO)CA correlates the backbone $^{1}H_N$ and ^{15}N nuclei of residue *i* with the C_α nuclei of residue *i*–1, whereas in the 3D HNCA experiment, the magnetization is transferred from the $^{1}H_N$ (*via* ^{15}N nuclei) of residue *i* to the C_α nuclei of residues *i* and *i*–1 (Figure 1.18). Using the 3D HN(CO)CA, C_α peaks in the (^{1}H, ^{13}C) plane (i.e., $C_{\alpha, i-1}$) can be aligned to peaks in the same plane of the 3D HNCA. These peaks are anchored to peaks in the 2D ^{1}H-^{15}N HSQC because of the common backbone amide $^{1}H_N$ and ^{15}N chemical shifts in both experiments. The 3D HNCA spectrum provides additional C_α peaks (i.e., $C_{\alpha, i}$) that can be matched to C_α peaks in different (^{1}H, ^{13}C) planes in the 3D HN(CO)CA (i.e., in another spin system). Aligning the 3D HN(CO)CA to the 3D HNCA spectrum identifies additional C_α chemical shifts (i.e., $C_{\alpha, i+1}$). Following this “sequential walk”, exemplarily shown in Figure 3.30 in Section 3.3, provides sequential assignment information for the $^{1}H_N$, ^{15}N and C_α resonances of the connected amino acids that can be translated to the 2D ^{1}H-^{15}N HSQC spectrum. Similarly, the 3D HNCO/HN(CA)CO combination records C' chemical shifts instead of C_α chemical shifts (Figure 1.18). These additional chemical shifts can be crucial when the “sequential walk” is performed for proteins larger than 25 kDa because several potentially matching C_α chemical shifts may be found, thus creating assignment ambiguity, and the correct match can be confirmed with

matching C' chemical shifts. As aforementioned, in addition to C_α and C' chemical shifts, C_β chemical shifts can be obtained by recording the 3D HN(CO)CACB/HNCACB combination (Figure 1.18), which can be used in combination with the C_α chemical shift to provide amino acid type information.

Because the type of some amino acids cannot be gained unambiguously using only C_α and C_β chemical shifts, experiments such as the ^{15}N TOCSY-HSQC can be recorded to gain chemical shifts of side chain ^{1}H nuclei. However, this experiment does not work with deuterated samples which are required for large protein. Thus, having only C_α and C_β chemical shifts available, amino acids with unique C_α and C_β chemical shifts such as serine, threonine, glycine, or alanine work as promising starting points to initially identify the position in the protein sequence. Dependent on the size of the protein, a unique combination of amino acids in the sequence needs to be found to unambiguously determine the exact position in the protein sequence and for large proteins, it is possible that more than four amino acids are required. Longer stretches of linked spin systems reduce any potential assignment ambiguity. Often, a complete “sequential walk” through the backbone using a combination of these experiments is not possible because of the presence of proline residues which do not contain amide protons and therefore, do not give rise to peaks in the 2D ^{1}H-^{15}N HSQC and related 3D experiments. Additionally, it is possible that particular correlations are absent because of line-broadening (i.e., fast transverse relaxation rate, R_2) caused by chemical exchange.

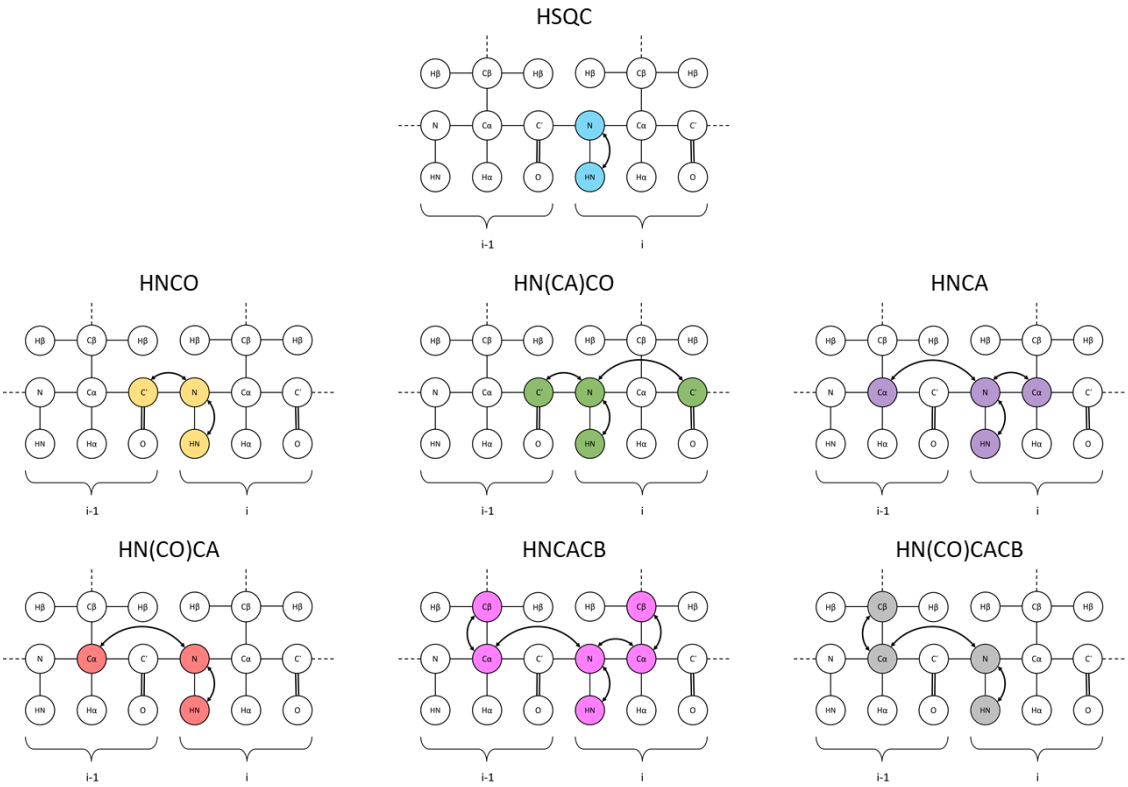


Figure 1.18: Overview of NMR experiments used to obtain sequence-specific backbone assignments, which can be used to assign the 2D ^{1}H-^{15}N HSQC spectrum (top).
Magnetization that is transferred among nuclei in each experiment is highlighted in different colors.

The abovementioned 3D triple-resonance experiments are generally suitable for obtaining the sequence-specific assignments of uniform [$^{13}C/^{15}N$]-labeled proteins no larger than 25 kDa. Two main challenges exist for studying large proteins by multi-dimensional triple-resonance NMR spectroscopy. These are (i) the low signal-to-noise (for example by line-broadening) due to the increased relaxation rates caused by the slow overall tumbling (leading to higher τ_{cor} values) of the protein and (ii) insufficient spectral resolution because of the large number of signals. Approaches to overcome these issues to enable the study of large proteins by NMR have been developed and are outlined below.

Perdeueration of proteins improves the resolution and sensitivity of NMR experiments. By substituting ^{1}H for ^{2}H in proteins the rate of dipole-dipole relaxation of the observed protons is reduced because the gyromagnetic ratio of deuterium (D) is 6.5 times smaller than that of proton (i.e., $\gamma_D \sim 1/6.5\ \gamma_H$, and the relaxation rates are scaled proportional to $(\gamma_D/\gamma_H)^2 \sim 0.02$; Figure 1.19). Therefore, complete deuteration of all but exchangeable H_N protons improves the sensitivity of heteronuclear triple-resonance NMR experiments used to obtain sequence-specific backbone assignments (Figure 1.19). Amide groups containing ^{2}H atoms, will usually back-exchange to ^{1}H when the protein is stored in an aqueous buffer prepared using H_2O. The improvements for these experiments arise from: (i) all nuclei, especially ^{13}C, relax more slowly thus allowing more magnetization to be transferred between *J*-coupled nuclei; (ii) removal of $J_{HNH\alpha}$ couplings; and (iii) the slower relaxation of $^{1}H_N$ nuclei yields sharper, more intense signals. Moreover, the ^{13}C chemical shift evolution periods can be extended because of the slower relaxation of ^{13}C nuclei, thereby providing higher resolution in the indirect dimensions, and allowing highly sensitive constant-time experiments, which yield poor signal-to-noise with fully protonated large proteins because of the long delays required to refocus unwanted J_{CC} couplings.

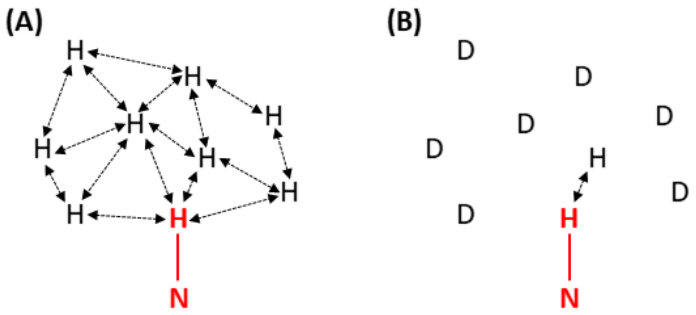


Figure 1.19: Principle of improved relaxation properties in deuterated protein.

(**A**) Protons of amino groups (red) build dipole-dipole interactions with other protons, resulting in spin diffusion and increased relaxation. (**B**) The lower gyromagnetic ratio of deuterium (γD ~1/6.5 γH) and the slower relaxation rate ((γD/γH)2 ~0.02) when compared to ^{1}H minimizes the dipole-dipole interactions. Therefore, the amide signal relaxes slower and leads to narrowed linewidths and higher intensities. Figure adapted from [137].

Although deuteration extends the molecular size range of proteins accessible for NMR studies, transverse ^{15}N relaxation is negligibly affected by deuteration because it is dominated by the heteronuclear dipole-dipole interaction with the attached amide proton. Unfavorable ^{15}N relaxation that hampers straightforward acquisition of high sensitivity 3D triple-resonance experiments of large proteins can be overcome by using transverse relaxation-optimized spectroscopy (TROSY; [138, 139]), which reduces the transverse relaxation rates of ^{15}N and $^{1}H^N$ in the amide groups by using constructive interference between the H^N–N dipolar coupling and either the ^{15}N CSA or ^{1}H CSA, respectively. TROSY-type optimization is achieved by preserving unique coherence transfer pathways with distinct transverse relaxation properties (Figure 1.20). TROSY benefits triple-resonance experiments because part of the gain achieved

with TROSY stems from the reduced transverse relaxation rates of ^{15}N nuclei. Thus, implementation of the ^{1}H-^{15}N TROSY principle in standard 3D triple-resonance NMR experiments combined with uniform ^{2}H/^{13}C/^{15}N-labeling of proteins enables backbone sequential assignments of larger proteins. Additionally, the TROSY effect increases at higher magnetic fields (greater than 600 MHz ^{1}H resonance frequency) with an optimum for the narrowest linewidth at 900 MHz and a sensitivity maximum around 1.2 GHz [140].

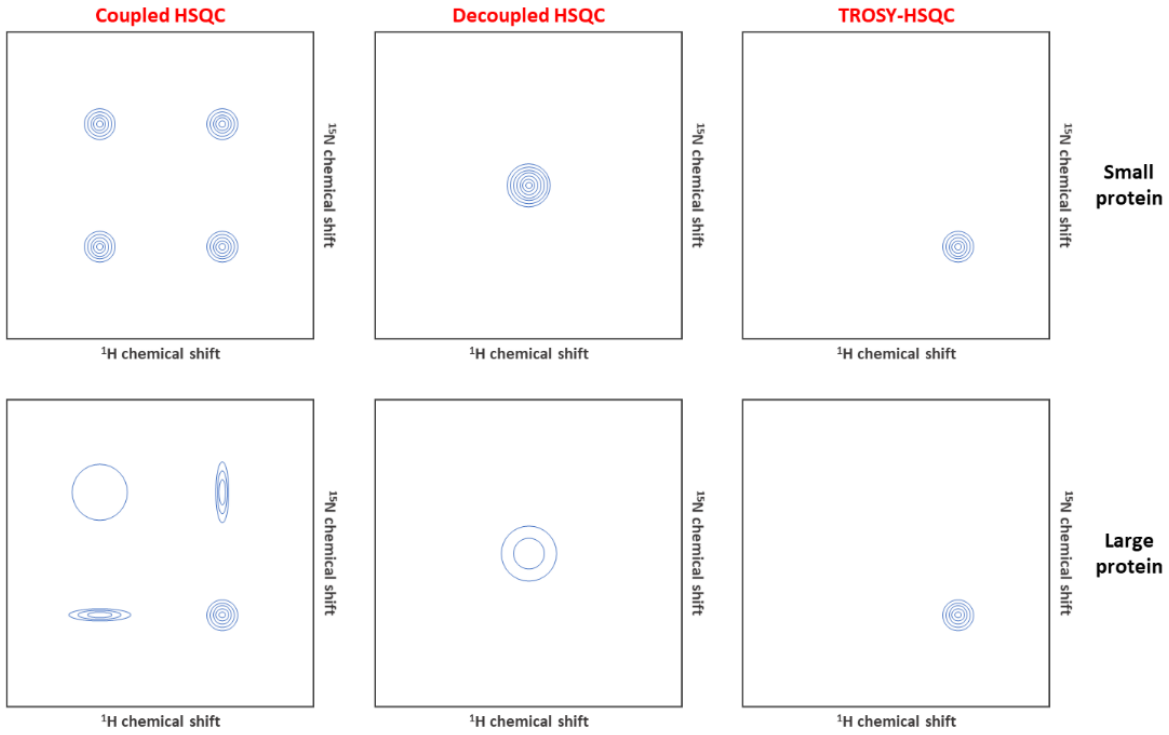


Figure 1.20: Schematic representation of the TROSY principle.
Left: in coupled HSQCs the components of H^N–N dipolar coupling and either the ^{15}N CSA or ^{1}H CSA are visible but will be averaged when decoupling is present (middle). Selecting the narrowest component using the TROSY principle (right) results in a narrowed peak when used for larger proteins (bottom). Thus, the resolution and the sensitivity can be increased when the TROSY principle is used. Figure adapted from [141].

As described above, line-broadening effects caused by large proteins can be partly circumvented by ^{2}H-labeling and TROSY-based experiments. These strategies increase the maximal resolution that can be achieved for a protein by approaching the natural linewidth of the signal, i.e., $1/\pi T_2$. However, this improvement can only be exploited if the digital resolution in the spectrum matches the linewidth [133]. Interferograms in indirect dimensions (t_1, t_2, ..., t_i) are traditionally recorded by uniform sampling evolution delays at increments Δt_1 of 1/SW (where SW is the sweep width). Therefore, to gain higher digital resolution with a constant SW, more increments have to be recorded. In particular, at higher magnetic fields with wider sweep widths (in Hz), the increments are smaller, and consequently more points have to be recorded to gain the required resolution. For this reason, the benefits of deuteration and introduction of the TROSY principle into 3D NMR experiments leads to acquisition of long experiments to obtain the desired resolution, with this increase being exponential as the dimensionality of the experiment increases. To overcome this issue, non-uniform sampling (NUS) has been introduced to avoid extended experimental times by skipping points in the sampling schedule (Figure 1.21; [142-144]). The sampling factor describes the number of non-uniform sampled points divided by traditionally uniform sampled points in the indirect dimensions, respectively, and can be decreased at higher dimensional experiments (factor of 5 for every additional

dimension). Sampling factors of 2–4% can be used for 3D experiments. Thus, many points are skipped during the acquisition and the experiment is considerably shorter than the standard uniform sampling experiment. To generate a random sampling schedule the Poisson gap schedule was developed with minimized gap sizes and less gaps at the beginning and the end of the sampling [145]. Skipping points in the sampling schedule would normally lead to artifacts in the spectrum during processing but are addressed by the iterative soft threshold (ist) algorithm that is used to reconstruct (i.e., it calculates the missing points) the data prior to Fourier-transforming. Overall, because of the dramatically reduced experimental time, experiments recorded with NUS feature an improved resolution, because further sampling in the indirect dimensions is possible, and enhanced signal-to-noise when compared to traditionally recorded spectra with the same experimental time.

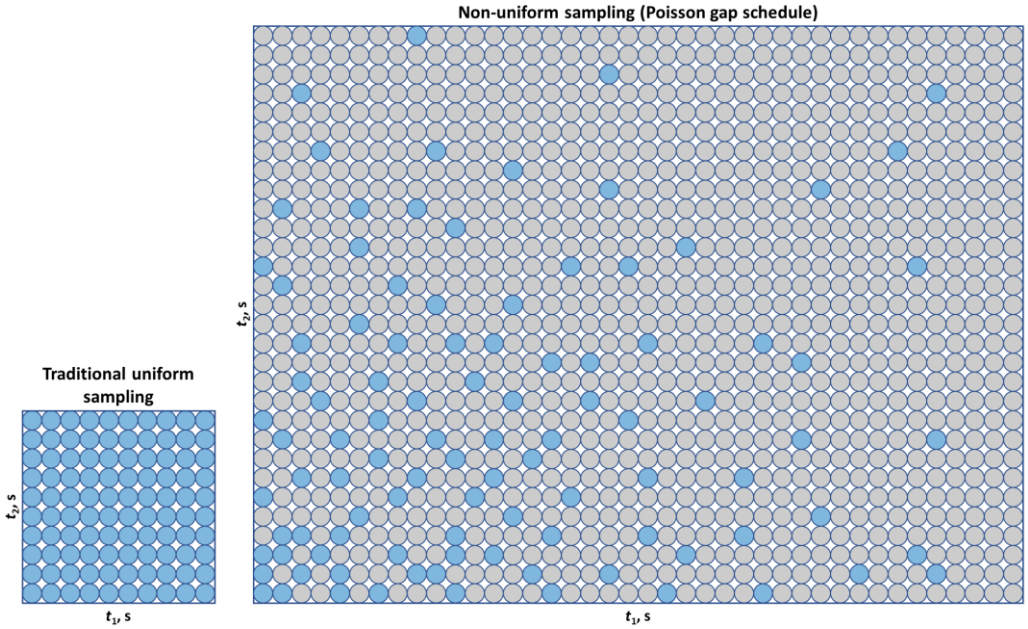


Figure 1.21: Schematic representation of the t_1 and t_2 sampling in the traditional uniform (left) and the non-uniform sampling (right, NUS).

Blue represents points, which are recorded during the experiment, while grey points are skipped in NUS. In the example, a Poisson gap schedule with a sampling factor of 8.3% was used with the same experimental time leading to longer t_1 and t_2 values being recorded and therefore, an increase in resolution in both indirect dimensions.

In some cases, it is not possible to find assignments for all peaks in the 2D ^{1}H-^{15}N HSQC, for example due to unfavorable relaxation properties causing a loss of signal in 3D experiments. Therefore, it might not be possible to identify the binding pocket for hit fragments in the process of FBDD. Nevertheless, it can be useful to record 2D ^{1}H-^{15}N HSQCs in FBDD. If two fragments interact and cause similar peaks to change chemical shift in the 2D ^{1}H-^{15}N HSQC spectrum, this indicates that these two fragments bind to a similar region of the protein. This qualifies the hit fragments for further development with either the linking or merging approach. However, without structural information it is difficult to do so, and peak assignments facilitate this process greatly. A further benefit of recording 2D ^{1}H-^{15}N HSQC spectra even without assignment information is the validation of hit fragments. Fragments that bind the protein give rise to changes in chemical shifts for particular peaks, whereas PAINS that do not bind a specific binding pocket, but bind non-specifically, can be excluded in further drug development.

1.11. Structural investigations of proteins using X-ray crystallography

For determining a protein 3D structure, prevailing methods currently in use are cryo-electron microscopy (cryo-EM), NMR spectroscopy and X-ray crystallography. Although cryo-EM is rapidly advancing and an increasing proportion of structures is getting solved using this technique, it still accounts for less than 5% of all entries in the PDB. While about 8% of the structures are determined with NMR, almost 90% are assigned to X-ray, which thus turns out to be the dominating technique for solving protein structures [146].

The first step in X-ray structure determination is the crystallization of the target protein. Starting from a protein sample with high purity, it is necessary to find conditions in which the protein is able to crystallize. The process of crystal formation requires a thermodynamic driving force, which is associated with supersaturation of the protein solution. In general, the solubility of a protein is affected by many factors, such as temperature, pH, and ionic strength; chemicals reducing protein solubility are commonly referred to as precipitants. Thus, the degree of supersaturation can be adjusted by modifying the protein concentration and / or the concentration of precipitants. At lower concentrations, the sample is in an undersaturated state, in which it forms a clear solution. Crossing the solubility line in the phase diagram, supersaturation is reached, and the system first enters the metastable zone, in which the solution stays homogeneous unless seeds are present (Figure 1.22A). At even higher concentrations, beyond the decomposition line, the nucleation and precipitation zones are reached. In contrast to precipitation, in which the protein yields a mixture of irregular aggregates, nucleation refers to the formation of more ordered oligomers which can grow to bigger homogeneous protein crystals. Crystal growth lowers the protein concentrations in solution until the system is equilibrated with grown crystals and protein in solution (Figure 1.22B).

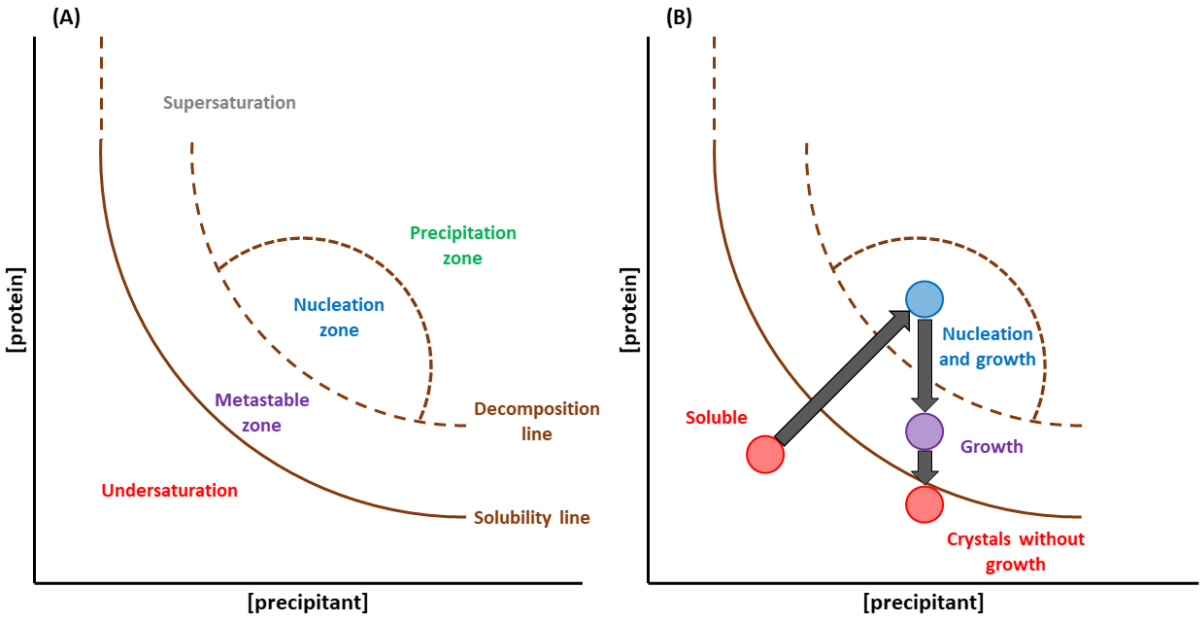


Figure 1.22: Phase diagram for crystallization of proteins.

(**A**) At low protein and precipitant concentrations, the system is in undersaturation. Increase leads to supersaturation and the metastable, nucleation, and precipitation zones can be reached by crossing the solubility and decomposition lines. (**B**) If the system reaches the nucleation zone, nuclei can be formed, and crystals start growing until the system is equilibrated with grown crystals and protein in solution.

This process is dependent on collision of protein molecules in favorable orientations, making crystallization a stochastic process. As a result, the time required for nucleation and crystal growth to occur can vary even between crystallization drops with identical conditions. Sometimes, *de-novo* nucleation is difficult to achieve practically, or nucleation rates are too high to be useful. In these cases, the nucleation step can be skipped by addition of microseeds to a metastable solution, which implies that lower protein and / or precipitant concentrations can be used. Interestingly, seeds do not need to be produced from crystals grown under similar conditions, and may even be derived from an unrelated protein, pointing to a role of non-specific surface effects [147]. Because the exact conditions for nucleation to occur are not readily predictable, in practice a large number (hundreds to thousands) of different conditions are tested to achieve protein crystals.

Independent of the conditions to be used for screening, the crystallization technique itself can be varied, potentially yielding different results. The most common approach is the vapor-diffusion method, which can be realized in hanging-drop or sitting-drop geometry. In principle this technique works by letting a mixture of protein solution and screening solution equilibrate (*via* the vapor phase) with a reservoir of the pure screening solution in a closed system. Because the reservoir osmotically absorbs water, the water in the protein drop diffuses, leading to a decrease in volume of the drop. Thus, both protein and precipitant concentrations slowly increase, and so does the probability of nucleation. Once the first protein crystals can be observed in initial screening conditions, a fine screening might help to improve their quality.

Ideally, all protein molecules build a regular 3D lattice, with the unit cell representing the smallest volume element which repeats periodically in all three dimensions. Therefore, the unit cell which is characterized by lengths a, b, c and angles α, β, γ represents the structure of the entire crystal. In the presence of additional crystallographic symmetry, the unique volume further reduces to the so-called asymmetric unit. Based on the associated metric restrictions to the lattice parameters, six primitive lattice types (corresponding to the six crystal systems) can be defined: triclinic, monoclinic, orthorhombic, trigonal or hexagonal, tetragonal, and cubic. Centering operations give rise to an additional eight lattice types, resulting in a total of 14 so-called Bravais lattices. Considering all possible combinations of crystallographic symmetry elements including centering operations, 230 unique symmetry groups, termed space groups, can be defined. However, the chiral nature of protein molecules limits the permissible symmetry operations to those preserving handedness, i.e., pure rotations and screw operations (coupling rotation and translation), in addition to the lattice translations. Out of the 230 space groups, only 65 (known as the Sohncke space groups) are thus compatible with protein molecules. Point groups, on the other hand, are symmetry groups not including any translational elements. In total, there are 32 unique point groups, of which eleven are chirality-preserving. Note that non-chiral point groups also play a role for describing symmetry properties in protein crystallography, an example being the diffraction pattern, which – in the absence of anomalous scattering – appears centrosymmetric.

To record a diffraction dataset, protein crystals are usually cooled down to 100 K with cryogenic nitrogen, which significantly reduces radiation damage. In this context, it is important to note that the solution should be cryo-protected to prevent water molecules from forming ice and thereby damaging the protein crystal. In the next step the sample gets irradiated with an X-ray beam and the scattering caused by electrons can be detected. Because of the regularity of the lattice and the interference of the scattered photons, a diffraction pattern consisting of distinct spots is observed. A simple way to visualize the conditions leading to a diffraction signal is the so-called Ewald construction. Basically, a signal can be observed on the X-ray detector when

a point of the reciprocal lattice, which is identified by the Miller indices *h*, *k* and *l*, matches the surface of the Ewald sphere (Figure 1.23). When rotating the crystal and thereby the reciprocal lattice, new matches will arise, and additional data are gained. Dependent on the space group, the crystal has to be rotated through a certain angular range to achieve a complete data set.

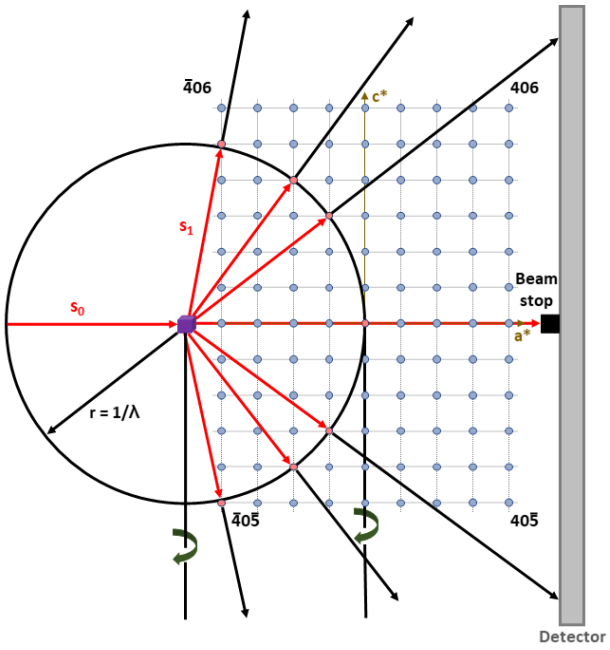


Figure 1.23: Schematic representation of the Ewald construction and the reciprocal lattice during the diffraction experiment.

The crystal (violet) in the center of the Ewald sphere with radius 1/λ (where λ is the wavelength of the beam) is irradiated from the left with an X-ray beam (red) with wave vector $\mathbf{s}_0$, giving rise to diffracted waves with vectors $\mathbf{s}_1$. The scattering vector **s** (not shown) is calculated by subtracting the wave vectors $\mathbf{s}_1$-$\mathbf{s}_0$. Using the distal intersection point of the primary beam and the Ewald sphere as the origin, the scattering vectors **s** representing potential diffraction signals define the reciprocal lattice (gray dots). A signal can actually be observed if the Ewald sphere touches reciprocal lattice points (red dots) and the diffracted beam reaches the detector. To obtain additional signals, the crystal together with the respective reciprocal lattice is rotated (indicated by green arrows) and new matches are realized that can be detected. Because most photons are not scattered the beam stop (black square) is required to prevent the detector from overexposure.

Following data integration and scaling, the structure factor amplitudes can be derived from the intensities of the diffraction signals. Importantly, the information about the phase angles associated with structure factors is lost in this process, giving rise to the so-called phase problem: calculation of an electron density requires both structure factor amplitudes and phases, so the latter need to be retrieved indirectly. In the past, additional phasing experiments were often required to solve the phase problem. Nowadays, as many more protein structures have become available, it is sufficient in most cases to perform a molecular replacement. In this approach, phases are calculated from a similar protein structure (with candidates identified by sequence alignment) after appropriate placement into the actual asymmetric unit. Thus, an initial electron density can be calculated by performing a Fourier transformation, using hybrid structure factors based on experimentally determined amplitudes and model-derived phases. In iterative cycles of automated refinement and interactive modification, the model of the protein is adapted to the experimental data and the electron density, respectively, the latter again using the phases of the current model. One issue with the molecular replacement approach is the model bias: the phases, which strongly dominate the appearance of an electron density, are not determined *de novo* but are based on a previously built structure, so certain errors in the model may be difficult to spot in the density. In addition to optimized Fourier coefficients for map calculation, the use of geometry restraints (such as bond lengths and

angles) helps to alleviate this issue. In practice there are different indicators used to assess the quality of the model, such as the R values, main chain (Ramachandran plot) and side chain torsion angles, and correlation with electron density. The process of determining a protein structure *via* X-ray crystallography, starting with production of the protein, and resulting in a database-quality model, is depicted in Figure 1.24.

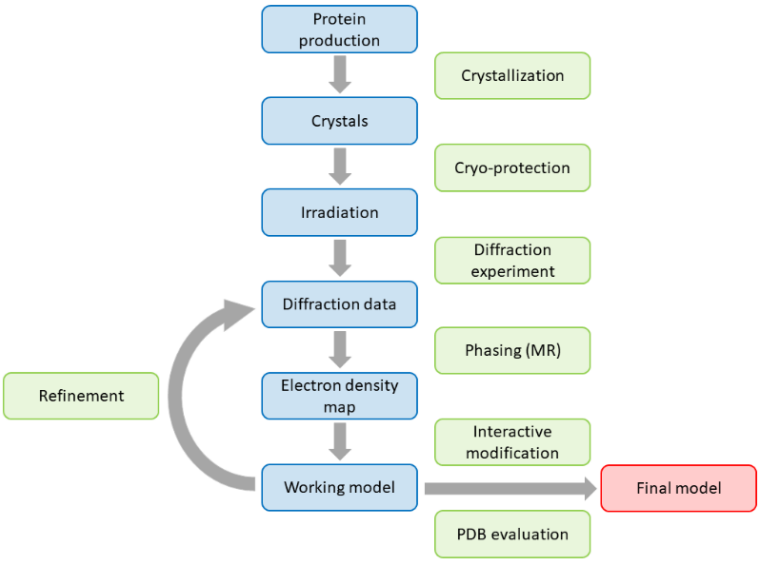


Figure 1.24: Flow chart of the process to determine a protein structure using X-ray crystallography.

1.12. Aims of the project

In this thesis there are basically three aims to achieve: (i) ^{19}F NMR fragment screening was performed against His-tagged *Ec*PBP3ΔTM as part of an FBDD project. Therefore, 1D ^{19}F NMR experiments were established and automated, and the results from screening of mono- and tri-fluorinated fragment libraries analyzed. Identified hit fragments were validated and checked by orthogonal methods. (ii) With the aim to locate the binding pockets of the fragments identified in the ^{19}F NMR screening, multi-dimensional NMR experiments were performed to obtain sequence-specific protein backbone assignments of the 2D ^{1}H-^{15}N TROSY-HSQC spectrum. Because of the size of *Ec*PBP3ΔTM (approximately 60 kDa), a novel truncated construct containing the catalytic TPd plus the adjacent linker subdomain (*Ec*TPd*; approximately 40 kDa) was established by the cooperation partner AiCuris and used for the experiments. (iii) Finally, novel inhibitors of *E. coli* and *P. aeruginosa* PBP3 were investigated. First, AIC499, developed in-house by AiCuris, binds covalently in the active site of PBPs and was structurally investigated in complex with PBP3 from *E. coli* and *P. aeruginosa*. To address a common issue regarding the flexible part of the *E. coli* PBP3, mainly in the n-PBd, the truncated *Ec*TPd* construct was crystallized in the absence and presence of AIC499, and the structures were determined. Resolved PBP3 protein structures were published in Freischem *et al.* [48]. Secondly, pyrrolidine-2,3-diones derivatives (also referred to as Cluster J compounds) were studied at AiCuris using an HTS approach against PBP3 from *P. aeruginosa.* To validate the binding with orthogonal techniques, K_d values were determined in this work using SPR spectroscopy. The development of a new fluorescence assay for read-out, the discovery of the pyrrolidine-2,3-dione derivatives as inhibitors against PBP3 from *P. aeruginosa* and the determination of K_d values was published in López-Pérez *et al.* [85].

2. Material and Methods

2.1. Material

NMR spectrometers (Table 2.1), software (Table 2.2), instruments (Table 2.3), chemicals (Table 2.4), materials (Table 2.5) and crystallization kits (Table 2.6) used in this research project are tabulated.

Table 2.1: NMR instruments used in this thesis.

All instruments were supplied by Bruker (Karlsruhe, Germany).

^{1}H resonance frequency [MHz]	Field [T]	Designation	Probe
600	14.1	B600-F	5 mm CryoProbe Prodigy BBO
600	14.1	B600-O	5 mm cryo-TCI H-/C/N-D
600	14.1	B600-M	5 mm cryo-QCI H-P/C/N-D
700	15.6	B700	5 mm cryo-TCI H-/C/N-D
900	21.1	B900	5 mm cryo-TCI H-/C/N-D

Table 2.2: Software used in this thesis.

Software	Supplier / reference
IconNMR	Bruker, Karlsruhe, Germany
TopSpin (v. 3.5, v. 4.0.8)	Bruker, Karlsruhe, Germany
NMRPipe / NMRDraw	[148]
CcpNMR analysis	[149]
XDS	[150]
STARANISO	[151]
MOLREP	[152]
phenix.refine	[153]
COOT	[154]
PyMol	[155]
LigPlot+	[156]
TALOS-N	[157]
Biacore T200 Evaluation Software (version 3.2)	Cytiva, Vancouver, Canada
GraphPad Prism (version 7)	GraphPad Software, San Diego, USA

Table 2.3: Instruments used in this thesis.

Instrument	Supplier
Biacore T200	Cytiva, Vancouver, Canada
Benchtop centrifuge 5417R	Eppendorf, Hamburg, Germany
Lambda 25 UV/VIS spectrometer	Perkin Elmer, Waltham, USA
SampleCase (24-position auto sampler)	Bruker, Karlsruhe, Germany
Spectrofluorometer QuantaMaster 7	HORIBA Scientific, Oberursel, Germany
J-1100 CD spectrometer	JASCO, Pfungstadt, Germany
Pipetting robot	Tecan, Männedorf, Switzerland

Table 2.4: Chemicals used in this thesis.

Chemical	Supplier
D_2O	Eurisotop, Saint-Aubin Cedex, France
Na-trimethylsilylpropanesulfonate (DSS)	Merck, Darmstadt, Germany
Trifluoro acetic acid (TFA)	AppliChem, Darmstadt, Germany
Fluorobenzene	Merck, Darmstadt, Germany
1,4-Difluorobenzene	Merck, Darmstadt, Germany
KF	Merck, Darmstadt, Germany
KH_2PO_4	AppliChem, Darmstadt, Germany
Na_2HPO_4	AppliChem, Darmstadt, Germany
NaCl	AppliChem, Darmstadt, Germany
Tris-HCl	AppliChem, Darmstadt, Germany
Glycerol	AppliChem, Darmstadt, Germany
Dimethyl sulfoxide (DMSO)	Merck, Darmstadt, Germany
20% (w/v) γ-PGA (Na^+ form, LM)	Molecular Dimensions, Altamonte Springs, USA
4-Bromopyrazole	Merck, Darmstadt, Germany
4-Iodopyrazole	Merck, Darmstadt, Germany
Methanol-d_4 (99.8%)	Merck, Darmstadt, Germany
1-ethyl-3-(3-dimethylaminopropyl)carbodiimide/N-hydroxysuccinimide (EDC/NHS)	Merck, Darmstadt, Germany
ethanolamine	Merck, Darmstadt, Germany
KCl	AppliChem, Darmstadt, Germany
Tween-20	AppliChem, Darmstadt, Germany
DMSO-d_6 (99.8%)	Merck, Darmstadt, Germany

Table 2.5: Materials used in this thesis.

Material	Supplier
0.5 ml Amicon ULTRA 10k	Merck, Darmstadt, Germany
Standard 5 mm NMR tube	Wilmad-LabGlass, Vineland, USA
Shigemi 5 mm NMR tube	Shigemi Co., Ltd., Tokyo, Japan
Coaxial insert for 5 mm NMR tube (capillary)	Wilmad-LabGlass, Vineland, USA
CM5 chip	Cytiva, Vancouver, Canada
96-well crystallization plate	Greiner Bio-One, Kremsmünster, Austria
Seal foil	Greiner Bio-One, Kremsmünster, Austria
Polystyrene 96-well plate	Merck, Darmstadt, Germany
Micro reaction tubes	Eppendorf, Hamburg, Germany

Table 2.6: Used crystallization kits to find conditions for crystalizing proteins.

Crystallization kit	Supplier	Application
Crystallization Low Ionic Kit for Proteins	Merck, Darmstadt, Germany	Initial screening
Crystallization PEG Grid Screening Kit	Merck, Darmstadt, Germany	Initial screening
NeXtal JCSG Core Suite I	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
NeXtal JCSG Core Suite II	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
NeXtal JCSG Core Suite III	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
NeXtal JCSG Core Suite IV	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
MIDAS 1 + 2	Molecular Dimensions, Altamonte Springs, USA	Initial screening
NeXtal PEGs Suite	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
NeXtal PEGs II Suite	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
NeXtal MPD Suite	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
NeXtal Anions Suite	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
The PGA Screen	Molecular Dimensions, Altamonte Springs, USA	Initial screening
NeXtal $AmSO_4$ Suite	NeXtal Biotechnologies, Trust Drive, Holland	Initial screening
JBScreen Wizard	Jena Bioscience, Jena, Germany	Initial screening
Silver bullet	Hampton Research, Aliso Viejo, USA	Additive screening
Additive Screen HT	Hampton Research, Aliso Viejo, USA	Additive screening
JBScreen Plus HTS	Jena Bioscience, Jena, Germany	Additive screening

2.2. PBP3 cloning, expression and purification

Protein samples were produced by Dr. Immanuel Grimm and Christian Dilk and delivered by our cooperation partner, AiCuris Anti-infective Cures AG (Wuppertal, Germany).

Constructs lacking the *N*-terminal transmembrane anchor were used for preparation of PBP3 from *E. coli* (*Ec*PBP3ΔTM; 49–588 aa) and *P. aeruginosa* (*Pa*PBP3ΔTM; 40–563 aa). The *E. coli* TPd construct (*Ec*TPd*) was derived from the *Ec*PBP3ΔTM construct and includes the n-PBd linker subdomain; regions K88–E163 and P204–Q227 of the n-PBd were replaced with GGG linkers. The sequences of the proteins used are shown in Table 2.7. All proteins were expressed with *N*-terminal His-tag (for *Ec*PBP3 constructs) or *C*-terminal His-tag (for *Pa*PBP3ΔTM) in *E. coli* BL21(DE3) and the target protein purified by Ni^{2+}-NTA affinity chromatography and size exclusion chromatography. A thrombin cleavage site inserted in the

sequence was utilized to remove the His-tag, and the quality of the purified protein was examined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Constructs derived from *Ec*PBP3ΔTM were stored in the standard *Ec*PBP3 buffer (15 mM KH_2PO_4, 2 mM Na_2HPO_4, 100 mM NaCl, pH 6.0) at 4 °C, whereas *Pa*PBP3ΔTM was stored in 10 mM Tris-HCl, 200 mM NaCl, 20% glycerol (pH 7.5; referred as *Pa*PBP3 standard buffer) at room temperature (approximately 21 °C).

Table 2.7: Amino acid sequences of the PBP3 constructs derived from *E. coli* and *P. aeruginosa*.

The thrombin cleavage site sequence is shown in red letters, whereas the GGG-linkers in the *Ec*TPd* construct are colored cyan. Numbers next to the sequence indicate the sequential number of the last amino acid in this row.

Designation	Original amino acids	Sequence
His-tagged *Ec*PBP3ΔTM	S49–S588	MGSSHHHHHHSSGLVPRGSHMASMSPDMLVKEGDMRSLRVQQVSTSRGMITDRSGRPLAV 84 SVPVKAIWADPKEVHDAGGISVGDRWKALANALNIPLDQLSARINANPKGRFIYLARQVN 144 PDMADYIKKLKLPGIHLREESRRYYPSGEVTAHLIGFTNVDSQGIEGVEKSFDKWLTGQP 204 GERIVRKDRYGRVIEDISSTDSQAAHNLALSIDERLQALVYRELNNAVAFNKAESGSAVL 264 VDVNTGEVLAMANSPSYNPNNLSGTPKEAMRNRTITDVFEPGSTVKPMVVMTALQRGVVR 324 ENSVLNTIPYRINGHEIKDVARYSELTLTGVLQKSSNVGVSKLALAMPSSALVDTYSRFG 384 LGKATNLGLVGERSGLYPQKQRWSDIERATFSFGYGLMVTPLQLARVYATIGSYGIYRPL 444 SITKVDPPVPGERVFPESIVRTVVHMMESVALPGGGGVKAAIKGYRIAIKTGTAKKVGPD 504 GRYINKYIAYTAGVAPASQPRFALVVVINDPQAGKYYGGAVSAPVFGAIMGGVLRTMNIE 564 PDALTTGDKNEFVINQGEGTGGRS 588
*Ec*PBP3ΔTM	S49–S588	GSHMASMSPDMLVKEGDMRSLRVQQVSTSRGMITDRSGRPLAVSVPVKAIWADPKEVHDA 101 GGISVGDRWKALANALNIPLDQLSARINANPKGRFIYLARQVNPDMADYIKKLKLPGIHL 161 REESRRYYPSGEVTAHLIGFTNVDSQGIEGVEKSFDKWLTGQPGERIVRKDRYGRVIEDI 221 SSTDSQAAHNLALSIDERLQALVYRELNNAVAFNKAESGSAVLVDVNTGEVLAMANSPSY 281 NPNNLSGTPKEAMRNRTITDVFEPGSTVKPMVVMTALQRGVVRENSVLNTIPYRINGHEI 341 KDVARYSELTLTGVLQKSSNVGVSKLALAMPSSALVDTYSRFGLGKATNLGLVGERSGLY 401 PQKQRWSDIERATFSFGYGLMVTPLQLARVYATIGSYGIYRPLSITKVDPPVPGERVFPE 461 SIVRTVVHMMESVALPGGGGVKAAIKGYRIAIKTGTAKKVGPDGRYINKYIAYTAGVAPA 521 SQPRFALVVVINDPQAGKYYGGAVSAPVFGAIMGGVLRTMNIEPDALTTGDKNEFVINQG 581 EGTGGRS 588
His-tagged *Ec*TPd*	S68–V88/E164–Q203/A228–T570	HHHHHHSSGLVPRGSTSRGMITDRSGRPLAVSVPVGGGESRRYYPSGEVTAHLIGFTNVD 185 SQGIEGVEKSFDKWLTGQGGGAAHNLALSIDERLQALVYRELNNAVAFNKAESGSAVLVD 266 VNTGEVLAMANSPSYNPNNLSGTPKEAMRNRTITDVFEPGSTVKPMVVMTALQRGVVREN 326 SVLNTIPYRINGHEIKDVARYSELTLTGVLQKSSNVGVSKLALAMPSSALVDTYSRFGLG 386 KATNLGLVGERSGLYPQKQRWSDIERATFSFGYGLMVTPLQLARVYATIGSYGIYRPLSI 446 TKVDPPVPGERVFPESIVRTVVHMMESVALPGGGGVKAAIKGYRIAIKTGTAKKVGPDGR 506 YINKYIAYTAGVAPASQPRFALVVVINDPQAGKYYGGAVSAPVFGAIMGGVLRTMNIEPD 566 ALTT 570
*Ec*TPd*	S68–V88/E164–Q203/A228–T570	GSTSRGMITDRSGRPLAVSVPVGGGESRRYYPSGEVTAHLIGFTNVDSQGIEGVEKSFDK 198 WLTGQGGGAAHNLALSIDERLQALVYRELNNAVAFNKAESGSAVLVDVNTGEVLAMANSP 279 SYNPNNLSGTPKEAMRNRTITDVFEPGSTVKPMVVMTALQRGVVRENSVLNTIPYRINGH 339 EIKDVARYSELTLTGVLQKSSNVGVSKLALAMPSSALVDTYSRFGLGKATNLGLVGERSG 399 LYPQKQRWSDIERATFSFGYGLMVTPLQLARVYATIGSYGIYRPLSITKVDPPVPGERVF 459 PESIVRTVVHMMESVALPGGGGVKAAIKGYRIAIKTGTAKKVGPDGRYINKYIAYTAGVA 519 PASQPRFALVVVINDPQAGKYYGGAVSAPVFGAIMGGVLRTMNIEPDALTT 570
*Pa*PBP3ΔTM	D40–A563	MDHDFLKGQGDARSVRHIAIPAHRGLITDRNGEPLAVSTPVTTLWANPKELMTAKERWPQ 98 LAAALGQDTKLFADRIEQNAEREFIYLVRGLTPEQGEGVIALKVPGVYSIEEFRRFYPAG 158 EVVAHAVGFTDVDDRGREGIELAFDEWLAGVPGKRQVLKDRRGRVIKDVQVTKNAKPGKT 218 LALSIDLRLQYLAHRELRNALLENGAKAGSLVIMDVKTGEILAMTNQPTYNPNNRRNLQP 278 AAMRNRAMIDVFEPGSTVKPFSMSAALASGRWKPSDIVDVYPGTLQIGRYTIRDVSRNSR 338 QLDLTGILIKSSNVGISKIAFDIGAESIYSVMQQVGLGQDTGLGFPGERVGNLPNHRKWP 398 KAETATLAYGYGLSVTAIQLAHAYAALANDGKSVPLSMTRVDRVPDGVQVISPEVASTVQ 458 GMLQQVVEAQGGVFRAQVPGYHAAGKSGTARKVSVGTKGYRENAYRSLFAGFAPATDPRI 518 AMVVVIDEPSKAGYFGGLVSAPVFSKVMAGALRLMNVPPDNLPTAKLVPRGS 563

2.3. Optimizing the 1D ^{19}F NMR experiment for FBDD screening

^{19}F NMR screening against His-tagged *Ec*PBP3ΔTM was performed as part of the FBDD project initiated by AiCuris. The His-tag was hypothesized to be unstructured and not a target of a specific interaction with fragments and it was expected to increase the overall global volume of PBP3, leading to an increase in the rotational correlation time (τ_{cor}) of the protein when compared with a construct without the His-tag. This increase in τ_{cor} may increase the rate of relaxation of the ^{19}F signals for fragments that bind the protein. Thus, the His-tag was not cleaved in the purification protocol for protein material used in the ^{19}F NMR FBS to potentially enhance the line-broadening of peaks arising from fragment hits.

The 1D ^{19}F NMR experiment was established prior to screening all fragments in the library. All experiments were recorded on the B600-F NMR spectrometer at 27 °C. External chemical shift referencing was examined during the process of establishing the experiments. Here, capillaries containing the reference standard for the respective sample were inserted into 5 mm standard NMR tubes. Because the sample volume was partly displaced by the capillary, the signals of the fragments were reduced. Moreover, only three capillaries were available, thus, hampering the automation process. Therefore, internal chemical shift referencing was examined by recording 1D ^{1}H NMR spectra of samples with and without the referencing standard. No changes in the quality of the protein nor fragments in a test pool were observed, and the addition of the protein did not cause a change in the chemical shift of the reference compound. Thus, and because of the simplified sample preparation, internal chemical shift referencing with 5 µM trifluoroacetic acid (TFA) and 120 µM fluorobenzene was used for tri- and mono-fluorinated fragments, respectively.

The standard 1D ^{19}F NMR experiment was modified to remove the very broad signal (approximately 90 ppm; Figure 3.1) arising from Teflon present in the probe head. Introduction of a spin echo delay (D4) of 8 ms (i.e., Δ–180°–Δ, where Δ = 4 ms) into the pulse sequence following the first 90° ^{19}F excitation pulse was found to be effective for removing the fast-relaxing Teflon signal without causing a noticeable change in the intensity of the signals arising from the fragments. Furthermore, first order phasing of 1D ^{19}F spectra was removed by inserting delays D8 and D9 into the pulse program [158]. The correction factor:

$$\mathrm{D8} = \frac{|\mathrm{phcor}_1|}{180^\circ} \mathrm{x}\, d_\mathrm{w}\ [\mu\mathrm{sec}] \qquad (2)$$

was used to correct for negative phcor_1 values, where phcor_1 is the first order phase correction in degrees observed during the initial experiment where D8 = D9 = 0 µsec and d_w is the dwell time in µsec, whereas for positive phcor_1 values D9 was adjusted by:

$$\mathrm{D9} = \frac{|\mathrm{phcor}_1|}{180^\circ} \mathrm{x}\, d_\mathrm{w}\ [\mu\mathrm{sec}] \qquad (3).$$

Note that D9 was coded into the pulse program but not used because the required first order phase correction was negative. The final schematic pulse sequence that was used in the ^{19}F NMR screens is shown in Figure 2.1.

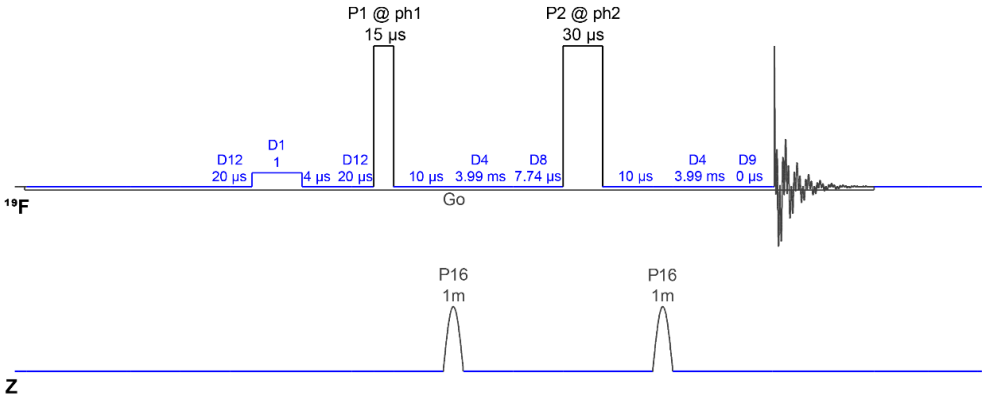


Figure 2.1: Pulse sequence for 1D ^{19}F NMR experiments used for screening tri-fluorinated fragments.
Note that in this example D9 was not required, because the initially observed first order phase correction was negative.

2.4. First screen to establish the protocol

For the ^{19}F NMR screen, 206 tri-fluorinated and 225 mono-fluorinated fragments from the BIONET Fluorine Fragment Library supplied by KeyOrganics (Camelford, UK) were pooled in tri-fluorinated-only and mono-fluorinated-only fragment pools. Fragments in this library are soluble in aqueous solutions at concentrations >100 µM and the purity was reported to be ≥95%. Additionally, generating this library filters were applied to remove toxic and reactive groups, PAINS and fragments likely to form aggregates. Additionally, a few di-fluorinated fragments were purchased from the fragment library, which were treated as mono-fluorinated fragments in further analysis based on chemical shifts of these fragments being similar to those of mono-fluorinated fragments (Figure 2.2). Note that the ^{19}F chemical shift range for fragments is considerably larger than chemical shift ranges for other nuclei, which can be used effectively to avoid assignment ambiguities caused by crowding or overlap of peaks in NMR spectra, as seen, for example, in 1D ^{1}H NMR spectra of fragment pools. Fragment pools contained 20 fragments each and fragments for each pool were selected to maximize peak separation, i.e., differences in chemical shift, based on the chemical shifts provided by the supplier. Nonetheless, chemical shifts of the fragments supplied by the company were determined in DMSO and thus, differences in chemical shifts between those provided by the supplier of the fragment library and observed in the aqueous buffer used in the screen were expected.

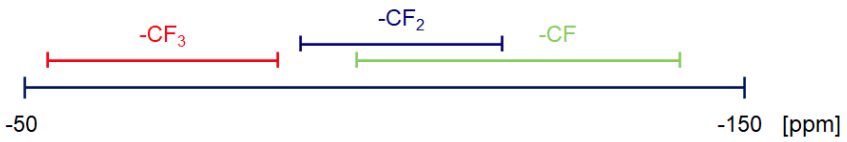


Figure 2.2: Chemical shift ranges for tri- (-CF_3, red), di- (-CF_2, blue) and mono-fluorinated (-CF, green) fragments used in this screening.
The fluorine atom of all mono-fluorinated fragments is attached to aromatic ring systems (Table 3.1, Table 3.2 and Table 3.3 as examples).

Each fragment pool was 550 µl, containing 30 µM of each fragment, 30 µM His-tagged *Ec*PBP3ΔTM and 5% (v/v) D_2O. These samples were placed in standard 5 mm NMR tubes. For tri-fluorinated fragment pools, 5 µM TFA was added as an internal chemical shift reference of –75.39 ppm, as reported in Pan Shi *et al*. [159]. For pools containing mono-fluorinated fragment, 120 µM fluorobenzene was used as reference (–113.6 ppm), which was close to the

measured average chemical shift of all samples (–113.6016 ± 0.0049 ppm) and is close to the chemical shift reported for this compound in D_2O (–113.73 ppm; [160]). Small differences in the reference chemical shift can arise from differences in temperature, pH and salt concentration.

Tri-fluorinated fragment experiments were recorded with 3072 scans, a sweep width of 34.95 ppm (19737 Hz), an acquisition time of 0.83 s, 32768 points in the time domain and an offset of –68 ppm, whereas spectra of mono-fluorinated fragments were recorded with 6144 scans, a sweep width of 46.61 ppm (26316 Hz), an acquisition time of 1.25 s, 65536 points in the time domain and an offset of –118.5 ppm. Additionally, for each sample a 1D ^{1}H control spectrum was recorded to be able to identify potential non-fluorinated contaminates and to compare signals arising from the protein, thus confirming that the equivalent protein concentration was used in each experiment. In this experiment 128 scans were used with a sweep width of 15.02 ppm (9014 Hz), an acquisition time of 0.91 s, 16384 points in the time domain and an offset of 4.7 ppm, i.e., H_2O signal. In the automated process using IconNMR, the first sample was inserted and tuning, matching and phasing the lock signal was performed manually to ensure an optimal result (Figure 2.3). After shimming in z and z^2, standard 1D ^{1}H control spectra were recorded prior to the 1D ^{19}F experiments. Subsequent samples were automatically loaded from the SampleCase auto sampler into the magnet and a 5 min delay was applied for temperature equilibration before the experiments were recorded as described above.

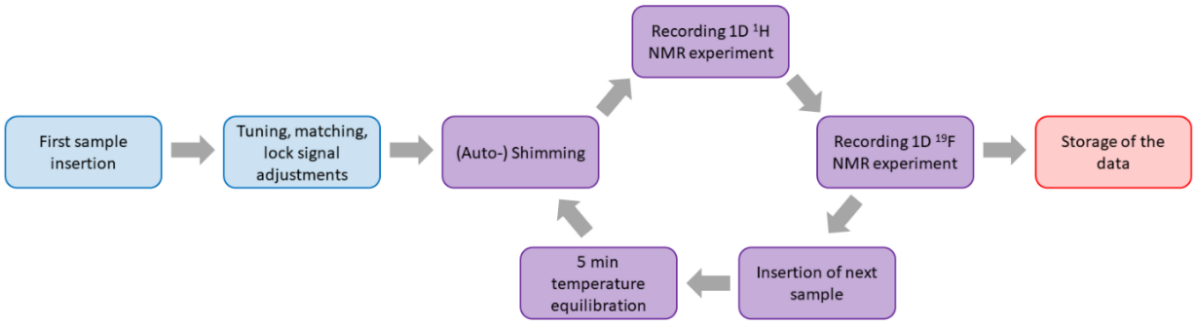


Figure 2.3: Steps in the automated process of recording 1D ^{1}H and ^{19}F NMR experiments for the screening.
The automation (shaded in violet) is performed by the Bruker software IconNMR. Note, that shimming was performed manually for the first sample, but repeated automatically by IconNMR for subsequent samples.

2.4.1. Analysis of the ^{19}F NMR screens

After processing the 1D ^{19}F spectra using an exponential apodization function with a line-broadening factor of 1 Hz in TopSpin version 4.0.8, the chemical shift of 1D ^{19}F spectra were referenced using the reference compound and a manual baseline correction with a 5th polynomial function was performed. The peaks were picked and the chemical shifts, intensities and integrals were determined in TopSpin. For the integration of TFA the region between –75.35 and –75.45 ppm was integrated, whereas for fluorobenzene the region between –113.56 and –113.64 ppm was used.

Intensities and integral of the peaks from the referencing compounds in related spectra with and without the protein were used to calculate a correction factor that accounted for any differences between spectra. The correction factor was used to calculate the corrected intensities that were used to calculate the loss of intensity relative to the initial intensity derived from the fragment peak without protein. Changes in chemical shift after the addition of the protein were calculated by multiplying the difference between both peaks (in ppm) with the base frequency to obtain changes in Hz.

2.4.2. Assignment and hit validation

Because of differences in the observed chemical shifts and the supplier-provided chemical shifts (measured in DMSO) for particular fragments, it was not possible to assign all peaks unambiguously to fragments used in the pools of the initial screen. Thus, unassigned peaks were assigned to all possible fragments with potential matching chemical shifts and an additional screen was performed for assignment validation featuring smaller pool sizes to reduce spectral overlap between several fragments. Initially, four pools that included 7 or 8 fragments potentially assigned to scores 1 to 4 and one additional fragment that did not show any changes after the addition of the protein were prepared. The screening and the analysis were repeated as described in the previous Sections 2.4 and 2.4.1. In the analysis, the non-binding control fragment was used to calculate the correction factor, which was used to correct for changes in the intensities between spectra with and without the protein. Based on the results, 17 hit fragments were selected to create eight additional pools with 2 to 4 fragments, including the non-binding control fragment. The pool size was reduced to reduce competitive binding among fragments. Data acquisition and analysis were performed as described above.

2.5. Second screen

In the first screen NMR signals arising from monofluorinated fragments appeared as multiplets because of ^{1}H-^{19}F couplings, which could not be decoupled with the NMR probe available. Such active couplings reduced the intensity of the monofluorinated fragments significantly and complicated analysis of the spectra. Thus, in the second ^{19}F NMR screen, mono-fluorinated fragments were excluded. A total of 475 tri-fluorinated fragments supplied by Enamine (Kyiv, Ukraine) were pooled in 30 pools, leading to pool sizes of 15 or 16 fragments. The design of the library was modified with the aim to increase the solubility of the fragments in aqueous buffer to >1 mM and all tri-fluorinated fragments that passed the filters (solubility, aggregation, reactive and unstable compounds, stability, “rule of three” compliance) were included. The purity of all fragments was examined before delivery and chemical shift values recorded in H_2O were provided. The pool size in the second screen was reduced by 20% compared with the pool size of the first initial screen to avoid spectral overlap and difficulties with making unambiguous assignments. Samples were prepared in standard NMR tubes with a total volume of 550 µl, including fragment and protein concentrations of 30 µM, 5% (v/v) D_2O for the lock signal and 5 µM TFA for chemical shift referencing. 1D ^{19}F NMR experiments were recorded with 3072 scans, an offset of –68 ppm and a sweep width of 33.95 ppm. The temperature was set to 27 °C and 1D ^{1}H experiments were recorded as control experiments for every sample. Data acquisition and analysis were performed as described for the 1st ^{19}F fragment screen.

2.6. Experiments to measure kinetic and competitive binding data

2.6.1. K_d value determination

An estimate of the K_d of the interaction between the hit fragment 5N-395S and *Ec*PBP3 was determined. Because of the poor solubility of 5N-395S in aqueous buffer (<500 µM), a protein titration was performed with constant fragment concentration. A His-tagged *Ec*PBP3ΔTM sample in the standard protein buffer was concentrated to 978 µM using a 0.5 ml Amicon ULTRA 10k and centrifugation at 4 °C and 14,000 x *g*. Two samples were prepared containing 5 µM TFA, 5% (v/v) D_2O, 60 µM 5N-395S and 0 or 830 µM protein. 1D ^{19}F NMR spectra of both these samples were recorded and then used to prepare the next samples containing the second highest and the lowest protein concentrations. Because only the protein concentration differed between both samples, all subsequent samples prepared contained 5 µM TFA, 5% (v/v) D_2O and 60 µM 5N-395S. In total, 1D ^{19}F spectra of samples containing protein concentrations of 0, 15, 30, 50, 75, 110, 150, 230, 350, 475, 650 and 830 µM were recorded. The chemical shifts and intensities of the peaks in the 1D ^{19}F spectra were determined and the K_d was obtained from the observed changes in ^{19}F chemical shifts ($\Delta\delta_{obs}$) assuming a one-site binding model and using the following equation:

$$\Delta\delta_{\mathrm{obs}} = \frac{1}{2}\Delta\delta_{\mathrm{max}}\left[1 + X + \frac{K_{\mathrm{d}}}{[\mathrm{L_0}]} - \sqrt{\left(1 + X + \frac{K_{\mathrm{d}}}{[\mathrm{L_0}]}\right)^2 - 4X}\right] \quad (4),$$

where $[L_0]$ is the total ligand concentration, $\Delta\delta_{max}$ is the chemical shift difference between the free-state chemical shift and the bound-state chemical shift of the fragment and X the molar ratio of protein and ligand [161]. For data fitting, a python script, written by Dr. Michael Schmitz (School of Chemical Sciences, The University of Auckland), was used to obtain the K_d value. Because of solubility issues, $\Delta\delta_{max}$ was not obtained experimentally and thus, was included in the fitting routine as a fitted parameter. In the case of low protein concentrations, errors in chemical shifts were set as the digital resolution of the spectrum (0.5 Hz). Because of strong line-broadening effects, an accurate determination of $\Delta\delta_{obs}$ was not possible at higher protein concentrations and thus, chemical shift errors were determined by taking the inverted intensities multiplied by a factor of 1000.

2.6.2. AIC499 competitive binding

As shown by Vulpetti and coworkers, known inhibitors of the target protein can be used to gain additional information about the binding of fragments [162]. In competitive binding studies a molecule with high affinity, e.g., nM range, is included in the sample with fragments and protein. Observation of no change in the chemical shift and intensities of a peak arising from a fragment indicates that this fragment does not bind in the same location as the high affinity molecule and is not affected by a conformational change of the protein that may occur upon interacting with the ligand that binds with high affinity. In contrast, binding of a high affinity molecule can cause the chemical shift and intensity of the peak arising from the fragment to change to the free-state chemical shift. This indicates that the fragment and high affinity ligand bind at the same binding site, or the binding of the high affinity ligand causes a conformational change that prohibits binding of the fragment to the protein.

The monobactam AIC499 was used to perform competitive binding studies (chemical structure of the AIC499 molecule is shown in Figure 1.11). AIC499 binds covalently to the active site of

His-tagged *Ec*PBP3ΔTM. In these experiments, 30 µM fragments in the pools used in the hit validation of the second screen and 5N-395S and 3S-528S were added to samples containing 5 µM TFA, 5% (v/v) D_2O, 30 µM His-tagged *Ec*PBP3ΔTM and 60 µM AIC499. Covalent binding of AIC499 to the catalytic site should inhibit fragments that bind at this site. 1D ^{19}F NMR experiments were recorded as described above for tri-fluorinated fragments. A correction factor was applied as described in the Section 2.4.1 based on the intensity of the TFA peak and the control non-binding fragment present in the pools. Peak intensity differences in the spectra between free ligand spectra and after the addition of His-tagged *Ec*PBP3ΔTM (Δ_{PBP3}) and between free ligand spectra and after the addition of His-tagged *Ec*PBP3ΔTM and AIC499 (Δ_{AIC499}) were calculated. The competition factor F_{comp} was calculated by dividing Δ_{AIC499} by Δ_{PBP3}. No influence of AIC499 on the interaction between protein and fragment yielded values of 100%. In contrast, a value of 0% indicates that the fragment was completely displaced from the protein when AIC499 was present. If the binding between protein and fragment is enhanced F_{comp} is greater than 100%.

2.6.3. Chemical shift perturbation analysis

Chemical shift perturbation analysis was performed by recording 2D ^{1}H-^{15}N TROSY-HSQC experiments at 37 °C. The experiments were recorded on the B900 NMR spectrometer, and data was processed with NMRPipe and analyzed with CcpNMR analysis. Initially, changes in chemical shift of resonances caused by the addition of DMSO were examined by carrying out a titration with only DMSO. The concentration of ^{15}N-labelled *Ec*TPd* was maintained at 100 µM while the DMSO concentration was increased (i.e., 0, 0.19, 0.475, 0.95, 1.9, 3.8, 6.65, and 9.5% (v/v)) and 2D ^{1}H-^{15}N TROSY-HSQC experiments were recorded at each DMSO concentration. Using small increments in DMSO concentration enabled tracking of the changes in the chemical shift of resonances. 2D ^{1}H-^{15}N TROSY-HSQC experiments were recorded with 16 scans, 3072 x 312 complex points and sweep widths of 16.327 (^{1}H) x 30.0 (^{15}N) ppm and offsets of 4.702 (^{1}H) and 118.3 (^{15}N) ppm were used. For hit fragments identified in the ^{19}F NMR screening, 50 µM *Ec*TPd* was supplemented with 1 mM of each respective fragment (final DMSO concentration: 1.9%) and experiments were recorded with 52 scans, 3072 x 312 points in the time domain and a sweep width of 16.327 x 30.0 ppm. For fragment FD-0035, which has a reported solubility limit of 5 mM in aqueous buffer, 5 mM was used with 50 µM *Ec*TPd* in one additional experiment (final DMSO concentration: 9.5%). Chemical shift perturbation analysis with Cluster J compounds was carried out using 50 µM Cluster J compound with 25 µM *Ec*TPd* (final DMSO concentration: 0.475%). 2D ^{1}H-^{15}N TROSY-HSQC experiments were recorded with 192 scans, 3072 x 312 complex points and sweep widths of 16.327 (^{1}H) x 30.0 (^{15}N) ppm.

For the analysis, the weighted average ^{1}H and ^{15}N chemical shift difference $\Delta\delta_{av}$ was calculated with:

$$\Delta\delta_{av} = \sqrt{\Delta\delta_H{}^2 + \frac{\Delta\delta_N{}^2}{6}} \quad (5),$$

where $\Delta\delta_H$ is the difference in ^{1}H chemical shift with and without ligand and $\Delta\delta_N$ is the difference in ^{15}N chemical shifts with and without ligand [163, 164]. $\Delta\delta_{av}$ was plotted against the designated peak number to visualize changes in chemical shift for each peak. Additionally, 3σ of all peaks was used as the threshold value to identify peaks that showed the largest changes.

2.7. Heteronuclear NMR experiments

2.7.1. HSQC experiments

2D ^{1}H-^{15}N HSQC experiments were recorded with different samples on the B700 and B900 NMR instruments. 238 µM ^{15}N-labeled His-tagged *Ec*PBP3ΔTM was used in standard buffer, which was supplemented with 5% (v/v) D_2O and DSS, and experiments were recorded at temperatures ranging from 25 to 42 °C. These buffer conditions were used in all subsequent experiments at 37 °C. For the 2D ^{1}H-^{15}N HSQC experiment 128 scans were used with a sweep width of 15.9 (^{1}H) x 30.0 (^{15}N) ppm, 2048 (^{1}H) x 256 (^{15}N) complex points and an offset of 4.7 (^{1}H) x 117.5 (^{15}N) ppm (Figure 3.25). The TROSY principle was introduced into the pulse sequence and a 2D ^{1}H-^{15}N TROSY-HSQC was recoded using 174 µM ^{15}N-labeled *Ec*PBP3ΔTM with 208 scans, a sweep width of 15.9 (^{1}H) x 52.0 (^{15}N) ppm, 3072 (^{1}H) x 640 (^{15}N) complex points and an offset of 4.7 (^{1}H) x 107.5 (^{15}N) ppm (Figure 3.26). The pulse sequence was used to record a 2D ^{1}H-^{15}N TROSY-HSQC using 111 µM [^{2}H, ^{15}N]-labeled *Ec*PBP3ΔTM with 280 scans, a sweep width of 15.9 (^{1}H) x 30.0 (^{15}N) ppm, 3072 (^{1}H) x 512 (^{15}N) complex points and an offset of 4.7 (^{1}H) x 117.5 (^{15}N) ppm (Figure 3.27), and a 2D ^{1}H-^{15}N TROSY-HSQC using 60 µM [^{2}H, ^{13}C, ^{15}N]-labeled *Ec*TPd* with 64 scans, a sweep width of 15.9 (^{1}H) x 30.0 (^{15}N) ppm, 3072 (^{1}H) x 512 (^{15}N) complex points and an offset of 4.7 (^{1}H) x 117.5 (^{15}N) ppm (Figure 3.28).

2.7.2. Experiments for sequence-specific backbone assignments

Backbone $^{1}H_N$, ^{15}N, $^{13}C_\alpha$, $^{13}C_\beta$ and ^{13}C' sequence-specific assignments were determined using 60 to 506 µM [^{2}H, ^{13}C, ^{15}N]-labelled *Ec*TPd* samples (300 µL in Shigemi tubes) that were prepared in standard protein buffer supplemented with 5% (v/v) D_2O as the lock signal. The temperature was set to 37 °C, which was calibrated using perdeuterated methanol and DSS was used as an internal ^{1}H chemical shift referencing agent (0 ppm) [165]. ^{13}C and ^{15}N chemical shifts were referenced with the IUPAC-IUB recommended chemical shift referencing ratios [166, 167]. All experiments were recorded on the B700 and B900 NMR spectrometers as TROSY-version and for 3D heteronuclear experiments NUS was used [168]. Selected acquisition parameters of the NMR experiments recorded are listed in Table 2.8. Recorded 2D ^{1}H-^{15}N TROSY-HSQC spectra were processed in one script using NMRPipe (Figure A1). Phase corrections were determined using NMRDraw and spectra were analyzed using CcpNMR analysis. In contrast, NUS acquired 3D multi-dimensional triple resonance experiments were reconstructed using several processing steps with specifically designed scripts, as outlined in Figure 2.4.

Figure 2.4: Steps during processing of multi-dimensional NMR experiments recorded with NUS.
The names of the scripts used to perform the respective steps are given in brackets. Exemplary scripts for processing a 3D TROSY-HNCACO are shown in Figures A2 to A6.

Table 2.8: Multi-dimensional NMR experiments used to obtain backbone sequence-specific assignments.

An experiment marked with an asterisk indicates that the experiment was recorded several times and added together using addNMR, which is part of the NMRPipe package. The number of asterisks indicates the number of times the experiment was recorded.

Experiment	Protein concen-tration [µM]	Num-ber of scans	Experi-mental time [h]	Sweep width [ppm]	Complex points	Offset [ppm]	Field (^{1}H resonance frequency) [MHz]
2D ^{1}H-^{15}N TROSY-HSQC	60	64	16	15.9 (^{1}H) x 30.0 (^{15}N)	1536 (^{1}H) x 256 (^{15}N)	4.701 (^{1}H) x 117.5 (^{15}N)	700
3D TROSY-HNCO (NUS)	60	128	60	15.9 (^{1}H) x 30.0 (^{15}N) x 11.0 (^{13}C)	1024 (^{1}H) x 52 (^{15}N) x 48 (^{13}C)	4.701 (^{1}H) x 117.5 (^{15}N) x 173.81 (^{13}C)	700
3D TROSY-HNCA (NUS)	506	96	80	16.3 (^{1}H) x 30.0 (^{15}N) x 26.0 (^{13}C)	1536 (^{1}H) x 67 (^{15}N) x 64 (^{13}C)	4.701 (^{1}H) x 118.3 (^{15}N) x 53.42 (^{13}C)	900
3D TROSY-HN(CA)CO (NUS) **	236	512	163	15.9 (^{1}H) x 30.0 (^{15}N) x 12.0 (^{13}C)	1024 (^{1}H) x 51 (^{15}N) x 32 (^{13}C)	4.702 (^{1}H) x 117.5 (^{15}N) x 173.81 (^{13}C)	700
3D TROSY-HN(CO)CA (NUS)	490	64	31	15.9 (^{1}H) x 30.0 (^{15}N) x 26.0 (^{13}C)	1024 (^{1}H) x 52 (^{15}N) x 48 (^{13}C)	4.701 (^{1}H) x 117.5 (^{15}N) x 53.31 (^{13}C)	700
3D TROSY-HN(CO)CACB (NUS)	490	96	63	15.9 (^{1}H) x 30.0 (^{15}N) x 52.0 (^{13}C)	1024 (^{1}H) x 52 (^{15}N) x 64 (^{13}C)	4.701 (^{1}H) x 117.5 (^{15}N) x 40.31 (^{13}C)	700
3D TROSY-HNCACB (NUS) ***	236	256	140	15.9 (^{1}H) x 30.0 (^{15}N) x 63.9 (^{13}C)	1024 (^{1}H) x 44 (^{15}N) x 64 (^{13}C)	4.702 (^{1}H) x 117.5 (^{15}N) x 40.31 (^{13}C)	700

First data stored in the Bruker file format need to be converted into the NMRPipe file format. This step was performed by the fid.com script (Figure A2). Secondly, addNMR was used for some experiments that were recorded multiple times and then added together using the following command (experiments added together are shown in Table 2.8): *add NMR -in1 nus%04d.fid -in2 nus%04d.fid -out add/nus%04d.fid -verb*. Repeat acquisition of datasets was performed to increase the signal-to-noise ratio. Alternatively, the number of scans can be increased; however, increasing the acquisition time was not desirable because the sample conditions in the magnet invariably changed, e.g., the presence of an air bubble at the top of the sample in a Shigemi tube adversely affected sample homogeneity, even with automated *z* and z^2 shimming. The ^{1}H dimension was then processed using the ft1.com script (Figure A3). The zero-order phase correction in the ^{1}H dimension was determined using NMRDraw and the ^{1}H dimension was reprocessed with this phase determined and used in the ft1.com script. Next, the script run.local, which contains the following line was used to run the ist.csh script (Figure A4) in parallel, that performs the reconstruction: *parallel -j 100% './ist.csh {} > /dev/null; echo {}' ::: yzx/nus*.nus*. Because the istHMS program stores data in the phf file format, the phf2pipe.com script was used to convert the data into the xyz file format, which is achieved using the following command:

xyz2pipe -in yzx_ist/nus%04d.phf | phf2pipe -user 1 -xproj xz.ft1 -yproj yz.ft1 | pipe2xyz -out rec/nus%04d.ft1

The ft23.com (Figure A5) and the ft3.com (Figure A6) scripts were used to process the ^{15}N and ^{13}C dimensions. Linear prediction (LP function (fn) in ft23.com and ft3.com scripts), was used to extend the time-domain data (i.e., improve the resolution) of the indirect dimensions. Linear prediction of either the indirect ^{13}C or ^{15}N dimension should be carried out when the other two dimensions are Fourier transformed to the frequency domain [148, 169]. Thus, for initial linear prediction of the ^{13}C dimension, Fourier transformation of the ^{15}N dimension without linear prediction was performed. After processing the ^{13}C dimension, a Hilbert-transformation, and an inverse Fourier-transformation on the ^{15}N dimension was performed to convert the frequency domain back to the time domain and then the ft3.com script was used to linear predict in the ^{15}N dimension and Fourier transform this dimension into the frequency domain. Furthermore, a quadratic sine-bell function (fn SP) was used for apodization and zero-filling (fn ZF) was implied to expand the number of acquired points in the time domain. Thus, the digital resolution was improved in all dimensions using sine-bell and zero-filling functions. Finally, the command *proj3D.tcl -in ./rec/nus%04d.ft3* was used to run the in-house TCL script that creates 2D projections of the spectrum and *pipe2azara nus%04d.ft3 20200904_TPd_PBP3 _HNCACO_add_fid_1* was used to convert the data file format into the azara file format required for loading, visualizing and analyzing spectra in CcpNMR analysis.

To quantify the improvement of various 2D ^{1}H-^{15}N HSQC spectra using different protein constructs, the average linewidths at half height of the peaks in the spectra were determined. Peaks were selected over the whole spectrum by using the peak picking tool in CcpNMR analysis [149]. Peak arising from degradation products were not peak picked. The linewidths of the peaks in the ^{1}H and ^{15}N dimensions were exported and averaged.

2.8. Experiments performed on Cluster J compounds

2.8.1. 1D ^{1}H NMR experiments

Because precipitate occurred when recording 2D ^{1}H-^{15}N TROSY-HSQC spectra in the chemical shift perturbation analysis and because of the low solubility in aqueous solutions, the solubility of Cluster J compounds was examined by recording 1D ^{1}H NMR experiments on the B600-M and B600-O NMR spectrometers at varying concentrations of Cluster J compounds from 10, 50 and 100 µM and DMSO concentrations from 0.95, 9.5 and 95% at 37 °C. Cluster J compound 36 showed a comparably strong affinity towards *Ec*PBP3 in biochemical assays and was dissolved in DMSO-d_6. Therefore, compound 39 was used as an exemplary compound to check solubility without the presence of a strong residual DMSO signal in the NMR spectrum. A slightly modified version of the Bruker pulse program zgesgp was used to record the spectra with 1024 scans and 16384 points with a sweep width of 15.9828 ppm at 37 °C. Additionally, spectra of samples containing 100 µM Cluster J compound 36 in the presence of 1, 2 and 10 µM *Ec*TPd* were recorded in 0.95% DMSO-d_6, 100 µM DSS and standard protein buffer with the same experimental settings to further investigate the type of interaction between the compound and the protein.

2.8.2. Fluorescence experiments

A specific interaction between Cluster J compound 35 and *Ec*TPd* was examined by fluorescence spectroscopy. This Cluster J compound was chosen because it showed the highest affinity against *Ec*PPB3 in biochemical assays performed by Dr. Arancha López-Pérez at AiCuris. Samples were prepared with 4 µM *Ec*TPd* and compound concentrations ranging from 0 to 100 µM. Spectra were recorded by exciting the samples with λ_{ex} = 270 nm and emission was detected between 290 and 500 nm. The fluorescence signal was determined by integrating the range between 330 and 350 nm. For analysis, these integrals were plotted against the concentration of the Cluster J compound.

2.8.3. SPR experiments

SPR experiments were recorded by using a Biacore T200 instrument and the temperature was internally held at 25 °C. Purified *Pa*PBPΔTM was coated on a CM5 chip, which was activated with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide/N-hydroxysuccinimide (EDC/NHS) before and blocked with ethanolamine after *Pa*PBPΔTM immobilization to eliminate residual amine-reactive NHS ester groups. The achieved protein amount bound on the surface of the CM5 chip corresponded to approximately 11.820 RU. For every analyte two concentration series between 3.25 and 200 µM were prepared in running buffer (13 mM NaCl, 257 µM KCl, 171 µM KH_2PO_4, 950 µM Na_2HPO_4, 0.05% Tween-20 and 5% DMSO; pH 7.4) in a polystyrene 96-well plate. During the injection the baseline was obtained by running buffer for 300 s, the analytes allowed to associate for 100 s and to dissociate for 300 s with a flow rate of 30 µl/min. A 30 s washing step with 50% DMSO was used after every injection to regenerate the chip. In the final step, ten different DMSO concentrations ranging from 4.0% to 5.8% were used as the internal solvent correction. Steady state analysis for a 1:1 interaction was performed with the internal Biacore T200 Evaluation Software using equation:

$$R_{\mathrm{eq}} = \frac{c * R_{\mathrm{max}}}{K_{\mathrm{d}} + \mathrm{c}} + R\mathrm{I} \qquad (6),$$

where R_{eq} is the steady state binding level, c the concentration of the analyte (i.e., Cluster J compound), R_{max} the analyte binding capacity of the surface and RI the Bulk refractive index contribution in the sample, which serves as an offset on the R_{eq}-axis. Figures were prepared with GraphPad Prism using the points of one concentration series.

2.9. X-ray crystallography experiments

In this thesis, several protein constructs were crystallized by the vapor diffusion method including His-tagged *Ec*PBP3ΔTM, *Ec*PBP3ΔTM, His-tagged *Ec*TPd*, *Ec*TPd* and *Pa*PBP3ΔTM (Table 2.7). Additionally, co-crystallization experiments with AIC499 and Cluster J compounds were performed. For the crystallization of the proteins, the sitting drop approach was used in all cases mixing 0.5 µl of the protein solution with 0.5 µl of the reservoir on 96 well crystallization plates using a pipetting robot from Tecan. As reservoir solution 80 µl were used and the set-ups were stored at 20, 10 or 4 °C. The commercial crystallization kits used are summarized in Table 2.6 in Section 2.1.

For co-crystallization experiments, 0.5 mM AIC499 or 0.5 mM of the Cluster J compounds dissolved in protein buffer were added to the protein solution, prior to mixing with the reservoir solution. Cluster J compounds were observed to precipitate because of their low solubility (< 200 µM). This issue was partly overcome by mixing the sample gently to ensure similar Cluster J concentrations in the final set up. However, it is not possible to provide equal concentrations of Cluster J in the set ups, because the precipitate settles over time. Nevertheless, it was possible to visually see the precipitate in every drop, therefore ensuring that the concentration of dissolved Cluster J compounds was at the maximum in every set up. To improve the visual and diffraction quality of the crystals, fine screenings were performed by varying the concentrations of chemicals in the original reservoir solution, the temperature and the protein concentration. Moreover, panels of additives (shown in Table 2.6 in Section 2.1) were tested together with the original reservoir solution at a ratio of 1:9. Because adding microseeds to the crystallization solution can improve crystal growth, microseeds were generated from previously found crystals and 0.25 µl were added to the drops [147]. Visually identified crystals were flash-cooled to –173.15 °C, stored in liquid nitrogen and shipped to the beamlines summarized in Table 2.9. In some cases, the solutions, in which the crystals were found, had to be cryoprotected by appropriate additives. Therefore, DMSO, polyethylene glycol 8000 (PEG8000), glycerol, sodium sulfate, ethylene glycol, 1,2-propanediol, sodium formate or 2-methyl-2,4-pentanediol (MPD), included in the cryo-protection kit CryoProtX supplied by Molecular Dimensions (Altamonte Springs, USA), were used. Soaking with such additives was performed after preparation of the crystals and prior to flash-cooling. 250 mM bromo-pyrazole and iodo-pyrazole were used in soaking experiments (5 to 10 min) to identify hydrophobic pockets on the surface of the protein and potentially generate anomalous diffraction data for SAD phasing [170]. Additionally, the heavy metal complex Ta_6Br_{12} was soaked into the crystals by adding solid material to the reservoir solution and incubating crystals in this modified reservoir solution for approximately 4 h. To generate structural data of the fragments, which were identified in the ^{19}F NMR screening, 5 mM 3N-528S or 5N-395S were incubated with protein crystals for 5 to 10 min, respectively. In total 183 96-well crystallization experiments were prepared and diffraction data of 398 crystals were collected. Only crystallization conditions and structural data that was published are presented in the results (Section 3.4).

Crystals of *Ec*PBP3ΔTM:AIC499 were observed within a week in 3% (w/v) dextran sulfate (M-5000), 0.1 M sodium cacodylate, 5% (w/v) PEG 8000 and 30% (v/v) MPD at pH 6.5 with a protein concentration of 10 mg/ml in the presence of 500 µM AIC499. Crystals of *Ec*TPd* were observed within a week in 0.1 M sodium cacodylate, 5% (w/v) PEG 8000, 20% (v/v) MPD, 0.2% (w/v) betaine, 0.2% (w/v) L-glutamic acid, 0.2% (w/v) L-proline, 0.2% (w/v) taurine, 0.2% (w/v) trimethlyamine N-oxide and 0.02 M HEPES at pH 6.5 with a protein concentration of 14.4 mg/ml. The *Ec*TPd*:AIC499-complex crystallized in 0.1 M MES and 25% (w/v) PEG 1000 at pH 6.5 with a protein concentration of 10 mg/ml in the presence of 500 µM AIC499 after one

week. The structures derived from *Ec*PBP3 were determined by molecular replacement using an appropriately modified version of the published *Ec*PBP3ΔTM structure (PDB entry 4BJP, [47]).

*Pa*PBP3ΔTM crystals were observed in 0.1 M Na_2SO_4 and 24% (w/v) polyvinylpyrrolidone with a protein concentration of 8 mg/ml (apo crystal form 1). In the presence of 500 µM AIC499 and with a protein concentration of 14 mg/ml, *Pa*PBP3ΔTM:AIC499 crystals formed in 20% (v/v) Jeffamine® M-2070 and 20% (v/v) DMSO and were cryoprotected by adding 1 µl 80% (v/v) glycerol to the crystallization drop.

In co-crystallization experiments using Cluster J compounds, two additional *Pa*PBP3ΔTM crystal forms were observed after four to six weeks in 0.1 M MES, 5% (w/v) PEG 3000, 30% (v/v) PEG 200, pH 6.0 and 0.2 M potassium/sodium tartrate, 20% (w/v) PEG 3350 using 9 mg/ml *Pa*PBP3ΔTM and 500 µM of Cluster J compound 34 dissolved in DMSO. The respective structures were found to not contain the Cluster J compound and were therefore treated as de-facto apo structures, and these crystalline states were designated apo crystal forms 2 and 3. For determining the *Pa*PBP3ΔTM and *Pa*PBP3ΔTM:AIC499 structures, a model derived from PDB entry 3OC2 was used in molecular replacement [50].

Table 2.9: Beamlines used to collect diffraction data of protein crystals.

Designation	Operator	Location
P11	Deutsches Elektronen-Synchrotron (DESY)	Hamburg, Germany
P13	European Molecular Biology Laboratory (EMBL)	Hamburg, Germany
P14	European Molecular Biology Laboratory (EMBL)	Hamburg, Germany
ID23	European Synchrotron Radiation Facility (ESRF)	Grenoble, France

Raw data were processed with XDS, followed by anisotropy treatment in STARANISO using default settings. Following molecular replacement using MOLREP, models were iteratively improved by reciprocal-space refinement in phenix.refine, including TLS (translation, libration, screw) parameterization with grouping suggested on the basis of refined B-factors, and interactive rebuilding using COOT. Data collection and refinement statistics are provided in Table 3.6 and Table 3.7. All figures showing structural data were prepared using PyMol and LigPlot+ graphs were created using the superposition option to ensure consistent orientation, and thresholds for plotting non-covalent interactions were modified as indicated in the respective legends. Note that LigPlot+ only uses atom coordinates, therefore, ignoring conformational restrictions to plot the environment of the ligand.

3. Results

3.1. ^{19}F NMR screening for FBDD targeting PBP3 from *E. coli*

3.1.1. Optimizing the 1D ^{19}F NMR experiment

Initial 1D ^{19}F NMR experiments using a small fragment pool sample were recorded using the *zg* pulse program from the Bruker pulse sequence library. This experiment is simply a 90° pulse applied at the ^{19}F offset, which was typically –68 ppm for tri-fluorinated fragments, followed by acquisition of the free induction decay. In initial 1D ^{19}F experiments, a broad baseline roll was observed in the spectrum, which was postulated to arise from polytetrafluoroethylene (PTFE, also known as Teflon) present in the probe head [171]. After widening the sweep width and shifting the offset to –100 ppm a clear set of broad signals were observed with a center at approximately –180 ppm (Figure 3.1). Therefore, the standard 90°–acquire pulse sequence was edited to include a spin-echo element (Figure 2.1). The introduction of the spin-echo in the pulse scheme facilitated transverse relaxation of the Teflon signal while not dramatically affecting the intensities of signals arising from the fragments. Spin-echo delays (D4) of 2, 4 and 10 ms were tested with 4 ms found to be optimal, because the signal arising from the Teflon was no longer detected and no measurable changes in the intensities of peaks arising from the fragments were observed. This delay was used in all subsequent experiments, leading to flat baselines (e.g., Figure 3.9 for pool 6 containing mono-fluorinated fragments).

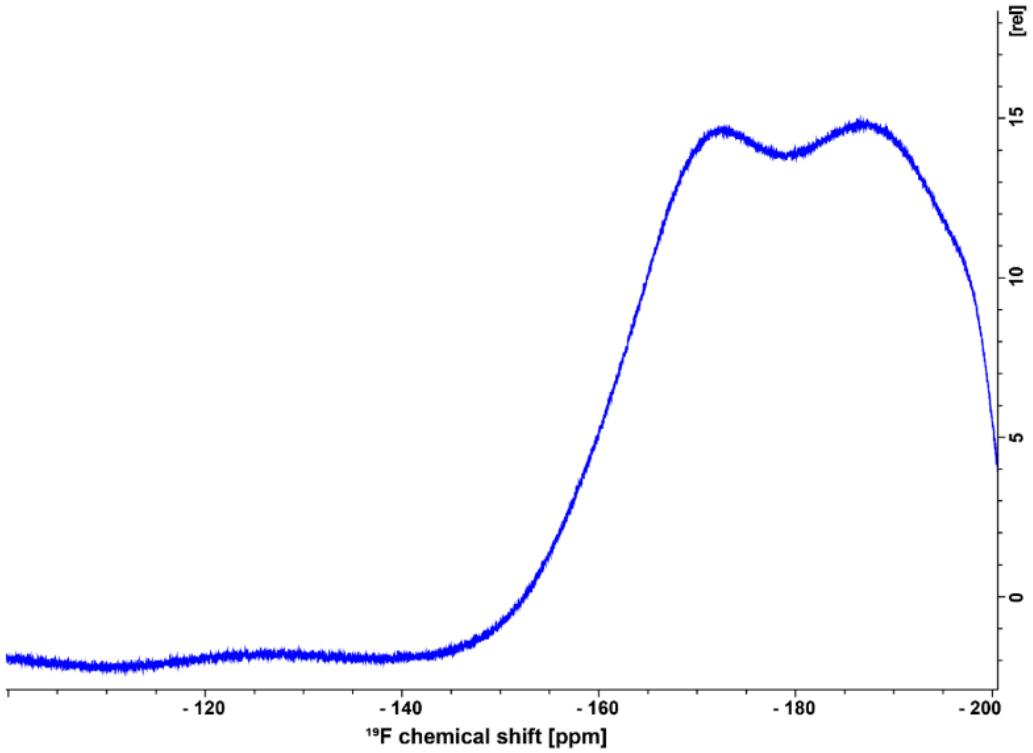


Figure 3.1: Broad signal arising from Teflon.
This signal was observed in the 1D ^{19}F NMR spectrum between –180 and –120 ppm using a standard 90°–acquire pulse experiment. These signals were not observed upon introducing a spin-echo scheme into the pulse program with a spin-echo delay of 4 ms.

Additionally, it was not possible to phase the spectrum easily to give in-phase signals because of first-order phase corrections (Figure 3.2). Delays that accounted for the first-order phase correction were added to the pulse program. This was achieved by adding a delay before or after the 180° refocusing pulse (D8 and D9, respectively). The length of this additional delay was calculated by multiplying the dwell time with the ratio of the first order phase correction and 180° (Equations 2 and 3 in Section 2.3; [158]).

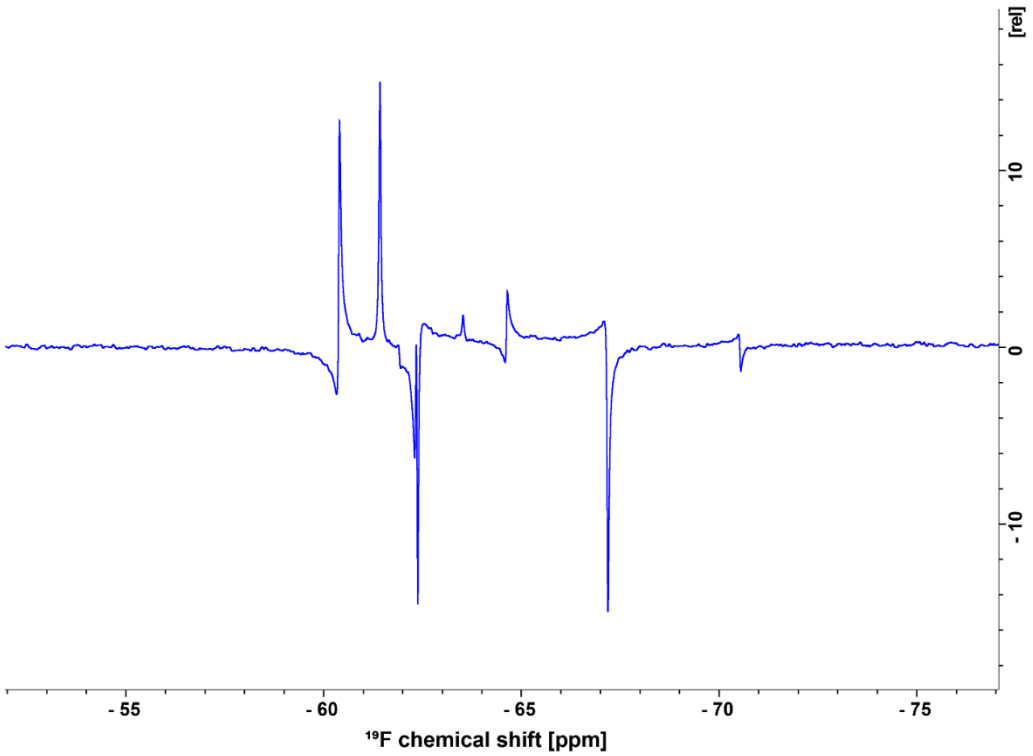


Figure 3.2: An initial 1D ^{19}F NMR spectrum with ten fragments showing issues with phasing.
The first-order phase was corrected by including delay D8 before the refocusing 180° pulse used in the spin-echo element.

Subsequently, a suitable approach to reference the ^{19}F NMR spectra using reference compounds was investigated. Initially, a coaxial small-volume NMR insert (1.26-mm inner diameter and 2.0-mm outer diameter) loaded with a reference compound was inserted into a 5-mm standard NMR tube containing a solution of fragments to avoid the reference compound potentially interacting with the protein or fragments. However, inserting coaxial inserts into all NMR samples was considered impractical for automated screening. Furthermore, inserting a capillary decreases the sample volume within the radiofrequency-coil region by displacing sample volume, thereby reducing signal intensities. For tri-fluorinated fragments, 5 µM TFA was used as a reference with a chemical shift of –75.39 ppm, whereas for mono-fluorinated fragments 120 µM fluorobenzene was used with a reference chemical shift of –113.6 ppm [172]. The fluorobenzene ^{19}F signal splits into a septet due to J_{FH} couplings. This splitting reduced the peak intensity and thus a concentration of 120 µM was required to ensure sufficient signal-to-noise. No interactions between the reference compounds and protein were observed by NMR, e.g., a change in the chemical shift of the reference compound signal (Figure 3.3). Interactions of the reference compound with fragments were unlikely. Nevertheless, any such interactions were filtered by way of double-checking fragment hits during the second round of screening hit fragments.

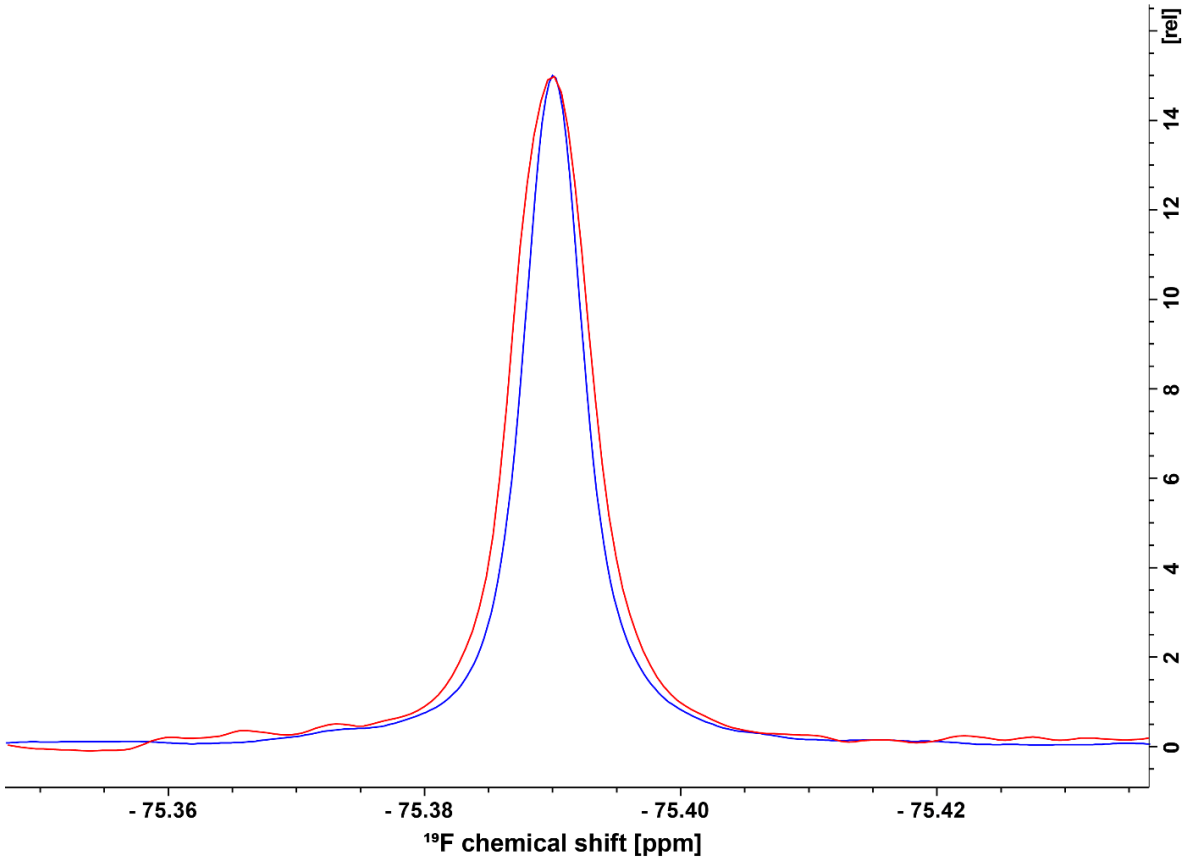


Figure 3.3: Peak arising from TFA in the absence (blue) and presence (red) of His-tagged *Ec*PBP3ΔTM.
Only small differences in linewidth (linewidth at half-height = 0.89 Hz) and in chemical shift (0.04 ppm; 2.3 Hz) were observed, most likely because of small differences in field homogeneity, and/or sample conditions; however, weak non-specific interactions with the protein cannot be excluded, which would likely lead to the slight broadening and shift of the peak.

Next, 1D ^{19}F NMR spectra of fragments in pool sizes of 10, 15 and 20 fragments were recorded to establish the pool size used in the FBS. Pool sizes of up to 50 fragments per pool have been reported; however, in the test pools of 20 fragments some overlap was observed [98]. Nevertheless, 20 fragments per pool were chosen to reduce the NMR time required to screen an entire library of fragments (see Figure 3.4 and Figure 3.9 for exemplary spectra) while retaining adequate resolution.

Lastly, the optimal fragment concentration to use in the screening experiments was determined. Based on information provided by the supplier of the fragment library, the solubility of the fragments should be >200 µM in aqueous buffers. Three samples containing the 20-fragment test pool at 30, 90 and 270 µM were prepared and NMR spectra recorded to determine solubility limits under the buffer conditions used for fragment screening. No difference in peak intensities were observed between the 30 and 90 µM fragment samples (after accounting for the three-fold difference in concentration). Thus, all fragments examined were soluble at 90 µM. In contrast, peak intensities for the 270 µM sample were weaker (after accounting for the concentration differences) when compared with those of the 30 and 90 µM samples, indicating that fragments were partially soluble at 270 µM under the conditions used. Fragments prepared at 30 µM gave sufficiently good signal-to-noise in an experimental time of 2 h, and thus, this concentration was used for fragment screening, which is in keeping with the low µM range (18–50 µM) used for reported fragment screens [98, 112, 162, 173, 174]. Typically, in ^{19}F NMF FBS, a 50 to 100-fold excess of the protein can be used to detect weak

binders [98]. However, to increase the observable changes upon protein addition and to avoid competitive binding between several fragments in the pool, equimolar concentrations of 30 µM His-tagged *Ec*PBP3ΔTM and fragments were used in the following screening experiments. To avoid differences in chemical shift and intensity of the peak arising from the reference compound caused by non-specific interaction, the protein and fragment concentrations were kept constant in all screening experiments.

3.1.2. First FBS used to establish the protocol

3.1.2.1. Tri-fluorinated fragments

In the first FBS by ^{19}F NMR, 206 tri-fluorinated and 225 mono-fluorinated fragments from KeyOrganics were screened against His-tagged *Ec*PBP3ΔTM. Samples were prepared with a pool size of 20 fragments and concentrations of 30 µM for each fragment. Additionally, for tri-fluorinated fragment pools, 5 µM TFA and 30 µM *Ec*PBP3-His were used and chemical shifts were recorded with an offset of –68 ppm and a sweep width of approximately 34 ppm (Figure 3.4). Spectra were referenced to the chemical shift and intensity of the TFA signal. The referencing of the intensity among spectra recorded ensured that minor differences caused by small differences in field homogeneity were accounted for.

In the first step, the number of peaks in the spectrum was determined (i.e., should be 21 peaks). ^{19}F chemical shifts for each fragment in DMSO were provided by the supplier (KeyOrganics). Using these chemical shifts to assign the recorded spectra was challenging because spectra were recorded in aqueous buffer and therefore, the chemical shifts for fragment samples did not always match closely to the supplier-provided chemical shifts (Table 3.1). This was particularly noticeable in crowded regions of the spectra (boxed region of spectrum in Figure 3.4). Thus, at this stage no attempt was made to assign all peaks unambiguously in fragment-only spectra. For example, fragments EC-0718, FS-1143, 9B-057, 11J-336S, and FS-2038 have a supplier-reported chemical shift range -62.2 and -62.5 ppm and measured values between -62.230 and -62.486 ppm, thus, making unambiguous assignments in this region of the spectrum challenging without further analysis (Table 3.1).

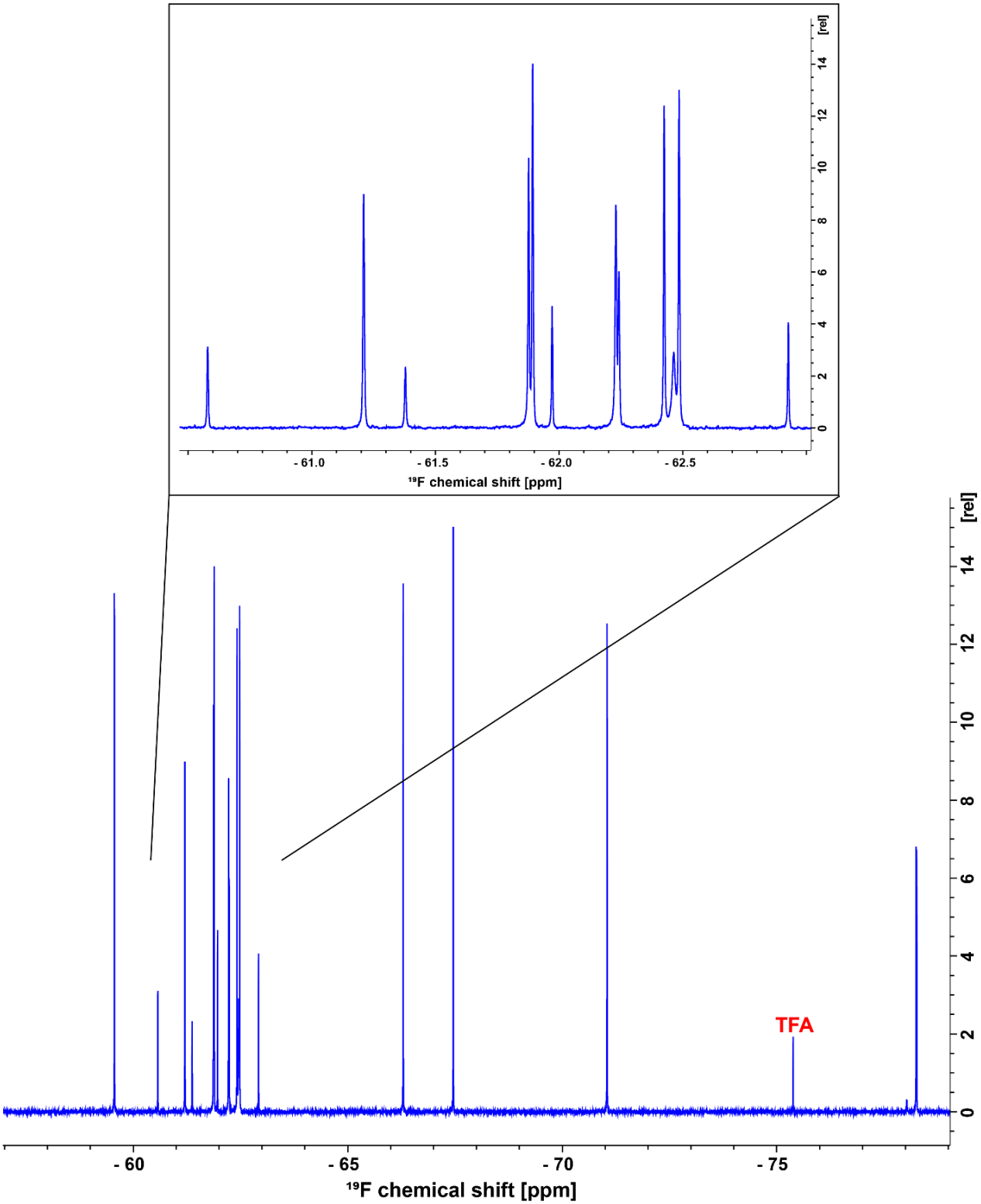
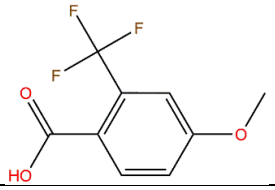
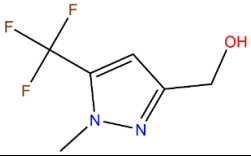
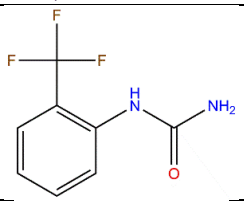
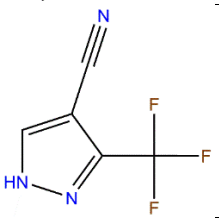
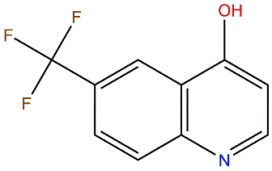
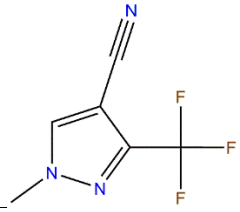
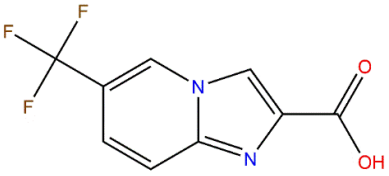
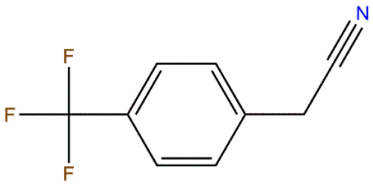


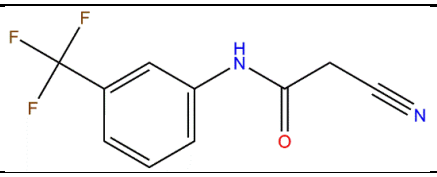
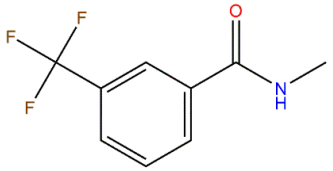
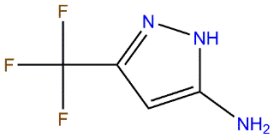
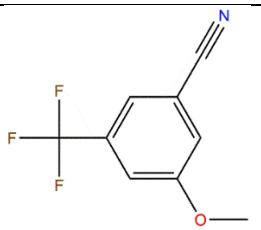
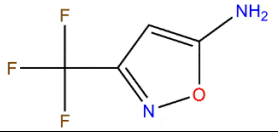
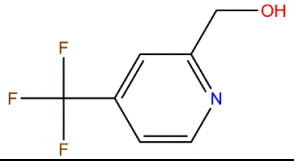
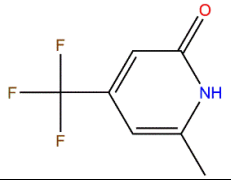
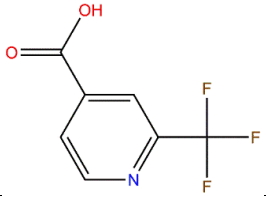
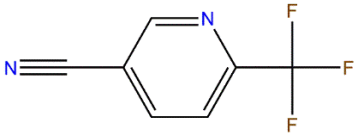
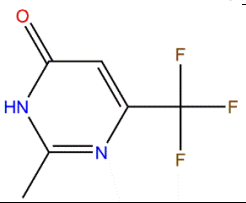
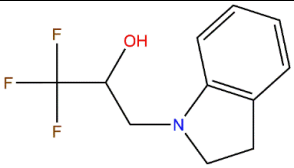
Figure 3.4: 1D ^{19}F NMR spectrum of 20 tri-fluorinated fragments recorded during the first screen.

The spectrum of the sample (30 µM pool 7, supplied by KeyOrganics) was recorded with 3072 scans, 32768 points in the time domain, a sweep width of approximately 34 ppm and an offset at –68.0 ppm. The spectrum was referenced to the TFA (5 µM; labeled) signal at –75.39 ppm. The concentration of each fragment was 30 µM. Expansion of the region between –60.5 and –63.0 ppm is presented in the inset. It shows that peaks with similar chemical shifts prohibited unambiguous assignment of all signals in this crowded region of the 1D ^{19}F spectrum. In particular, this is clearly seen for peaks at approximately –62.5 ppm. The signal-to-noise ratios were ~50 to 300.

Table 3.1: Comparison between the observed and supplied chemical shifts of tri-fluorinated fragments in pool 7 from the first screen.

In this example, the closest matching observed chemical shifts were assigned to the peaks arising from the fragments. In some cases, peaks from fragments were not assigned, whereas some peaks were not found at the supplier-provided chemical shifts.

Observed chemical shift [ppm]	Supplied chemical shift [ppm]	Fragment ID	Chemical structure
-58.775	–	–	–
-59.568	-59.6	FS-2089	
-60.581	-60.6	AS-1017	
-61.211	-61.2	BS-4139	
-61.380	-61.5	7H-076	
-61.878	-61.9	FS-1145	
-61.894	–	–	–
-61.973	-62.0	GF-0702	
-62.231	-62.2	EC-0718	
-62.243	-62.3	FS-1143	

-62.426	-62.4	9B-057	
-62.465	-62.5	11J-336S	
-62.486	-62.5	FS-2038	
-62.928	-62.9	PS-6942	
-63.187	-64.2	FD-0252	
–	-64.8	EE-0755	
-66.300	-66.3	12T-0332	
-67.468	-67.5	FS-2591	
–	-68.2	FS-1207	
-71.052	-70.8	BS-3211	
-78.264	-78.3	6P-094	

Initially, and to circumvent the abovementioned issues with assigning signals, spectra with and without the protein were compared without making assignments to expedite the screening process. Changes in the 1D ^{19}F NMR spectrum upon addition of 30 µM His-tagged *Ec*PBP3ΔTM were broadly grouped into three sets of observations:

(i) No change in the peak intensity and chemical shift showed that the fragment corresponding to this signal does not interact with the protein (Figure 3.5A).

(ii) A small decrease in peak intensity (ΔI) with a small change in the chemical shift ($\Delta\delta$; Figure 3.5B). This observation indicated that only a minute fraction of the fragment population is in the protein-bound state.

(iii) Larger ΔI and $\Delta\delta$ indicated an interaction with the protein (Figure 3.5 C and D). The chemical shift change represents the weighted chemical shift of the free- and bound-state chemical shifts. The decrease in peak intensity arises from an increase in the ^{19}F transverse relaxation rate R_2, because of the population of fragment in the protein-bound state (i.e., bound-state R_2, which is dependent on τ_{res} of the fragment bound to the protein over a defined range of τ_{res}) and the exchange contribution (i.e., $k_{ex} = k_{on} + k_{off}$) deriving from the difference in the isotropic chemical shift between free and bound state isotropic chemical shifts (e.g., k_{ex} regime is fast-to-intermediate on the ^{19}F chemical shift timescale) [126-128, 175].

After visually observing the different types of changes, a scoring system was developed to classify fragments after the addition of His-tagged *Ec*PBP3ΔTM. In total, five scores from 1 to 5 were established using the standard deviation (σ) of the changes in intensity and chemical shifts of all fragments, which were determined to be 16.9% and 3.6 Hz, respectively (Figure 3.6). Mean values were 12.5% and 2.4 Hz, respectively. In the group with a score of 1, fragments were pooled and showed changes of at least 10 Hz (corresponding to approximately 3σ) with a reduction in intensity of more than 30% in the presence of the protein when compared with the corresponding signal in the absence of the protein (corresponding to approximately 2σ). With these boundaries the proportion of score 1 fragments was adjusted to fit with the expected hit rate reported in literature, i.e., between 2% to 7% [112, 174, 176]. An additional boundary for the chemical shift was defined to be 3 Hz corresponding to approximately the σ and 5% and 15% decreases in intensity, which corresponded to approximately 1σ and 0.3σ, respectively. These additional scores were applied to identify fragments with high but not the highest observed changes. In contrast to this scoring scheme representing peaks that showed different behaviors after the addition of the protein, for example, in the literature a loss of 25% of the intensity was considered to be a hit [174].

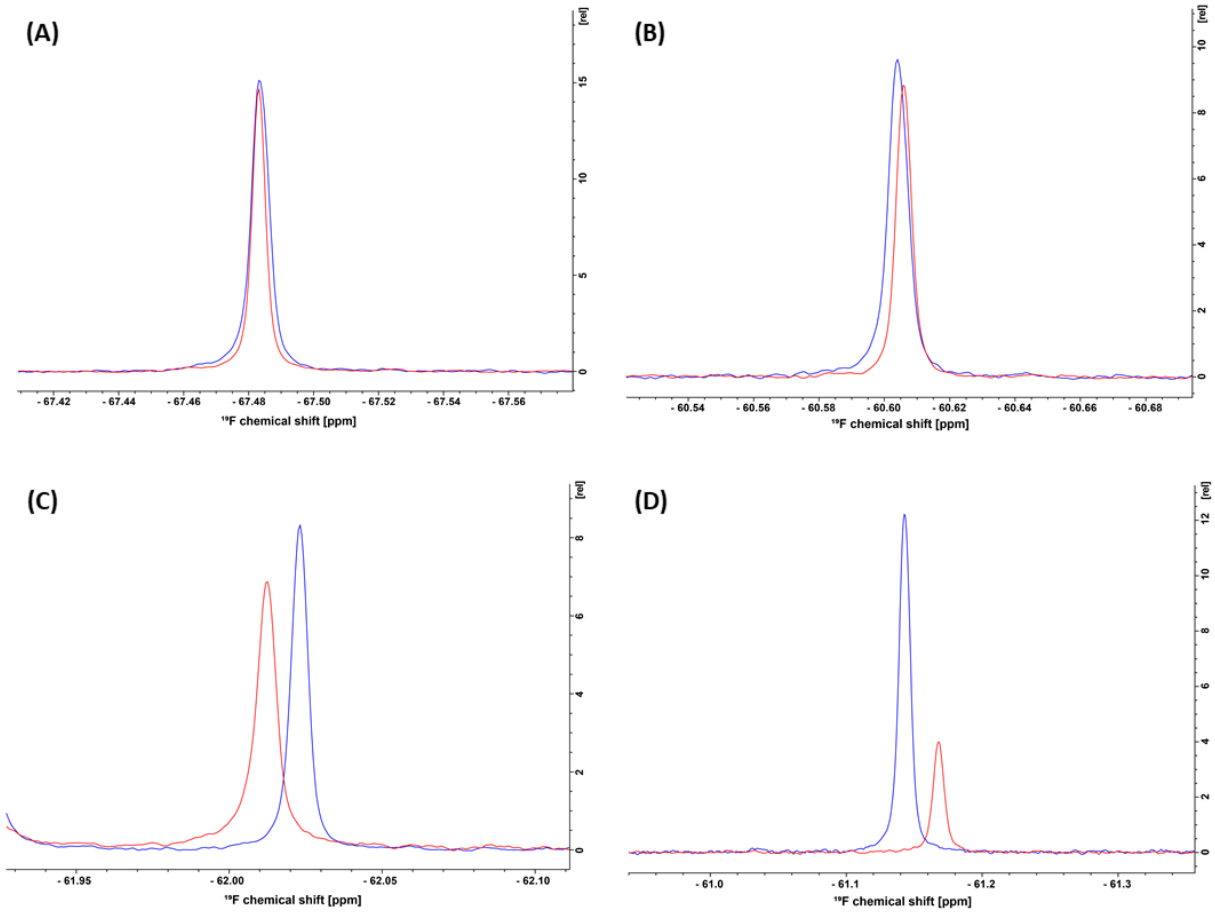


Figure 3.5: Examples of changes in chemical shift and intensity for peaks arising from particular fragments in the absence (blue) and presence (red) of His-tagged *Ec*PBP3ΔTM.

(**A**) Very slight change in the chemical shift and intensity of the signal arising from a fragment in the presence of His-tagged *Ec*PBP3ΔTM, indicating that this fragment does not interact or interacts very weakly with the protein. Fragments that yielded these types of changes to the spectrum when the protein was present were classed as non-binders and excluded from further analysis. (**B**) Only a small change in intensity and chemical shift is observed, indicating that a very small fraction of the fragment population is in the protein-bound state and that this fragment binds very weakly to the protein. These fragments have low affinities and were excluded in further analysis. Peaks arising from fragments that interact with the protein, show changes in their intensities because of the increased R_2 arising from the ligand in the protein-bound state and fast-to-intermediate chemical exchange on the ^{19}F chemical shift timescale (see Equation 1 in Section 1.9.2). Furthermore, the chemical shift represents the population average of the free- and protein bound-state populations of the fragment. Smaller $\Delta\delta$ and ΔI are shown in (**C**), whereas the largest $\Delta\delta$ and ΔI observed are shown in (**D**).

In total, 4% (8) of the fragments were classified as score 1, 3% (7) as score 2, 17% (36) as score 3, 7% (14) as score 4 and 68% (141) as score 5. The distribution of $\Delta\delta$ and ΔI is shown in Figure 3.7.

Assignment of peaks with scores 1 to 2 and selected peaks with scores 3 and 4 was initially performed (i.e., 22 out of 206) using the supplier-provided chemical shifts. However, supplier-provided chemical shifts did not always match with the observed chemical shifts, which prohibited unambiguous identification of all 22 potential hit fragments. Moreover, most peaks assigned to scores 1 and 2 were weaker in intensity when compared with the majority of other peaks, raising the possibility that these peaks represented impurities or degraded fragments, or fragments with poor solubility in the buffer used (Figure 3.8). To overcome assignment ambiguity, 31 fragments were selected based on supplier-provided chemical shifts matching closely on of the 22 chemical shifts obtained for peaks that represent potential hit fragments and an assignment validation screen was performed (Section 3.1.2.3).

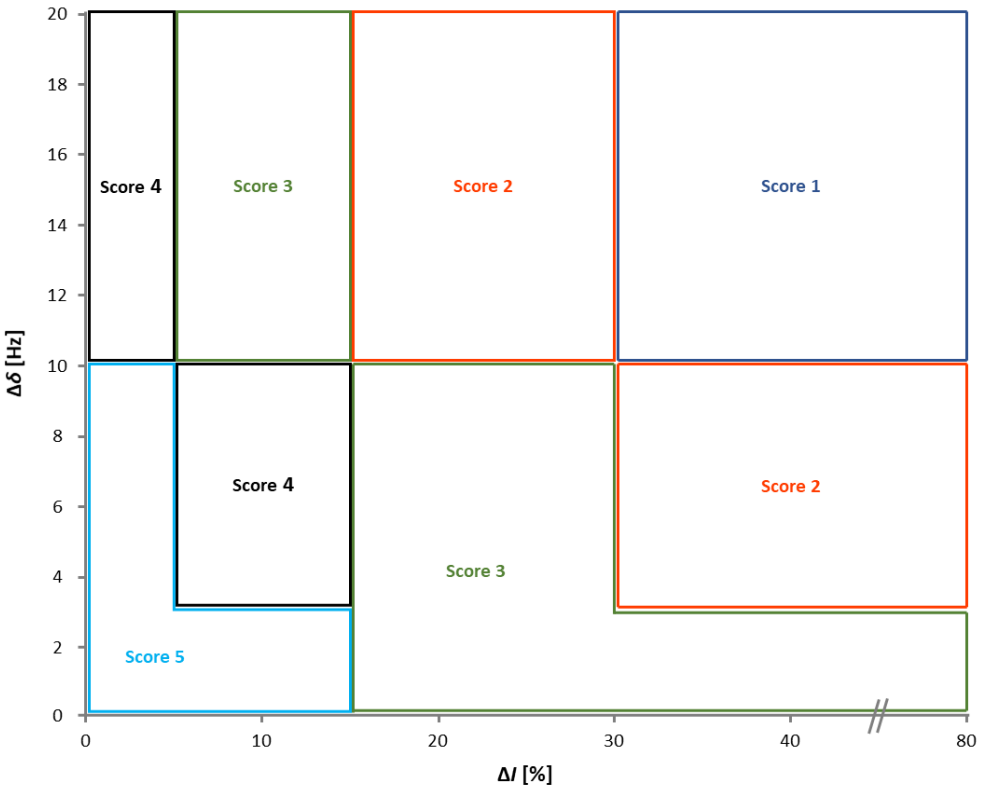


Figure 3.6: Developed scoring system for the first ^{19}F fragment library screened.
The observed reduction in peak intensity and change in chemical shift of peaks were used to categorize interactions between fragments and the His-tagged *Ec*PBP3ΔTM. This system was applied to mono- and tri-fluorinated fragments in the first performed screen performed and was based on standard deviation values.

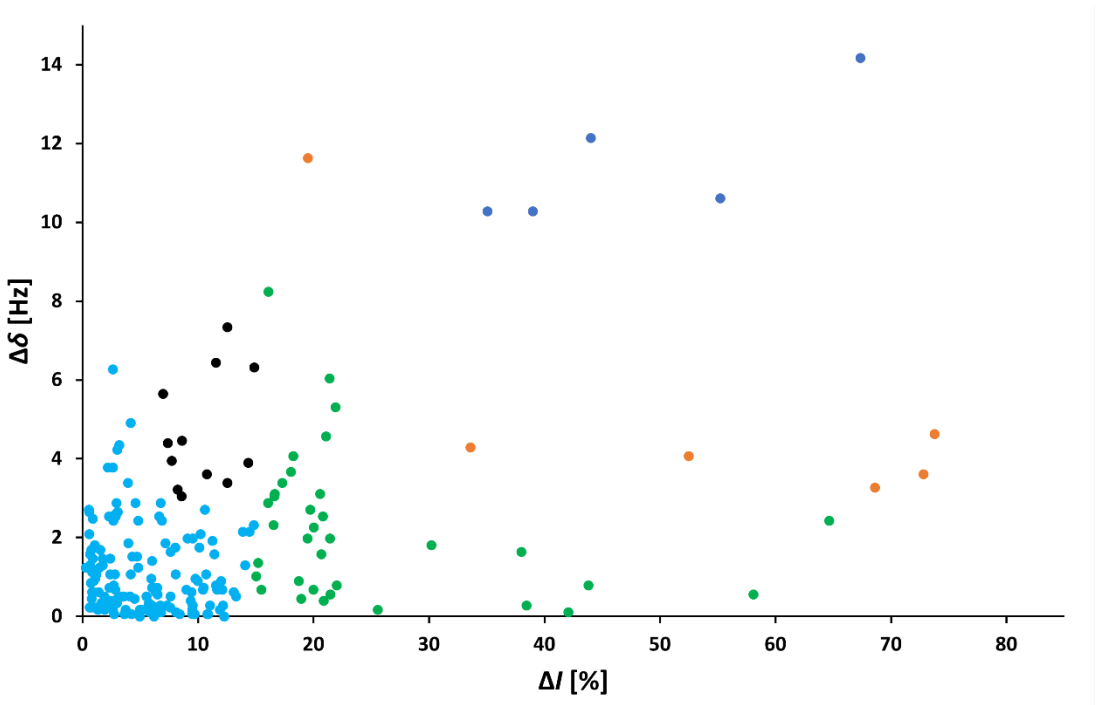


Figure 3.7: Result of the first screen of tri-fluorinated fragments.
Peaks with scores 1 to 5 are colored dark blue, orange, green, black and light blue, respectively.

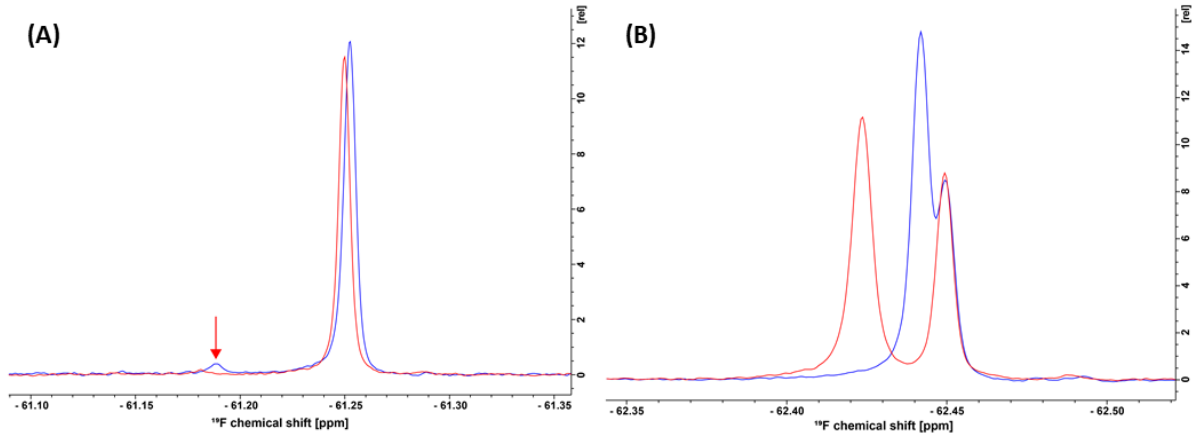


Figure 3.8: Examples of peaks assigned to score 1 and 2 not unambiguously assigned to fragments.
(**A**) Relatively weak peak (indicated by the red arrow) was assigned to score 1, but no fragment with matching chemical shift was found. This peak most likely arises from an impurity or degraded fragment. (**B**) One of the two peaks showed a change in chemical shift and intensity after the addition of His-tagged *Ec*PBP3ΔTM and was classified as score 2. Because of the similar chemical shifts, it is not possible to unambiguously assign the peak to a specific fragment with only these spectra.

3.1.2.2. Mono-fluorinated fragments

Pools containing 20 mono-fluorinated fragments per pool with concentrations of 30 µM were prepared. 1D ^{19}F NMR spectra were recorded for each pool in the absence and presence of 30 µM of His-tagged *Ec*PBP3ΔTM (Figure 3.9). Fluorobenzene at 120 µM was used as the chemical shift and intensity reference.

As shown in the inset in Figure 3.9, a major issue was the presence of active heteronuclear scalar couplings between ^{19}F and ^{1}H nuclei. Thus, signals were split into multiplets, thereby reducing the peak intensities. In the absence of NMR hardware to ^{1}H decouple, the number of scans was increased by a factor of two to give a $\sqrt{2}$ theoretical improvement of the signal-to-noise ratio. Additionally, because of the reduced number of ^{19}F nuclei in mono-fluorinated fragments, the respective signals were weaker when compared with the signals of tri-fluorinated fragments. Although obtaining intensities similar to those observed in NMR spectra for the tri-fluorinated fragments would be possible by recording the experiments with more scans, such an approach would be prohibitive because the acquisition time would have increased dramatically (approximately 100 times to obtain comparable signal-to-noise ratios). Increasing the concentration of the mono-fluorinated fragments to improve the signal-to-noise was considered. However, potential solubility issues of the fragments and the use of conditions that did not match the tri-fluorinated screening conditions precluded the use of higher fragment concentrations. Greater spectral dispersion was observed for NMR data recorded on mono-fluorinated fragments when compared with those recorded on tri-fluorinated fragments, which facilitated the assignment process of spectra recorded on monofluorinated fragments. Nevertheless, complete assignment of all spectra was not possible because some expected peaks were absent, and additional peaks were present that did not match the chemical shifts provided by the supplier (Table 3.2). These issues were probably caused by poor solubility of these fragments and the differences in the solution conditions used.

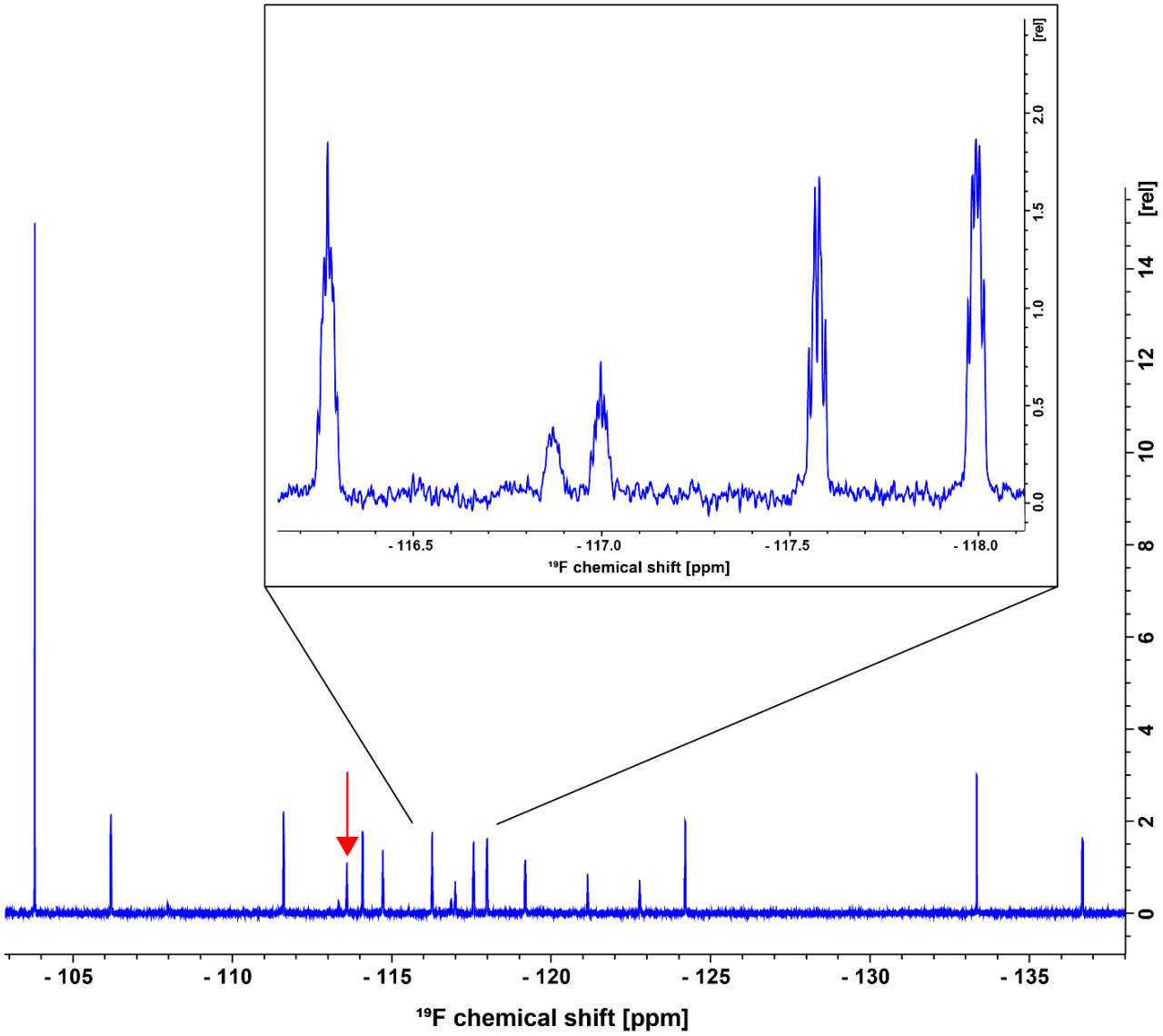
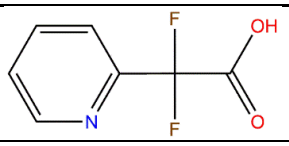
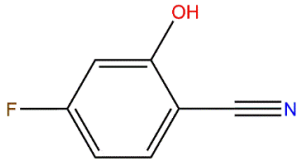
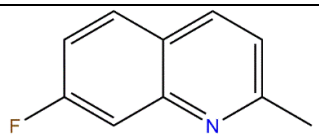
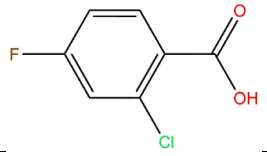
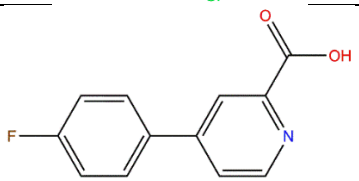
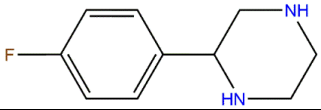
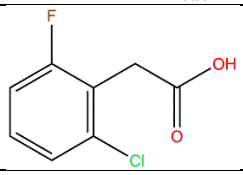
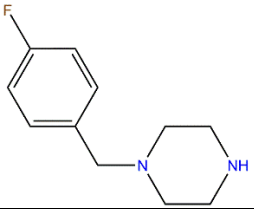
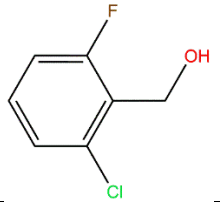
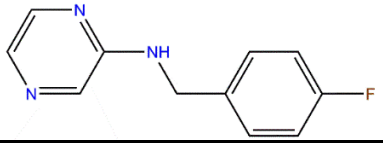


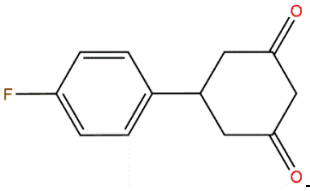
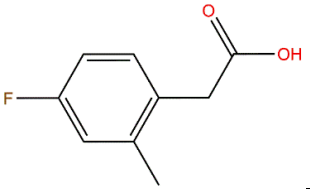
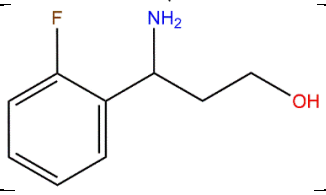
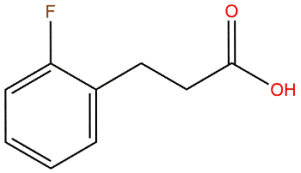
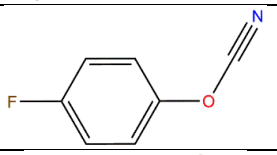
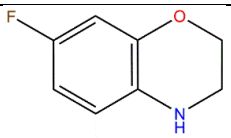
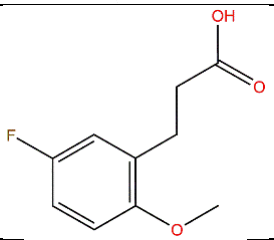
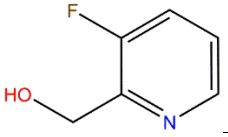
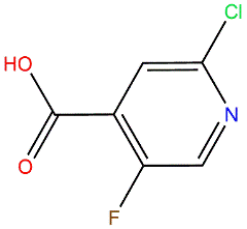
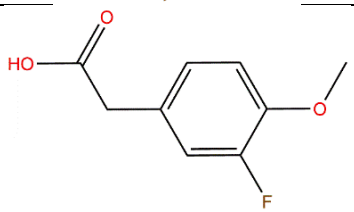
Figure 3.9: Sample 1D ^{19}F NMR spectrum of 20 mono-fluorinated fragments.

The spectrum of the sample (30 µM pool 6, supplied by KeyOrganics) was recorded with 6144 scans, 65536 points in the time domain, a sweep width of approximately 47 ppm and an offset at –118.5 ppm and fluorobenzene (indicated by the red arrow) with a chemical shift at –113.6 ppm was used as a chemical shift reference and for normalizing intensities. The peak arising from the di-fluorinated fragment at approximately –104.0 ppm has a relatively strong intensity when compared with intensities of other peaks in the spectrum. Expansion of the region between –116.1 and –118.1 ppm of the spectrum is provided in the inset, where peaks are split because of scalar couplings between ^{19}F and ^{1}H nuclei. To increase the signal-to-noise ratio, the number of scans was doubled from 3072 to 6144 when compared with spectra of tri-fluorinated fragments, yielding a theoretical improvement of $\sqrt{2}$ in signal-to-noise. The experimental time was approximately 4 h with a signal-to-noise ratio between 4 and 20.

Table 3.2: Comparison of chemical shifts between observed and supplied chemical shifts for mono-fluorinated fragments in pool 6.

In some cases, fragments were not assigned. Furthermore, some peaks were not observed at the chemical shifts provided by the supplier.

Observed chemical shift [ppm]	Supplied chemical shift [ppm]	Fragment ID	Structure
-103.811	-103.9	TS-02367	
-106.199	–	–	–
-107.987	-107.7	BH-0414	
–	-109.3	PS-6328	
-111.614	-111.6	BS-3950	
–	-112.7	TS-03104	
-113.338	-113.6	FS-1802	
-114.100	-114.1	BS-4026	
-114.726	-114.9	FS-1578	
-115.535	-115.6	GF-0090	
-116.273	-116.3	MS-0440	

-116.871	–	–	–
-116.998	-117.1	6H-024	
-117.600	-117.6	FS-2029	
-118.004	-118.1	10W-0838	
-119.197	-119.2	MS-3534	
-121.149	-121.2	7C-016	
-122.779	-122.8	MD-0717	
-124.200	-124.3	PS-6836	
–	-126.1	GF-0219	
-133.361	-133.4	CS-4413	
-136.700	-136.7	FS-1683	

Spectra of the mono-fluorinated fragment pools were recorded in the presence and absence of His-tagged *Ec*PBP3ΔTM with 6144 scans (approximately 4 h per experiment) and the data analyzed. Chemical shifts and intensities of multiplets were taken from the center line of the multiplets and for doublets, the average intensity was determined using both lines of the doublet. Similar to assignment of spectra recorded on tri-fluorinated fragments (Section 3.1.2.1), unambiguous assignment of all peaks was difficult because of differences between supplier-provided and observed chemical shifts. Thus, peaks were first assigned to different scores in the same scoring system that was developed for tri-fluorinated fragments prior to assigning them to fragments (Figure 3.6). No peak was assigned a score of 1, whereas 1% (2), 5% (11), and 6% (13) of the fragments scored as scores 2, 3 and 4, respectively. 88% (200) of the peaks showed no-to-minor changes and were classified as score 5 (Figure 3.10).

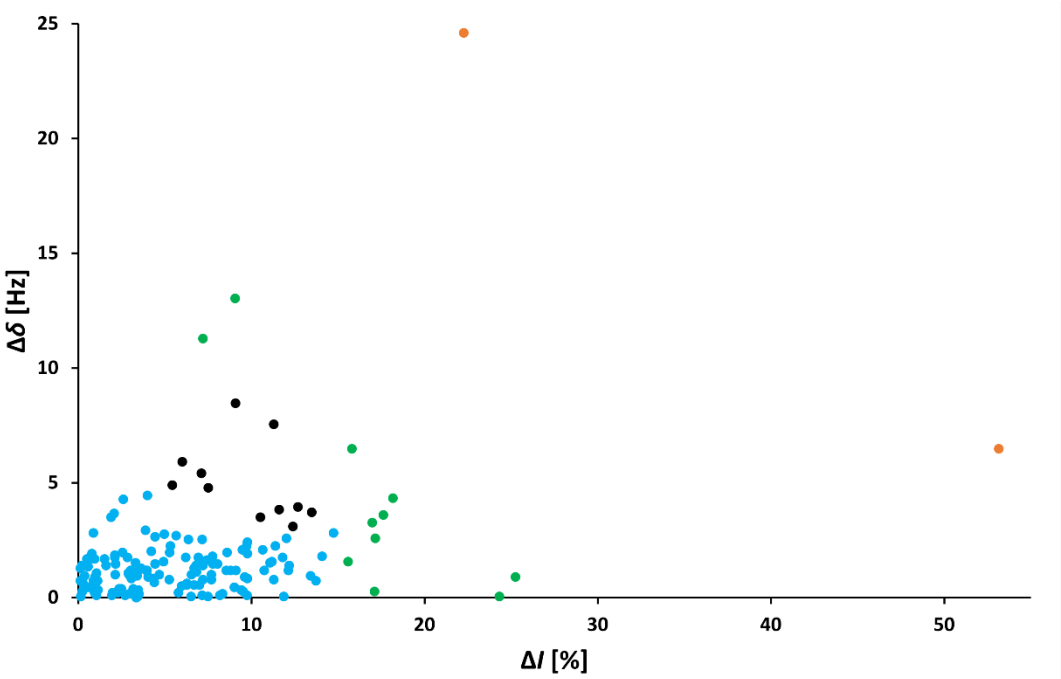


Figure 3.10: Result of the first screening of mono-fluorinated fragments.
Peaks with a score 2 to 5 are colored orange, green, black, and light blue, respectively. No mono-fluorinated fragments were classified as score 1.

The two peaks classified as score 2 showed remarkably larger Δ*δ* or Δ*I* when compared with Δ*δ* or Δ*I* for other peaks (Figure 3.10). With signal-to-noise ratios of 3.6 and 4.9, respectively, these two peaks were relatively weak when compared with the signal-to-noise ratios of other peaks (up to ~20). Therefore, smaller changes in the intensity appeared as relatively large (Figure 3.11). Moreover, based on the weak peak intensities and lacking chemical shift information, these peaks are probably impurities with poor solubility in the buffer used. Thus, they may be hydrophobic and bind the protein non-specifically causing the chemical shift to change. Additionally, it was not possible to unambiguously assign these two peaks to corresponding fragments because no chemical shift provided by the supplier matched either of the chemical shifts of these two peaks. Thus, based on these aforementioned issues, mono-fluorinated fragments were not further investigated. Only assignments of hits identified in the 1st screen of tri-fluorinated fragments were validated, and a second library of tri-fluorinated fragments was screened, which featured fragments with higher solubility, purity and chemical shifts determined in an aqueous buffer that was comparable to the one used in this study.

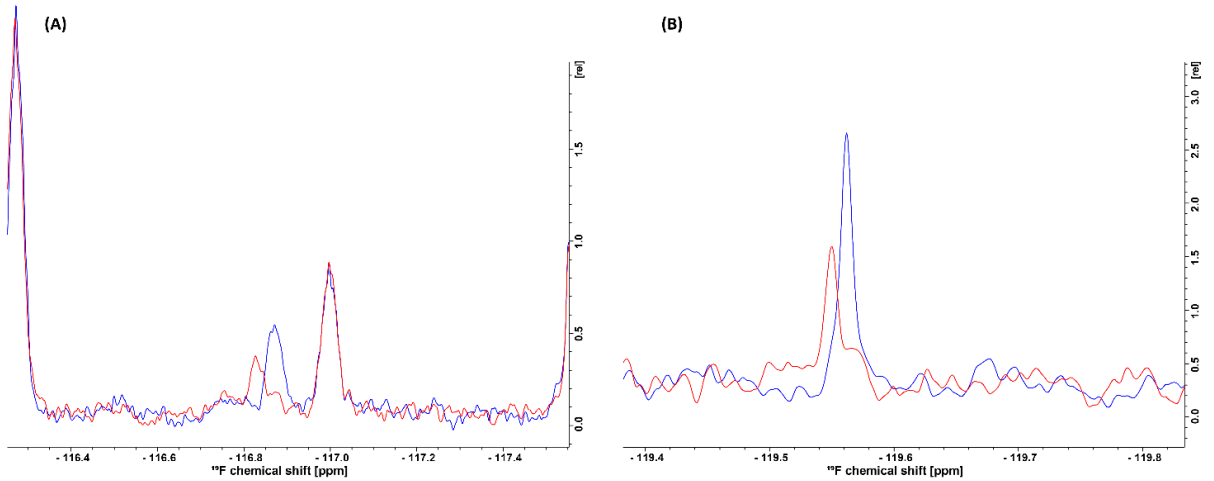


Figure 3.11: Peaks of mono-fluorinated fragments that were classified as score 2.

The spectra in the absence and presence of His-tagged *Ec*PBP3ΔTM are shown in blue and red, respectively. (**A**) The peak at approximately –116.8 ppm was found in pool 6 of mono-fluorinated fragments but was not unambiguously assigned to any fragment in the pool. (**B**) The peak was observed in spectra recorded for mono-fluorinated fragments in pool 3. The signal-to-noise ratios are relatively weak for these peaks (3.6 and 4.9, respectively).

3.1.2.3. Assignment validation

Assignment of all peaks in 1D ^{19}F NMR spectra to specific tri-fluorinated fragments was not possible because the available ^{19}F chemical shifts of the fragments differed from the chemical shifts observed in the experiments and peaks in crowded regions of the spectrum could not be assigned unambiguously (Figure 3.8 and Table 3.1). Presumably, differences in solution conditions and temperature were responsible for these differences in chemical shifts. Therefore, an additional screening round was performed to confirm assignment of peaks that showed the largest $\Delta\delta$ and ΔI (peaks assigned to score 1 and 2 and a selected number of peaks assigned to score 3 and 4). Thirty-one tri-fluorinated fragments were selected to assign 22 peaks and grouped in four pools with 8 or 9 fragments in each pool, including one additional fragment that did not interact with His-tagged *Ec*PBP3ΔTM. This fragment was used as a negative control and standard for referencing. The concentrations of the fragments and the protein were 30 µM and 5 µM TFA was added. 1D ^{19}F NMR experiments were recorded as described for the first screening of tri-fluorinated fragments. As expected, the peak arising from the control fragment did not show any changes when His-tagged *Ec*PBP3ΔTM was added and was used with the peak arising from TFA to reference the spectrum and account for any field inhomogeneities. The resulting changes in chemical shift and intensities are shown in Figure 3.12

In this assignment validation step, only one peak was assigned a score of 1. Four peaks were scored 2 and four and five peaks were scored 3 and 4, respectively. After this screening round it was possible to unambiguously assign most of the peaks that showed the largest changes in chemical shift and intensities. However, the peak with a score of 1 was not assigned because this peak was weak and probably represents an impurity or degraded fragment. To further validate the hits and avoid competitive binding of fragments to the protein, 17 hit fragments with scores 1 to 4 identified in the assignment validation step and still unambiguously assigned fragments were pooled in eight new pools with 2 to 4 fragments. The same non-binding fragment was also included as a negative control and referencing standard.

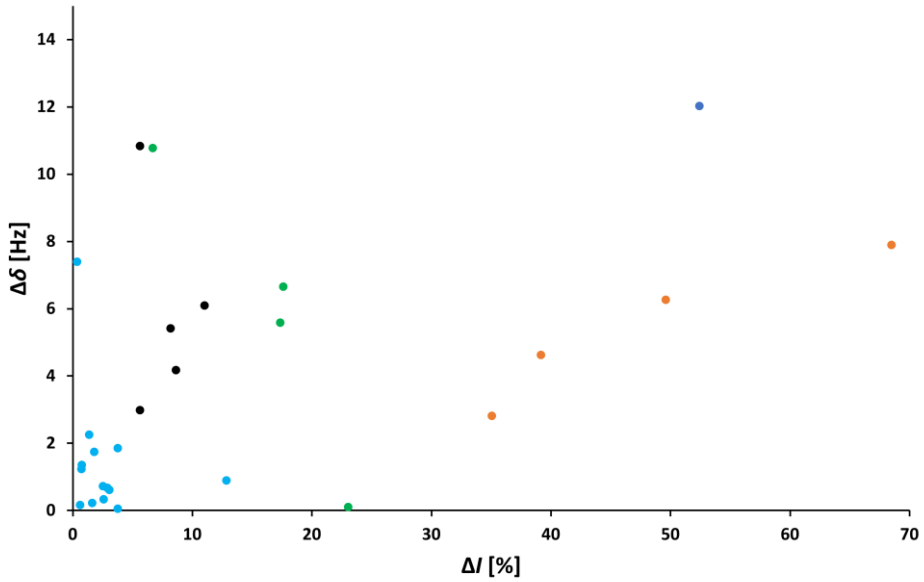


Figure 3.12: Result of the assignment validation of 30 selected tri-fluorinated fragments.
In this set of NMR experiments, pools of 8 to 9 fragments were prepared and a non-binding fragment was added as a negative control.

3.1.2.4. Hit validation

1D ^{19}F NMR experiments of the 17 fragments were recorded as described for the first screening of tri-fluorinated fragments and the peaks arising from the TFA and the non-binding fragment were used to reference the spectrum. Interestingly, not all changes in the intensity and chemical shift in the presence of the protein for these 17 fragments were reproducible (compare Figure 3.13 with Figure 3.7 and Figure 3.12). The changes between the initial and validation screening rounds are presented in Table 4.1 for four fragments that were assigned to scores 2 and 3 after the hit validation screen.

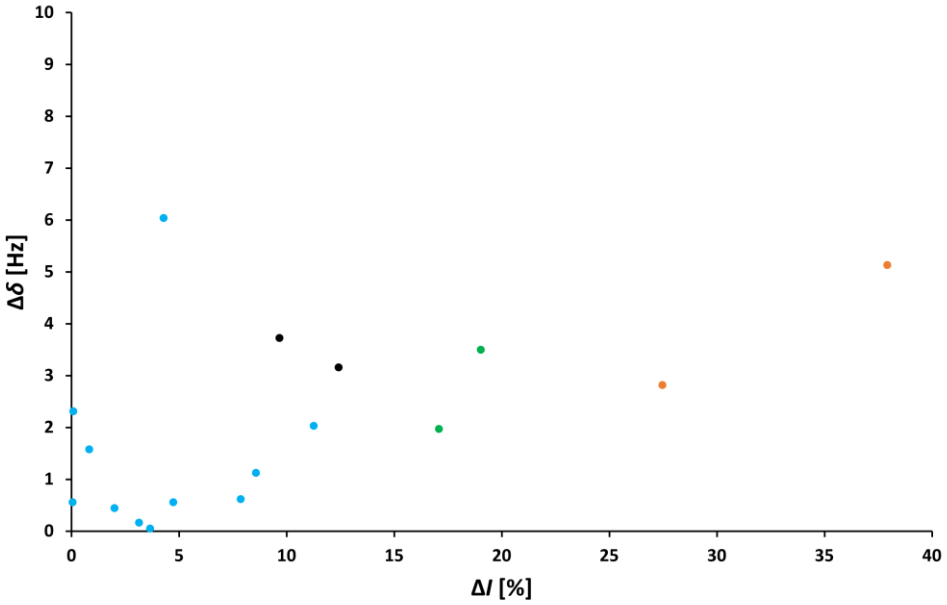


Figure 3.13: Hit validation of selected tri-fluorinated fragments.
5N-395S (right peak) and 3N-528S (left peak) were classified as score 2 and showed the biggest changes after the addition of His-tagged *Ec*PBP3ΔTM.

Nevertheless, similar changes after adding the protein were observed for fragments 5N-395S and 3N-528S (Table 4.1). The chemical structures of these two fragments share a similar scaffold with the only difference being two chemical moieties at positions 2 and 4, respectively (Figure 3.14). Consequently, it was possible to identify two chemically similar fragments that interacted with His-tagged *Ec*PBP3ΔTM. Those changes in the 1D ^{19}F NMR spectra for these two fragments were reproducible in the performed screens.

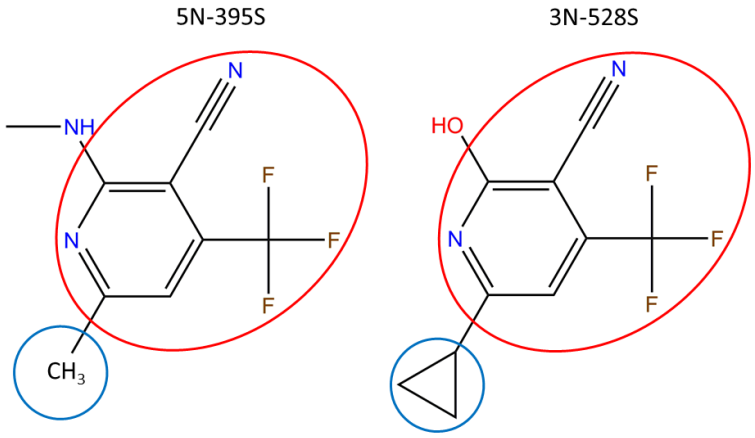


Figure 3.14: Comparison between chemical structures of 5N-395S and 3N-528S.
These hit fragments were identified to interact with His-tagged *Ec*PBP3ΔTM in the screening of tri-fluorinated fragments. Both structures share the same core scaffold (red) containing a pyridine ring with a nitrile group at position 3 and a trifluoromethyl group at position 4. Additionally, a hydrophobic group at position 6 was identified (blue). Only the secondary amide or hydroxy group at position 2, respectively, can be considered slightly different.

Overall, two fragments with score 2 and two fragments with score 3 were identified to interact with His-tagged *Ec*PBP3ΔTM (Figure 3.13). Those fragments were chosen for additional analysis and in-house experiments performed by Dr. Immanuel Grimm at AiCuris. Taking all tri-fluorinated fragments into account, the final hit rate of the first screening was 1.9%, which is similar to the hit rate reported by Vulpetti *et al.* (2%; [176]).

3.1.3. Second FBS to identify hit fragments

Using a ^{19}F fragment library from Enamine, 475 tri-fluorinated fragments were pooled into 30 pools, giving a pool size of 15 or 16 fragments. Fragments in this library feature a higher solubility of >1 mM in aqueous buffer, and the chemical shifts were determined in H_2O. Therefore, the assignment and solubility issues regarding the fragment library that were observed during the first screen were addressed. Using a smaller pool size reduced potential overlap, which facilitated assignment of the peaks to respective fragments (see example for pool 13 in Figure 3.15). For this 1D ^{19}F NMR experiment, the fragment and TFA concentration were kept at 30 µM and 5 µM, respectively, and 30 µM His-tagged *Ec*PBP3ΔTM was used to identify interacting fragments. Chemical shift values that were determined by the supplier are summarized in Table 3.3 with the observed chemical shift values and the structure of the fragments in the pool. 1D ^{19}F NMR experiments were recorded with the same settings as reported for the first screen of tri-fluorinated fragments.

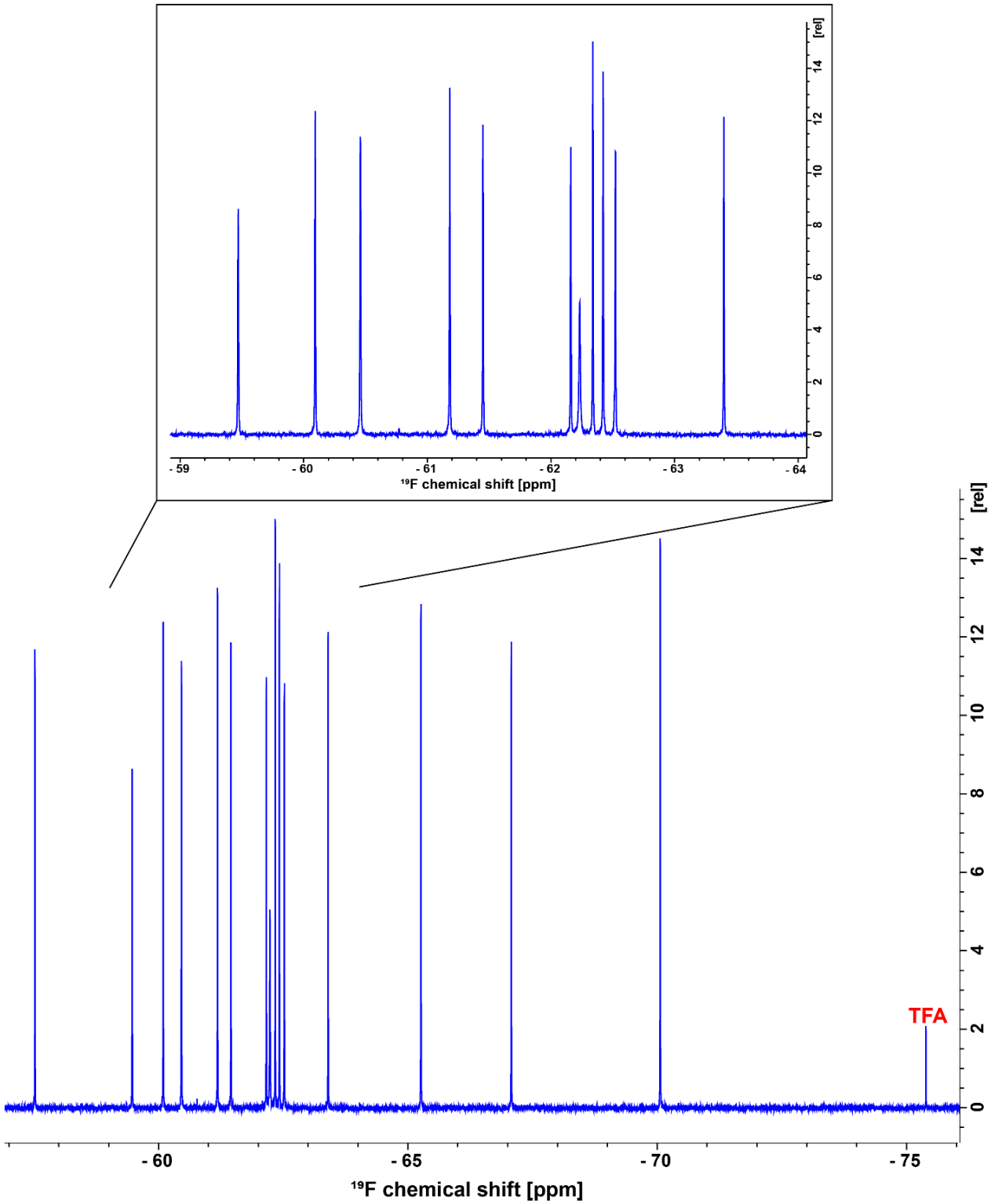
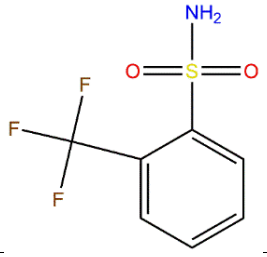
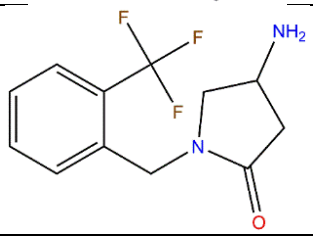
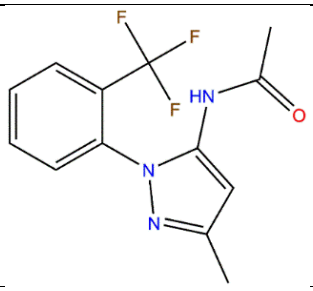
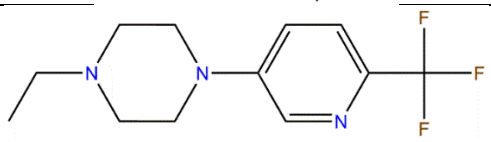
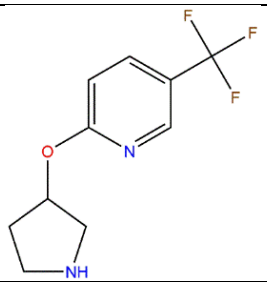
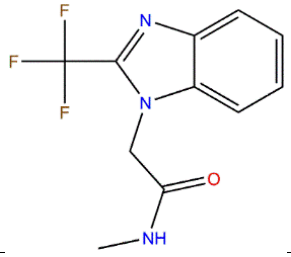
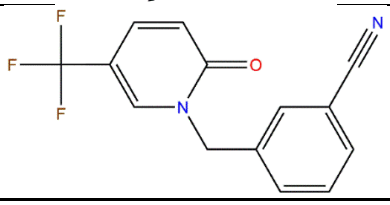


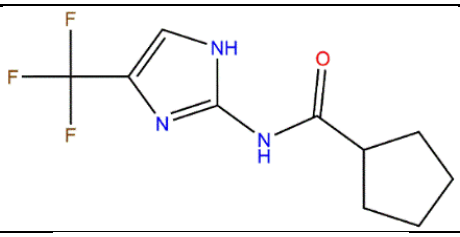
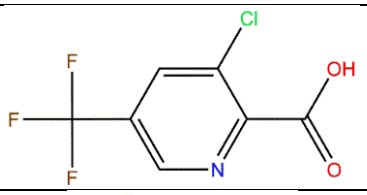
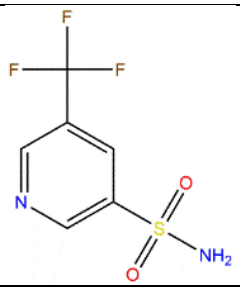
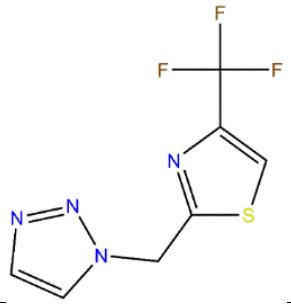
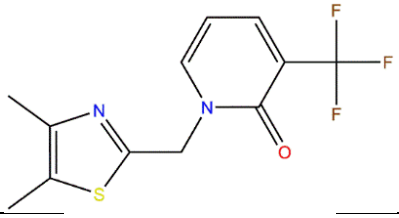
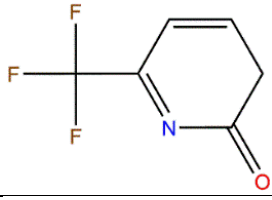
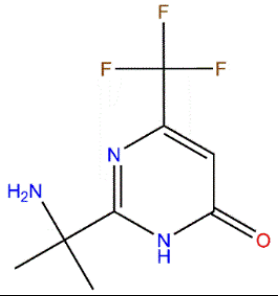
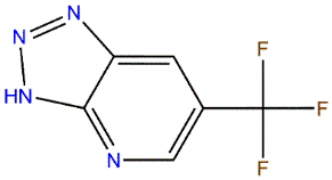
Figure 3.15: 1D ^{19}F spectrum of 15 tri-fluorinated fragments recorded during the second screen.

The spectrum of the sample (pool 13, supplied by Enamine) was referenced to the TFA (5 µM; labeled) signal at –75.39 ppm. The concentration of each fragment was 30 µM. The expanded region between –58.9 and –64.1 ppm is presented in the inset. In this spectrum, all peaks are completely resolved and no overlap was observed, which facilitated the assignment process. Nevertheless, crowded regions of the spectra, e.g., in this case at ~–62.4 ppm, prohibited unambiguous assignment of the peaks.

Table 3.3: Chemical shifts comparison of tri-fluorinated fragments in pool 13 of the second screen between the observed and supplier-provided chemical shifts.

Note that the peak at –60.095 ppm might belong to AIC252665, which may be folded in the spectrum.

Observed chemical shift [ppm]	Supplied chemical shift [ppm]	Fragment ID	Structure
-57.523	-57.20	AIC252519	
-59.472	-59.60	AIC252924	
-60.095	–	–	–
-60.461	-60.51	AIC252918	
-61.184	-61.05	AIC252570	
-61.452	-61.44	AIC252781	
-62.162	-61.90	AIC252645	
-62.235	-62.18	AIC252707	

-62.342	-62.30	AIC252673	
-62.424	-62.44	AIC252902	
-62.524	-62.61	AIC252739	
-63.402	-63.40	AIC252667	
-65.263	-65.33	AIC252547	
-67.075	-67.31	AIC252766	
-70.061	-70.47	AIC252648	
–	-119.50	AIC252665	

Because of the more accurate supplier-provided chemical shift values and the reduced pool size, in the second screen using a library of tri-fluorinated fragments supplied by Enamine, it was easier to unambiguously assign the peaks to the corresponding fragments when compared with assigning signals in the first screen. Nonetheless, it was still not possible to unambiguously assign all peaks. For this reason, similar to the first screen, the changes in chemical shift and intensity of the peaks were initially analyzed without all assignment information. Because of the increased number of fragments, σ differed to the first screen of tri-fluorinated fragments. The average change in chemical shift was 1.3 Hz with σ = 1.6 Hz, whereas the average loss of intensity was 4.6% with σ = 7.3%. Therefore, the boundaries to classify the observed changes of the peaks were adjusted slightly (Figure 3.16). Approximately 3σ corresponds to a loss of 25% intensity or a change of 5 Hz in chemical shift, whereas a reduction of 15% in intensity and a change of 3 Hz in chemical shift are approximately 2σ.

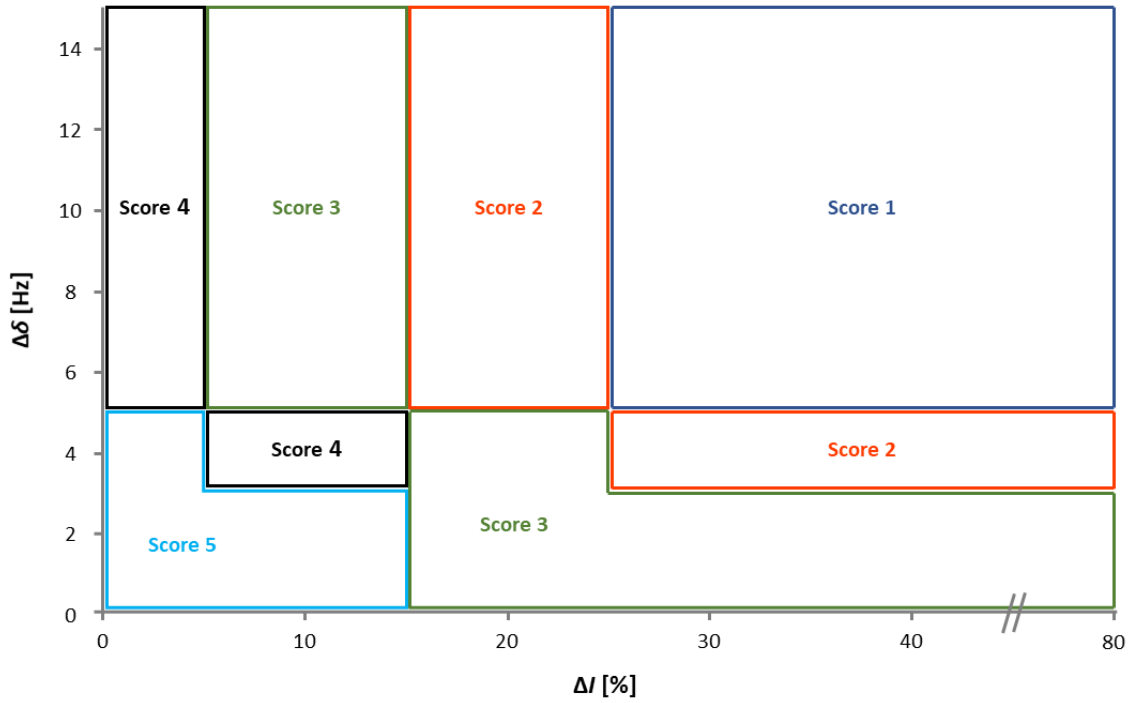


Figure 3.16: Scoring system for the second ^{19}F NMR FBS.
The observed reductions in intensity and changes in chemical shift were used to categorize interactions between fragments and the His-tagged *Ec*PBP3ΔTM.

In this screen, a total of 502 peaks were observed for 475 fragments in the library. This observation suggested that impurities were present in the fragment pools. Applying the adjusted scoring system of the second screen resulted in 0.6% (3) of the peaks with score 1 and 1.6% (8), 4.6% (23), 3.6% (18) and 89.6% (450) peaks for scores 2 to 5, respectively (Figure 3.17).

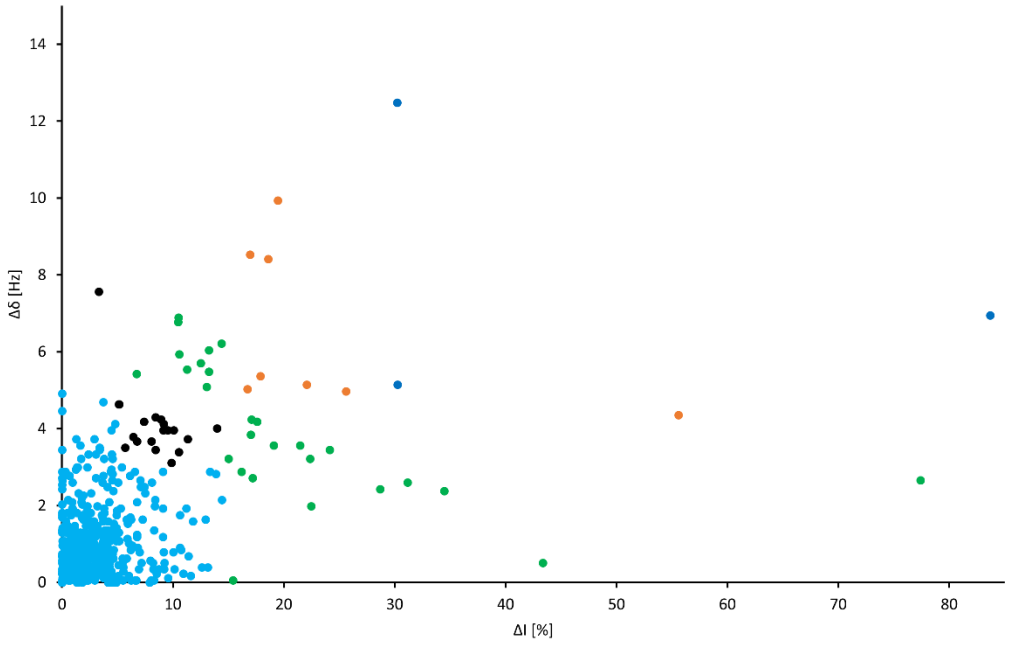


Figure 3.17: Result of the second screen of tri-fluorinated fragments.
Peaks with scores 1 to 5 are colored dark blue, orange, green, black and light blue, respectively.

In contrast to the result of the first screen, an assignment validation step was not required because the supplier-provided chemical shift information was more accurate and the overlap between peaks was reduced by recording NMR data with smaller pool sizes. Thus, the hit and assignment validation steps were combined. In this screening round, nineteen peaks (assigned to score 1 and 2 and a selected number of peaks assigned to score 3) were investigated by selecting thirty-three fragments. Pool sizes of five or six fragments were used to reduce competitive binding and potential overlap in the spectra. 1D ^{19}F NMR experiments with 5 µM TFA, 30 µM fragments, and 30 µM His-tagged *Ec*PBP3ΔTM were performed, as reported previously. In this validation screen, two, eleven, six and one peaks were assigned to fragments and classified as scores 1 to 4, respectively (Figure 3.18). Taking all 475 fragments used in this library into account, 0.4%, 2.3%, 1.3%, 0.2% and 95.8% of the fragments were assigned as scores 1 to 5, respectively. Fragments assigned to scores 1 and 2 were further investigated as hits by orthogonal methods at AiCuris by Dr. Immanuel Grimm. All results are summarized in Table 4.1 and Table 4.2.

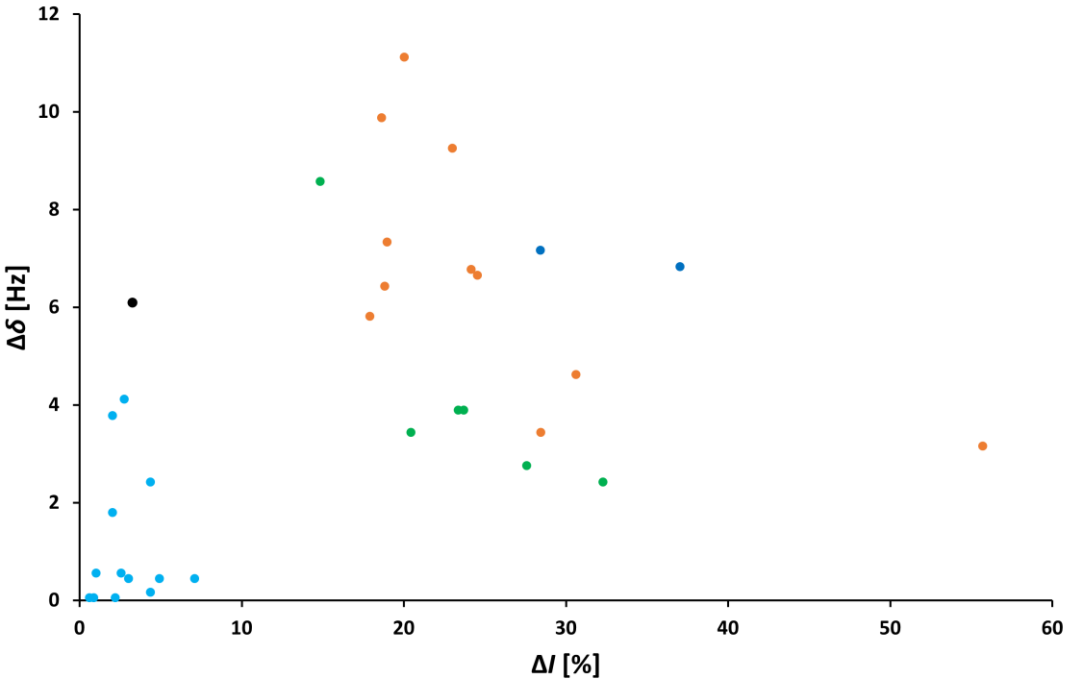


Figure 3.18: Validation of the hit fragments that were identified in the second screen.
Peaks with scores 1 to 5 are colored dark blue, orange, green, black and light blue, respectively.

3.2. Experiments for characterizing hit fragments

3.2.1. Hit fragment K_d value determination

The K_d of the interaction between fragment 5N-395S and His-tagged *Ec*PBP3ΔTM was examined by measuring ^{19}F NMR spectra of the fragment. This fragment was selected because the largest changes in chemical shift and intensity were observed after adding 30 µM His-tagged *Ec*PBP3ΔTM in the first screen. Limited production of the protein prohibited NMR K_d determination of other fragments identified to interact with *Ec*PBP3ΔTM. In determining the K_d, the protein concentration was increased from 0 to 830 µM with a fixed fragment concentration of 60 µM (Figure 3.19). Higher protein and fragment concentrations were not examined because of protein and ligand solubility issues at concentrations above those used in this assay.

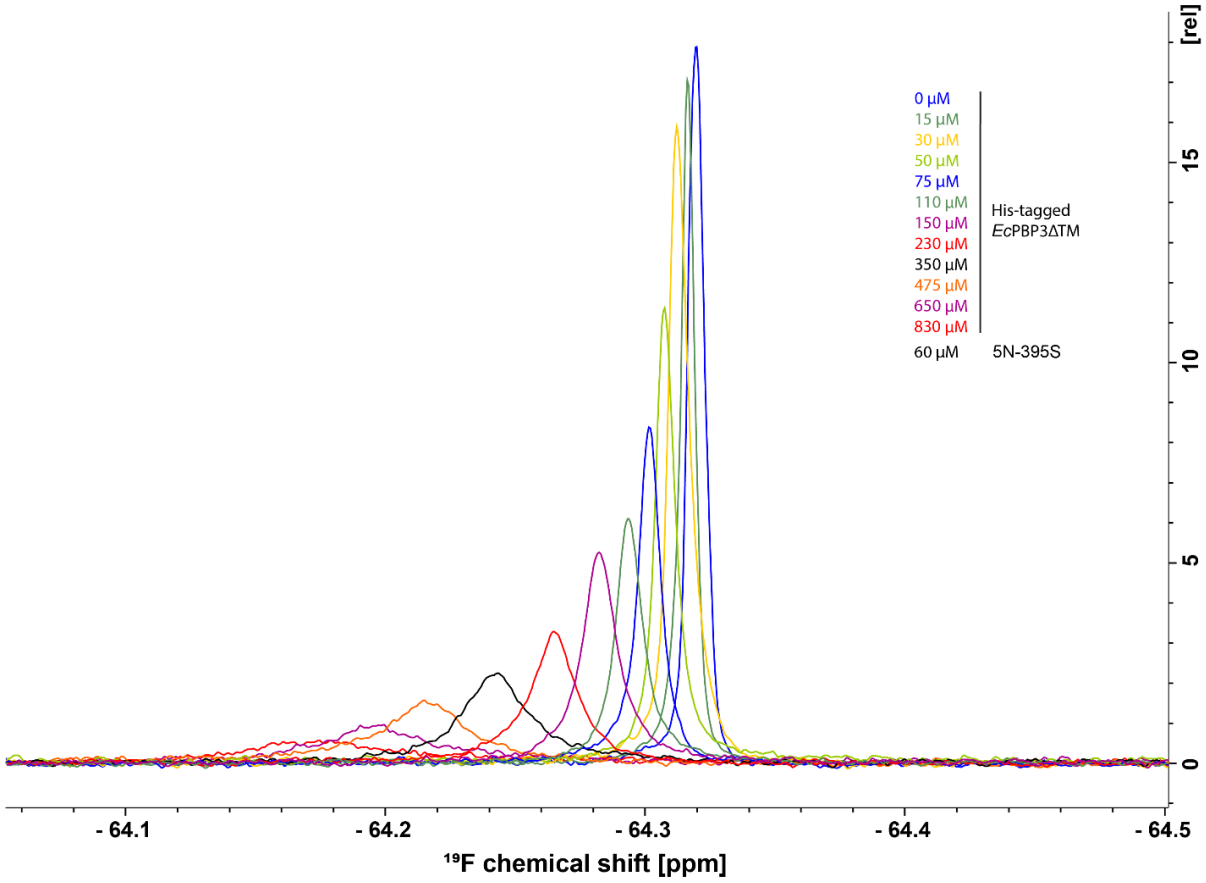


Figure 3.19: Titration of His-tagged *Ec*PBP3ΔTM against 5N-395S to determine the K_d.

His-tagged *Ec*PBP3ΔTM was titrated against 60 µM 5N-395S, which showed the biggest changes in intensities and chemical shift after the addition of the protein during the first screen. Extensive line-broadening and change in the chemical shift of the ^{19}F signal representing 5N-395S were observed as the concentration of the protein increased, indicating chemical exchange on the fast-to-intermediate ^{19}F chemical shift timescale and an increased τ_{cor} for the protein-bound fragment. The absence of the bound-state chemical shift δ_{bound} precludes accurate determination of the K_d.

The ^{19}F chemical shifts were used to calculate $\Delta\delta_{obs}$ for each concentration and $\Delta\delta_{max}$ of 0.39 ppm (i.e., δ_{bound} = –63.97 ppm) and K_d of 1.23 ± 0.54 mM were obtained assuming a one-site binding model using Equation 4. As shown in Figure 3.20, the plateau is not reached for the protein concentration range used because the K_d is higher than the highest protein concentration used. Consequently, these values are estimates and not accurate. However, an estimate of the K_d of the interaction between the identified hit fragment and His-tagged *Ec*PBP3ΔTM was obtained using this fitting routine. This K_d value is in the expected affinity range for small fragments, i.e., 100 µM and 10 mM, and confirms the identification of a fragment that is interacting with His-tagged *Ec*PBP3ΔTM with weak affinity [86].

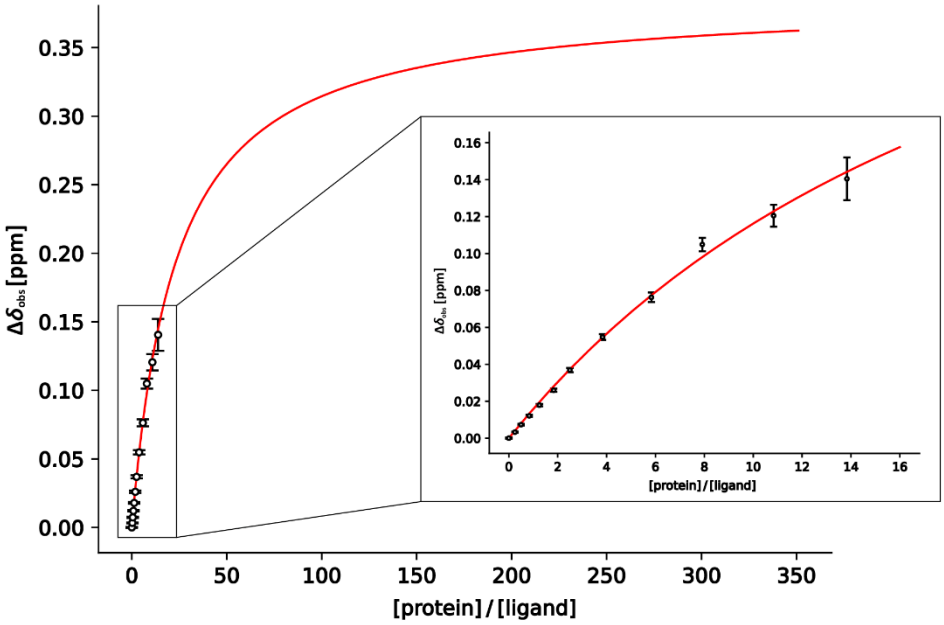


Figure 3.20: Changes in ^{19}F chemical shifts were used to determine kinetic values of the interaction between 5N-395S fragment and His-tagged *Ec*PBP3ΔTM.

Observed changes in chemical shift ($\Delta\delta_{obs}$) were fitted against the molar ratio of His-tagged *Ec*PBP3ΔTM and 5N-395S concentrations. The data were fitted using Equation 4 (red line) to determine $\Delta\delta_{max}$ (0.39 ppm) and K_d (1.23 ± 0.54 mM) values. Because the used concentrations are below the K_d, the plateau of the sigmoidal curve is not reached (boxed plot). Thus, the $\Delta\delta_{max}$ and K_d values are not determined accurately.

3.2.2. AIC499 competition binding

In another validation step, a competitive binding assay was used to determine whether the fragment hits bind to the same site as AIC499, which binds covalently to the active site of PBP3 (Figure 3.38). Therefore, binding of AIC499 to the active site prohibits binding of weak-affinity fragments to the same site. As reported by Vulpetti and coworkers [162], there are basically two expected results after the addition of AIC499 to the NMR sample containing PBP3 and a hit fragment: (i) the peak corresponding to the fragment has a chemical shift and intensity that do not change (Figure 3.21A; ligand likely binds at a site that differs to the binding site of AIC499) or (ii) the peak arising from the fragment has a chemical shift and intensity that matches (closely) the chemical shift and intensity of the signal arising from the free ligand (Figure 3.21B; ligand is displaced by AIC499). In some cases, the presence of AIC499 caused a shift of the chemical shift and peak intensity back towards the free-state chemical shift and intensity (Figure 3.21 C and D). These observations suggest binding of the fragment was partially impeded by conformational change and/or the presence of more than one binding site for the fragment. As shown in Figure 3.21E, addition of AIC499 appears to enhance binding of the fragment, suggesting a conformational change to a fragment binding site enhances affinity. In one case it was possible to observe two peaks after the addition of AIC499 (Figure 3.21F). The chemical shift of one of the peaks matches closely the free ligand state with reduced intensity, whereas the other peak is severely line-broadened and at different chemical shift to that observed when only the protein is present. This result has been observed for ^{19}F-labeled fragment binding to target proteins and arises because only one enantiomer of the fragment, which is a racemate in the sample, interacts with the AIC499:His-tagged *Ec*PBP3ΔTM complex; thus, AIC499 acts as a chiral agent [175].

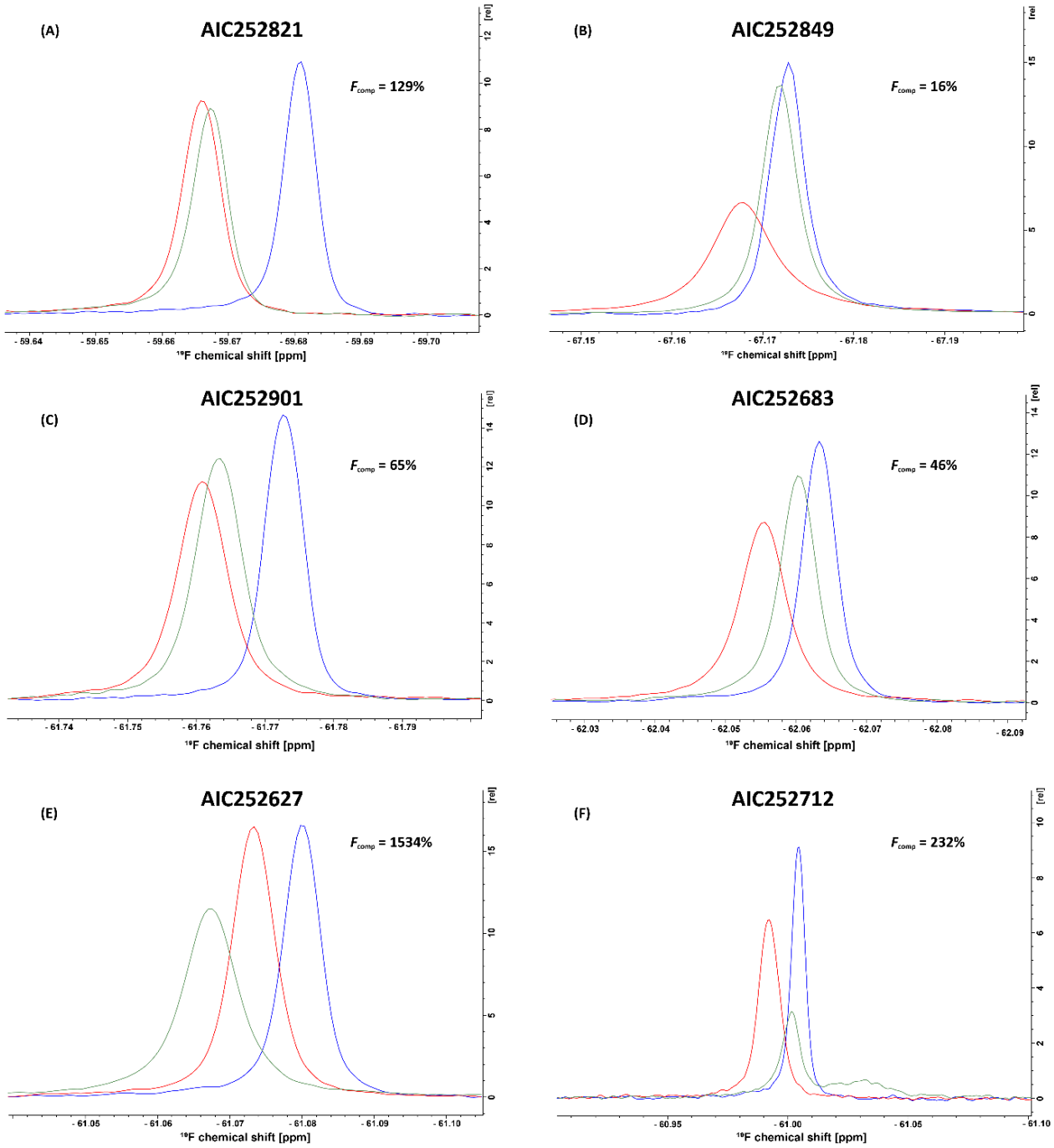


Figure 3.21: Competition binding with 60 µM AIC499 yields different results.

Spectra of 30 µM fragments recorded in the absence of His-tagged *Ec*PBP3ΔTM are shown in blue, whereas spectra recorded in the presence of 30 µM His-tagged *Ec*PBP3ΔTM are shown in red. Spectra of 30 µM fragments recorded in the presence of both 30 µM His-tagged *Ec*PBP3ΔTM and 60 µM AIC499 are shown in green. (**A**) Fragment binding to His-tagged *Ec*PBP3ΔTM is not affected when AIC499 is present, indicating that the fragment does not bind to the same site on His-tagged *Ec*PBP3ΔTM as AIC499 or is affected by conformational change. (**B**) In contrast, AIC499 binding to His-tagged *Ec*PBP3ΔTM hampers the interaction between the fragment and His-tagged *Ec*PBP3ΔTM, suggesting the fragment binds a site that AIC499 also binds or interaction of AIC499 with His-tagged *Ec*PBP3ΔTM causes an allosteric effect that essentially inhibits fragment binding to His-tagged *Ec*PBP3ΔTM. (**C**) and (**D**) The affinity between the protein and fragment is reduced in the presence of AIC499 because of an allosteric effect or fragment binding is impeded by AIC499. (**E**) The binding affinity of the fragment is enhanced in the presence of AIC499. It is also possible that the fragment interacts with bound AIC499, leading to an increase in affinity. (**F**) AIC499 acts as a chiral agent, leading to one enantiomer of the fragment, which is a racemate in the sample, interacting stronger with the AIC499-His-tagged *Ec*PBP3ΔTM complex than the other enantiomer.

Interestingly, the broadened peak does not match the chemical shift of the peak when only the protein is present, indicating that a different higher affinity binding site is available when AIC499 is bound to His-tagged *Ec*PBP3ΔTM. Because no control experiment of fragment and AIC499 without protein being present in the sample was recorded, it cannot be excluded, although unlikely, that AIC499 is interacting directly with fragments in the latter two examples.
Because of the different observations the competition factor F_{comp} was calculated, which is the difference of the intensity between the AIC499 competition and free ligand state divided by the difference of bound and free ligand states. 100% indicates that AIC499 does not affect the interaction between the fragment and the protein, whereas 0% indicates that AIC499 completely replaces the fragment. Values above 100% indicate that the AIC499 enhances binding of the fragment. F_{comp} values are summarized in Table 4.2.

3.2.3. Chemical shift perturbation analysis of fragments identified in the ^{19}F NMR screening

After identifying fragments that interact with His-tagged *Ec*PBP3ΔTM, a chemical shift perturbation assay was performed with fragments found in the second screen to validate their interaction with *Ec*PBP3. For this approach, 2D ^{1}H-^{15}N TROSY-HSQC spectra of the target protein were recorded in the absence and presence of fragments. Shifting peaks indicate an interaction of the fragment with the protein, and can for example, be used to identify the binding pocket (if sequence-specific assignment information is available) or calculate K_d values (when a titration is performed). For small fragments that interact with the protein, it is expected that only a few peaks in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum show a change in chemical shift in the presence of the fragment, when there is only one specific binding site, albeit of relatively low affinity (i.e., µM - mM). Fragments binding the TPd rather than the n-PBd are more likely to inhibit the catalytic activity of the PBP3. Thus, the analysis was performed using the *Ec*TPd* construct. The fragments were prepared in DMSO, which affects buffer conditions and therefore, can cause buffer-induced chemical shift changes. Therefore, reference spectra were recorded with DMSO present at different concentrations. Spectra of samples containing 50 µM *Ec*TPd* and 1 mM fragments with a final DMSO concentration of 1.9% were then recorded. Examples of changes in the 2D ^{1}H-^{15}N TROSY-HSQC caused by the addition of fragment AIC252626 are shown in Figure 3.22.

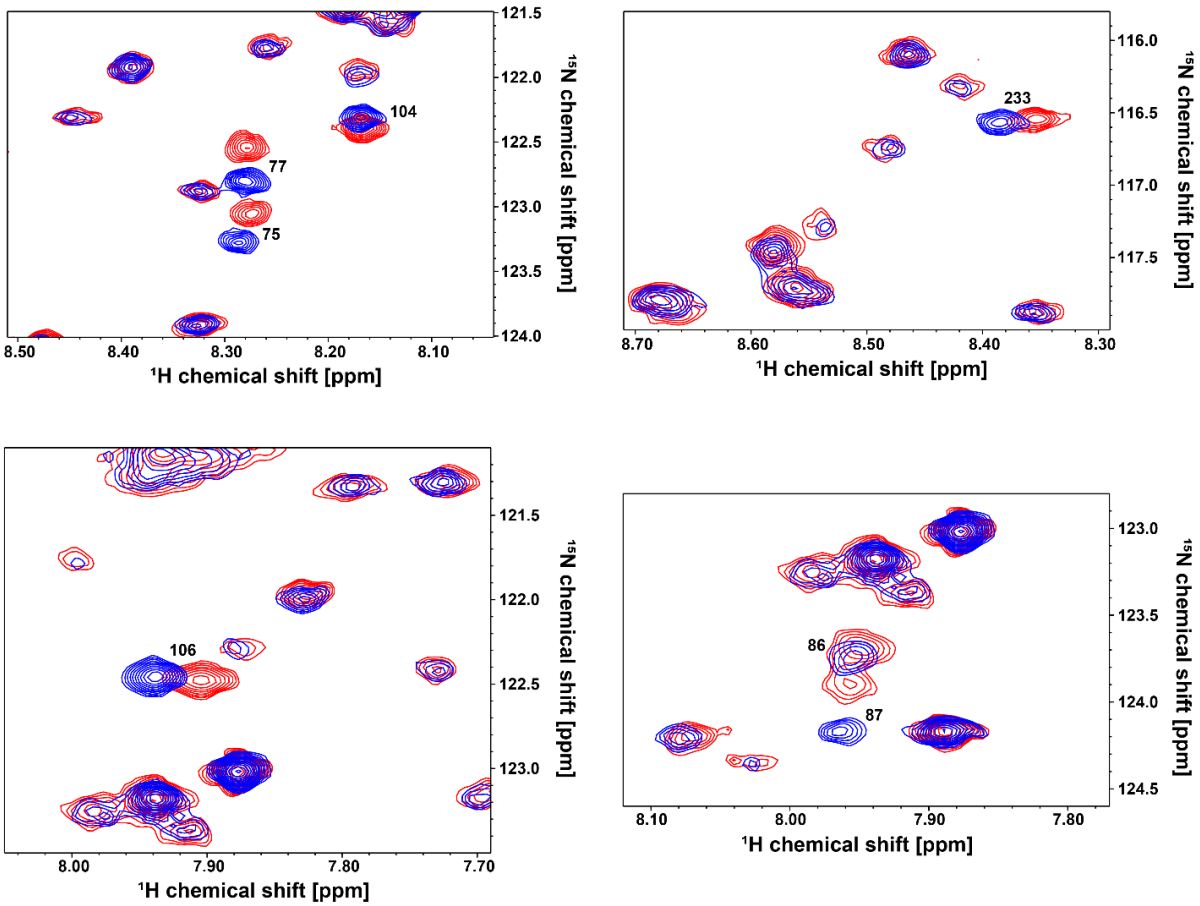


Figure 3.22: Regions of the 2D ^{1}H-^{15}N TROSY-HSQC of 50 µM *Ec*TPd* in the presence and absence of AIC252626.

The regions of the 2D 1H-15N TROSY-HSQC recorded in 1.9% (v/v) DMSO are shown in blue, whereas the spectrum in the presence of 1 mM AIC252626 in 1.9% (v/v) DMSO are shown in red. Peaks that showed a clear change in chemical shift after the addition of the fragment are labeled with their respective peak identifier number.

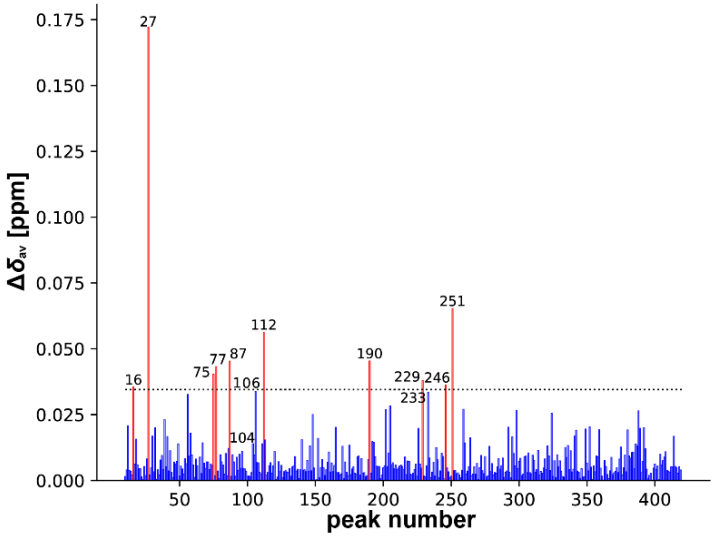


Figure 3.23: Chemical shift perturbation analysis of AIC252626.

$\Delta\delta_{av}$ was calculated following Equation 5. 3σ was used as the threshold (dotted line) to identify peaks that showed the largest changes in chemical shift (red). These peaks represent residues that are likely involved in interaction with the fragment. Numbers above the bars indicate specific peak identifier numbers.

CcpNMR analysis was used to obtain the ^{1}H and ^{15}N chemical shifts of all peaks, which were assigned to specific peak identifier numbers (i.e., 10–419) because assignment information for all peaks was not available. These assignments were propagated to the peaks in the spectra of samples containing the fragments, and the weighted average ^{1}H and ^{15}N chemical shift difference $\Delta\delta_{av}$ was calculated per peak using Equation 5 (Figure 3.23 for an example, and Figure A7 and Figure A8). The standard derivation of all peaks was calculated and 3σ was used as the threshold value to identify peaks that showed the largest changes.

The analysis was not performed as reported in Schumann *et al.* where peaks showing changes above a 3σ threshold were iteratively excluded from calculating σ [177, 178]. In this analysis the corrected σ_0 was used as a cutoff to identify residues specifically interacting with the ligand. Using this approach would have led to many peaks showing changes above the threshold and made it impossible to identify residues specifically interacting with the hit fragments. This analysis was performed for a protein-ligand complex in the low µM range and thus, is not comparable with the interaction between fragments and proteins (expected in the low mM range). For this reason, a fixed threshold value should be considered as well. For example, 0.06 or 0.01 ppm were used as threshold values in Stricht *et al*. and Morrison *et al*., respectively [164, 179]. However, only six out of fourteen fragments had $\Delta\delta_{av}$ values above 0.06 ppm. In contrast, many peaks showed changes with $\Delta\delta_{av}$ values of > 0.01 ppm for every fragment examined (more than half of the peaks in some cases). Thus, 3σ was used as a threshold $\Delta\delta_{av}$ value to identify peaks that showed the largest changes after the addition of fragments.

With this approach, the maximal number of peaks showing changes above the threshold was observed for AIC252728 (32), the minimal number of peaks showing such changes was found for AIC252626 (10), and on average for all fragments examined the number of peaks that had values above the 3σ threshold was 22.

Comparison of profiles like the one presented in Figure 3.23 may reveal common binding sites for fragments. Following this approach, for example, peaks 39 and 295 shifted greater than 3σ above the $\Delta\delta_{av}$ for ten and eleven of the 14 fragments examined, respectively. However, both peaks are weak compared to other peaks and no assignment information was available. Thus, because no structural information is available and the fragments may have different binding pockets, this approach may lead to wrong assumptions and identify peaks that are more affected by changes in the sample conditions. Nevertheless, it was also possible to find peaks that showed changes for some of the fragments where assignment information was available. In this context, peak 37 and 104 were found to be shifted after the addition of AIC252490, AIC2552728, AIC252732, and AIC252901. Interestingly, those peaks were assigned to neighboring amino acids H179 (peak 104) and E180 (peak 37) of *Ec*TPd* and the largest $\Delta\delta_{av}$ was observed for peak 104 ($\Delta\delta_{av}$ > 0.07 ppm) upon addition of the latter three fragments to the protein sample. Furthermore, peak 218 represents residue G178 and addition of AIC252732 and AIC252901 yielded $\Delta\delta_{av}$ values slightly above 3σ.

3.3. Sequence-specific backbone assignments

As described in Section 1.10, sequence-specific backbone assignments of 2D ^{1}H-^{15}N HSQCs can be useful in the process of FBDD. To obtain assignments of the 2D ^{1}H-^{15}N HSQC spectrum of *Ec*PBP3, the first step was to establish a suitable sample.

NMR experiments at different temperatures ranging from 25 to 42 °C revealed that the His-tagged *Ec*PBP3ΔTM was stable up to 37 °C. Temperatures above 37 °C resulted in protein precipitation at the concentrations used for NMR experiments, whereas temperatures below 37 °C had noticeably broader linewidths because of a slower τ_{cor} of the protein. Additionally, stability tests of the protein were performed at AiCuris. After storing the protein at 37 °C the sample was analyzed by SDS-PAGE and the intensities of the bands compared. The protein was stable for several month at 37 °C. Furthermore, circular dichroism (CD) spectroscopy experiments revealed, that His-tagged *Ec*PBP3ΔTM is stable up to 50 °C and has a midpoint of thermal unfolding (T_m) of 59 °C (Figure 3.24). For the abovementioned reasons, all NMR experiments were recorded at 37 °C.

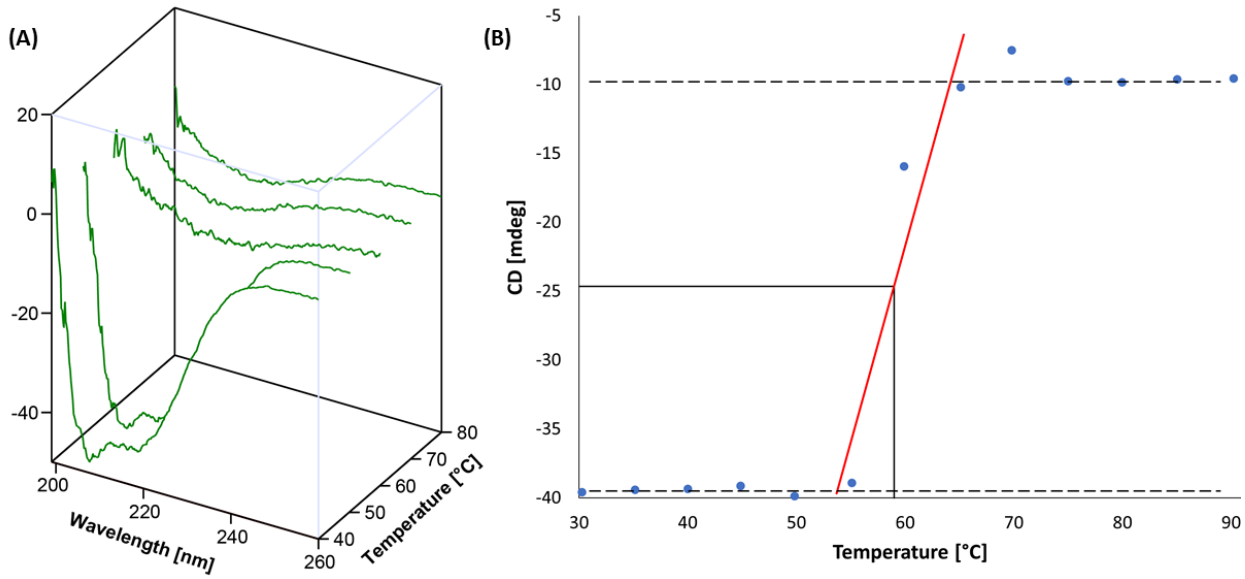


Figure 3.24: Temperature stability tests performed by circular dichroism (CD) spectroscopy.
(**A**) CD spectra recorded at different temperatures were used to monitor the thermal stability of His-tagged *Ec*PBP3ΔTM (i.e., at which temperature the protein starts to unfold). (**B**) Plotting the CD signal at 222 nm allows to roughly estimate T_m = 59 °C. For this, T_m can graphically be determined by matching the slope (red) with the average between upper and lower plateaus (dotted lines).

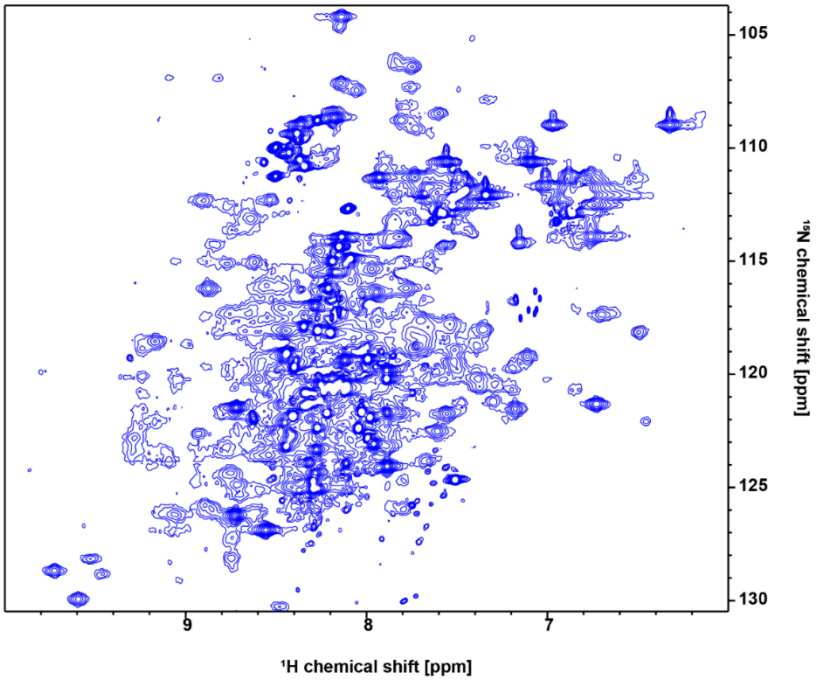


Figure 3.25: 2D ^{1}H-^{15}N HSQC of the 61 kDa His-tagged *Ec*PBP3ΔTM.
The protein concentration was 238 µM and is ^{15}N-labeled. The experiment was recorded with 128 scans at 37 °C and at a ^{1}H field strength of 700 MHz. Weak, sharp peaks within the ^{1}H random coil region (between 8.00 and 8.43 ppm; [167]) indicate the presence of degradation products or impurities.

Peaks in the 2D ^{1}H-^{15}N HSQC (Figure 3.25) showed significant overlap and broad average linewidths at half height of 167 and 76 Hz in ^{1}H and ^{15}N dimensions, respectively, which were measured (Table 3.4), in the 2D ^{1}H-^{15}N HSQC spectrum of His-tagged *Ec*PBP3ΔTM. The peaks were picked to match the number of expected peaks; although peak picking in the central region between 8.00 and 8.43 ppm was challenging. Moreover, the presence of a large number of signals in this region, which is the typical ^{1}H chemical shift range for unstructured proteins or segments of proteins, indicated that a sizeable proportion of the protein was disordered or loosely structured [167]. This observed overlap in the 2D ^{1}H-^{15}N HSQC indicates that this construct was not suitable for recording 3D heteronuclear NMR experiments for obtaining backbone assignments or chemical shift perturbation analysis. Additionally, weaker peaks around 8 ppm (^{1}H) and 130 ppm (^{15}N) indicate the presence of protein fragments caused by sample degradation. A potential major issue that exacerbated the poor quality of the spectrum was the presence of the His-tag, which is probably highly flexible and leads to intense signals in the spectrum at random coil chemical shifts. Additionally, the His-tag may adopt a conformation that leads to an increase in τ_{cor}. Thus, a construct that enabled cleavage of the His-Tag was designed. Moreover, the 2D ^{1}H-^{15}N TROSY-HSQC experiment was used to take advantage of the benefits of TROSY (Section 1.10). The 2D ^{1}H-^{15}N TROSY-HSQC spectrum of *Ec*PBP3ΔTM is shown in Figure 3.26.

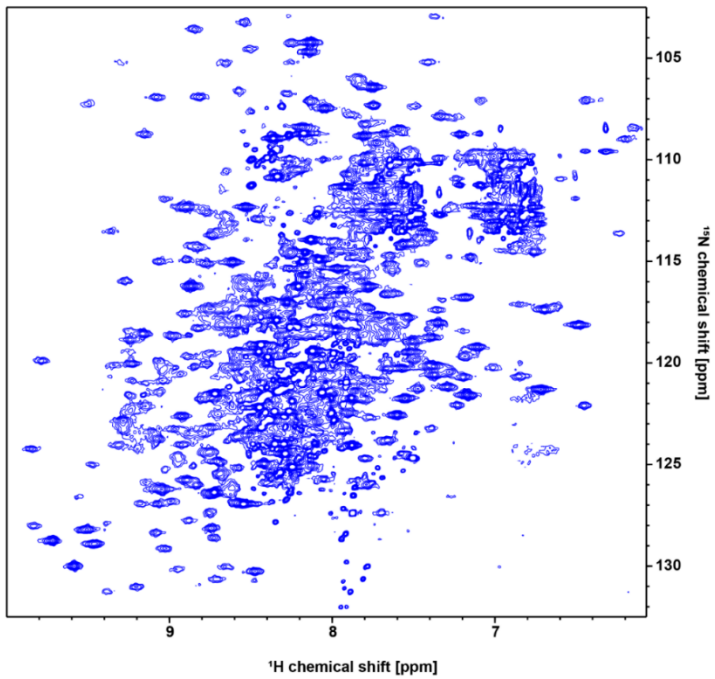


Figure 3.26: 2D ^{1}H-^{15}N TROSY-HSQC spectrum of 60 kDa *Ec*PBP3ΔTM.
The ^{15}N-labelled protein concentration was 174 µM. The experiment was recorded with 208 scans at a temperature of 37 °C and a ^{1}H field strength of 700 MHz.

Clearly an improvement in the 2D ^{1}H-^{15}N HSQC spectrum is observed by removing the His-tag and using the TROSY element to detect the narrowest component of the multiplet (Figure 1.20). The observed average linewidth at half height decreased to 61.5 Hz and 33.8 Hz in the ^{1}H and ^{15}N dimension, respectively (Table 3.4). Signals from the His-tag are no longer observed in the central region of the spectrum and weaker signals that were absent in Figure 3.26 are visible and likely belong to regions of structured parts of the protein, such as the peak at approximately 9.5 ppm (^{1}H) / 108 ppm (^{15}N). The degradation, which was observed in the previous sample, was still present in the sample without the His-Tag. However, the spectrum still has an overly crowded central region. Therefore, to narrow the peak linewidth further, a [^{2}H,^{15}N]-labeled *Ec*PBP3ΔTM sample without the His-Tag was produced to take advantage of the benefits of deuteration, as described in the Section 1.10.

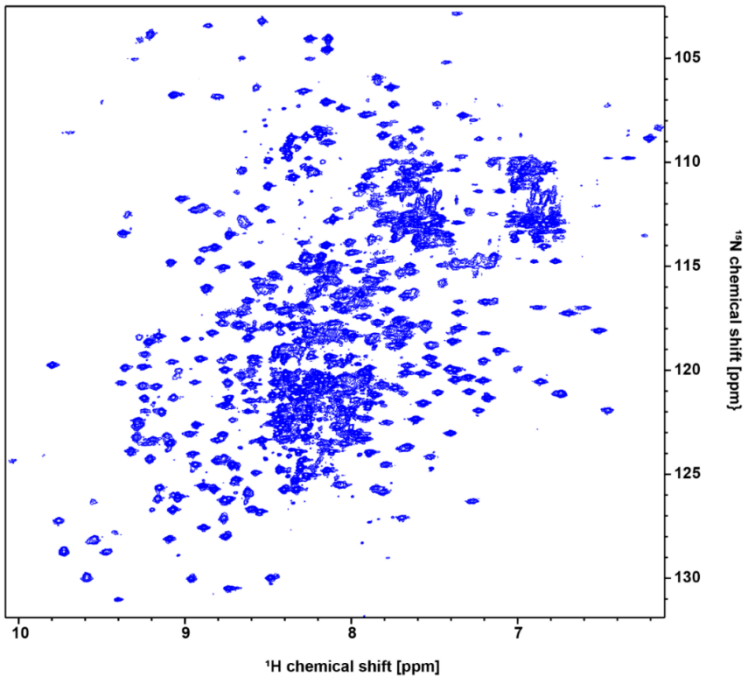


Figure 3.27: 2D ^{1}H-^{15}N TROSY-HSQC spectrum of [^{2}H,^{15}N]-labeled 60 kDa *Ec*PBP3ΔTM.
The protein does not contain the His-Tag and the sample is 111 µM. The experiment was recorded with 280 scans at 37 °C and a ^{1}H field strength of 700 MHz.

Deuteration yielded a further decrease in linewidths of signals in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum (Figure 3.27). The average linewidth at half peak height narrowed to 26.9 Hz and 14.4 Hz in the ^{1}H and ^{15}N dimension, respectively (Table 3.4). As a result, more peaks were resolved when compared with Figure 3.26. Nevertheless, significant resonance overlap remained in the central region of the spectrum, presumably because of the n-PBd, which was hypothesized to be highly flexible based on high B-factors and the partial absence of electron density in the crystal structure (PDB entry 4BJP; [47]). Therefore, a truncated version of the *Ec*PBP3 was constructed that contained only the structural parts of the TPd (*Ec*TPd*), resulting in a globular, 44 kDa protein (Figure 1.5). Compared to a previously published structure, the linker subdomain was kept as part of the TPd because this subdomain is required to stabilize other parts of the TPd. *Ec*TPd* was prepared as a non-His-Tag [^{2}H, ^{15}N]-labeled protein and the 2D ^{1}H-^{15}N TROSY-HSQC is shown in Figure 3.28. The 2D ^{1}H-^{15}N TROSY-HSQC spectrum shows a well disperse set of signals with linewidths of 20.3 Hz and 9.7 Hz in the ^{1}H and ^{15}N dimension, respectively (Table 3.4). Clearly, removing the n-PBd yielded a reduced set of peaks in the central region of the *Ec*TPd* 2D ^{1}H-^{15}N TROSY-HSQC spectrum, supporting the abovementioned hypothesis that the n-PBd of *Ec*PBP3ΔTM is dynamic and loosely structured. Moreover, no indication of protein degradation was observed, suggesting that *Ec*TPd* is stable under the experimental conditions used. An overlay between 2D ^{1}H-^{15}N TROSY-HSQC spectra of uniformly [^{2}H, ^{13}C, ^{15}N]-labeled *Ec*TPd* and *Ec*PBP3ΔTM revealed that most peaks in the spectrum of *Ec*TPd* matched peaks in the corresponding spectrum of *Ec*PBP3ΔTM (Figure 3.29), indicating that the TPd and linker subdomain in both proteins adopt a similar 3D fold. The *Ec*TPd* construct was used to record 3D heteronuclear NMR experiments to obtain backbone assignments.

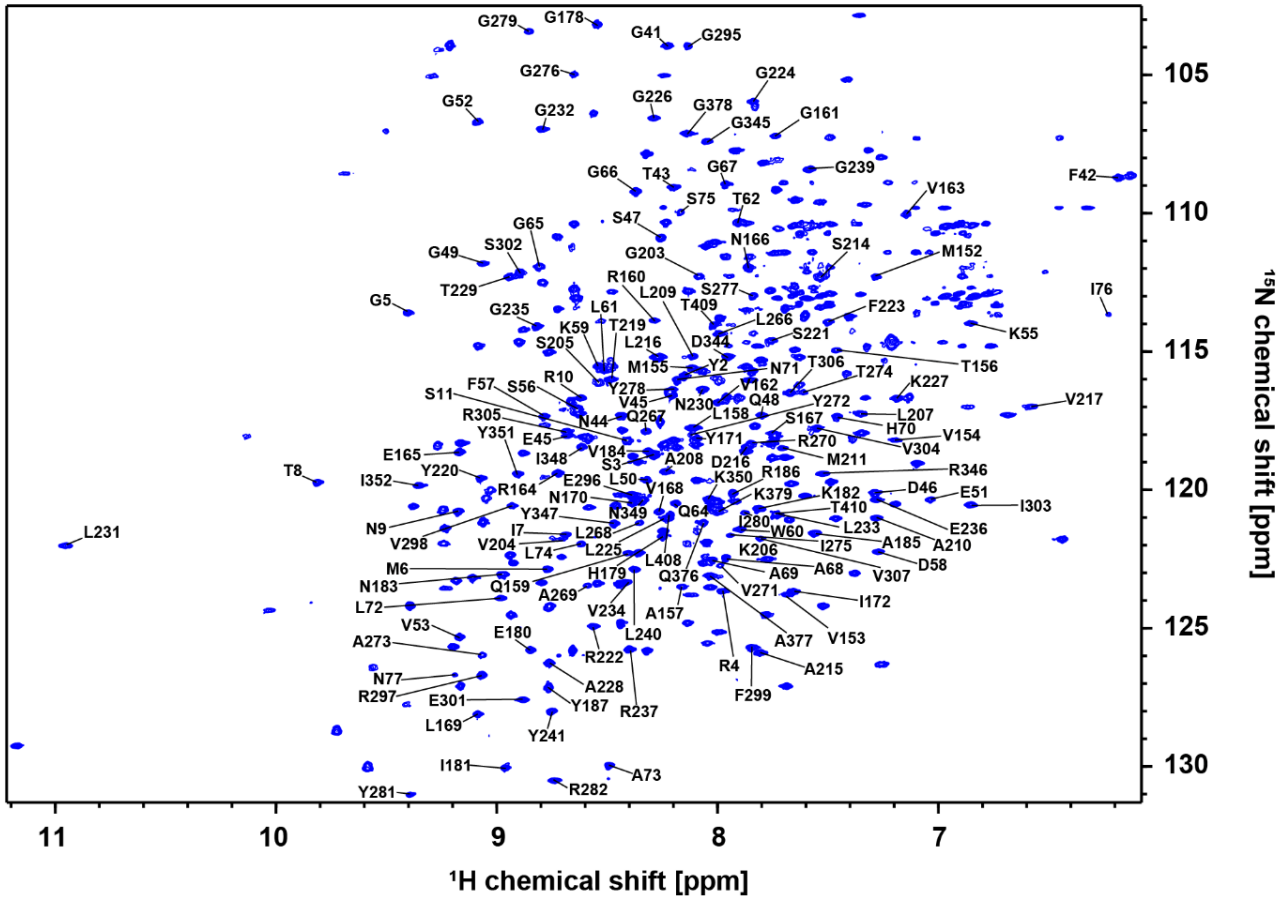


Figure 3.28: 2D ^{1}H-^{15}N TROSY-HSQC of [^{2}H, ^{13}C, ^{15}N]-labeled *Ec*TPd*.

The protein concentration was 60 µM and the experiment was recorded with 64 scans at 37 °C and a ^{1}H field strength of 700 MHz. Due to technical reasons, peaks in the 2D ^{1}H-^{15}N TROSY-HSQC are assigned to the amino acid sequence of the *Ec*TPd* construct (sequence in Table 2.7).

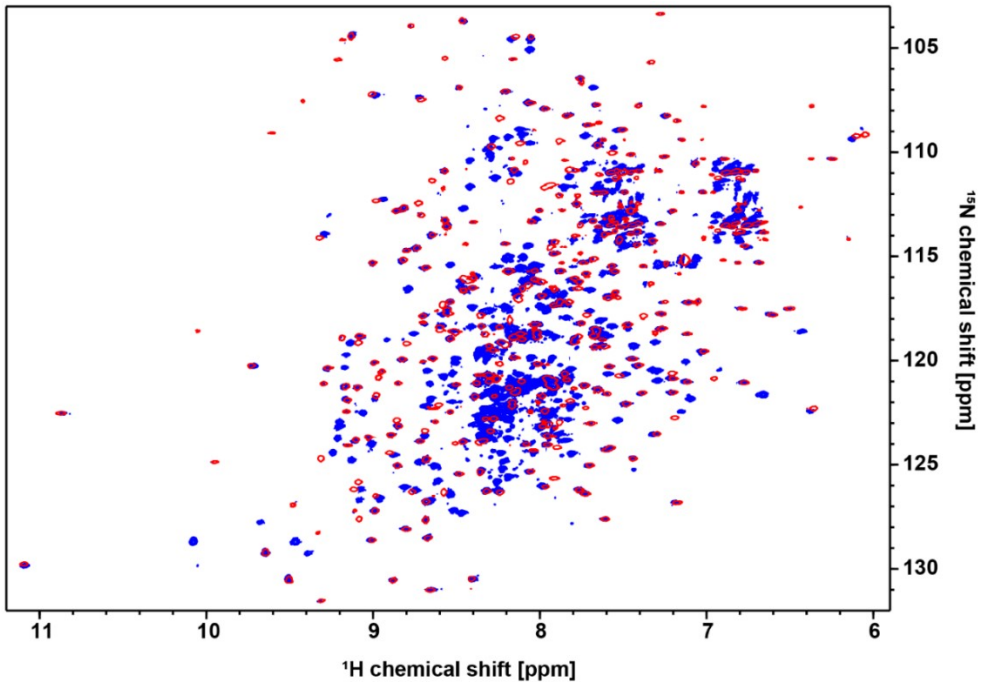


Figure 3.29: Investigation of structural similarities between *Ec*PBP3ΔTM and *Ec*TPd* by NMR.

The overlay of 2D ^{1}H-^{15}N TROSY-HSQC NMR spectra of uniformly [^{2}H, ^{13}C, ^{15}N]-labeled *Ec*PBP3ΔTM (111 µM; blue) and *Ec*TPd* (60 µM; red) demonstrates the structural similarity of both proteins in solution. Both spectra were recorded at 700 MHz and 37 °C.

As described previously, the averaged linewidths at half peak height were determined to estimate the quality of the spectra. Clearly, the linewidths decreased at every optimization step (Table 3.4). In total, starting with the ^{15}N-labeled His-tagged *Ec*PBP3ΔTM sample 2D ^{1}H-^{15}N HSQC the linewidths were improved by factor of approximately 8 when examining the ([^{2}H, ^{13}C, ^{15}N]-labeled *Ec*TPd* sample 2D ^{1}H-^{15}N TROSY-HSQC.

Table 3.4: Average linewidths at half height for peaks in 2D ^{1}H-^{15}N HSQC and 2D ^{1}H-^{15}N TROSY-HSQC spectra of the various *Ec*PBP3 constructs.

The peaks were picked in CcpNMR analysis [149]. Expected peak numbers were calculated using the protein sequence (number of all amino acids plus two times number of Asn and Gln plus number of Trp minus number of Pro minus one for the *N*-terminal residue) and are shown in brackets.

Protein	Spectrum	Number of backbone peaks	^{1}H Line-width [Hz]	^{15}N Line-width [Hz]
His-tagged *Ec*PBP3ΔTM (^{15}N-labeled)	2D ^{1}H-^{15}N HSQC (Figure 3.25)	604 [619]	166.91	75.64
*Ec*PBP3ΔTM (^{15}N-labeled)	2D ^{1}H-^{15}N TROSY-HSQC (Figure 3.26)	758 [603]	61.53	33.80
*Ec*PBP3ΔTM ([^{2}H, ^{15}N]-labeled)	2D ^{1}H-^{15}N TROSY-HSQC (Figure 3.27)	635 [603]	26.87	14.41
*Ec*TPd* ([^{2}H, ^{13}C, ^{15}N]-labeled)	2D ^{1}H-^{15}N TROSY-HSQC (Figure 3.28)	441 [446]	20.28	9.73

After ensuring that it is possible to record good quality 2D ^{1}H-^{15}N HSQC spectra, the *Ec*TPd* construct was used to record 3D heteronuclear NMR experiments to assign the 2D ^{1}H-^{15}N TROSY-HSQC. The 3D spectra HNCO, and HN(CA)CO were recorded to obtain carbonyl ^{13}C' chemical shifts, whereas the HNCA, HN(CO)CA, HNCACB and HN(CO)CACB spectra provided $^{13}C_{\alpha}$, and $^{13}C_{\beta}$ chemical shift information. The TROSY principle and NUS were used for these 3D experiments. In some cases, the quality of the spectra was increased by adding several experiments together to increase the signal-to-noise ratio. This approach was beneficial when compared with simply increasing the number of scans because sample homogeneity was observed to change slowly over time (e.g., air bubble in the sample). To avoid loss of field homogeneity the sample was checked for potential air bubbles frequently and re-shimmed. As described in Section 1.10, the “sequential walk” was used to assign sequence-specifically 40% of the peaks in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum of *Ec*TPd* (Table A1; Figure 3.28 and 3.30).

The most challenging task in the process of the assignment was the identification of good starting points. Often, it was possible to find a “sequential walk” (Figure 3.30) for a couple of peak sets but unambiguous assignment of peaks representing $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ nuclei was not possible to sequence-specifically assign these. Additionally, weak peak intensities and missing peaks hampered the sequence-specific assignment process. For example, as shown in Figure 3.30B the absence of an expected peak for the ^{13}C' chemical shift of residue T274 in the 3D HNCACO experiment is indicated by a dotted line. The expected peak cannot be observed, even though the 3D HNCACO was recorded three times using TROSY and NUS and the three spectra added together during processing. Additionally, there are 24 prolines in the 410 amino acid sequence of the *Ec*TPd* construct (approximately 6% of all amino acids), which contributed to the challenges associated with making sequence-specific assignments. These two issues hampered the identification of starting points in the spectra

TALOS-N and the sequence-specifically assigned $^{1}H_N$, ^{15}N, $^{13}C'$, $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts were used to calculate the secondary structure elements [157]. The predicted secondary structures based on the assignment match closely the secondary structure of the protein taken from the crystal structure (PDB entry 7ONO; [48]), thus supporting that the assignments made are correct.

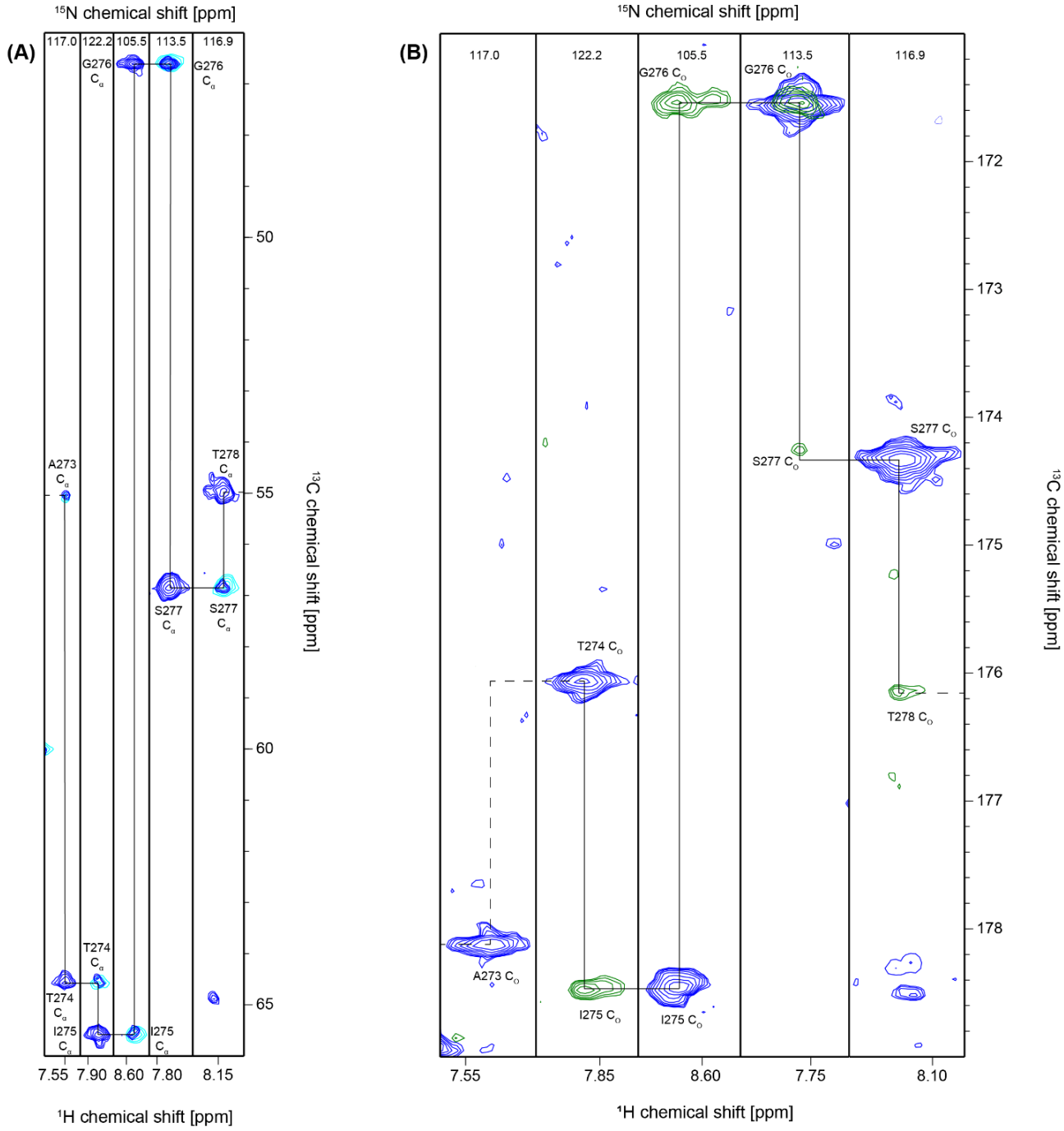


Figure 3.30: An example of a "sequential walk" to assign $^{13}C\alpha$ and $^{13}C'$ chemical shifts of amino acids A273 to T278 of *Ec*TPd*.

(**A**) The 3D HNCA (blue) and 3D HNCOCA (cyan) provided a "sequential walk" for $^{13}C\alpha$ chemical shifts. Note that the HNCOCACB was also used to determine the $^{13}C\beta$ chemical shift and type of amino acid but is not displayed. (**B**) The 3D HNCO (blue) and 3D HNCACO (green) provide the "sequential walk" using $^{13}C'$ chemical shifts. The lines indicate the connection between the "sequential walk", whereas dotted lines indicate the theoretical connection that was not observed because of missing peaks. Because of the unambiguous 3D HNCA / 3D HNCOCA combination the assignment can still be made.

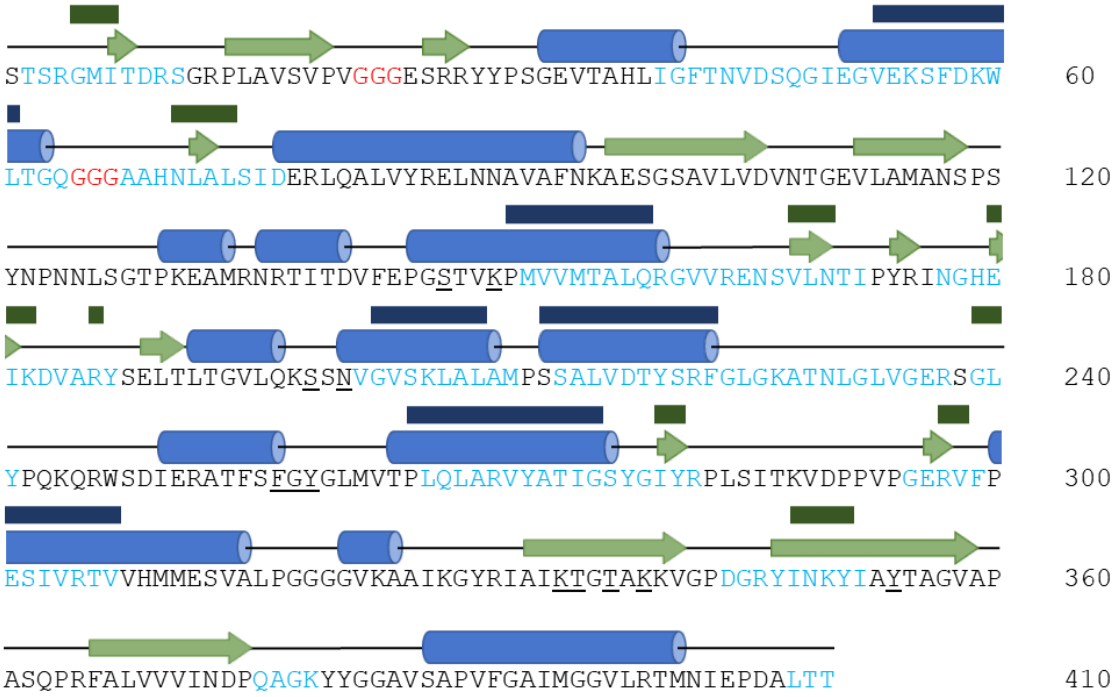


Figure 3.31: *Ec*TPd* sequence to be sequence-specifically assigned to peaks in the 2D ^{1}H-^{15}N TROSY-HSQC.

The GGG linkers that were introduced to remove parts of the n-PBd are shown in red, whereas assigned amino acids are highlighted in light blue and amino acids in the active site are underlined. Blue cylinders and green arrows above the sequence indicate α-helices and β-sheets found in PDB entry 7ONO, respectively [48]. Additionally, secondary structure elements calculated by TALOS-N based on backbone chemical shift assignment information are indicated with dark blue and dark green bars above the secondary structure elements for α-helices and β-sheets, respectively [157]. In total, 40% (318) of the 795 expected backbone hydrogen or nitrogen chemical shifts in the 2D ^{1}H-^{15}N TROSY-HSQC were assigned (Table A1).

3.4. Cluster J compounds investigation

After the identification of the interaction between PBP3 and Cluster J compounds, crystal soaking experiment were performed with the *Pa*PBP3ΔTM and compound 34. Unfortunately, it was not possible to find electron density for the compound and the resulting data sets were published as *de facto* apo-models (Section 3.5.2.1). For this reason, perturbation shift NMR experiments were performed using 25 µM ^{15}N-labeled *Ec*TPd* and 50 µM of the Cluster J compounds to show a specific interaction between PBP3 from *E. coli* and obtain kinetic information about the binding site of the Cluster J compounds. Because K_d values for Cluster J compounds were also available for *Ec*PBP3ΔTM, the already established *Ec*TPd* sample was used for these experiments. The DMSO concentration was adjusted to be 0.475%. In these experiments, which were recorded for 24 h, respectively, changes in the chemical shift of peaks when compared with a sample containing 0.475% DMSO were only small, and large changes in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum were not observed (Figure 3.32). Additionally, sample precipitation was observed during NMR experiments. For investigating the emergence of the precipitant, samples with the protein and DMSO were prepared, and 2D ^{1}H-^{15}N HSQC experiments recorded over a time of one week. No changes were observed, indicating that DMSO does not cause precipitation. In another experiment, the Cluster J compounds were mixed in the sample buffer and stored at 37 °C. Like the protein-only samples, no visual precipitate was observed after incubating for several days. Varying the DMSO concentration up to 95% DMSO did not have any effect.

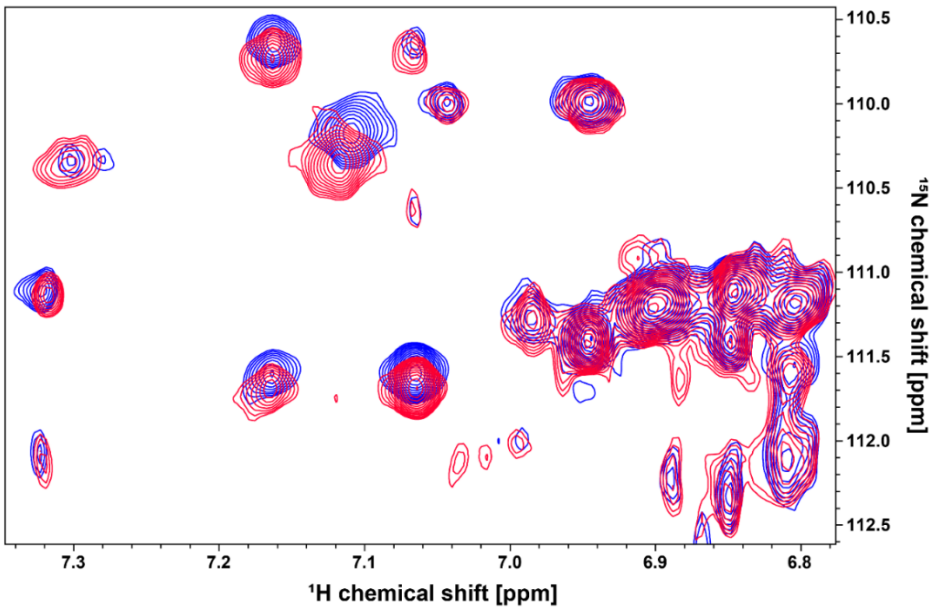


Figure 3.32: Region of the *Ec*TPd* 2D ^{1}H-^{15}N TROSY-HSQC with and without 200 µM Cluster J compound 34.

The spectrum without the compound (blue) shows minor changes after the addition of 200 µM Cluster J (red). A compound concentration of 200 µM was used in initial tests, but because of the poor solubility the concentrations for other Cluster J compounds were reduced to 50 µM in further experiments.

To determine what the precipitate represented, additional 1D ^{1}H NMR experiments of Cluster J compound 36 dissolved in DMSO were recorded. In series of experiments over one day using different Cluster J and DMSO concentrations, spectra of the compound, dissolved in the protein buffer were compared. No change was observed. In accordance with the absence of precipitate, it was concluded that the Cluster J compounds are soluble and stable at concentrations up to 100 µM with DMSO concentrations of 0.475%.

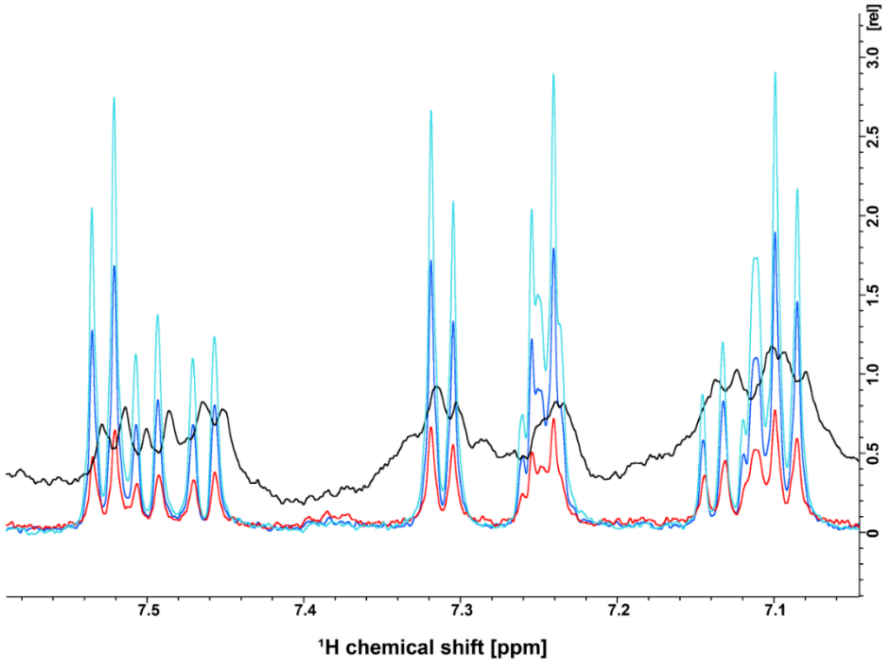


Figure 3.33: Aromatic region of a 1D ^{1}H NMR spectrum of 100 µM Cluster J (compound 36) without and with different concentrations of *Ec*TPd*.

The spectrum without protein is shown in cyan, whereas blue, red and black represent 1 µM, 2 µM and 10 µM *Ec*TPd*, respectively.

Following from the solubility and stability tests, the *Ec*TPd* was added to the sample and the spectra were recorded. After the addition of 1 µM protein the signal intensities were lowered by approximately 40% and by approximately 80% when adding 2 µM protein (Figure 3.33). After adding 10 µM *Ec*TPd* it is possible to see the signal derived by the protein, whereas peaks representing the compound were severely line-broadened.

The peak intensities are reduced because the bound Cluster J compound adopts the R_2 of the protein and possible chemical exchange, which can be assumed to be fast-to-intermediate on the ^{1}H NMR timescale. However, because precipitate occurred in the samples during acquisition, it was concluded that the Cluster J compound initially bound the *Ec*TPd* specifically but is also able to bind to the protein non-specifically through hydrophobic interactions, causing the protein-compound complex to precipitate.

Thus, with the performed 1D ^{1}H NMR experiment, it was not possible to define the specific interactions between Cluster J compounds and *Ec*PBP3, and therefore, fluorescence experiments were recorded varying the concentration Cluster J compound, while the protein concentration was maintained at 4 µM. The experiments were only performed with Cluster J compound 35. In the resulting spectra (Figure 3.34), Raman scattering was observed with a λ_{ex} λ_{em} difference of approximately 30 nm and the peak resulting from protein fluorescence was determined at 340 nm.

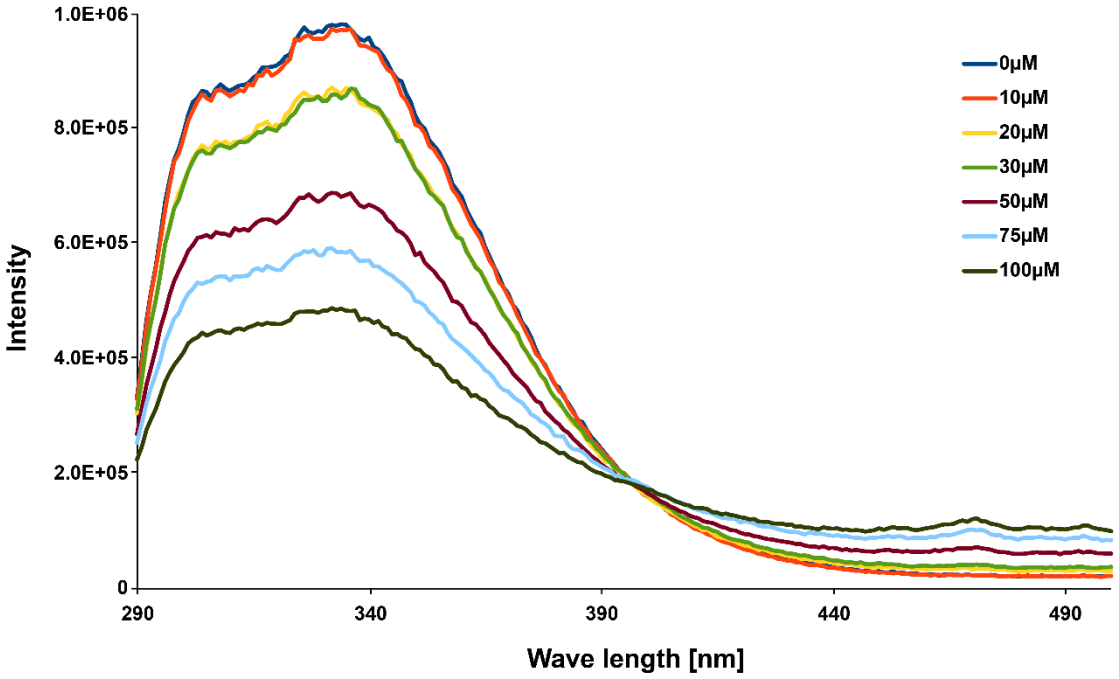


Figure 3.34: Fluorescence experiment with *Ec*TPd* and Cluster J compound 35.
Cluster J concentrations were varied from 0 (dark blue), 10 (red), 20 (yellow), 30 (green), 50 (brown), 75 (light blue), and 100 µM (dark green), whereas *Ec*TPd* was kept at 4 µM. The samples were excited with λ_{ex} = 270 nm.

The integrals between 325 and 350 nm were calculated and plotted against the Cluster J concentrations to evaluate binding between the compound and *Ec*TPd* (Figure 3.35).

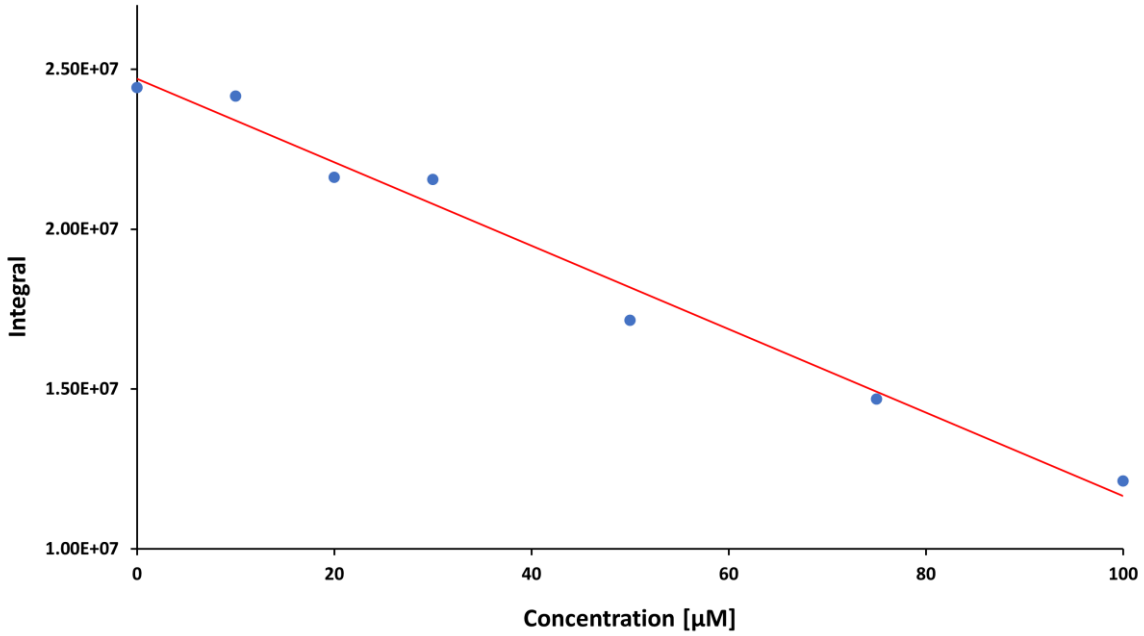


Figure 3.35: Plot of integrals derived from the fluorescence signals against Cluster J concentrations.
The integrals were determined between 325 and 350 nm. It was possible to fit a linear trend line (red) with R^2 = 0.98 to the data.

Because of the linear regression fit, the fluorescence data do not suggest a specific interaction between compound 35 and *Ec*TPd*. Unlike the previous reported NMR data, no aggregation was observed, but probably because of the low protein concentration used in the experiments, i.e., no visible aggregation was observed in the sample.

Overall, the NMR and fluorescence data indicate that interaction of Cluster J compounds with *Ec*TPd* cause aggregation of the protein-ligand complex. To avoid aggregation in the sample, SPR was used to show a specific interaction. The protein was coated on a CM5 chip to yield a protein amount of 11.820 RU. For every analyte, two concentration series between 3.25 µM and 200 µM were recorded. The resulting graphs after an internal solvent correction showed a specific interaction enabled determination of the K_d values, which were calculated to be between 2.73 µM and 470 µM and were published in *A. López-Pérez et al.* (Figure 3.36 and Table 3.5; [85]). In the case of compound 36, the sensorgrams did not provide a sufficient quality to determine the K_d value.

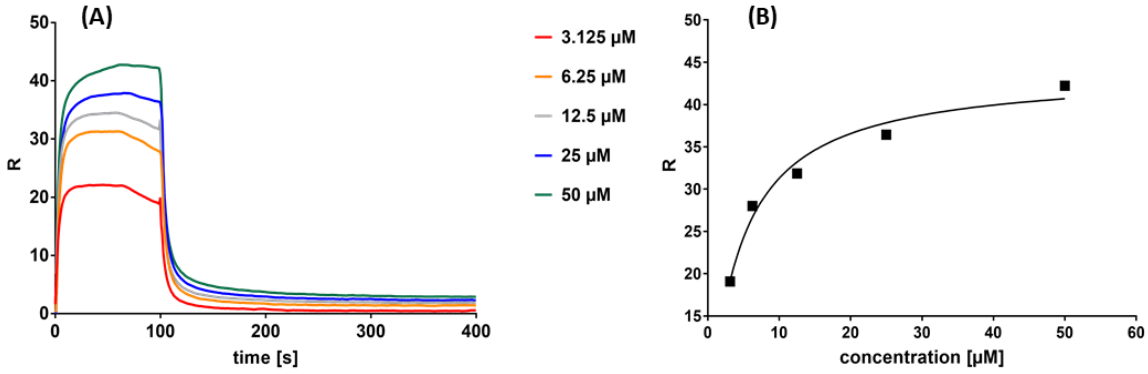


Figure 3.36: Exemplary SPR results of Cluster J compound 35.
The sensorgrams (**A**) were plotted against the Cluster J concentrations (**B**) to determine the K_d value, which is summed up with other Cluster J compounds in Table 3.5.

Table 3.5: Determined kinetic values of Cluster J compounds.

An optimized fluorescence-based assay and a competition binding assay using bocillin FL were used to determine the IC_{50} values with *Pa*PBP3 and *Ec*PBP3, measured by Dr. Arancha López-Pérez and published in *A. López-Pérez et al.* [85]. The K_d values for *Pa*PBP3 were determined using SPR.

Compound	IC_{50} (*Pa*PBP3)		IC_{50} (*Ec*PBP3)		K_d (*Pa*PBP3)
	S2d [µM]	*Bocillin FL [µM]*	*S2d [µM]*	*Bocillin FL [µM]*	*SPR [µM]*
34	14 ± 9	4 ± 9	22	17	6.44
35	3 ± 1	0.6 ± 2	6	5	9.82
36	>100	>100	>100	>100	n.a.
37	16 ± 8	17 ± 3	>100	>100	454
38	100 ± 11	117 ± 4	117	42	62.7
40	8 ± 4	4 ± 2	23	15	22.9

3.5. X-ray crystallography provides structural insights on PBP3 structures from *E. coli* and *P. aeruginosa* and the AIC499 binding mode

X-ray data obtained during this work were published in the following paper: Interaction Mode of the Novel Monobactam AIC499 Targeting Penicillin Binding Protein 3 of Gram-Negative Bacteria; [48]. The complete research paper can be found online at https://doi.org/10.3390/biom11071057. See Section 6 for more information about contributions.

3.5.1. Structures of *E. coli* PBP3

3.5.1.1. *E. coli* PBP3 apo protein

Apo-*Ec*PBP3ΔTM was readily crystallized, but crystals yielded weak and anisotropic diffraction data extending to a maximum resolution of 4 Å, which prevented determination of its 3D structure. It was reasoned that the *N*-terminal part of the protein comprising the n-PBd may prevent formation of a highly ordered crystal lattice because of its flexibility. This view is supported by a crystal structure published previously (PDB entry 4BJP) that features very high B-factors and partial absence of electron density, particularly in the head and anchor subdomains (nomenclature according to [180]; also refer to Figure 1.5 and Figure 3.39). For this reason, a truncated version of *Ec*PBP3ΔTM (termed *Ec*TPd*) was designed that included the catalytical TPd and replaced other segments with tri-glycine linkers. In contrast to a previously published *Ec*TPd structure (PDB entry 6HZQ; [49]), the truncation was defined strictly based on tertiary fold, ensuring that the closely apposed linker subdomain (S68-V88, E164-Q203) was included.

To support the crystallography work spectra of uniformly [^{2}H, ^{13}C, ^{15}N]-labeled *Ec*PBP3ΔTM and *Ec*TPd* were recorded and overlayed (Figure 3.29), indicating that the TPd and linker subdomain in the two proteins adopt a similar 3D fold. After showing that the solution structure of the *Ec*PBP3 TPd is largely unaffected by the truncation, the *Ec*TPd* protein was crystallized. Indeed, these crystals were found to exhibit considerably improved diffraction quality (including

reduced anisotropy) when compared with those obtained using *Ec*PBP3ΔTM, with useful data extending to a resolution of approximately 2.3 Å. The final *Ec*TPd* model, featuring hexagonal space group P 6_2 2 2 with one molecule per asymmetric unit, contains residues T69-T569 of *Ec*PBP3 (Table 3.6). Additionally, soaking experiments with 250 mM bromo- and iodo-pyrazole, and Ta_6Br_{12} were performed but lead to poor diffraction.

Table 3.6: Data collection and refinement statistics of *Ec*PBP3 structures.

Values in parentheses refer to the highest-resolution shell. [a] Conventional definition using spherical shells. [b] Calculated with respect to an ellipsoidal portion of reciprocal space fitted to the cut-off surface, as defined in STARANISO.

	Ec*TPd	***Ec*TPd*:AIC499**	***Ec*PBP3ΔTM:AIC499**
PDB entry	7ONO	7ONN	7ONW
Data collection			
Beamline	DESY P11	DESY P11	ESRF ID23
Wavelength [Å]	1.0332	1.0332	0.97242
Space group	P 6_2 2 2	P 6_2 2 2	P 6_4 2 2
Cell dimensions			
a, b, c [Å]	109.3, 109.3, 143.2	110.4, 110.4, 142.1	106.9, 106.9, 285.8
α, β, γ [°]	90, 90, 120	90, 90, 120	90, 90, 120
Resolution range [Å]	47.74–2.30 (2.60–2.30)	47.80–1.92 (2.22–1.92)	48.64–2.70 (3.03–2.70)
$CC_{1/2}$ [%]	99.9 (84.0)	99.9 (78.3)	99.9 (79.3)
R_{meas} [%]	10.3 (145.6)	6.6 (160.2)	10.1 (121.1)
I/σ	20.7 (2.4)	24.2 (2.3)	15.4 (2.0)
Completeness [%][a]	40.8 (6.7)	45.7 (6.5)	51.1 (9.0)
Ellipsoidal Completeness [%][b]	93.9 (74.6)	95.0 (74.3)	92.2 (76.1)
Refinement			
Resolution range [Å]	47.7–2.3	47.8–1.9	48.6–2.7
No. unique reflections	9457	18221	13994
No. protein atoms	3031	3022	3516
No. ligand atoms	5	135	71
No. water molecules	16	60	6
R_{work} [%]	23.00	22.33	23.35
R_{free} [%]	28.20	26.93	27.07
RMSD			
Bond lengths [Å]	0.001	0.005	0.002
Bond angles [°]	0.403	1.031	0.672
Mean B factor [Å^2]	50.49	41.23	64.13
Ramachandran plot			
Favored [%]	95.52	93.98	95.01
Allowed [%]	4.48	6.02	4.99
Outliers [%]	0	0	0
Z-score	–2.45	–1.72	–2.82

Comparison of the *Ec*TPd* structure (Figure 3.37B, dark blue) with the *Ec*PBP3ΔTM structure published previously (PDB entry 4BJP [47]; Figure 3.37A, gold) confirms that the 3D fold is mostly identical; 324 common C_α atoms superimpose with an overall RMS distance of 0.39 Å. As expected, notable differences are observed close to the truncation sites, i.e., the V88–G_3–

E164 and Q203–G_3–A228 regions, which are in contact with the β5-α11 and α9-α10 segments, respectively, of symmetry equivalent molecules (designation of PBP3 secondary structure elements according to [180]). These lattice interactions, in turn, cause the α9-α10 backbone to shift outwards by up to 2 Å; the β5-α11 region has not been traced in the *Ec*PBP3ΔTM structure, but electron density indicates a relative displacement on the order of 5 Å. Additional differences in the vicinity of the catalytic center concern the β2b-β2c-β2d region as well as the β3-β4 hairpin. While the former is probably affected by a lattice contact chiefly mediated by the adjacent α5-α6 linker and the segment preceding helix α8, contacting their respective equivalents in a symmetry mate, the latter appears to be similarly influenced by the cognate region in a neighboring copy, with distances suggesting a repulsive electrostatic interaction. Since all of these moderate alterations are far from the truncation sites and can be readily explained by differences in lattice contacts, we conclude that the structure of *Ec*TPd* is largely representative of the *Ec*PBP3ΔTM protein, in agreement with our observations using NMR spectroscopy.

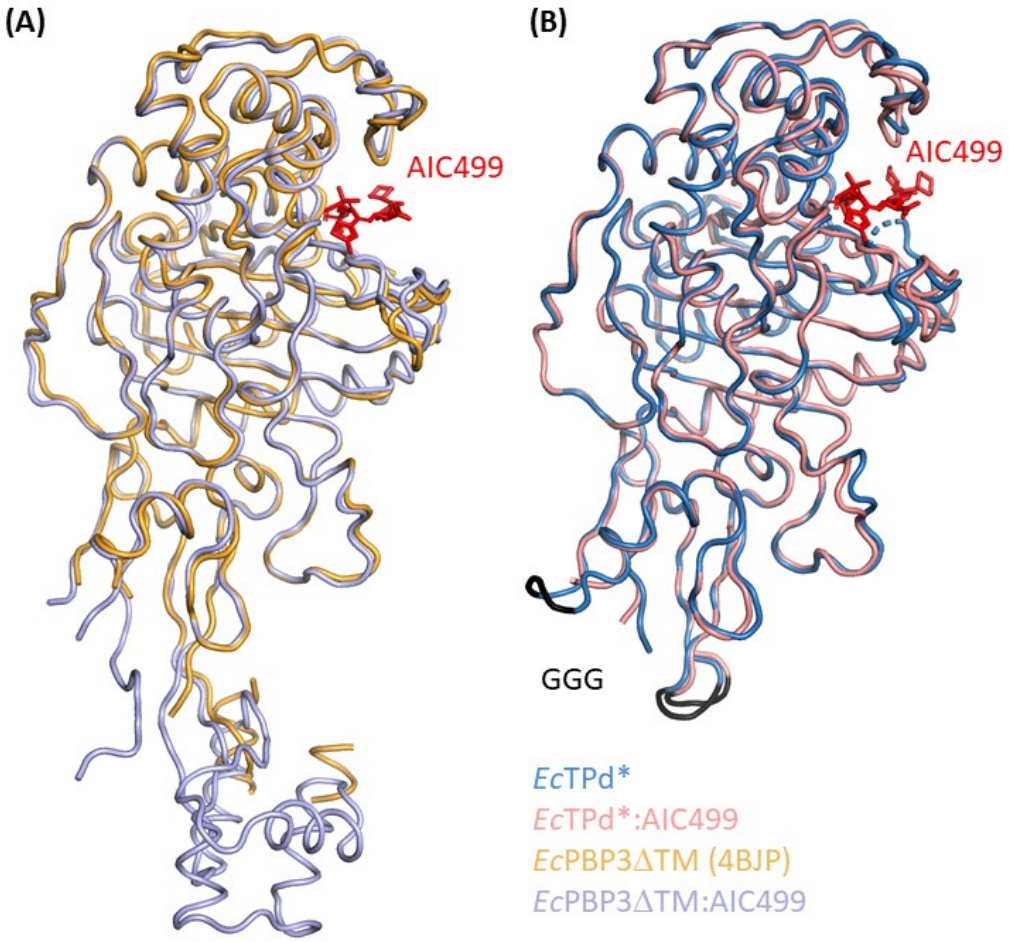


Figure 3.37: X-ray structures of *Ec*PBP3 determined in the absence and presence of AIC499 (red stick model).

(**A**) The published apo-*Ec*PBP3ΔTM structure (PDB entry 4BJP [47], gold) is used for superposition with the AIC499-complexed *Ec*PBP3ΔTM (blue, this study). (**B**) The structure of apo-*Ec*TPd* is shown in dark blue, while the complex with AIC499 is colored salmon. The GGG linkers replacing the removed segments are shown in black.

3.5.1.2. Effect of AIC499 binding on *E. coli* PBP3

The *Ec*PBP3ΔTM and *Ec*TPd* proteins were crystallized after pre-incubation with the experimental β-lactam AIC499. In the case of the truncated version, the presence of the compound left the space group (P 6_2 2 2) and packing unchanged. It did, however, impart a notable improvement in useable resolution, along with slight changes in lattice constants. For the *Ec*PBP3ΔTM:AIC499 complex, we observed space group P 6_4 2 2, which differs from the symmetry reported previously for the apo protein (PDB entry 4BJP [47], space group P 6_1 2 2); again, diffraction quality was clearly improved by the presence of the ligand but was still inferior to *Ec*TPd*:AIC499 (Table 3.6). To understand the effects of structural optimization leading to the final AIC499 compound, the environment of the five functional groups addressed during the process (Figure 1.11) was analyzed in the crystal structures of *Ec*TPd* and *Ec*PBP3ΔTM complexes. Since there were only negligible differences between the two models (RMS distance: 0.37 Å for 336 equivalent C_α positions), we will focus on the binding of AIC499 to *Ec*TPd* (Figure 3.37B and Figure 3.38) because (i) electron density was clearly more informative in this structure, and (ii) despite differences in crystallization conditions, the complex has been crystallized in the same space group as the apo form, minimizing spurious differences caused by non-conserved lattice contacts.

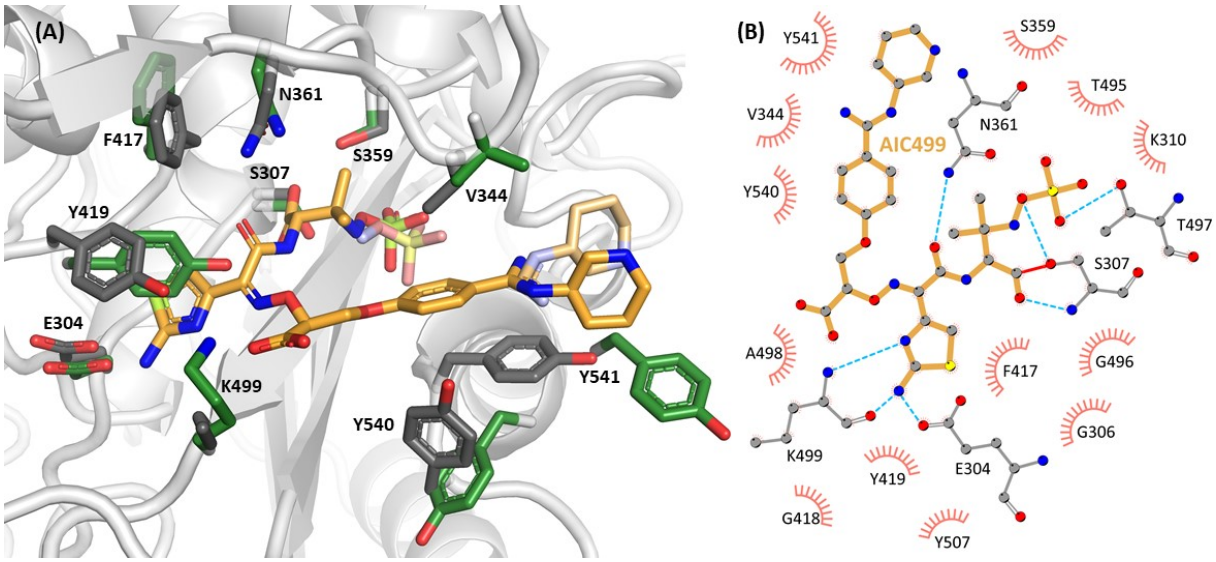


Figure 3.38: Structure of AIC499 within the active site of *Ec*TPd*.
(**A**) 3D representation of the covalently bound ligand together with the most relevant interacting side chains (gray) and their counterparts in the apo structure (green). The second conformation of the amidine-based head group as well as the terminal sulfate moiety are shown in lighter color. Note that K499 has been truncated in the complex structure because of missing electron density. (**B**) LigPlot+ representation of individual contacts for one conformer. Hydrogen bonds (cyan) are plotted for donor-acceptor distances between 2.3 Å and 3.2 Å, while hydrophobic interactions (salmon) have distances between 3.0 Å and 4.0 Å. A complete list of distances between protein side chains and the AIC499 compound is provided in Table A2.

As expected for a β-lactam antibiotic, AIC499 is found as a covalent acyl-enzyme intermediate, with the carboxyl group of the hydrolyzed lactam forming an ester bond with the hydroxyl of the catalytic S307. The compound adopts a U-shaped overall conformation, and its presence correlates with several conformational changes in the protein environment, mostly regarding the β2b-β2c-β2d, β3-β4 and β5-α11 regions, which contribute to the upper lobe, the bottom, and the lower lobe, respectively, of the binding cleft.

The largest differences between apo and complex structures can be observed in the β5-α11 region, in particular residues K539-A544. Being poorly ordered and not completely resolved in

the native protein, this segment is well-structured when AIC499 is bound in the active site. Indeed, ordering of the β5-α11 loop has been observed regularly as a result of PBP-β-lactam interaction [50]. Residues Y540 and Y541 appear to be particularly relevant here, because they re-orient towards the ligand and, together with Y511 from strand β4, shield the binding pocket for the R^1 substituent of the compound as an “aromatic wall”. These hydrophobic interactions centered on the phenyl ring of the head group are accompanied by a notable reduction of B-factors in the side chains involved.

Strands β3 and β4 are slightly shifted towards the core of the domain, relative to the remainder of the central β-sheet, with the protruding twisted hairpin being displaced in the opposite direction. Interactions between residues in the β3 strand and AIC499 include coordination of the terminal sulfate moiety (R^2) by the T497 side chain, rotation of the T495-G496 peptide plane because of steric interference, and formation of a hydrogen bond between the K499 carbonyl and the primary amine of the amino-thiazole group (R^4). A less favorable bond is potentially formed between the K499 amide nitrogen and either of the two amines in group R^4. Interestingly, the K499 side chain does not interact with the carboxyl group of the AIC499 linker region even though the apo structure suggests this side chain is in a favorable position for such an interaction. In fact, the side chain becomes less ordered in the presence of the compound, possibly alternating between various hydrogen bonding partners, which prohibits accurate modeling. The impact of AIC499 on strand β4 is mediated mostly by hydrophobic contacts. In addition to Y511 mentioned above, side chains of Y507 and Y514 are both involved in aromatic clusters. Y507 apparently moves in concert with Y419 in the proximal α8-α9 loop, which is displaced by and aligns parallel to the AIC499 amino-thiazole moiety, forming a π-π stack. Y514, on the other hand, propagates its displacement by the β3 strand on to F303 in the β2-α2 linker; the flipped side chain of F303 engages in a hydrophobic cluster with I512 and the aliphatic portion of K500, thus linking back to the β3-β4 region. Remarkably, while the side chain of S307 is involved in the acyl-enzyme intermediate and E304 establishes an important hydrogen bond with the amino-thiazole moiety, the main chain of the 300s region is only moderately affected by compound binding. Together with helix α8, the whole upper lobe comprising the β2b-β2c-β2d region and the adjacent helices α4 and α5 move slightly towards the active site, allowing some side chains to interact with AIC499. In particular, V344 between strands β2c and β2d appears to play an important role in stabilizing AIC499 by hydrophobic interactions. On the one hand it is in Van der Waals (VdW) contact to the phenyl ring of the amidine group (R^1), thus forming the counterpart of the aromatic wall on the opposite face of the compound. On the other hand, it also favorably interacts with one of the methyl groups at the C-4 position of the former β-lactam ring. Furthermore, the side chains of S359 and N361 preceding helix α5 are hydrogen-bonded to the acyl-enzyme ester oxygen and carboxylamide oxygen atoms, respectively, of AIC499.

The piperidine moiety of the head group R^1 does not engage in strong interactions with the protein and is modeled in two alternate conformations, with one edge of the ring in VdW contact with the Y541 side chain. Thus, this part of AIC499 does not seem to be very important for stabilizing the bound-state.

Certain structural differences between apo-*Ec*TPd* and *Ec*TPd*:AIC499 in regions distant from the active center of *Ec*PBP3 were noted. In addition to the truncation sites, this also concerns neighboring segments with enhanced flexibility, specifically the *N*-terminus, the β5n-β6n loop, and the vicinity of helix α1b. These alterations can be explained by the extensive conformational change in the β5-α11 region, which is in direct contact with the V88–G_3–E164 hairpin of a symmetry-equivalent copy, together with differences in crystallization conditions,

resulting in the presence of a PEG molecule linking the V88–G_3–E164 region to the *N*-terminal segment in the complex but not in the apo structure. The region preceding and including helix α1b is remarkable because it shifts more than 2 Å away from its conformation in the apo structure, which exceeds the differences in its immediate neighborhood, without itself being involved in a lattice contact. These differences might be an additive result of long-range conformational triggers propagated from the ligand binding site on the one hand and the "truncation pole" on the other.

3.5.2. Structures of *P. aeruginosa* PBP3

3.5.2.1. *P. aeruginosa* PBP3 apo protein

Remarkably, three distinct crystal forms were identified for *Pa*PBP3ΔTM, featuring different unit cells and lattice contacts, but all belonging to space group C 1 2 1. Two of those crystals were used for soaking experiments using Cluster J compound 34, but because no electron density was found for this compound the data are treated as *de facto* apo-crystals. Additionally, soaking experiments with 250 mM bromo- and iodo-pyrazole, Ta_6Br_{12}, and the identified hit fragments 3N-528S and 5N-395S were performed but yielded poor diffraction. Diffraction quality of *Pa*PBP3ΔTM crystals was generally superior when compared to *Ec*PBP3ΔTM crystals, with moderate anisotropy and usable resolutions of 2.2 Å (crystal form 1), 1.8 Å (crystal form 2) and 1.9 Å (crystal form 3; for details refer to Table 3.7).

As expected, superposition of the three models reveals a very similar overall structure (Figure 3.39A). The RMS distance between corresponding C_α positions was calculated between 0.50 Å and 0.65 Å, but values decrease to approximately 0.30 Å if only the TPase fold (R62-S76, R152-A187, and K217-A563) is considered. In fact, the head subdomain (T77-R152) appears to largely bend as a rigid body, leading to slightly different orientations. These can be explained by packing effects; in addition to a lattice contact shared by all three structures (α1n-α2n region and α3n helix with the α2-β2a loop and the proximal α10 helix of a symmetry mate), crystal forms 1 and 3 feature unique interactions involving, among others, the α3n helix and β3n-α4n regions, respectively, contacting the β2e-β2f and α2-β2a segments, or the β2b-β2c hairpin, of neighboring molecules.

Table 3.7: Data collection and refinement statistics of *Pa*PBP3 structures reported (cf, crystal form).

Values in parentheses refer to the highest-resolution shell. [a] Conventional definition using spherical shells. [b] Calculated with respect to an ellipsoidal portion of reciprocal space fitted to the cut-off surface, as defined in STARANISO. Low values for *Pa*PBP3ΔTM cf2 and cf3 are a consequence of a rugged cut-off surface complicating the determination of a meaningful ellipsoid.

	***Pa*PBP3ΔTM (cf 1)**	***Pa*PBP3ΔTM (cf 2)**	***Pa*PBP3ΔTM (cf 3)**	***Pa*PBP3 ΔTM:AIC499**
PDB entry	7ONX	7ONY	7ONZ	7ONK
Data collection				
Beamline	DESY P11	EMBL P13	EMBL P13	DESY P11
Wavelength [Å]	1.0332	0.9999	0.9999	1.0332
Space group	C 1 2 1	C 1 2 1	C 1 2 1	P 2_1 2_1 2_1
Cell dimensions				
a, b, c [Å]	110.6, 82.2, 91.4	104.1, 125.0, 74.2	151.5, 37.5, 82.8	81.0, 91.1, 148.4
α, β, γ [°]	90, 116.3, 90	90, 122.5, 90	90, 112.6, 90	90, 90, 90
Resolution range [Å]	44.40–2.16 (2.34–2.16)	39.55–1.77 (1.97–1.77)	40.75–1.86 (2.09–1.86)	46.92–1.73 (1.90–1.73)
$CC_{1/2}$ [%]	99.6 (47.1)	99.8 (64.1)	99.8 (77.2)	99.9 (62.4)
R_{meas} [%]	11.1 (101.5)	7.0 (115.9)	7.1 (96.6)	9.4 (92.5)
I/σ	7.9 (1.4)	13.9 (1.8)	12.3 (1.6)	10.9 (1.6)
Completeness [%][a]	66.6 (15.4)	55.5 (10.4)	51.8 (8.8)	74.2 (15.3)
Ellipsoidal Completeness [%][b]	91.4 (57.8)	70.4 (4.4)	74.8 (3.6)	95.9 (64.8)
Refinement				
Resolution range [Å]	44.40–2.16	39.6–1.8	40.8–1.9	46.9–1.7
No. unique reflections	26398	42915	19031	85337
No. protein atoms	3609	3841	3668	7803
No. ligand atoms	32	118	12	260
No. water molecules	146	274	134	750
R_{work} [%]	18.17	17.98	20.91	17.34
R_{free} [%]	21.52	21.34	25.85	21.52
RMSD				
Bond lengths [Å]	0.003	0.004	0.003	0.007
Bond angles [°]	0.579	0.666	0.601	1.037
Mean B factor [$Å^2$]	42.63	34.16	29.60	23.60
Ramachandran plot				
Favored [%]	95.98	96.79	96.30	97.72
Allowed [%]	4.02	3.21	3.70	2.28
Outliers [%]	0	0	0	0
Z-score	–1.20	–1.07	–0.93	–0.09

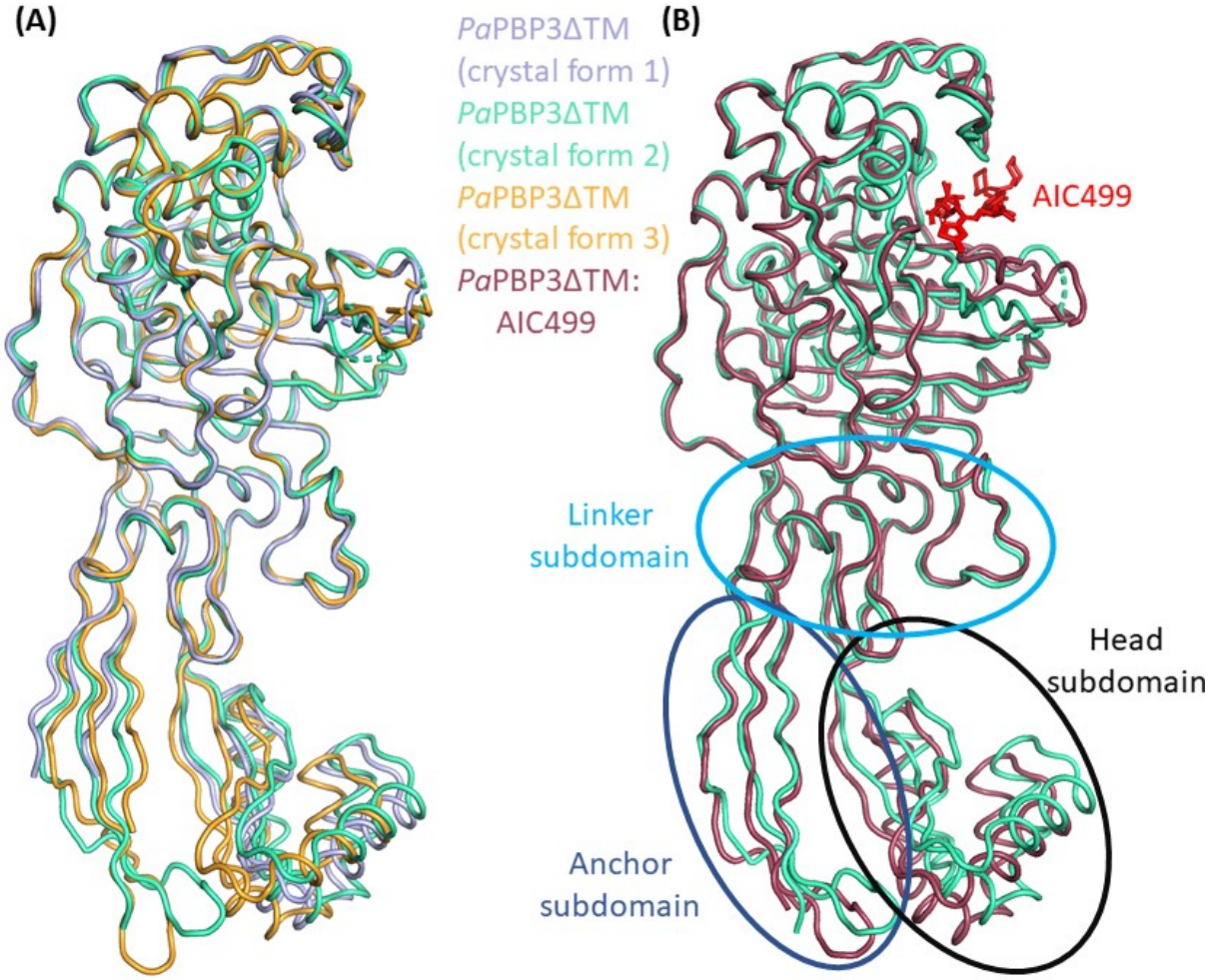


Figure 3.39: X-ray structures of *Pa*PBP3 determined in the absence and presence of AIC499 (red stick model).

(**A**) Crystal form 1 (blue), crystal form 2 (green) and crystal form 3 (bright orange) of *Pa*PBP3ΔTM feature different unit cells, leading to slightly different orientations predominantly in the n-PBd. (**B**) The structure of apo-*Pa*PBP3ΔTM crystal form 2 (green) is used for superposition with *Pa*PBP3ΔTM:AIC499 chain B (dark red). Additionally, the head, anchor and linker subdomains of the n-PBd are highlighted with black, dark blue and light blue ellipses, respectively.

The anchor subdomain is non-contiguous, consisting of the *N*-terminal segment (A50-H61) and a long β-hairpin (G188-P215), and despite its spatial proximity moves independently of the head domain. This domain features particularly high flexibility, as evidenced by large B-factors and the difficulties of consistent tracing. In fact, the final models for crystal forms 2 and 3 lack several side chains in this region as well as small portions of the extreme *N*-terminus because of missing electron density, whereas in crystal form 1 larger parts of the backbone could not be traced. Again, the differences can be explained by packing effects.

In the TPd, the largest differences between the three apo-*Pa*PBP3ΔTM structures concern the β3-β4 and β5-α11 regions, both of which show indications of high flexibility. The β3-β4 hairpin is generally non-contiguous in electron density, preventing it from being modeled completely. In crystal forms 1 and 2 eight and six residues are not resolved, respectively, while in crystal form 3 nine residues are missing. In contrast to the structure in crystal forms 1 and 2, the protruding part of the β3-β4 region of crystal form 3 seems to bend towards the active site. This difference correlates with a lattice contact established with helix α1 that is exclusive to form 3. Based on previously solved apo-*Pa*PBP3ΔTM structures (PDB entries 3OC2, 3PBN and 6HZR), the β3-β4 hairpin was thought to be disordered and to be stabilized upon a ligand

(or substrate) binding to the active site [49-51]. Our observations indicate that the protruding β3-β4 segment is at least partly structured in the apo protein, even if not engaged in lattice contacts.

While it was not possible to completely model the β3-β4 loop in any of the three crystal forms, the β5-α11 region was traced at least in crystal form 1. In contrast, two and five residues, respectively, are missing in the other two forms. This observation might be related to a packing effect caused by the head subdomain of a neighbor molecule restricting mobility of the β5-α11 region in form 1 (indirectly via helix α1) to a bent conformation. In contrast, this segment appears to adopt a more extended structure in crystal form 2, as judged by the resolved portion. Comparison with published *Pa*PBP3ΔTM apo structures confirms the notion of enhanced conformational freedom of the β5-α11 loop. While it is bent towards the active site in crystal form 1, it is orientated in the opposite direction in the 6HZR structure and is essentially absent in 3OC2. The 3PBN structure adopts what may be considered an intermediate state, since the β5-α11 loop is not bent strongly in either direction; however, this is again the result of a lattice contact with a neighboring head subdomain.

3.5.2.2. Effect of AIC499 binding on *P. aeruginosa* PBP3

In addition to investigation of the apo structure, *Pa*PBP3ΔTM was crystallized in the presence of AIC499; crystals belonged to space group P 2_1 2_1 2_1 with two copies per asymmetric unit and yielded diffraction data extending to a resolution of 1.7 Å (Table 3.7). As expected, the AIC499 complex structure displays high overall similarity to the apo version, with overall RMS distances (chain B) of 0.55 Å, 0.58 Å and 0.57 Å w.r.t. crystal forms 1, 2 and 3, respectively (Figure 3.39B). The two molecules in the asymmetric unit differ from each other and from the apo structures in the orientation of the head and anchor subdomains relative to the TPd. Similar to the variation among the apo-*Pa*PBP3ΔTM structures, this effect can be explained by slight bending of β-strands dominating these extended folds, induced by different lattice environments. Specifically, in addition to the highly favorable contact described above for all three apo structures, chain A displays additional extensive interactions of its *N*-terminal part (strands β1n and β9n, α4n-β4n loop) with helix α5n in the linker subdomain, the α6-β2e loop and helix α1 of symmetry-related copies, while in chain B a contact of strand β9n with the β2h-β2i segment of a symmetry mate is noteworthy. Within the asymmetric unit, the extended *C*-terminus of chain A is in contact with the head and linker subdomains of chain B; the reciprocal interaction is not observed.

As described above for the complex with the *E. coli* protein, the AIC499 molecule is covalently associated with *Pa*PBP3ΔTM via the catalytic serine (S294) side chain. In the active site environment, the β3-β4 loop as well as the β5-α11 loop appear quite flexible, despite the presence of the ligand, as evidenced by high B-factors and often discontinuous electron density. Nevertheless, we note that strands β3 and β4 clearly bend towards the active site, supporting some interactions of their side chains with the AIC499 molecule, most importantly R489 (see below). In contrast, the β5-α11 loop is ordered in chain B only and bends in a similar direction to that observed in apo form 2. More importantly, helix α11 is *N*-terminally extended by more than one turn in both chains, which again allows additional contacts with the ligand.

The overall conformation and interactions of AIC499 in its complex with *Pa*PBP3ΔTM resemble those described above for the *Ec*PBP3ΔTM adduct (Figure 3.40). The terminal sulfate group (R^2) is hydrogen bonded to the side chains of K484, S485 and T487. The nitrogen of the former

β-lactam ring forms a hydrogen bond with the hydroxyl moiety of S349, whereas the ligand amide group interacts with the carbonyl of T487 on the one hand and with the side chain amide of N351 on the other. The amino-thiazole (R^4) is well stabilized, engaging in three hydrogen bonds with the side chain of E291 (specifically, one of its alternate conformations) and the backbone of R489. Unlike the situation in the *Ec*PBP3ΔTM:AIC499 complex, the amino-thiazole does not form a parallel π-π stack with Y409 (equivalent to Y419 in *Ec*PBP3) but a displaced T-stack. Additionally, Y407 (equivalent to F417 in *Ec*PBP3) is close to the aromatic rings of Y409 and the amino-thiazole, establishing an aromatic network for further stabilization. The carboxyl group of the linker (R^5) forms hydrogen bonds to the guanidinium group of R489, and the phenyl ring of the head group (R^1) is close to the aliphatic side chain of V333. On the opposite face of this ring, another aromatic network is formed: the R^1 phenyl ring stacks with Y532, which is further stabilized by parallel π-π stacking with Y503. Additionally, C_β of F533 appears close to the aromatic ring of AIC499, but the electron density was too weak to build the remainder of the side chain. Interestingly, with an occupancy of 40% it was possible to model an alternative set of correlated side chain conformations for R489 and Y503, located in the β3 and β4 strand, respectively. Due to a clash of the alternate Y503 rotamer with Y532, the β5-α11 loop must be displaced as well; however, electron density was not conclusive as to the respective conformer. Similar to the complex with *Ec*PBP3ΔTM, the AIC499 head group (R^1) appears with two conformations in both copies present in the asymmetric unit. In general, the piperidine moiety is devoid of strong interactions with the protein; while in chain B V333 at a VdW distance from either variant is the only notable contact, one of the conformers in chain A orients towards Y532.

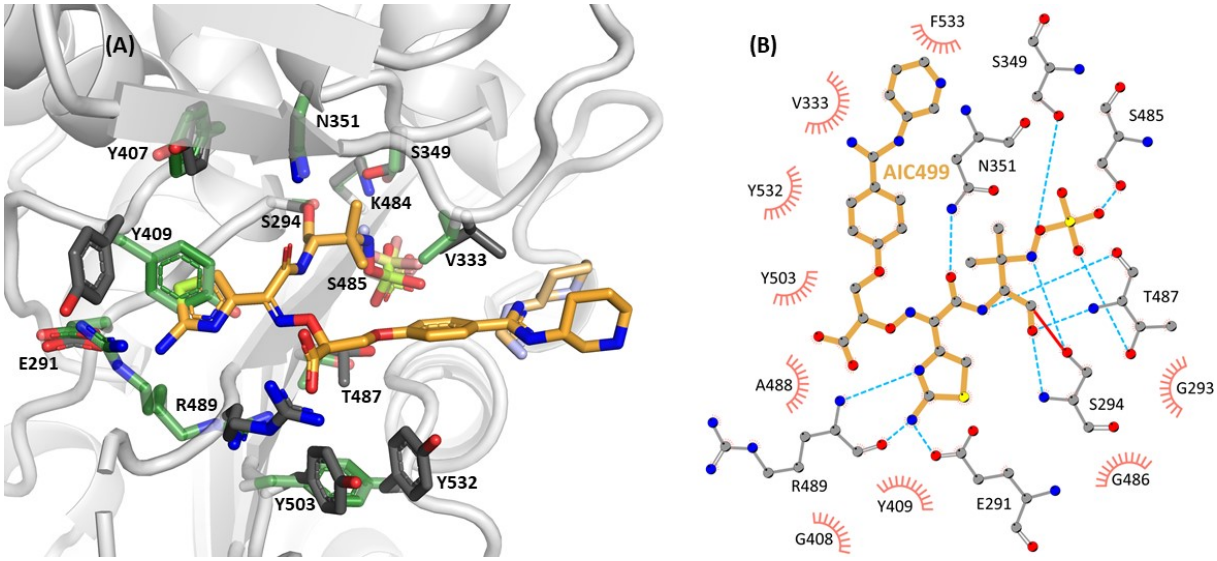


Figure 3.40: Structure of AIC499 within the active site of *Pa*PBP3ΔTM.

(**A**) 3D representation of the covalently bound ligand together with the most relevant interacting side chains (gray) and their counterparts in the apo structure (crystal form 2, green). Alternative conformations of the amidine-based head group and the terminal sulfate moiety are shown in lighter color. (**B**) LigPlot+ representation of individual contacts. Hydrogen bonds (cyan) are plotted for donor-acceptor distances between 2.3 Å and 3.2 Å, while hydrophobic interactions (salmon) have distances between 3.0 Å and 4.0 Å. A complete list of distances between protein side chains and the AIC499 compound is provided in Table A2.

4. Discussion

4.1. ^{19}F NMR screen for FBDD targeting PBP3 from *E. coli*

In this project, PBP3 was selected as target structure to develop new classes of antibiotics and a FBS study was performed to cover a greater chemical space when compared with that covered by HTS (Section 1.9). ^{19}F NMR was used as the screening technique because fragments typically give rise to only one peak in the NMR spectrum with a high sensitivity; thus, lowering the concentrations used (Section 1.9.2). Therefore, more fragments can be used in each pool when compared to ^{1}H NMR leading to a higher throughput. In total, 17 fragments (hit rate: 1.9%) were identified in two FBS to potentially bind His-tagged *Ec*PBP3ΔTM and may be used in further drug development stages.

4.1.1. Establishing the protocol for ^{19}F NMR FBS

The first FBS was performed to establish the ^{19}F NMR screening protocol and identify fragments that bound PBP3 from *E.coli*, with expected affinities between 100 µM and 10 mM [86]. 1D ^{19}F NMR spectra of 206 tri-fluorinated and 225 mono-fluorinated fragments in pools of 20 fragments were screened against His-tagged *Ec*PBP3ΔTM (Figure 3.4 and Figure 3.9). Unlike the tri-fluorinated fragments, which showed singlets, active $^{n}J_{FH}$ scalar couplings resulted in multiplets for the mono-fluorinated fragments, leading to significantly lower signal-to-noise ratio of 4 to 20 when compared with the signal-to-noise ratios of 50 to 300 for signals in spectra recorded on tri-fluorinated fragments. With appropriate ^{1}H decoupling these multiplets would not be observed; however, a probe-head equipped with the B600-F instrument uses the same channel for ^{1}H and ^{19}F and thus, ^{1}H decoupling during ^{19}F acquisition was not possible. In addition to the weak intensities observed, only two relatively weak peaks (signal-to-noise of 3.6 and 4.9, respectively) were assigned to score 2 (using the same scoring system used for the tri-fluorinated fragments) and no assignment information was available for these peaks (Figure 3.11). These issues made analysis of the monofluorinated NMR data challenging and consequently, the mono-fluorinated fragments were excluded from the analysis and subsequent fragment selection. Excluding monofluorinated fragments may have reduced the chemical space sampled and consequently the number and type of potential hit fragments. However, besides the fluorinated chemical groups, the mono- and tri-fluorinated fragments were found to share similar structures (example shown in Figure 4.1) and therefore, the absence of the mono-fluorinated fragments did not lead to a significant reduction in the chemical space sampled, although ^{19}F atoms may interact with the protein and have an impact on the affinity.

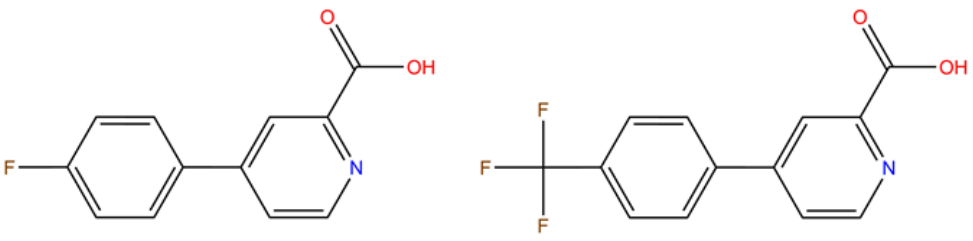


Figure 4.1: Structural similarities between mono- and tri-fluorinated fragments.
In this example, the ^{19}F atom of the mono-fluorinated fragment (left) was exchanged by a tri-fluoromethly group (right). The remaining scaffold is identical.

Peak changes in 1D ^{19}F spectra of fragments screened against His-tagged *Ec*PBP3ΔTM were scored from 1 to 5, with scoring based on changes to chemical shift and intensity of the peaks arising from these fragments (Figure 3.6). Boundaries, as defined by σ, used to establish the scoring system were chosen to match reported hit rates (approximately 2% to 7%; [112, 174, 176]). Unlike a fixed threshold of 25% loss of intensity, which was reported in literature, the scoring system was developed to (i) also account for changes in chemical shift, and (ii) identify potential hit fragments where the corresponding peaks showed smaller changes, confirming the reproducibility of score 1 peaks [174]. Because of differences in the observed and supplier-provided chemical shifts, which were determined in DMSO, it was not possible to assign unambiguously all peaks in the 1D ^{19}F NMR spectra (Table 3.1). Consequently, peaks were initially classified to respective scores 1 to 5, and additional screenings featuring smaller pool sizes were performed to validate the assignments and the interaction with His-tagged *Ec*PBP3ΔTM (Sections 3.1.2.3 and 3.1.2.4).

Interestingly, the number of fragments corresponding to peaks that scored 1 to 3 decreased in the assignment and hit validations rounds (Figure 3.7, Figure 3.12 and Figure 3.13). Initially, eight peaks were defined as score 1, seven as score 2 and 36 as score 3. After the assignment validation step, only one, four and four fragments were assigned to scores 1, 2 and 3, respectively, and following hit validation only two fragments remained as score 2 and two fragments with score 3.

The inability to reproduce larger Δδ and ΔI when using smaller pool sizes can be explained as follows. First, it was not possible to unambiguously assign all peaks to fragments in the initial screen. Thus, up to three fragments were selected for one peak showing significant changes after the addition of His-tagged *Ec*PBP3ΔTM (i.e., 31 fragments were selected to assign 22 peaks). Logically, only one of the three fragments is associated with this peak. Secondly, weak peaks that showed relatively large changes after the addition of His-tagged *Ec*PBP3ΔTM and were initially assigned to score 1 may have represented impurities, byproducts of fragment degradation, or aggregated fragments in aqueous solutions. Thus, these hydrophobic molecules give rise to peaks with weak intensities in the 1D ^{19}F spectra and may non-specifically bind hydrophobic pockets on the surface of the protein, leading to greater changes in chemical shift and intensity. Consequently, these peaks were initially assigned as hits, but it was not possible to identify the respective fragment. Thus, fragments with similar supplier-provided chemical shifts to the chemical shift of these peaks were selected and screened in the assignment and validation rounds. Noticeably, changes observed upon the addition of His-tagged *Ec*PBP3ΔTM for these relatively weak peaks were not reproducible. Thus, these peaks most likely arise from impurities, degraded, or aggregated fragments and the assignment and hit validation screens worked to identify fragments showing the largest changes in 1D ^{19}F NMR spectra and no PAINS were selected as hit fragments. When aiming for further screening experiments using this library it may be effective to use a SPAM NMR filter to identifies impurities, degraded fragment components or aggregated fragments [112]. As described in Section 1.9, this approach is time-consuming and thus, was not considered prior to screening the libraries against His-tagged *Ec*PBP3ΔTM. Nonetheless, this set of experiments with the hits may further clarify identification of hit fragments.

Furthermore, large Δδ and ΔI for peaks assigned to fragments in the initial screen were not reproducible. As an example, Δδ and ΔI from FD-0035 in the presence of His-tagged *Ec*PBP3ΔTM scored 1 in the initial screen but in the validation screen, observed spectral changes were smaller (i.e., Δδ was 67% and 17% and ΔI was 14 Hz and 2 Hz in the initial screen and validation screen, respectively; Table 4.1). Because only the number of additional

fragments in the sample was reduced, it might be possible that the larger changes in the initial screen were caused by interacting fragments, i.e., other fragments in the sample were able to enhance the interaction between His-tagged *Ec*PBP3ΔTM and particular fragments. However, for FD-0035, it was possible to determine the K_d using MicroScale thermophoresis (MST; 1.4 mM; performed by Dr. Immanuel Grimm at AiCuris), thus, confirming an interaction with *Ec*PBP3. This enhancement may also be possible for other fragments in samples and following from this, larger changes were not reproducible when the number of fragments in pools was reduced and the composition of fragments in the pool differed.

Noteworthy, development of the score system was beneficial because, for example, it was possible to identify FD-0035 as a hit fragment, which showed a loss of peak intensity of 17% in the hit validation round and would have been excluded from the analysis following the threshold of 25% reported by Nagatoishi *et al.* [174]. As mentioned above, it was possible to determine the K_d, thus, FD-0035 was identified as a hit fragment. However, usage of different types of experiments, such as the ^{19}F T_2 Carr-Purcell-Meiboom-Gill (CPMG) NMR experiment, might simplify the assessment of changes in intensity. In this experiment, a T_2 filter is used to more easily identify hit fragments that interact with the protein because the intensity of a signal from a bound fragment decreases far more rapidly during the T_2 delay when compared with a fragment that does not interact with the protein. Additionally, the chemical exchange may have an impact on R_2 (Equation 1), potentially increasing the relaxation. Another NMR-based experiment that might help to assess the affinities of hit fragments is the chemical shift-anisotropy-based affinity ranking (CSAR) experiment [129]. With this experiment, relaxation data directly proportional to the binding affinities are gained by removing R_{ex} using high power spin lock pulses and dipolar relaxation effects when recording at different magnetic fields. Differences in CSA are accounted for by normalization. Furthermore, with high-quality relaxation data, it may be possible to determine K_d values with the CSAR approach.

Regardless of the non-reproducibility of peaks showing larger $\Delta\delta$ and ΔI, the final hit validation of tri-fluorinated fragments identified, four fragments with scores 2 and 3. Thus, approximately 1% of the 206 tri-fluorinated fragments used in this screen were assigned as scores 2 and 3, respectively, and with a hit rate of approximately 2% this result still matches reported hit rates [176]. Hit fragments 5N-395S and 3N-528S, which were assigned as score 2, share a similar chemical scaffold (Figure 3.14) and the changes of the peaks in the spectra were reproducible. This result indicates that the screen was successful because it was possible to gain similar and reproducible results for chemically similar fragments. Furthermore, it was possible to estimate K_d values for 5N-395S (~1.2 mM) performing a protein titration (Section 3.2) and FD-0035 (1.4 mM) by MST (by Dr. Immanuel Grimm at AiCuris), respectively. Because of the poor solubility of 5N-395S and 3N-528S (>500 µM in aqueous solutions) and the comparably high K_d value, it was not possible to determine the K_d value more accurately using MST. Nevertheless, 3N-528S has a similar chemical structure when compared with the chemical structure of 5N-395S and thus, it is likely that the K_d is in a similar range, which was expected to be between 100 µM and 10 mM [86].

In summary, it was possible to establish the experimental and analysis protocol for ^{19}F NMR FBS using libraries of fluorinated fragments and the limitations of the equipment were explored. Thus, a second screen was performed with slight adjustments (Section 4.1.2). Furthermore, additional experiments to gain details of the interaction between hit fragments and His-tagged *Ec*PBP3ΔTM were performed and are discussed in Sections 4.2 and 4.3.

4.1.2. Thirteen hit fragments were identified in the second FBS

In the second ^{19}F NMR FBS targeting His-tagged *Ec*PBP3ΔTM, a library of 475 tri-fluorinated fragments was supplied by Enamine. Based on the results in the first screen some adjustments were made (Section 2.5). However, unambiguous assignment of all peaks to respective fragments was still not possible (Table 3.3) and for some fragments assignments had to be validated by the hit validation screen (Section 3.1.3). Nonetheless, it was sufficient to perform only one additional screening round and combine the hit and assignment validations in one step.

Similar to the first screen, some weaker peaks were present in the spectra, showing relatively large changes after the addition of His-tagged *Ec*PBP3ΔTM. These peaks were assigned as potential hits (Figure 3.16). As in the first FBS, such weak peaks likely represent impurities, degraded fragments or fragments that have partially (or mostly) aggregated. Probably featuring a higher hydrophobicity, these molecules are less soluble in aqueous solutions, give rise to weaker peaks in the 1D ^{19}F spectrum and non-specifically bind hydrophobic pockets on the surface of the protein, thereby leading to greater changes in intensity and chemical shift. Consequently, assignment information was not available and fragments with similar chemical shifts were selected to reproduce the changes of those peaks in the hit validation round. For this reason, some changes observed in the first round of the second FBS were not reproducible in the validation round, which was performed with smaller pool sizes of 5 to 6 fragments. However, when compared to the first screen, a higher percentage of peaks showed larger changes in the hit validation round and were reproduced and assigned to higher scores. Subsequently, 0.4% (2), 2.3% (11), 1.3% (6), 0.2% (1) and 95.8% (455) peaks were assigned to scores 1 to 5, respectively (Figure 3.17). Defining scores 1, 2 and 3 as hits, the hit rate of the second screen (~4%) was two-fold higher than that of the first screen (~2%), but still matches reported hit rates [112, 174, 176]. In summary, four fragments identified in the first screen (scores 2 and 3) and 13 fragments in the second screen (scores 1 and 2) were assigned as hits and were further investigated in additional experiments to validate and characterize the interaction between fragment and His-tagged *Ec*PBP3ΔTM (discussed in Sections 4.2 and 4.3).

4.2. Experiments for characterizing hit fragments

After identifying 17 hit fragments in the first and second screen (Section 4.1), experiments were performed to further validate the binding and gain additional information on the mode of binding, including determining a K_d value using ^{19}F NMR, AIC499 competition binding and chemical shift perturbation analysis.

4.2.1. K_d value determination

The K_d and $\Delta\delta_{max}$ values for the interaction between 5N-395S and His-tagged *Ec*PBP3ΔTM were estimated by an NMR titration to be 1.23 ± 0.54 mM and –63.97 ppm, respectively (Figure 3.19 and Figure 3.20). Although not measured, similar K_d values in the low mM range are expected for fragments of similar size and chemical composition, such as 3N-528S (Figure 3.14). This K_d for 5N-395S is an estimate because the concentrations used in the titration were below the K_d value. Ideally, for obtaining suitable data sets for fitting and determination of the

K_d, ligand concentrations five times above the K_d should be used when working at a protein concentration 0.5 times the K_d [181]. For titration where a constant fragment concentration is used, the protein is treated as the “ligand” and the fragment as the “protein”. This approach was used successfully to obtain a reasonable estimate of the K_d for 5N-395S. Moreover, MST experiments performed by Dr. Immanuel Grimm at AiCuris yielded K_d values for identified hit fragments between 60 µM (AIC252712) and 1.4 mM (FD-0035) (Table 4.2). Due to the low solubility of 5N-395S and 3N-528S, it was not possible to determine the K_d by MST.

4.2.2. AIC499 competition binding

Competitive binding of the hit fragments using the covalent binder AIC499 (Section 1.7.1) was investigated and the competition factor F_{comp} calculated as described in Section 2.6.2 (results in Table 4.2). Different percentages may be related to different modes of fragment binding inhibition (Figure 4.2) or can be explained by multiple binding sites. For example, fragment AIC252821 is only slightly influenced by the addition of AIC499 (F_{comp} = 129%; Figure 3.21A). Thus, both molecules are unlikely to compete for the same binding site, and the fragment might bind to another binding pocket. In contrast, the binding of AIC252849 shown in Figure 3.21B is affected by the presence of AIC499 (F_{comp} = 16%) and might be explained by the mode of inhibition shown in Figure 4.2A. In this case, both molecules compete for the same binding site, which is blocked by covalently bound AIC499. Results shown in Figure 3.21 C (AIC252901) and D (AIC252683) indicate that the fragments are partly displaced by AIC499 (F_{comp} = 65% and 46%, respectively). Therefore, inhibition modes such as shown in Figure 4.2 B and C are plausible. In these cases, a part of the fragment’s binding site is blocked by the AIC499 and therefore, the affinities of the fragments toward His-tagged *Ec*PBP3ΔTM are reduced, leading to smaller changes observed in the spectrum after the addition of AIC499. However, it is possible that allosteric effects, depicted in Figure 4.2E, change the conformation of the binding pocket of the fragment, which may be distal from the active site. Thus, without structural information it is not possible to determine that these fragments are binding close to the active site, which would enable their use in merging or linking approaches in FBDD (Figure 1.16). In addition to the above-mentioned inhibition modes, multiple binding sites for one fragment have to be considered. Possibly, one of them might be blocked by AIC499, whereas the fragment is still able to interact with other binding sites. Thus, the F_{comp} value related to the blocked binding site would be close to 0%, whereas the F_{comp} value related to the unaffected binding site would be close to 100%. Because the observed peak in the 1D ^{19}F spectrum is an average of the peaks related to different binding sites, this inhibition mode would lead to F_{comp} values between 0% and 100%, as observed for AIC252901 and AIC252683.

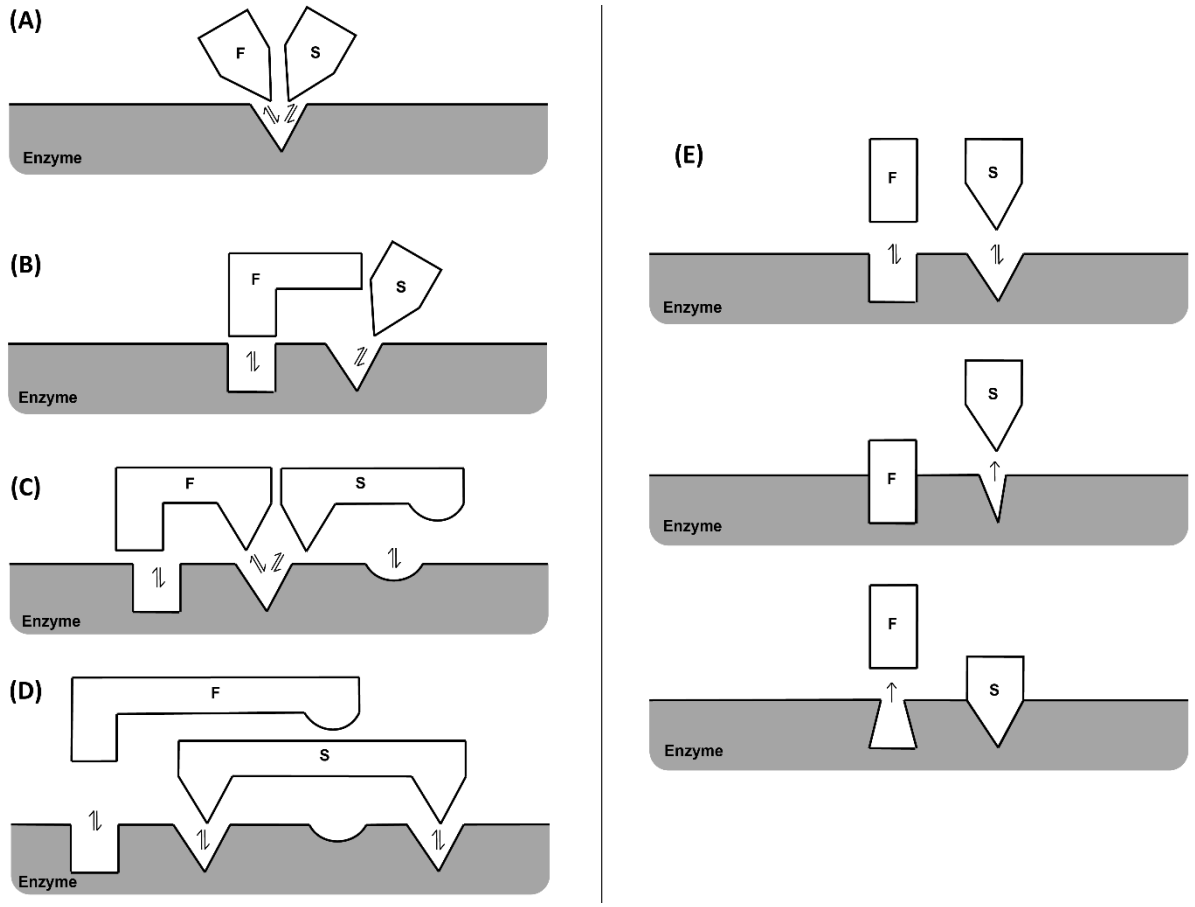


Figure 4.2: Schematic examples of mutual interference between one high affinity binder (i.e., substrate; S) and one fragment (F).

In context with the AIC499 competition binding, AIC499 is the high affinity binder and fragments feature low affinities. Note that AIC499 covalently binds to the active site and the turn-over constant K_3 (Figure 1.8A) is slow and therefore, can be neglected. (**A**) Substrate and fragment compete for the same binding site (active site in case of AIC499). (**B**) The fragment does not bind the same binding pocket but is sterically hindered by bound substrate. (**C**) The binding sites of substrate and fragment overlap. The affinities toward these sites are lowered because the surface that is available for binding is reduced. In case of covalently bound substrate, only the affinity of the fragment is lowered. (**D**) Binding sites of substrate and fragment overlap and only one of the molecules is able to bind the enzyme. (**E**) Allosteric inhibition by either the substrate or the fragment. Note that because AIC499 binds covalently, it is unlikely that AIC499 gets displaced from the active site upon fragment binding. Figure adapted from [182].

For AIC252627 and AIC252712 analysis, larger $\Delta\delta$ and ΔI were found in the presence of AIC499 (Figure 3.21E and F). Because of the relatively weak changes after the addition of His-tagged *Ec*PBP3ΔTM in the FBDD screen and because it was not possible to determine a K_d value using MST (performed by Dr. Immanuel Grimm at AiCuris), AIC252627 was not considered a hit fragment. However, further investigation might potentially reveal weak affinity to the catalytic site of *Ec*PBP3, which might be enhanced by the addition of AIC499. Thus, it might be possible to combine the scaffolds of AIC499 and AIC252627 to build a larger compound with beneficial properties (i.e., for example, better PBP3 inhibition, ADME properties). In the case of AIC252712, the presence of a second peak indicates that the exchange between the free and bound states might be in the slow-to-intermediate regime on the ^{19}F NMR timescale, indicating that the interaction in presence of AIC499 is comparably strong. This argument is supported by the K_d value of 60 µM (in absence of AIC499), which was determined using MST (performed by Dr. Immanuel Grimm at AiCuris). Larger changes in the 1D ^{19}F spectrum indicate that the interaction between *Ec*PBP3 and this fragment is further enhanced in the presence of AIC499. Furthermore, the additional peak in the spectrum

in presence of AIC499 had a unique chemical shift, indicating a different chemical environment, when compared with the chemical environment in the absence of AIC499. Thus, it is likely that AIC252712 binds *Ec*PBP3 in the active site in the vicinity of AIC499. In this context it is important to notice that this fragment has a chiral center (Table 4.2) and is a racemate in the sample. Thus, it might be possible that only one enantiomer is able to interact with AIC499 and AIC499 acts as a chiral agent [175]. However, this behavior was not observed in the presence of only the protein, which could also act as a chiral agent. Therefore, it might be possible that AIC252712 is directly interacting with AIC499. Similar to AIC252627, it might be possible to optimize the AIC499 scaffold using the chemical structure of AIC252712. Interestingly, in the chemical shift perturbation analysis, peak 190 was found to show a $\Delta\delta_{av}$ value above 0.06 ppm but no sequence assignment is available for this residue (Figure A7). For further development in the process of FBDD, details on the binding mode would be required. For this, NOE spectra of the protein might help to identify all residues involved in the interaction. Furthermore, STD NMR might give some information on the atoms of the fragments involved in the interaction with the protein. Soaking experiments using crystals of *Ec*PBP3ΔTM in complex with AIC499 might enhance the affinity and thus, it might be possible to identify electron density associated with the fragment. Additionally, docking studies and/or molecular modeling might help support experimental data and gain details about the binding mode.

4.2.3. Chemical shift perturbation analysis

Because of solubility issues only 14 of the identified hit fragments were examined in a chemical shift perturbation assay to identify potential common binding sites, which can be easily identified when assignment information is available (Section 1.9.1). Furthermore, the chemical shift perturbation analysis can validate fragment binding. Thus, 2D ^{1}H-^{15}N TROSY-HSQC spectra of *Ec*TPd* were recorded in the presence of the fragments (Section 3.2.3). Changes in the chemical shift of peaks after adding a fragment are exemplarily shown in Figure 3.22. $\Delta\delta_{av}$ was calculated for every peak, assigned to a specific number and 3σ was used as a threshold for peaks showing the largest changes (example in Figure 3.23). Using this threshold, an average of 22 peaks per fragment showed changes, with a range of 10 for AIC252626 and 32 for AIC252728.

Changes in the chemical shift of peak identifier numbers 104 and 37 were affected by fragments AIC252490, AIC2552728, AIC252732, and AIC252901, suggesting that the residues corresponding to these peaks are involved in recognition of these hit fragments. Assignment information for these two peaks was obtained from backbone sequence-specific assignments, representing residues H179 and E180 of *Ec*TPd*. Using the solved X-ray crystal structure (PDB entry 7ONO; [48]), H179 and E180 are located in the β2c strand (as part of the β2b-β2c-β2d region; Figure 1.5), which is in the vicinity of the catalytic site (Figure 4.3). Comparison of the chemical structures of the fragments (Table 4.2) shows that they all contain aromatic ring systems, which might form a π-stacking with the imidazole ring of H179. Furthermore, there are multiple H-bond donor and acceptor groups in the fragments. For example, H-bonds between nitrogen atoms of the fragments and the carboxyl group of E180, or backbone carbonyls may form.

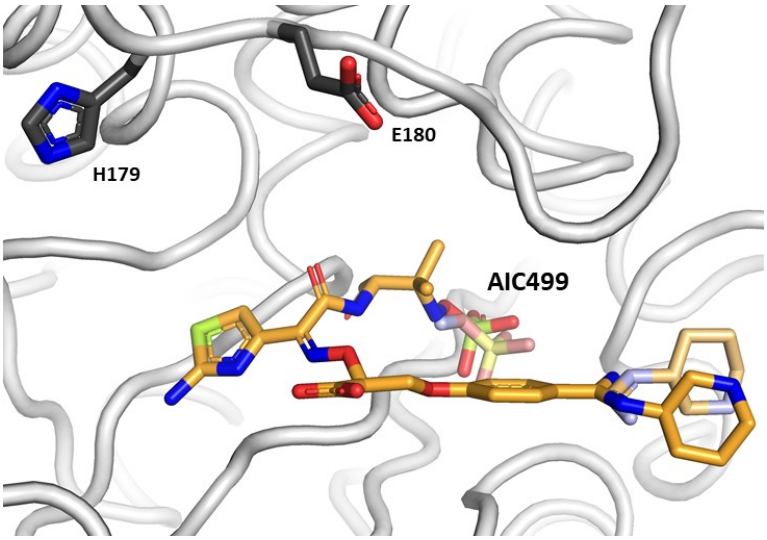


Figure 4.3: Residues H179 and E180 are close to the active site.
In the *Ec*TPd* structure (PDB entry 7ONN; [48]), the side chains of residues H179 and E180 are in close proximity to the active site. The distances between the imidazole ring of H179 and the amino-thiazole (R^4) of AIC499 and between the carboxyl groups of E180 and the linker (R^5) are approximately 12 Å and 11 Å, respectively. Fragments binding towards the active site might be linked with the carboxyl group of R^5 of AIC499 to build an enlarged molecule featuring higher affinities and optimized ADME properties.

Although these fragments were found to bind close to the catalytic site of *Ec*TPd*, F_{comp} between 19% and 65% indicate that these fragments are not or only partly competing with AIC499 in the catalytic site. Therefore, they might be considered for use in a linking approach in FBDD (Figure 1.16C). However, for further development of these hit fragments, structural data is required. As described in Section 4.2.2, NOE spectra, STD NMR experiments, or molecular modeling studies should provide further details on the binding mode.

4.3. Summary of ^{19}F NMR FBDD screening results

In the ^{19}F NMR screen, two libraries containing 681 tri-fluorinated fragments supplied from KeyOrganics and Enamine were screened against His-tagged *Ec*PBP3ΔTM, resulting in 17 identified hit fragments. Thus, the overall hit rate of approximately 2.5% matches reported hit rates [112, 174, 176]. Taking 225 mono-fluorinated fragments into account would lead to a hit rate of 1.9%, which is still matching reported hit rates. Similar $\Delta\delta$ and ΔI for peaks arising from fragments that are structurally similar (such as 5N-395S and 3N-528S, or AIC252849 and AIC252561) showed that with the screening experiments and the established analysis it was possible to reproduce the results and identify fragments with similar chemical structures. A comparison of hit fragment structures with structures of β-lactam antibiotics (Figure 1.10), revealed similarities. For example, the amino-thiazole ring of AIC252561 can be found in aztreonam and several cephalosporins. However, without any structural data it cannot be concluded that this fragment binds to PBP3 similarly to these antibiotics. In addition to the identification of similar chemical structures of hit fragments, peaks arising from impurities, degraded, or aggregated fragments where successfully excluded in the process of the screens by additional assignment and hit validation rounds.

In further investigation of the hit fragments, it was possible to obtain an approximate K_d value for the interaction between His-tagged *Ec*PBP3ΔTM and 5N-395S of 1.23 ± 0.54 mM fitting $\Delta\delta_{obs}$ against the ratio of protein and fragment concentrations and using Equation 4.

Furthermore, MST measurements were performed at AiCuris by Dr. Immanuel Grimm, in which K_d values between 60 µM and 1.4 mM were determined (Table 4.2). These affinities are in the expected range for fragments of 100 µM – 10 mM [86].

In competition binding studies, F_{comp} was used as an indicator for fragment displacement from the catalytic site by the covalent binder AIC499. For further development in FBDD, AIC252627 and AIC252712 are interesting candidates because their interaction with the His-tagged *Ec*PBP3ΔTM was enhanced upon binding of AIC499, indicated by F_{comp} values above 100%. In the case of AIC252712, it is possible that one enantiomer is strongly interacting with AIC499, whereas the other one remains in the free ligand state. Thus, the chemical structure might potentially be extended by the introduction of chemical groups found in these fragments, such as pyrazole or pyridine motifs. However, structural details on the mode of binding would be required. For this, soaking experiments using crystals of *Ec*PBP3ΔTM in complex with AIC499 may yield suitable crystals to solve the protein-bound states of AIC252627 and AIC252712.

Chemical shift perturbation analysis using the optimized *Ec*TPd* sample revealed that binding of AIC252490, AIC2552728, AIC252732 and AIC252901 involved residues H179 and E180, which are located on the β2c strand. With F_{comp} values between 19% and 65% the binding of AIC499 seems to affect binding of these fragments, probably similar to the inhibition modes shown in Figure 4.2 B and C. Additionally, K_d values between 0.75 and 1.22 mM determined by MST measurements at AiCuris by Dr. Immanuel Grimm (no K_d was available for AIC252490) confirm a reasonable affinity between those fragments and *Ec*PBP3. Thus, these fragments are promising starting scaffolds that can be used in growing or merging approaches (Figure 1.16). Dependent on the mode of binding linking with AIC499 might be another option but would lead to relatively large molecules, which might not be able to pass the outer membrane of the bacterial cell wall or be readily synthesized.

For the further development of hit fragments, structural data are required to determine a detailed mode of binding. For this, protein structures in complex with the fragments solved by X-ray crystallography would be optimal to obtain these details. Co-crystallization and soaking experiments can be used to produce protein crystals in complex with the fragments. However, due to their low affinity, crystals might not contain the fragments leading to no or insufficient electron density of the model, as observed for fragments identified in the first screen and Cluster J compounds. Furthermore, 3D NOESY NMR spectra of the protein might help to identify all residues involved in the interaction. Additionally, assigning more peaks in the 2D ^{1}H-^{15}N TROSY-HSQC might identify additional binding pockets in close proximity of the catalytic site of *Ec*PBP3. Improvement of the assignment process is discussed in Section 4.4. Furthermore, performing STD NMR experiments might gain some information on the atoms of the fragments involved in the interaction with the protein. Additionally, *in silico* methods, such as molecular modeling and docking studies with hit fragments are helpful to identify the exact mode of binding if no experimental structural data are available. With these computational studies, similar chemical structures of promising candidates used in an optimized virtual library can be used in an *in silico* screening.

Investigating the binding of hit fragments with PBP3s from other Gram-negative bacteria, such as *P. aeruginosa*, might be an approach to reduce the number of hit fragments, which may be used in further development. With this approach, the development of a broad-band antibiotic that inhibits PBP3s from various Gram-negative bacteria may be feasible.

The results of the initial and the hit validation screens of the 17 hit fragments are shown in Table 4.1 and calculated K_d and F_{comp} values are shown with the solubility in aqueous solutions and the structures of these fragments in Table 4.2.

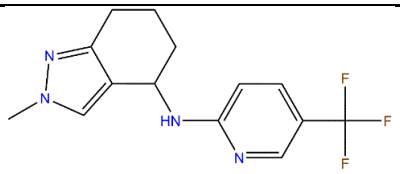
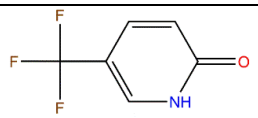
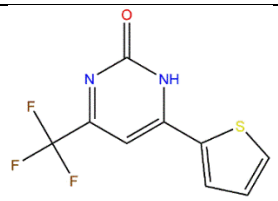
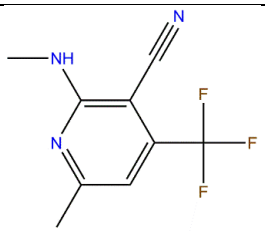
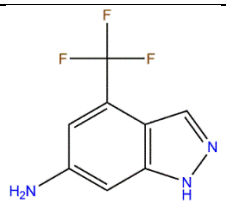
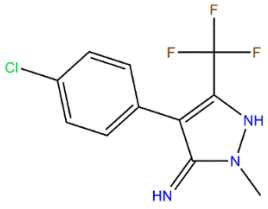
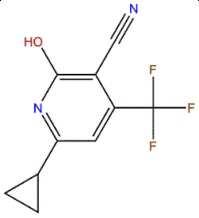
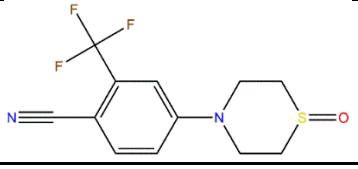
Table 4.1: Summary of the screening results (differences in chemical shift and peak intensity) for the 17 identified hit fragments.

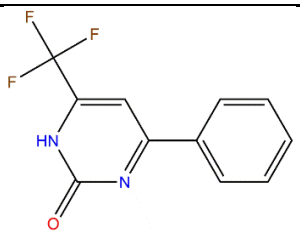
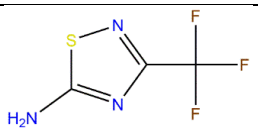
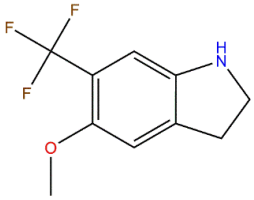
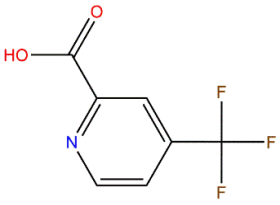
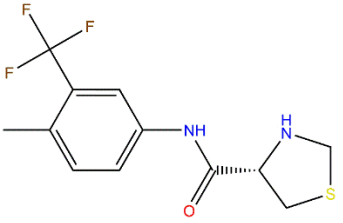
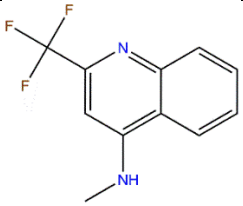
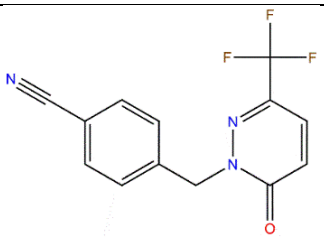
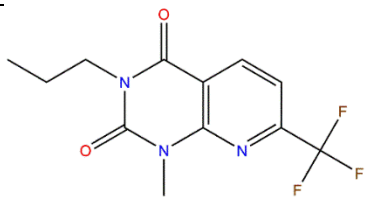
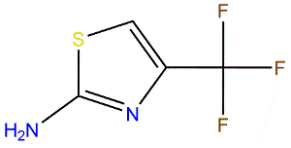
For fragments identified using the library from KeyOrganics, the data of the assignment validation are not shown. The score is based on the developed scoring system used for both libraries (Figure 3.6 and Figure 3.16).

Fragment	Library	Initial screen			Validation screen		
		Score	ΔI [%]	Δδ [Hz]	Score	ΔI [%]	Δδ [Hz]
AIC252712	Enamine	2	18	5	1	28	7
AIC252490	Enamine	1	29	13	2	28	3
AIC252773	Enamine	1	30	5	2	23	9
5N-395S	KeyOrganics	2	34	4	2	38	5
AIC252901	Enamine	2	17	5	2	25	7
AIC252728	Enamine	2	19	8	2	20	11
3N-528S	KeyOrganics	3	38	2	2	27	3
AIC252683	Enamine	3	21	4	2	31	5
AIC252721	Enamine	3	24	3	2	19	10
AIC252849	Enamine	3	34	2	2	56	3
FD-0035	KeyOrganics	1	67	14	3	17	2
FD-0739	KeyOrganics	1	47	19	3	19	4
AIC252732	Enamine	2	19	10	3	20	3
AIC252853	Enamine	3	17	4	3	23	4
AIC252626	Enamine	3	17	3	3	28	3
AIC252541	Enamine	3	19	4	3	24	4
AIC252561	Enamine	3	22	2	3	32	2

Table 4.2: Summary of kinetic and competitive information for the 17 identified hit fragments.

The solubility in aqueous buffer is provided by the respective supplier of the library. * The K_d value of 5N-395S was estimated by fitting $\Delta\delta_{obs}$ of 1D ^{19}F NMR experiments against the ratio of the concentrations (Figure 3.20). Other K_d values were determined by MST for selected fragments at AiCuris by Dr. Immanuel Grimm. F_{comp} was calculated with AIC499 blocking the active site. 100% indicates that the AIC499 does not have an impact on the interaction between the fragment and the protein, whereas 0% indicates, that AIC499 completely displaces or inhibits binding of the fragment. Values above 100% indicate that the AIC499 enhances protein binding of the fragment.

Fragment	Solubility [mM]	K_d [mM]	F_{comp} [%]	Structure
AIC252712	>1	0.06	232	
AIC252490	>1	1.22	24	
AIC252773	>1	0.42	27	
5N-395S	<0.5	1.23*	87	
AIC252901	>1	–	65	
AIC252728	>1	0.95	19	
3N-528S	<0.25	–	105	
AIC252683	>1	/	46	

AIC252721	>1	–	38	
AIC252849	>1	–	16	
FD-0035	>5	1.4	–	
FD-0739	>5	–	–	
AIC252732	>1	0.75	39	
AIC252853	>1	–	72	
AIC252626	>1	–	3	
AIC252541	>1	–	58	
AIC252561	>1	–	38	

4.4. Challenges associated with sequence-specific backbone assignments

Following optimization of the truncated *E. coli* PBP3 construct (i.e., *Ec*TPd*) to obtain a 2D ^{1}H-^{15}N TROSY-HSQC spectrum with satisfactory spectral dispersion of signals and narrow linewidths (Figure 3.28, Table 3.4), 3D heteronuclear NMR experiments were recorded (Figure 1.18 and Table 2.8) to obtain backbone sequence-specific assignments (Figure 3.30). Using combinations of 3D heteronuclear experiments (Figure 1.18 and Section 1.10), 40% of the 2D ^{1}H-^{15}N TROSY-HSQC spectrum was readily assigned (Figure 3.28 and Table A1). The assignments were confirmed by secondary structure analysis using TALOS-N, where the predictions match the secondary structures in the crystal structure (Figure 3.31). Obtaining more than 40% assignments was not possible with the spectra recorded because often peaks with matching chemical shifts were not found when performing the "sequential walk" to unambiguously identify the types of amino acid in spin systems with sufficient length (i.e., string of at least three to four residues). As described in Section 1.10, these stretches of linked spin systems are required to assign peaks in the 2D ^{1}H-^{15}N TROSY-HSQC to sequence-specific amino acids.

Mapping the assignments to the apo-*Ec*TPd* structure (PBD entry 7ONO; [48]), revealed that most of the assigned amino acids are located in loops and α-helices (Figure 4.4A). No assignments of residues in the central β-sheet were made. An explanation for the absence of assignment information for this core region is challenging. Nonetheless, it is possible that nuclei in this core region had unfavorable relaxation properties (e.g., line-broadening due to chemical exchange, unfavorable bond orientation relative to the main axis of the diffusion tensor) that preclude unambiguous assignment using the 3D experiments recorded, for example, missing peaks in 3D spectra.

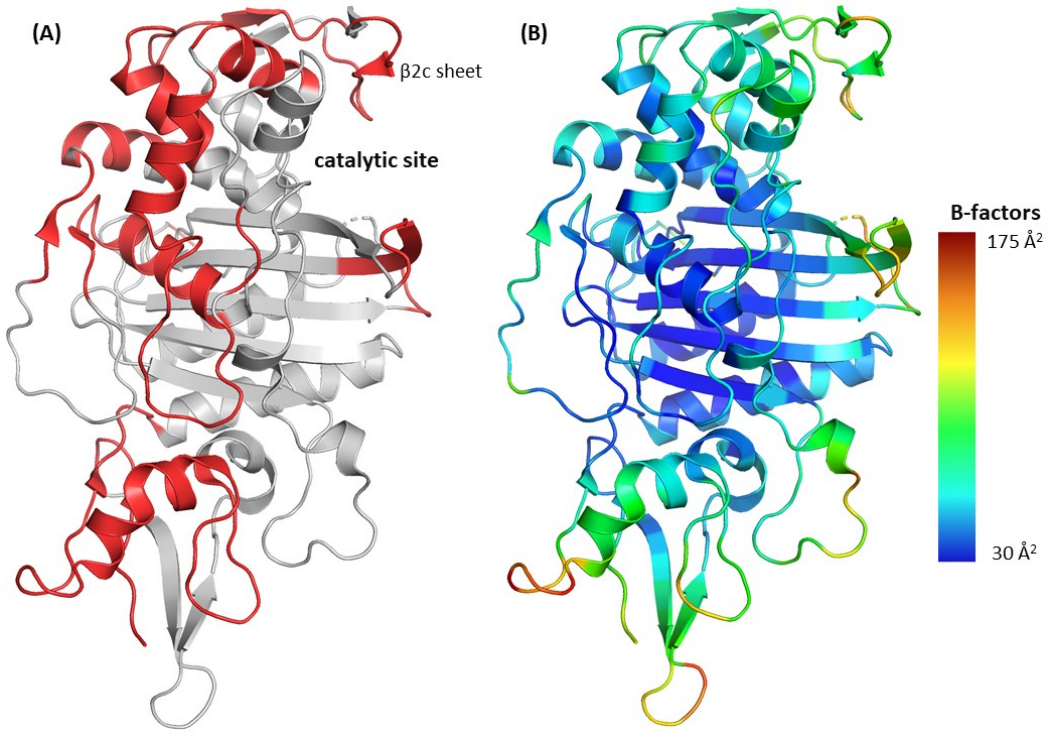


Figure 4.4: Sequence-specifically assigned residues and B-factors in the *Ec*TPd* structure.

PDB entry 7ONO has been used to prepare this figure [48]. (**A**) Residues that have been sequence-specifically assigned to peaks in the 2D ^{1}H-^{15}N TROSY-HSQC are shown in red. The catalytic site and the β2c sheet, which was found to bind some hit fragments (discussed in Section 4.2.3), are labeled. (**B**) Increasing RMS fluctuations of amino acids, which are represented by B-factors obtained from PDB entry 7ONO, were assigned to a color scheme from blue to green to red, where blue represents low and red high B-factors.

To explore the possible reasons for missing assignment information in the core region, RMS fluctuations, represented by B-factors, were calculated for *Ec*TPd* using PBD entry 7ONO (Figure 4.4 B; [48]). In general, the majority of the B-factors with higher values are observed for residues located in loop regions and lower values for most core residues. Interestingly, sequence-specific assignment information was obtained for a large proportion of residues with higher B-factors, indicating that assignment of the more flexible regions was easier because longer chains of linked amide groups with their corresponding C_α, C' and C_β chemical shifts were obtained during the assignment process. In this context, the calculated average backbone (i.e., C_α, C' and N_H atoms) B-factor of assigned residues is 74.3 Å^2, whereas the average B-factor of unassigned amino acids is 61.9 Å^2 (i.e., approximately 17% lower).

Missing assignment information of core residues may also be due to incomplete back-exchange of amide protons of these residues following preparation of deuterated protein material, leading to unobservable correlations in the 3D heteronuclear experiments for these residues. Furthermore, H-bonds found in β-sheets are stronger than those found in α-helices, thus slowing the process of amide exchange [183, 184]. However, the number of backbone amide correlations in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum was 441, which closely matches the expected number of 446 correlations based on the amino acid sequence (Figure 3.29). Nonetheless, differences in peak intensities with up to factor 100 occurred and the per residue solvent accessible surface area (SASA) was determined using PISA and the apo-*Ec*TPd* structure to ascertain whether differences in solvent accessibility between assigned and non-assigned amino acids existed [48, 185]. The average SASA of assigned amino acids was determined to be 47.7 Å^2, whereas the average SASA of unassigned amino acids is 38.5 Å^2, representing an average lower SASA value of approximately 20%. Following from this, incomplete back-exchange of amide protons after preparation of the recombinant [^{2}H, ^{13}C, ^{15}N]-labeled *Ec*TPd* sample as a potential issue for incomplete assignments may be excluded and presumably, weaker peaks in the 2D ^{1}H-^{15}N TROSY-HSQC spectrum are caused by unfavorable relaxation properties, causing issues when searching for starting points in the “sequential walk”. However, there are no data available to distinguish between weaker peaks caused by faster relaxation or caused by incomplete back-exchange. To ensure that the back-exchange is maximized in future samples, an additional step involving partial denaturation of the protein by increased temperature or a chemical denaturant (e.g., using guanidine hydrochloride) may be helpful.

In summary, in many cases it was not possible to follow the “sequential walk”, because no matching chemical shifts were found for spin systems containing three or more residues, presumably because peaks were below or equal to the noise threshold due to unfavorable relaxation properties. Using higher magnetic fields may resolve this issue. Thus, working at 1.2 GHz ^{1}H resonance frequency should increase the sensitivity and resolution for particular 3D heteronuclear experiments when combined with the TROSY principle. Additionally, the presence of 24 prolines in the protein sequence (approximately 6%) is relatively high when compared to the natural abundance of prolines in bacterial proteins (4.61%, [186]) and thus, impede the assignment process. Because prolines do not give rise to any signals in the 2D ^{1}H-^{15}N TROSY-HSQC and related 3D experiments, it is not possible to follow the “sequential walk” through prolines.

Selective amino acid [^{13}C, ^{15}N]-labeling may aid the assignment process. This can be achieved by growing the recombinant bacteria in media that only contains particular labeled amino acids. Although sequence-specific assignment information may not be obtained using this approach, signals arising in the 2D ^{1}H-^{15}N TROSY-HSQC, for example, can be amino acid specifically

assigned, which should complement the backbone assignment process using 3D heteronuclear NMR experiments. Moreover, specific ^{1}H-labeling of methyl groups of isoleucine, leucine or valine in combination with 3D Ile,Leu-(HM)CM(CGCBCA)NH and 3D Val-(HM)CM(CBCA)NH NMR experiments as described by Tugarinov *et al.*, or combinatorial labeling as described by Xun *et al.* should provide amino acid type assignments [187, 188]. Additionally, construction of expression constructs of particular PBP3 domains and thus selective [^{2}H, ^{13}C, ^{15}N]-labeling of these domains and re-incorporation with the remaining unlabeled PBP3 protein by either chemical or protein-catalyzed ligation reactions may be feasible [189]. Such domain labeling reduces the complexity of the NMR spectra significantly to facilitate sequence-specific assignment of the spectra. Such an approach is particularly useful when spectral overlap hampers unambiguous assignment of resonances and the NMR data assignments are made in the context of the full protein. Additional NMR experiments may also facilitate completion of the assignment process. Here, the 3D ^{1}H-^{15}N NOESY-HSQC should be helpful in combination with structural data obtained by the crystal structures to identify starting points in the protein sequence. For example, because amide-amide NOEs can be observed between residues *i* and *i*+2 and *i*+3 in α-helices, this approach may support other 3D experiments and facilitate the identification of starting points. Additionally, 4D experiments such as 4D TROSY-HNCACO and 4D TROSY-HNCOCA should aid assignments by reducing spectral overlap [190]. Furthermore, in combination with additional experiments, programs for automated assignments, such as MARS and Nexus, which are both implemented in the CcpNMR package, may facilitate assignment of the 3D NMR spectra [149, 191].

4.5. Cluster J compounds are suitable lead compounds for drug development

For the determination of specific interactions between PBP3 from *E. coli* and *P. aeruginosa* and the Cluster J compounds, various biophysical methods were used. First, chemical shift perturbation analysis was performed using *Ec*TPd* because Cluster J compounds showed inhibitory activity against *Ec*PBP3 (data not shown) and the NMR experiments for this analysis were established (Section 3.3). Unfortunately, observed changes in the 2D ^{1}H-^{15}N TROSY-HSQC spectra were inconclusive to carry out further analysis (Figure 3.32). Presumably, Cluster J concentrations of 50 µM that were used in the experiments were too low to cause larger changes in the spectrum. Additionally, some precipitant was observed during the experiments, which was initially hypothesized to be caused by the Cluster J compounds and their poor solubility. However, the solubility was shown to be >100 µM in the used buffer. Therefore, it was hypothesized that the precipitant arose from the formation of complexes between Cluster J compounds and *Ec*TPd*. This theory was confirmed by fluorescence experiments, where no specific 1:1 interaction between Cluster J compound 35 and *Ec*TPd* was observed (Figure 3.34 and Figure 3.35). Following from these results indicating non-specific binding, it was hypothesized that the Cluster J compounds are specifically interacting with PBP3, but due to some conformational changes the complex aggregates, which was observed as a precipitant. To confirm this hypothesis, SPR was used, because with this method the protein is immobilized on the surface of the CM5 chip and thus, is unlikely to readily aggregate. Because of the higher affinities when compared to *Ec*PBP3ΔTM, the SPR experiments were performed using *Pa*PBP3ΔTM, exemplarily shown for Cluster J compound 35 in Figure 3.36. The resulting K_d values between 2.7 µM and 470 µM are published in López-Pérez *et al.* and summarized in Table 3.5 [85].

4.6. X-ray crystallography provides structural insights into PBP3 structures from *E. coli* and *P. aeruginosa* and the AIC499 binding mode

X-ray crystallography studies were performed to obtain structural information about PBP3 from *E. coli* and *P. aeruginosa*. Additionally, soaking experiments were performed using bromo- and iodo-pyrazole to identify hydrophobic binding pockets on the surface of the protein [170], Ta_6Br_{12} to potentially improve the quality of the diffraction data and use the data for phasing purposes, and the identified hit fragments 3N-528S and 5N-395S to gain structural data, which are important in further steps of FBDD. Unfortunately, these soaking experiments remained unsuccessful, because the resulting sample yielded diffraction data with resolutions lower than 4 Å. In the case of Cluster J compound soaking experiments, the resolution of the diffraction experiment was sufficient for analysis, but it was not possible to identify the compounds in the electron density. Therefore, the structures were used as *de facto* apo structures featuring different unit cells and lattice contacts.

Besides the above-mentioned molecules that were used for soaking experiments, the novel β-lactam compound AIC499 was used in co-crystallization experiments to obtain structural insights in the binding mode and potentially suggest changes that can be performed on the molecule to improve the protein inhibition. In this context, the crystal structures of PBP3 in complex with AIC499 reveal that the extent of non-covalent interactions with AIC499 differs slightly between the *Pa*PBP3 and *Ec*PBP3 (Figure 3.38 and Figure 3.40). For example, a total of eleven hydrogen bonds are formed with *Pa*PBP3, whereas only seven are formed in the *Ec*PBP3 complex, which may reflect that *Pa*PBP3 was used as the target during optimization. In both cases, the amino-thiazole (R^4) is the functional group exhibiting the highest density of interactions with PBP3. Besides the hydrogen bonds established with residues E291 and R489 (E304 and K499 in *Ec*PBP3) by the nitrogen-containing half of the heterocycle, the less polar moiety engages in an aromatic network with the side chains of Y409 and Y407 (Y419 and F417). In a similar fashion, the phenyl ring of the head group (R^1) is in intimate contact with both proteins, being sandwiched between an aromatic cluster (Y511, Y540, Y541 in *Ec*PBP3; Y503, Y532, F533 in *Pa*PBP3) and the aliphatic V344/V333 side chain. The piperidine ring of the head group, on the other hand, seems to have relatively low impact on the binding of AIC499; in fact, it appears to be the most dynamic part with two discernible conformations in either complex. In contrast to Y541 in *Ec*PBP3, the homologous F533 in *Pa*PBP3 is disordered, and the piperidine moiety instead contacts Y532 or V333. While this modification was found to improve the pharmacodynamic and pharmacokinetic properties of AIC499 compared to its predecessors (data not shown), its benefit is thus not immediately obvious from the X-ray structures alone. It was hypothesized that the piperidine substituent may instead impact the kinetics or thermodynamics of the primordial Michaelis complex; clarification of this issue will require further experimentation, applying catalytically inactive PBP3 variants. Notable differences between *E. coli* and *P. aeruginosa* complexes are observed for the carboxyl moiety of the linker (R^5); in *Pa*PBP3 the side chain of R489 forms a salt bridge with this group, whereas the homologous K499 in *Ec*PBP3 is partly disordered. This is consistent with the notion that arginine is more versatile in establishing electrostatic interactions because of its geometric properties [192].

Direct comparison between AIC499 and aztreonam acyl-enzyme intermediates (Figure 4.5) reveals significant differences at the extremities of the molecules, whereas the core structure including the amino-thiazole moiety and the carboxyl group in the linker are conserved. Specifically, note the dramatic increase in contact area because of the R^1 head group of AIC499 (panels A and C), which mostly participates in hydrophobic interactions. Conversely,

replacement of the terminal sulfonate by sulfate (panels B and D) is likely to provide additional freedom for electrostatic interactions, but an equally relevant modification in this region may be the additional methyl substituent (R^3); in all AIC499 complexes investigated, one of these methyl groups contributes to the hydrophobic cluster orchestrated by the phenyl ring, effectively linking both branches of the molecule. Irrespective of the differences outlined above, binding of both compounds entails a significant increase in thermal stability of PBP3. In the case of *E. coli* PBP3 (Figure A9), the midpoint of thermal unfolding (T_m) determined via differential scanning fluorometry (DSF) increased from 58.6 to 63.4 °C for aztreonam and to 65.4 °C for AIC499, whereas for the *P. aeruginosa* protein (Figure A10) the value rises from 46 to 54 °C for both complexes. The temperature-dependent shifts in the 350 nm/330 nm fluorescence ratio were closely mirrored by increases in turbidity, supporting the view that they do not merely reflect local changes in the environment of aromatic residues, but they correspond to real unfolding transitions.

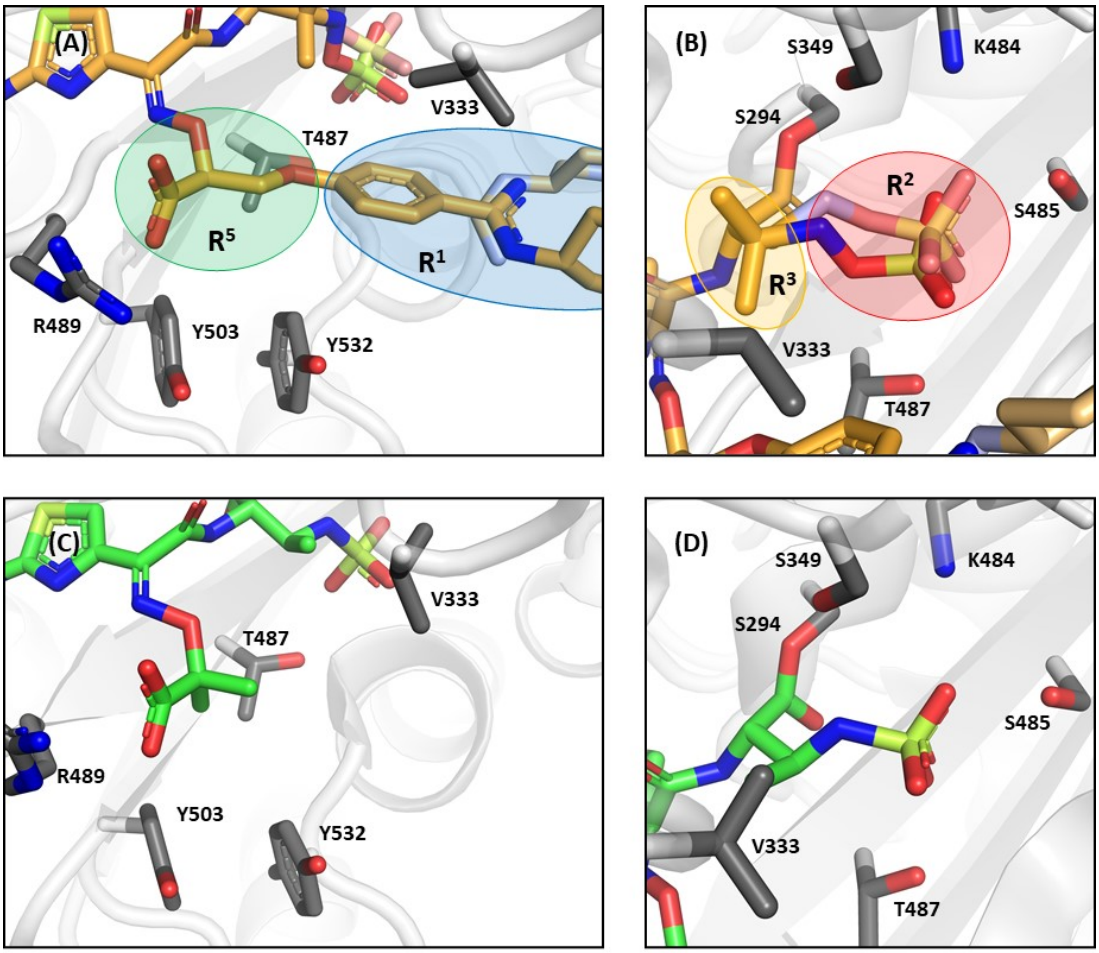


Figure 4.5: Comparison of aztreonam and AIC499 binding to *Pa*PBP3.

Panel (**A**) shows a close-up view of the linker (R^5; green) and head group (R^1; blue) of AIC499 in complex with *Pa*PBP3ΔTM (PDB entry 7ONK; [48]), while (**B**) focuses on the sulfate (R^2; red) and dimethyl group (R^3; orange). The analogous views of aztreonam in complex with *Pa*PBP3ΔTM (PDB entry 3PBS; [51]) are represented in (**C**) and (**D**), respectively. Refer to Table A3 for a complete list of distances between protein side chains and the aztreonam molecule.

The crystallographic investigation revealed a considerably higher flexibility of *Ec*PBP3 when compared with its *P. aeruginosa* orthologue, particularly in the extended *N*-terminal portions of the head and anchor subdomains. This observation is consistent with data published previously on the *E. coli* apo form [47]. The reasons for this difference are unclear but are likely

to reflect differences in the properties of one or more interaction partners in the multiprotein complex (the divisome) that PBP3 is chiefly involved in. To circumvent the drawbacks associated with enhanced dynamics in crystallographic studies, the newly introduced truncated construct for the *E. coli* protein (*Ec*TPd*) was used and yielded crystals with significantly improved diffraction quality. Notably, the new design differs from the one published previously by Bellini et al. (PDB entry 6HZQ; [49]). While those authors used a straightforward *N*-terminal truncation, keeping residues 234-588, the linker subdomain was included in the newly designed construct because it intimately interacts with the core TPase fold and is unlikely to be affected by enhanced dynamics. Since the linker subdomain is discontinuous, this involved removal of two internal sequence segments, in addition to the far *N*-terminus. Indeed, the structural data thus obtained for our truncated *Ec*PBP3 variant (*Ec*TPd*) allowed to trace the entire linker region, which is the vast majority of *N*-terminal residues included in the previous structure of soluble *Ec*PBP3ΔTM (Figure 4.6), yet avoiding the negative impact of excessive dynamics on crystal packing and diffraction quality.

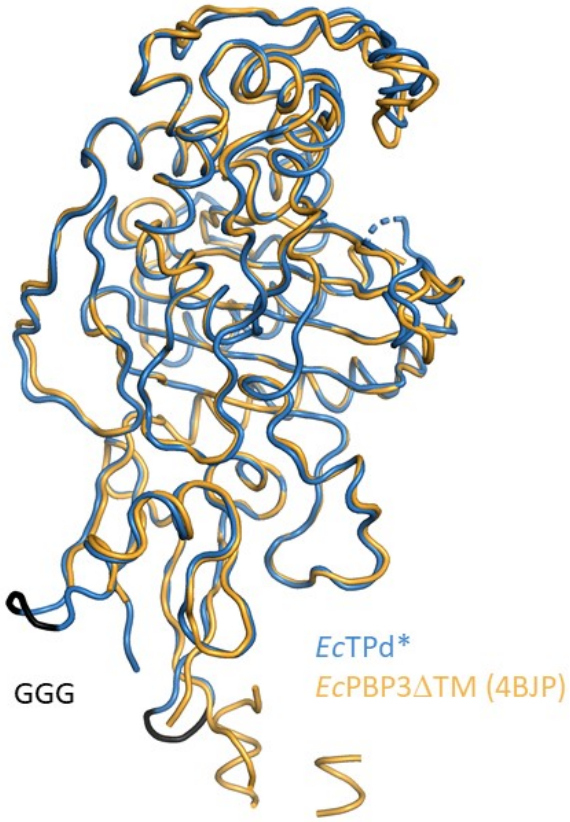


Figure 4.6: Superposition of the X-ray structures of *Ec*TPd* and *Ec*PBP3ΔTM.

PDB entries 7ONO (blue; [48]) and 4BJP (gold; [47]) were used to prepare this figure, demonstrating the similar coverage of ordered structure[47, 48]. The GGG linkers replacing the deleted segments in *Ec*TPd* are colored black.

Notably, the presence or absence of the linker subdomain has a significant influence on adjacent structural features of the TPase fold. For example, it was possible to resolve the region between P279 and R297, which is not included in the previous *Ec*TPd structure, presumably because it is stabilized by interactions with the linker region. In addition, residues

R297-T300 adopt a non-native extended conformation in PBP entry 6HZQ, instead of the partly helical structure observed in our model. This local restructuring appears to propagate into other parts of the protein, especially the segment connecting helices α6 and α8; the region G392-W407 is disordered in PDB entry 6HZQ, whereas in our structure it was modeled completely, including most of the side chains. Notably, the *Ec*TPd* structure is consistent with previous data for soluble *Ec*PBP3 (PBP entry 4BJP; [47]), confirming it closely reflects the native fold. In the vicinity of the active site the most notable differences are found in the β5-α11 region; it is entirely ordered and orientated towards the active site in PBP entry 6HZQ, resembling what is commonly found after ligand binding. The reasons for this discrepancy are unclear, given that this region is not restrained by lattice contacts. In contrast, the notoriously flexible parts of the β2b-β2c-β2d region and the β3-β4 hairpin appear more similar between the two *Ec*PBP3 structures. Taken together, these observations support the view that the linker region contributes significantly to the stable fold of the TPd of *Ec*PBP3 and, by extension, its orthologues in other species. Therefore, it appears advisable not to exclude this segment when designing truncated constructs for biophysical applications.

Regarding the *Pa*PBP3 orthologue, the newly identified crystal forms expand the repertoire of structural information on the apo protein, allowing for a more reliable assessment of protein dynamics and the impact of lattice interactions. For example, the β3-β4 region, while clearly showing enhanced dynamics, is traced to a larger extent than described previously, indicating a β-hairpin conformation that protrudes from the TPase fold. Previous structures of apo-*Pa*PBP3ΔTM, e.g., PDB entry 6HZR, were lacking coordinates for the β3-β4 loop, and it was assumed that this region is only stabilized upon acylation of the catalytic serine. For the β5-α11 region electron density suggests a high degree of conformational freedom with our crystal form 1 and PDB entry 6HZR representing two extremes of the conformational space sampled, i.e., the β5-α11 loop bends in opposite directions.

The structural data presented here may guide future efforts to further improve the properties of AIC499. One exciting option, inspired by the remarkable U-shape of protein-bound AIC499, is introduction of a covalent linkage or salt bridge between R^1 and R^3, resulting in a cyclized compound with enhanced rigidity. Such a modification may strongly affect PBP3 affinity and β-lactamase stability as well as impact pharmacokinetics. Furthermore, extending the scaffold of AIC499 might be an option using identified hit fragments. As discussed in Section 4.2, AIC252627 and AIC252712, which showed an enhanced interaction with the protein in presence of AIC499, and AIC252490, AIC2552728, AIC252732 and AIC252901, which were found to interact with residues in the β2c sheet in close proximity to the catalytic site, might be good candidates for linking, growing and/or merging chemical structures.

5. Conclusion

In this thesis ^{19}F NMR experiments were established as a tool to perform FBS using two libraries of fluorinated fragments. With this setup, it should be possible to perform analogous studies against other proteins and analyze the data in a similar fashion. Targeting *Ec*PBP3, identified 17 fragments out of 681 fragments screened (approximately 2.5%) as hits and these hits were confirmed by orthogonal methods. Furthermore, to gain additional structural information on binding fragments, NMR studies were performed with the aim to assign the 2D ^{1}H-^{15}N TROSY-HSQC of the PBP3 from *E. coli*. Because of the protein size of 61 kDa and severe signal overlap in the central region of the 2D ^{1}H-^{15}N HSQC spectrum, a truncated *Ec*TPd* construct was used to perform multi-dimensional NMR experiments. Approximately 40% of the peaks in the 2D ^{1}H-^{15}N TROSY-HSQC were assigned and assignments were confirmed by comparing the results from secondary structure analysis of the chemical shifts with the X-ray structure. Based on chemical shift perturbation assays, some of the hit fragments identified in the ^{19}F NMR screening interact with the β2b-β2c-β2d region of *Ec*PBP3, which is close to the active site; these fragments might be promising candidates in further drug development. In competitive binding studies two fragments were found to show enhanced binding in the presence of covalently bound AIC499. These fragments and the fragments found to bind in the β2b-β2c-β2d region might be used to extend the AIC499 scaffold to yield an enlarged molecule with beneficial properties.

Moreover, lead molecules at further stages in drug development were investigated. Cluster J compounds, which were initially found by performing an HTS approach with a newly designed fluorescence assay against *P. aeruginosa*, were studied by biophysical methods to obtain thermodynamic information. In SPR experiments, the K_d values were determined to be between 6.4 µM and 470 µM. Accounting for the results of other methods, it is hypothesized that in addition to specifically binding the protein with high affinity, non-specific binding occurs that causes aggregation of the protein-complex.

Finally, structural data on the mode of interaction between the monobactam AIC499 and the PBP3 from *P. aeruginosa* and *E. coli* were obtained by determining the crystal structures of apo proteins and AIC499 complexes. These data might help to improve the properties of the drug compound; for example, the U-shape of the AIC499 molecule might be used to create a cyclic molecule with improved properties. In this context, *Pa*PBP3ΔTM was found to crystallize in unit cells that have not been reported previously. Solving the structures revealed some novel insights; for example, the β3-β4 loop was assumed to be structured only upon binding a substrate or inhibitor but was shown here to be partly ordered in a β-hairpin conformation.

6. List of publications

During this thesis the following research papers were published:

- **López-Pérez, A.**, *et al.*, Discovery of Pyrrolidine-2, 3-diones as Novel Inhibitors of P. aeruginosa PBP3. Antibiotics, 2021. **10**(5): p. 529.

- **Freischem, S.**, *et al.*, Interaction Mode of the Novel Monobactam AIC499 Targeting Penicillin Binding Protein 3 of Gram-Negative Bacteria. Biomolecules, 2021. **11**(7): p. 1057.

The research paper Freischem *et al.* was used in this thesis with the following contributions:

- Conceptualization: All authors contributed to the conceptualization of the research paper.

- Methodology, formal analysis and investigation: All crystals were prepared by Stefan Freischem. X-ray crystallography data were recorded by Stefan Freischem with the supervision by Dr. Oliver Weiergräber. These data were analyzed by Stefan Freischem and Dr. Oliver Weiergräber. MST data were recorded by Dr. Immanuel Grimm and Dr. Arancha López-Pérez.

- Writing: The original draft was written by Stefan Freischem and Dr. Oliver Weiergräber and reviewed and edited by all authors prior to submission.

- Visualization: All figures were prepared by Stefan Freischem.

7. Acknowledgments

Throughout working on my dissertation, I received a lot of support and assistance.

First, I would like to thank the Institute of Biological Information Processing – Structural Biochemistry (IBI-7) for the opportunity to perform my PhD thesis as part of an industrial cooperation project. In particular, I would like to thank Prof. Dr. Dieter Willbold for founding, providing a lab and office space and examining my thesis.

Furthermore, I am very thankful for the support I received by my supervisors apl. Prof. Dr. Andrew Dingley and PD Dr. Oliver Weiergräber. They were always available for me in case of any question, taught me how to use the instruments, assisted me with the analysis and writing and were open for fruitful discussions.

Moreover, I would like to thank AiCuris for the opportunity to work with a pharmaceutical company. I am especially grateful to Dr. Immanuel Grimm who was taking over the project from AiCuris' side for the great discussions. He always supported me in internal decisions, performed additional experiments and provided protein samples. Our time at the Fragments 2018 conference in Cambridge was very enjoyable. Thanks to Christian Dilk, who worked in the laboratory, protein samples were always provided without any issues. Furthermore, I appreciate the important discussions with Dr. Arancha López-Pérez, Dr. Burkhard Klenke, Dr. Cuong Vuong, Dr. Daniela Rausch, Dr. Alexander Birkmann, Dr. Alastair Donald, Dr. Helmut Buschmann, Dr. Irith Wiegand and Dr. Kamal Kumar. Moreover, I would like to thank Dr. Thomas Goldner for the invitations to the summer and winter schools organized at AiCuris.

Regarding the practical work, I received sample support by Marc Sevenich with the SPR experiments, by Dr. Iman Abdollahzadeh with the fluorescence experiments, by Dr. Joachim Granzin with the crystal preparation, shipment and data collection and by Dr. Karen Hänel with general questions in the lab and always having an open ear.

I am grateful to the beamline scientists at the European Synchrotron Radiation Facility (ERSF; Grenoble, France) as well as at the Deutsches Elektronen-Synchrotron (DESY; Hamburg, Germany) for providing assistance with the use of beamlines.

Access to the Biomolecular NMR Centre, jointly run by Forschungszentrum Jülich and Heinrich-Heine-Universität Düsseldorf, is acknowledged. In this context, technical support with the NMR spectrometer by Kevin Bochinsky, Dr. Rudi Hartmann, and Dr. Matthias Stoldt was very much appreciated.

I would like to thank Dr. Sabine Willbold for the assistance and providing the NMR spectrometer at the Central Institute for Engineering, Electronics and Analytics – Analytics (ZEA-3) to perform the ^{19}F NMR screening.

Furthermore, I would like to thank all my other colleagues for the good work environment at the IBI-7 so that it was always a pleasure to come to Jülich.

Last, but not least, I would like to thank my partner, Anni, and my parents for their support, which I could always rely on.

8. Appendix

```
bruk2pipe -in ../ser -verb \
    -xN          3072          -yN          512                \
    -xT          1536          -yT          256                \
    -xMODE       DQD           -yMODE       Echo-Antiecho      \
    -xSW         11160.714     -ySW         2129.472           \
    -xOBS        700.30329225  -yOBS        70.9691208920025   \
    -xCAR        4.686         -yCAR        117.509            \
    -xLAB        HN            -yLAB        N                  \
    -ndim        2             -aq2D        States             \
    -out A.DAT -verb -ov
nmrPipe -in A.DAT -verb                                        \
| nmrPipe -fn POLY -time -auto                                 \
| nmrPipe -fn SP -off 0.35 -end 0.95 -pow 2.0 -c 0.500         \
| nmrPipe -fn ZF -size 4096                                    \
| nmrPipe -fn FT -verb                                         \
| nmrPipe -fn PS -p0 152.6 -p1 -24474.3035888671920 -di        \
| nmrPipe -fn TP                                               \
| nmrPipe -fn LP -fb -ord 16 -pred 128                         \
| nmrPipe -fn SP -off 0.35 -end 0.95 -pow 2.0 -c 0.500         \
| nmrPipe -fn ZF -size 2048                                    \
| nmrPipe -fn FT -verb                                         \
| nmrPipe -fn PS -p0 0.0 -p1 0.0 -di                           \
| nmrPipe -fn TP                                               \
| nmrPipe -fn MED -nw 400                                      \
| nmrPipe -fn SET -x1    1 -xn  200                            \
| nmrPipe -fn SET -x1 3897 -xn 4096                            \
| nmrPipe -fn EXT -x1 5.5ppm -xn 11.5ppm -sw                   \
| nmrPipe -out B.DAT -di -ov -verb
pipe2azara B.DAT 20190411_TPd_PBP3_EC_HSQC_3
```

Figure A1: Typical script used to process 2D ^{1}H-^{15}N TROSY-HSQC spectra.

In the first part of the script the file format is converted from the Bruker file format into the nmrPipe file format (bruk2pipe). For this, the actual size in file (xN / yN), the number of points in the time domain (xT / yT), the mode of accusation (xMODE / yMODE), the spectral width (xSW / ySW), the observed base frequency for 0 ppm (xOBS / yOBS), the offset (xCAR / yCAR), the axis label (xLAB / yLAB), and the number of dimensions (ndim) need to be defined. Once converted into the nmrPipe format, the first dimension (x, in this case ^{1}H) can be processed. First, a polynomial is subtracted from the time domain to correct for the solvent (-fn POLY). A shifted sine bell function (-fn SP) and a zero-filling (-fn ZF) are applied to optimize the quality of the data, before Fourier transformation is performed to convert the first dimension into the frequency domain and the data is phased (-fn PS). The transpose function (-fn TP) exchanges the x and y dimensions. Linear prediction (-fn LP) is performed on the ^{15}N dimension data. After repeating the steps, a median baseline correction (-fn MED) is applied, the first and last points of the data are set to constant (-fn SET) and the spectral region of interest is extracted (-fn EXT). Finally, the data are converted into the azara file format which can be displayed with CcpNMR analysis.

```
bruk2pipe -in ../ser                                                    \
  -aswap -decim 1792 -dspfvs 20 -grpdly 67.9841766357422                \
  -xN 2048              -yN 4                 -zN 163                   \
  -xT 1024              -yT 4                 -zT 163                   \
  -xMODE DQD            -yMODE Real           -zMODE Real               \
  -xSW 11160.714        -ySW 2129.472         -zSW 2113.271             \
  -xOBS 700.3032927     -yOBS 70.9691208920025 -zOBS 176.121181302841   \
  -xCAR 4.667           -yCAR 117.490         -zCAR 176.417             \
  -xLAB HN              -yLAB 15N             -zLAB CO                  \
  -ndim 3               -aq 2D States                                   \
| nmrPipe -fn MAC -macro $NMRTXT/bruk_ranceY.M -noRd -noWr              \
| pipe2xyz -x -out ./fid/nus%04d.fid -verb -ov -to 0
```

Figure A2: Exemplary fid.com script used to convert data stored in the ser file format to the xyz format.

The number of points recorded for the y dimension (yN) has to be four and is added later on in the reconstruction of the spectrum, while the points in the z dimension (zN) represents the number of points used in the nuslist. The nuslist contains sampled points in the acquisition that are shown as blue points in Figure 1.21. In this script the sweep widths (xSW, ySW, and zSW), the offset frequencies (xOBS, yOBS, and zOBS), and the offsets in ppm (xCAR, yCAR, and zCAR) are inputted. In the final step (last line) the data are converted to the xyz file format by the program pipe2xyz, which is required in the further processing.

```
xyz2pipe -in fid/nus%04d.fid -x                 \
| nmrPipe  -fn SP -off 0.35 -end 0.98 -pow 2.00 -c 0.500  \
| nmrPipe  -fn ZF -size 4096 -auto              \
| nmrPipe  -fn FT -verb                         \
| nmrPipe  -fn PS -p0 -39.6  -p1 0.0 -di        \
| pipe2xyz -ov -out yzx/nus%04d.nus -z
```

Figure A3: Exemplary ft1.com script used to process the ^{1}H dimension.

After performing a sine-bell function (-fn SP) and zero-filling (-fn ZF) on the FID, Fourier transformation (-fn FT) was performed, and the data phased (-fn PS) in the zeroth (p0) and first order (p1), and finally stored in the xyz file format.

```
set F = $1
set in = $F:t
set out = $F:t:r.phf
echo $in $out
istHMS -dim 2 -incr 1 -xN 102 -yN 64 -user 1  \
   -itr 400 -verb 0 -ref 0 -vlist ./nuslist \
   < ./yzx/${in} >! ./yzx_ist/$[30]
```

Figure A4: The ist.csh script is used for reconstruction of the data performed by the program istHMS.

xN and yN represent the number of scans used in the experiment for the ^{15}N and ^{13}C dimensions, respectively. The nuslist contains the points that are recorded in the experiment (see Figure 1.21).

```
xyz2pipe -in rec/nus%04d.ft1 -x                    \
| nmrPipe  -fn SP -off 0.50 -end 0.80 -pow 2.00 -c 0.500    \
| nmrPipe  -fn ZF -auto                            \
| nmrPipe  -fn FT -verb                            \
| nmrPipe  -fn PS -p0 180.0 -p1 0.0 -di            \
| nmrPipe  -fn TP                                  \
| nmrPipe  -fn LP -fb -ord 24 -pred 64             \
| nmrPipe  -fn SP -off 0.40 -end 1.00 -pow 2.00 -c 1.00    \
| nmrPipe  -fn ZF -size 256 -auto                  \
| nmrPipe  -fn FT -alt -verb                       \
| nmrPipe  -fn PS -p0 -119.8 -p1 151.4 -di         \
| nmrPipe  -fn TP                                  \
| nmrPipe  -fn HT -auto                            \
| nmrPipe  -fn PS -inv -p0 180.0 -p1 0.0           \
| nmrPipe  -fn FT -inv -verb                       \
| nmrPipe  -fn ZF -inv                             \
| nmrPipe  -fn SP -inv -off 0.50 -end 0.80 -pow 2.0 -c 0.500    \
| nmrPipe  -fn ZTP                                 \
> rec/data.pipe
pipe2xyz -in rec/data.pipe -out rec/nus%04d.ft2 -x
```

Figure A5: The ft23.com script is used to perform a linear prediction (-fn LP) on the ^{13}C dimension.

Previously, after performing a sine-bell function (-fn SP) and zero-filling (-fn ZF), a Fourier transformation (-fn FT) is performed on the 15N dimension. Using -fn TP rotates the dimensions so that in the ^{13}C dimension can be processed. Afterwards, the transpose function (-fn TP) is used to again to perform a Hilbert transformation (-fn HT) on the ^{15}N dimension, converting it back to the time domain.

```
xyz2pipe -z -in rec/nus%04d.ft2 -verb              \
| nmrPipe -fn LP -ps0-0 -ord 24 -pred 64           \
| nmrPipe -fn SP -off 0.40 -end 1.00 -pow 2.0 -c 0.500    \
| nmrPipe -fn ZF -size 256 -auto                   \
| nmrPipe -fn FT -verb                            \
| nmrPipe -fn PS -p0 0.0 -p1 0.0 -di              \
| nmrPipe -fn TP                                  \
| nmrPipe -fn MED -nw 200                         \
| nmrPipe -fn SET -x1    1 -xn  100               \
| nmrPipe -fn SET -x1 3997 -xn 4096               \
| nmrPipe -fn EXT -x1 5.5ppm -xn 11.0ppm -sw      \
| nmrPipe -fn TP                                  \
| nmrPipe -fn ZTP                                 \
| nmrPipe -fn TP                                  \
| pipe2xyz -x -out rec/nus%04d.ft3 -verb -ov
```

Figure A6: The ft3.com script is used to perform a linear prediction (-fn LP) on the ^{15}N dimension and convert the time domain to the frequency domain by Fourier transforming (-fn FT).

A sine-bell function (-fn SP) and zero-filling (-fn ZF) are applied beforehand and the phase is adjusted (-fn PS) afterwards. Before storing the data in the xyz file format a median baseline is applied (-fn MED), the important part of the ^{1}H dimension is extracted (-fn EXT), and the dimensions are transposed with the -fn TP and -fn ZTP functions.

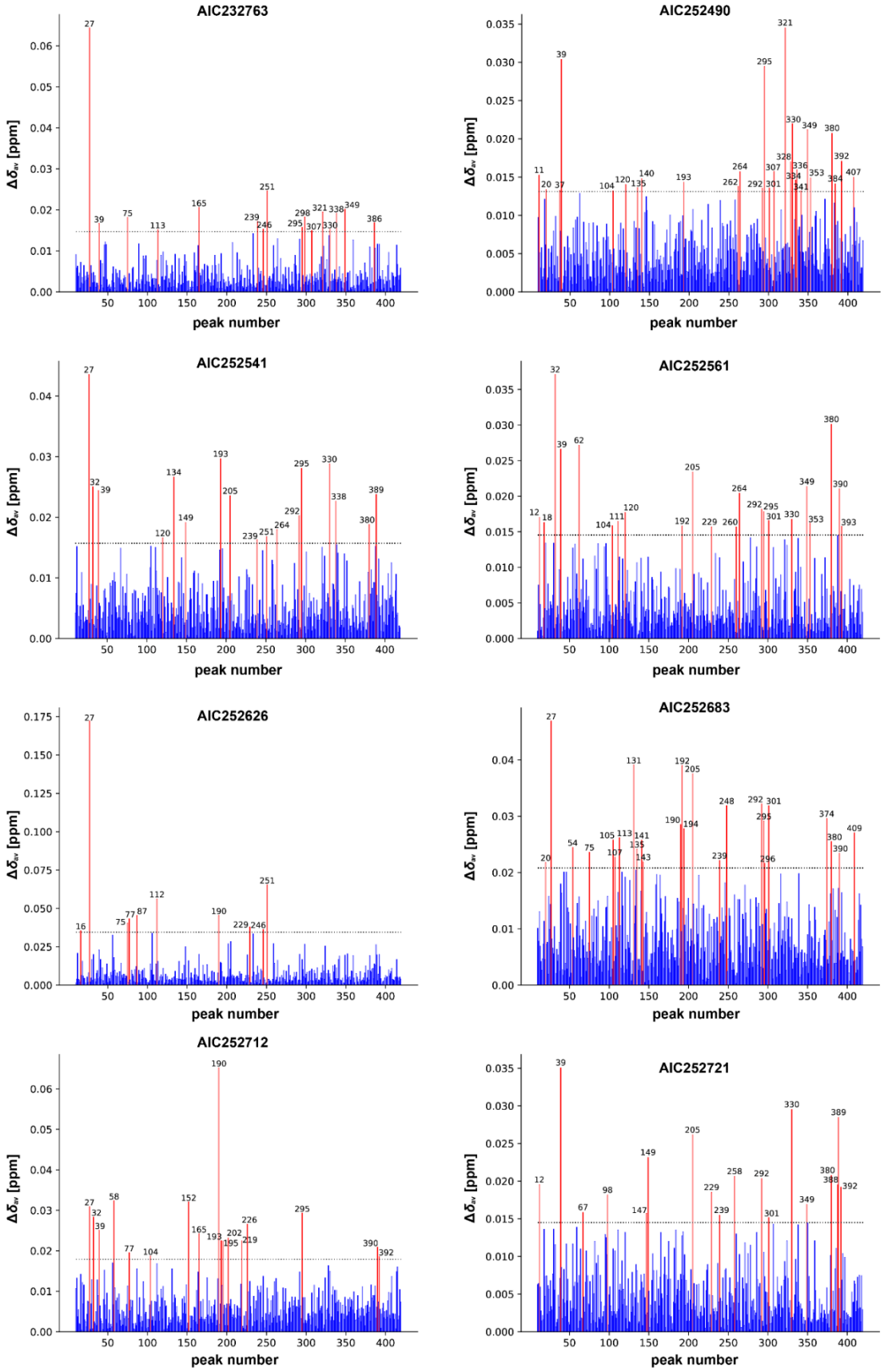


Figure A7: Perturbation shift analysis of AIC232763, AIC252490, AIC252541, AIC252561, AIC252626, AIC252683, AIC252712 and AIC252721.

$\Delta\delta_{av}$ was calculated using Equation 5. 3σ was used as the threshold (dotted line) to identify peaks that showed the largest changes in chemical shift (red). These peaks may represent residues that are involved in interaction with the fragment. Numbers above the bars indicate peak assignments.

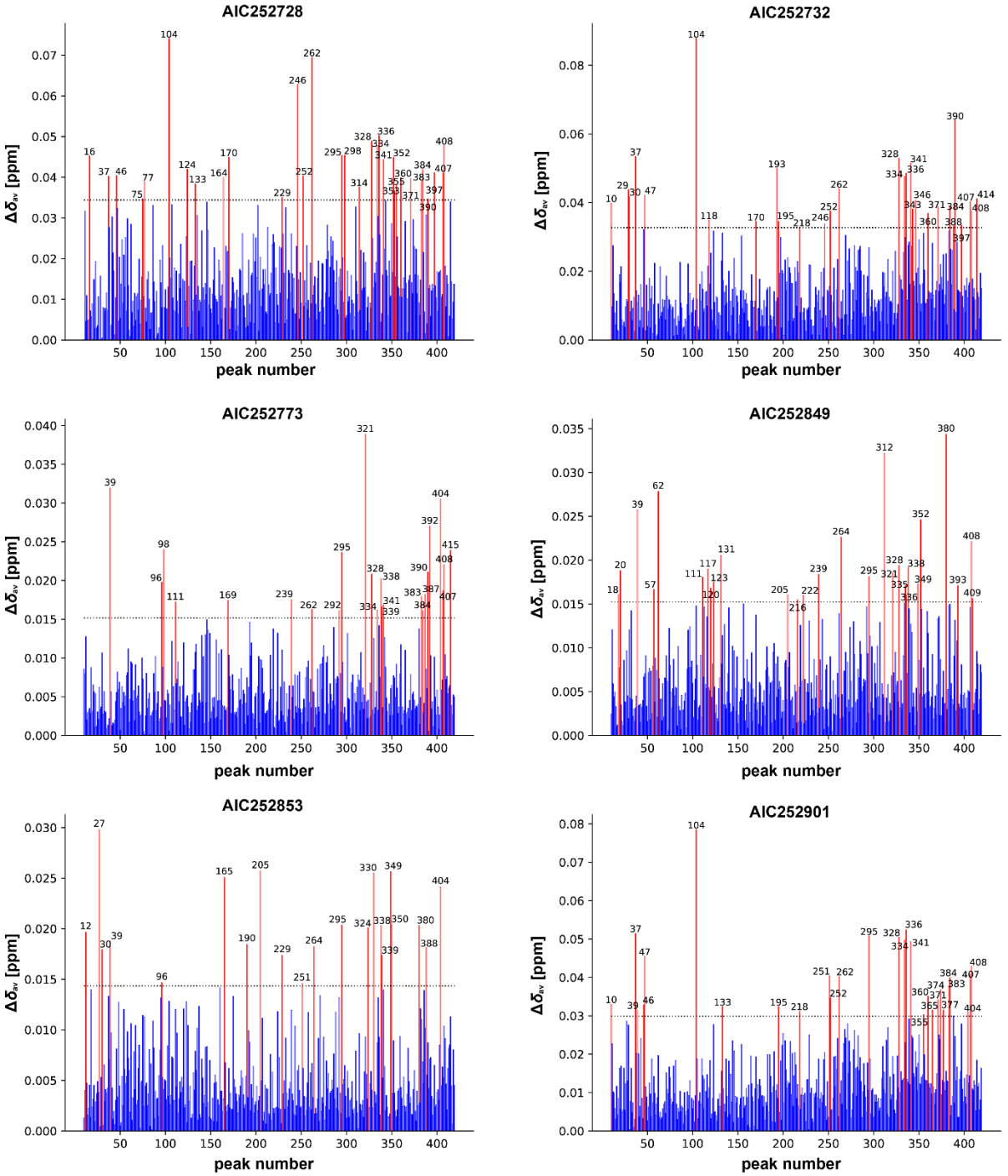


Figure A8: Perturbation shift analysis of AIC252728, AIC252732, AIC252773, AIC252849, AIC252853 and AIC252901.

$\Delta\delta_{av}$ was calculated using Equation 5. 3σ was used as the threshold (dotted line) to identify peaks that showed the largest changes in chemical shift (red). These peaks may represent residues that are involved in interaction with the fragment. Numbers above the bars indicate peak assignments.

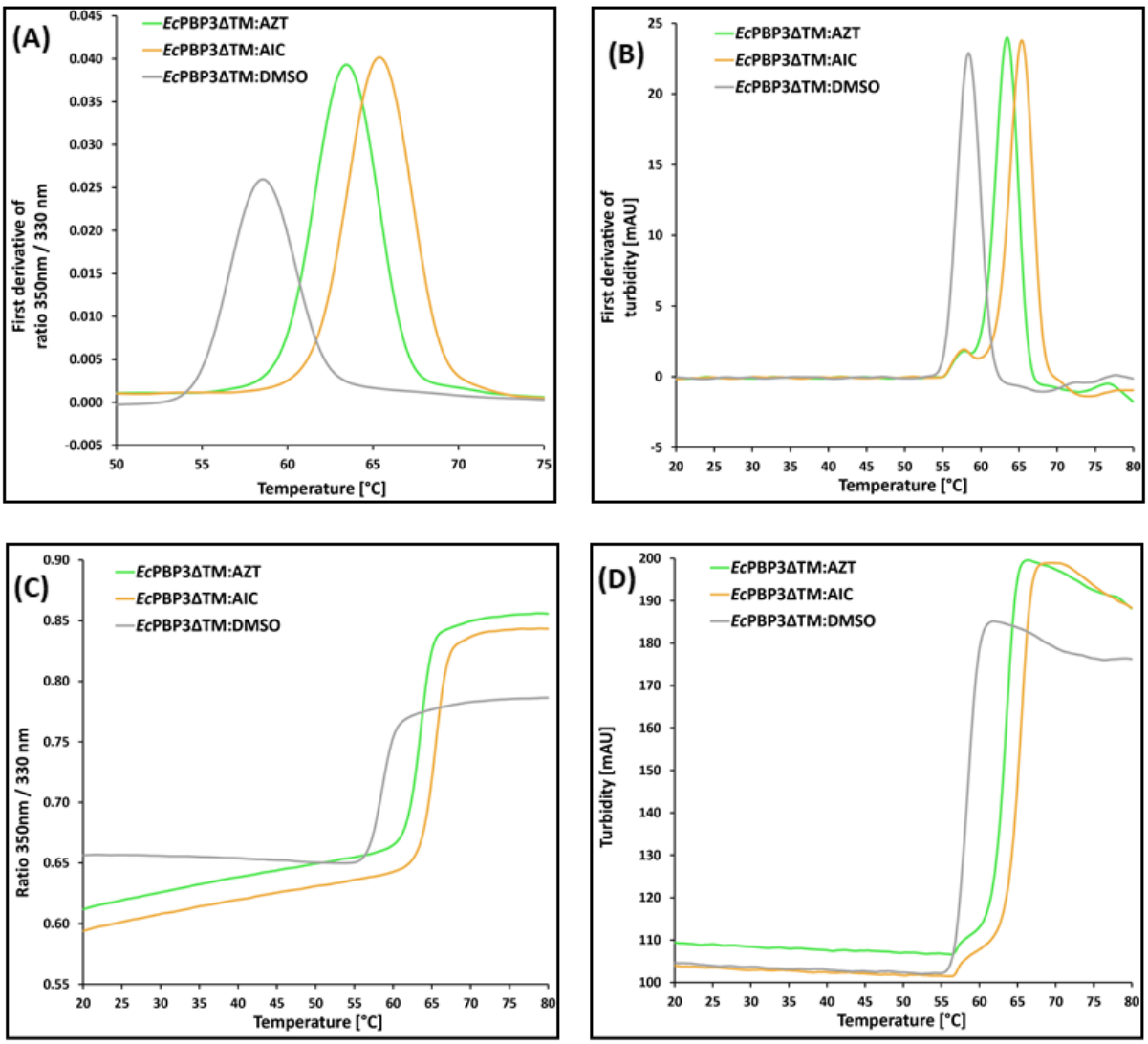


Figure A9: DSF data obtained with *Ec*PBP3ΔTM in complex with aztreonam or AIC499.

(**A**) The first derivative of the 350 nm/330 nm fluorescence ratio is plotted against the temperature. The control experiment with DMSO added to the protein (gray) yields a T_m of 58.6 °C, which increases to 63.4 and 65.4 °C after addition of 500 µM aztreonam (AZT, green) and AIC499 (AIC, gold), respectively. (**B**) The first derivative of the turbidity is plotted against the temperature and T_m derived from turbidity measurements were determined to be 58.4 °C for apo-*Ec*PBP3ΔTM, and 63.4 and 65.3 °C for *Ec*PBP3ΔTM in complex with aztreonam and AIC499, respectively. (**C**) and (**D**) show the raw data obtained for the ratio of 350 nm/330 nm and the turbidity. DSF data were recorded at AiCuris by Dr. Immanuel Grimm and Dr. Arancha López-Pérez.

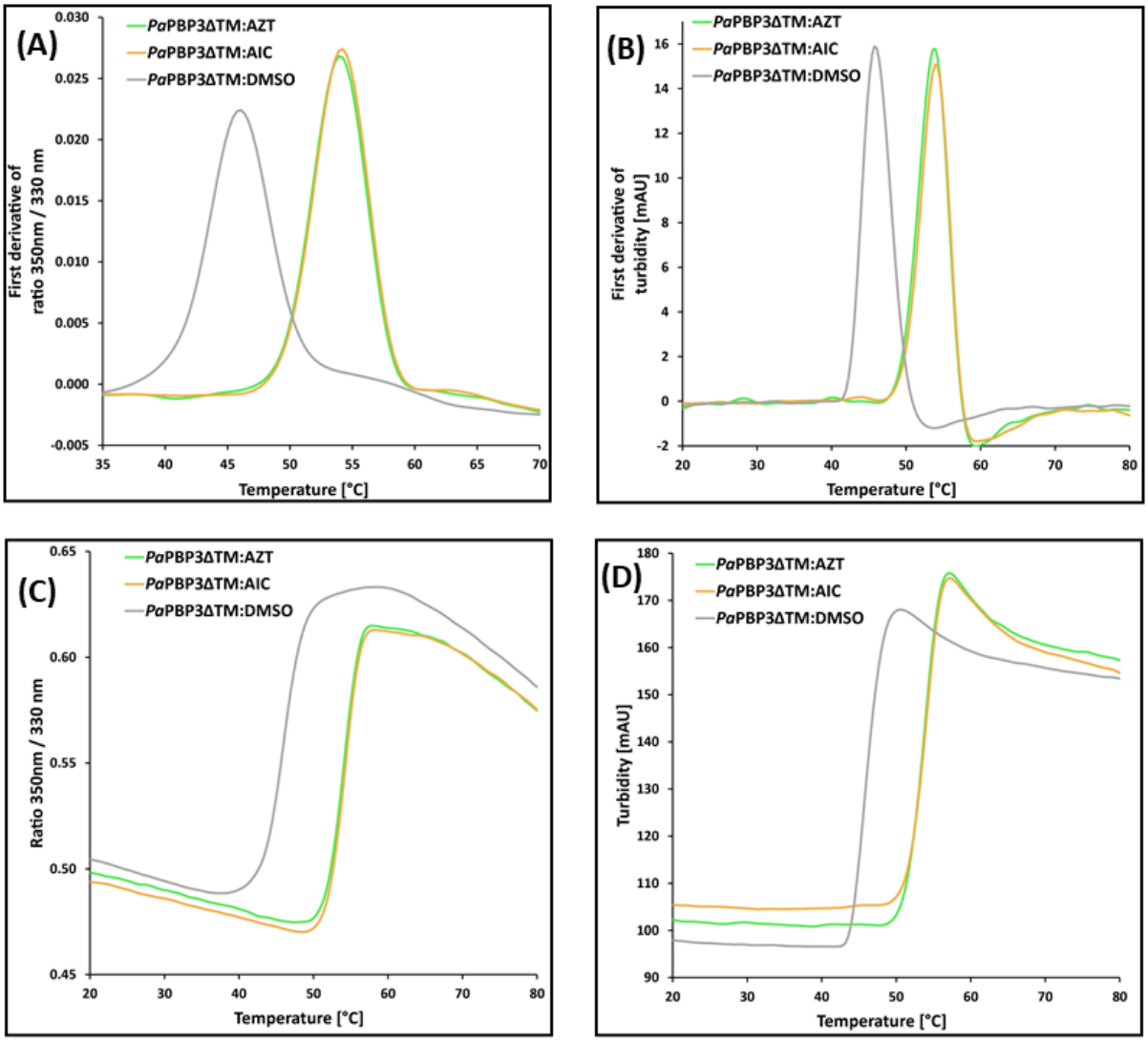


Figure A10: DSF data obtained with *Pa*PBP3ΔTM in complex with aztreonam or AIC499.

(**A**) The first derivative of the 350 nm/330 nm fluorescence ratio is plotted against the temperature. The control experiment with DMSO added to the protein (gray) yields a T_m of 46.0 °C, which increases to 54.0 and 54.1 °C after addition of 500 µM aztreonam (AZT, green) and AIC499 (AIC, gold), respectively. (**B**) The first derivative of the turbidity is plotted against the temperature and T_m derived from turbidity measurements were determined to be 45.9 °C for apo-*Pa*PBP3ΔTM, and 53.8 and 54.0 °C for *Pa*PBP3ΔTM in complex with aztreonam and AIC499, respectively. (**C**) and (**D**) show the raw data obtained for the ratio of 350 nm/330 nm and the turbidity. DSF data were recorded at AiCuris by Dr. Immanuel Grimm and Dr. Arancha López-Pérez.

Table A1: Assigned chemical shifts for amide $^{1}H_N$ and ^{15}N, $^{13}C'$, $^{13}C_\alpha$, and $^{13}C_\beta$ for *Ec*TPd* residues.

Due to technical reasons, peaks in the 2D ^{1}H-^{15}N TROSY-HSQC are assigned to the amino acid sequence of the *Ec*TPd* construct (sequence in Table 2.7).

Residue	Amide $^{1}H_N$ chemical shift [ppm]	Amide ^{15}N chemical shift [ppm]	$^{13}C'$ chemical shift [ppm]	$^{13}C_\alpha$ chemical shift [ppm]	$^{13}C_\beta$ chemical shift [ppm]
S1	–	–	174.79296	58.29748	63.53734
T2	8.07348	116.36673	174.44365	61.45377	17.46244
S3	8.21313	119.23693	173.24722	57.97295	63.69156
R4	7.90777	124.14792	176.62178	55.52621	31.57597
G5	9.33385	114.11989	173.28916	44.36154	–
M6	8.70123	123.31546	175.86983	55.12966	34.81273
I7	8.59102	122.1044	176.56135	59.74845	38.26164
T8	9.74588	120.25813	173.9368	59.36632	19.81374
D9	9.10694	121.32633	178.87119	52.04077	40.31652
R10	8.54311	117.17938	176.47454	58.85082	28.26134
S11	8.34486	118.73061	–	56.97658	–
I40	–	–	176.10194	62.38771	34.36994
G41	8.1494	104.4548	171.74973	43.7888	–
F42	6.11982	109.24945	174.88979	55.28396	38.30533
T43	8.13819	109.56543	174.58651	58.92038	19.36071
N44	8.37121	117.81015	178.38466	50.80143	39.07436
V45	8.13556	117.10281	176.11504	64.24916	30.8092
D46	7.22389	120.63077	175.39094	54.17248	40.51556
S47	8.18241	111.32295	172.80536	59.3986	61.4971
Q48	7.73389	117.82281	177.12144	53.62245	28.7735
G49	8.993	112.33182	173.08456	46.40766	–
I50	8.24723	120.12735	174.34821	60.04837	38.47211
E51	6.97275	120.88911	176.83549	53.8427	34.40991
G52	9.02675	107.24904	175.05772	45.56686	–
V53	9.09045	125.81799	177.35902	65.40241	31.54813
E54	8.6092	118.52221	177.7639	59.52592	28.1057
K55	6.78518	114.47908	179.31969	57.7684	30.80816
S56	8.57073	117.56795	175.87213	62.30599	–
F57	8.7098	117.83835	175.31659	57.03407	36.88838
D58	7.2067	122.76679	178.26152	59.96964	43.58471
K59	8.46675	116.01748	178.78617	59.04265	30.98059
W60	7.81888	121.92543	177.96203	59.3279	29.68555
L61	8.43904	116.19337	178.02913	56.07934	42.56776
T62	7.83134	110.78963	175.17685	62.2544	70.21405
G63	7.97466	111.68669	173.97781	44.99458	–
Q64	8.11346	120.97725	177.1167	55.51417	29.11849
G65	8.72987	112.40355	175.15039	45.79699	–
G66	8.29002	109.64658	174.82985	44.80117	–
G67	7.88906	109.40931	172.922	44.26177	–
A68	7.88542	122.97352	176.82206	51.1662	19.03985

A69	7.94211	123.00905	177.01099	51.71034	18.48807
H70	7.38773	117.90182	173.2632	54.19403	28.77949
N71	8.10412	116.51122	175.65091	51.69022	38.90686
L72	8.89478	124.39189	172.80436	53.93679	47.25002
A73	8.4128	130.47054	177.20541	49.78903	17.66772
L74	8.54188	122.4811	174.69455	53.09727	42.9
S75	8.09313	110.45704	175.57883	59.07585	62.15513
I76	6.16197	114.1433	173.67486	58.85359	38.04812
D77	9.12193	127.16323	173.87179	53.30717	–
P151	–	–	178.13893	66.50293	31.20674
M152	7.21758	112.80481	178.9271	58.67729	32.8036
V153	7.61335	124.28241	177.7207	65.76399	30.7227
V154	7.12328	118.6916	177.24557	67.02971	30.49617
M155	8.04552	116.07559	177.08088	59.27961	–
T156	7.39182	115.47462	174.46062	67.05683	68.49515
A157	8.08776	124.01865	179.71693	54.91004	19.08901
L158	8.03843	118.2355	180.68314	57.09281	41.47965
Q159	8.33192	122.77786	177.77943	59.11979	27.86059
R160	8.21599	114.36911	176.75789	56.06013	29.78394
G161	7.67174	107.67604	174.89223	45.74923	–
V162	7.92716	117.3575	175.20972	63.29622	31.76896
V163	7.07327	110.46947	173.36337	58.28081	35.03317
R164	8.6451	119.86022	177.05891	53.98659	31.96349
E165	9.1056	119.12609	175.56081	60.0397	28.37012
N166	7.7895	112.49504	176.03077	51.43756	37.31141
S167	7.66794	118.53473	171.69181	61.64176	63.49926
V168	8.18369	121.27461	175.62405	61.1053	32.59071
L169	9.00856	128.53976	175.12803	53.44378	42.71558
N170	8.31311	120.97829	176.87736	53.24784	37.91485
T171	8.03007	118.64459	173.37041	60.52358	65.75578
I172	7.58105	124.15053	174.72181	59.24779	–
N177	–	–	175.225	54.26354	37.86024
G178	8.46374	103.67929	173.75413	44.99824	–
H179	8.17232	122.27543	174.43595	54.92885	29.89409
E180	8.75442	126.2095	174.88028	56.74102	30.01637
I181	8.88235	130.50185	173.29811	59.17704	34.83183
K182	7.73709	121.18268	174.09851	54.21823	35.08259
D183	8.89079	123.54539	177.06273	52.82854	42.72197
V184	8.2402	119.05283	175.89466	64.76028	30.85477
A185	7.4824	122.05611	174.66294	49.82913	20.0204
R186	7.8552	120.57153	175.1987	55.53808	29.73543
Y187	8.68394	127.6083	–	56.44635	–
V202	–	–	176.28504	67.89758	–
G203	8.00736	112.75719	174.98245	46.92659	–
V204	8.6385	122.29356	177.86095	66.22655	30.55643

S205	8.46171	116.64538	175.85117	61.84364	–
K206	7.7082	123.01628	180.5387	59.26295	31.22882
L207	7.27576	117.70934	178.08941	56.98187	40.09701
A208	8.15892	119.83864	180.1072	54.90108	17.97533
L209	8.04082	115.693	177.15967	55.80225	39.49192
A210	7.21246	121.51106	176.00937	51.38467	18.61004
M211	7.47657	118.21795	171.32084	52.7127	–
S213	–	–	175.0751	60.50189	61.3379
S214	7.46983	112.82684	176.16742	58.87827	62.31535
A215	7.73236	126.32786	179.49396	55.22509	17.40864
L216	8.16119	115.656	177.75715	56.66506	39.75059
V217	6.52165	117.49325	179.08337	64.68457	31.13965
D218	7.79743	118.95155	178.27598	57.17403	41.09255
T219	8.40176	116.48862	175.81592	67.32368	67.93364
Y220	8.9992	120.08943	179.07028	58.77632	35.02943
S221	7.69269	115.09333	178.84013	61.18859	62.96906
R222	8.49661	125.4415	177.42948	59.19802	28.90366
F223	7.43133	114.44779	175.41968	58.55106	38.92989
G224	7.76403	106.44229	175.499	45.98138	–
L225	8.16155	121.5378	176.35599	55.62152	38.08139
G226	8.20536	107.05061	173.07174	44.89208	–
K227	7.11765	117.20023	176.7494	54.29043	33.38523
A228	8.68639	126.75173	180.03886	51.99778	17.96458
T229	8.86407	112.76369	175.26483	62.7767	67.12896
N230	7.99415	116.88814	174.99716	54.10539	37.66107
L231	10.86821	122.51689	178.70331	56.4713	40.6947
G232	8.72892	107.42325	175.37585	45.01384	–
L233	7.63973	121.3688	177.326	52.76923	42.28071
V234	8.3376	123.81477	177.14747	63.7655	30.56291
G235	8.74047	114.57154	173.68242	44.24739	–
E236	7.2126	120.8441	176.7338	57.43066	30.56682
R237	8.33341	126.26516	–	54.64809	–
S238	–	–	175.39619	56.85946	63.47816
G239	7.51384	108.87693	172.43499	44.4456	–
L240	8.30269	123.34915	174.84951	55.05039	43.77179
Y241	8.67319	128.49338	173.29177	55.2077	–
P265	–	–	177.8632	66.05614	30.34986
L266	7.93332	114.87822	178.21921	57.74866	41.30475
Q267	8.25497	118.39116	180.67451	59.29935	29.73811
L268	8.28063	121.70101	178.11077	56.86671	40.70909
A269	8.51992	123.95098	178.69361	55.2848	68.60684
R270	7.78423	118.83125	179.21467	58.93217	29.49317
V271	7.91827	123.25644	178.60609	65.90596	30.35074
Y272	8.04925	118.53273	177.48252	62.87093	36.51161
A273	8.99464	126.48242	178.08644	55.03297	16.93892

T274	7.54301	117.01103	176.06053	64.58203	67.76017
I275	7.87432	122.14256	178.43307	65.57566	36.65638
G276	8.57874	105.49905	171.58917	46.59854	–
S277	7.77043	113.50088	174.32201	56.83453	62.61571
Y278	8.13444	116.8345	175.25276	54.98183	35.03251
G279	8.78219	103.9284	172.76787	45.14077	–
I280	7.80058	121.34548	174.17174	59.38882	37.06293
Y281	9.32028	131.53272	174.81629	56.79388	40.49568
R282	8.65108	130.97436	172.0966	52.60271	–
P294	–	–	176.84064	63.53264	31.56518
G295	8.04975	104.47422	172.03919	44.72161	–
E296	8.31554	120.72682	175.44857	53.89285	32.21326
R297	8.99175	127.21002	176.77388	56.35183	26.67359
V298	8.86019	121.08078	174.63869	60.05796	32.96717
F299	7.77539	126.1619	173.12254	56.68428	–
P300	–	–	176.73331	62.97995	31.74325
E301	8.80658	128.05493	178.65989	58.52839	29.51101
S302	8.81958	112.64726	177.20868	60.90962	61.50954
I303	6.78476	121.05518	177.4666	63.6068	37.34657
V304	7.63391	119.00167	177.98945	67.2058	30.56468
R305	8.60838	118.40247	178.88022	60.01528	29.23835
T306	7.59675	116.96637	176.38432	66.46819	68.36623
V307	7.72829	122.25496	177.63318	66.06914	–
P343	–	–	176.62479	64.456	30.92718
D344	7.88812	115.68821	177.0285	52.49515	39.68575
G345	7.97682	107.88554	173.73871	44.88629	–
R346	7.44993	119.8716	176.92777	54.55395	30.1372
Y347	8.39184	121.71687	176.9549	59.06688	38.18878
I348	8.54182	118.92638	175.63692	59.21755	39.68327
N349	8.26366	120.89046	175.16424	52.51397	36.448
K350	7.97043	120.86271	174.80216	55.20251	37.05886
Y351	8.83366	119.94038	174.39334	57.58168	41.31321
I352	9.28206	120.39127	173.57743	60.09971	–
P375	–	–	177.03413	62.84466	31.37823
E376	7.98856	121.83631	176.15408	55.69328	28.86063
A377	7.96725	123.67009	177.64338	51.96889	18.59736
G378	8.03953	107.56328	174.43539	45.23244	–
K379	7.92767	121.23014	–	56.04605	–
A407	–	–	177.13307	51.52692	18.3588
L408	8.13121	121.38959	177.65195	54.89129	41.24849
T409	7.93951	114.48704	173.97052	61.08938	17.51778
T410	7.66219	121.36043	179.13929	62.74978	–

Table A2: Overview of the distances between atoms of AIC499 (center) and *Pa*PBP3ΔTM (chain B, left) or *Ec*TPd* (right).

The type of interaction as well as the functional group which the respective AIC499 atom belongs to are given in parentheses. See Figure 1 for the nomenclature of AIC499 atoms. Hydrogen bonds are included with donor-acceptor distances ranging from 2.3 to 3.2 Å, and hydrophobic and aromatic interactions are listed for distances between 3.0 and 4.5 Å. For distances differing between both conformers, the smaller value is given.

***Pa*PBP3ΔTM**		**AIC499**	***Ec*TPd***	
Atom	*Distance [Å]*	*Atom*	*Distance [Å]*	*Atom*
–	–	C2 (R^1)	3.6 (hydrophobic)	Y541 $C_{\varepsilon2}$
V333 $C_{\gamma1}$	3.5 (hydrophobic)	C4 (R^1)	–	–
V333 $C_{\gamma2}$	3.6 (hydrophobic)	C11 (R^1)	3.7 (hydrophobic)	V344 $C_{\gamma1}$
Y532 $C_{\varepsilon2}$	3.8 (aromatic)	C13 (R^1)	4.2 (aromatic) [b]	Y540 $C_{\varepsilon2}$
T487 C_{β}	3.5 (hydrophobic)	C14 (R^1)	4.2 (hydrophobic)	T497 C_{β}
F533 C_{β}	3.7 (aromatic) [a]	C15 (R^1)	4.0 (aromatic)	Y541 $C_{\delta1}$
T487 C_{β}	4.3 (hydrophobic)	C18 (R^5)	–	–
–	–	C19 (R^5)	4.1 (hydrophobic)	Y511 $C_{\varepsilon2}$
R489 $N_{\eta2}$	3.0 (H-bond)	O20 (R^5)	–	–
T487 O	3.2 (H-bond)	N23	–	–
R489 N	2.8 (H-bond)	N26 (R^4)	3.0 (H-bond)	K499 N
Y409 $C_{\delta1}$	3.8 (aromatic)	C27 (R^4)	3.6 (aromatic)	Y419 $C_{\varepsilon1}$
A488 C_{α}	3.8 (hydrophobic)	C27 (R^4)	3.6 (hydrophobic)	A498 C_{α}
E291 $O_{\varepsilon1}$	2.5 (H-bond)	N28 (R^4)	2.3 (H-bond)	E304 $O_{\varepsilon1}$
R489 O	2.5 (H-bond)	N28 (R^4)	2.8 (H-bond)	K499 O
–	–	N28 (R^4)	3.1 (H-bond)	K499 N
–	–	C30 (R^4)	4.4 (aromatic)	F417 $C_{\delta1}$
G293 C_{α}	3.8 (hydrophobic)	C30 (R^4)	3.8 (hydrophobic)	G306 C_{α}
N351 $N_{\delta2}$	2.8 (H-bond)	O32	3.0 (H-bond)	N361 $N_{\gamma2}$
T487 O	2.8 (H-bond)	N33	3.2 (H-bond)	T497 O
V333 $C_{\gamma2}$	3.6 (hydrophobic)	C36 (R^3)	–	–
T487 C_{β}	3.8 (hydrophobic)	C36 (R^3)	–	–
V333 $C_{\gamma2}$	3.8 (hydrophobic)	C37 (R^3)	3.6 (hydrophobic)	V344 $C_{\gamma1}$
S349 C_{β}	4.3 (hydrophobic)	C37 (R^3)	–	–
S294 O_{γ}	2.7 (H-bond)	N38 (β-lactam ring)	–	–
S349 O_{γ}	2.8 (H-bond)	N38 (β-lactam ring)	–	–
S294 N	2.7 (H-bond)	O40	2.5 (H-bond)	S307 N
T487 N	2.9 (H-bond)	O40	3.2 (H-bond)	T497 N
S294 O_{γ}	3.0 (H-bond)	O41 (R^2)	3.0 (H-bond)	S307 O_{γ}
K484 N_{ζ}	2.9 (H-bond)	O44 (R^2)	2.7 (H-bond)	T497 O_{γ}
S485 O_{γ}	2.4 (H-bond)	O44 (R^2)	–	–
T487 O_{γ}	2.5 (H-bond)	O45 (R^2)	3.1 (H-bond)	K494 N_{ζ}
T487 N	2.9 (H-bond)	O45 (R^2)	3.1 (H-bond)	T495 O

[a] Inferred from the C_{β} position, although the phenyl group is not resolved in the electron density. [b] An even shorter distance of 4.0 Å is found between AIC499 C12 (R^1) and *Ec*TPd* Y540 C_{ζ}.

Table A3: Overview of the distances between atoms of aztreonam and *Pa*PBP3.

PDB entry 3PBS was used to determine the distances and the type of interaction is given in parentheses [51]. Aztreonam atoms are named in accordance to PDB entry 3PBS, and equivalent atoms of AIC499 together with the respective functional groups are shown in parentheses. Hydrogen bonds are included with donor-acceptor distances ranging from 2.3 to 3.2 Å, and hydrophobic and aromatic interactions are listed for distances between 3.0 and 4.5 Å.

***Pa*PBP3**	***Distance [Å]***	**Aztreonam**
Atom		*Atom*
T487 $C_{\gamma 2}$	3.6 (hydrophobic)	C28 (C4; R^1) [c]
Y532 $C_{\varepsilon 2}$	4.4 (hydrophobic)	C28 (C13; R^1) [c]
Y503 $C_{\varepsilon 1}$	3.7 (hydrophobic)	C28 (C13; R^1) [c]
F533 C_{β}	3.9 (hydrophobic)	C27 (C15; R^1) [c]
R489 N_{ε}	3.1 (H-bond)	O31 (O20; R^5)
A488 C_{α}	4.2 (hydrophobic)	C25 (C25; R^4)
Y409 $C_{\delta 1}$	3.9 (aromatic)	C26 (C27; R^4)
A488 C_{α}	3.9 (hydrophobic)	C26 (C27; R^4)
E291 $O_{\varepsilon 1}$	2.9 (H-bond)	N16 (N28; R^4)
G293 C_{α}	4.0 (hydrophobic)	C25 (C30; R^4)
N351 $N_{\delta 2}$	2.9 (H-bond)	O10 (O32)
T487 O	3.0 (H-bond)	N13 (N33)
T487 C_{β}	4.3 (hydrophobic)	C18 (C35; R^3)
S294 C_{β}	4.4 (hydrophobic)	C18 (C35; R^3)
V333 $C_{\gamma 1}$	3.6 (hydrophobic)	C7 (C36; R^3)
S294 O_{γ}	3.1 (H-bond)	N12 (N38; β-lactam ring)
S349 O_{γ}	3.0 (H-bond)	N12 (N38; β-lactam ring)
G293 C_{α}	4.5 (hydrophobic)	C20 (C39)
S294 N	3.0 (H-bond)	O9 (O40)
T487 N	2.8 (H-bond)	O9 (O40)
K484 N_{ζ}	3.1 (H-bond)	O34 (O44; R^2)
S485 O_{γ}	2.7 (H-bond)	O34 (O44; R^2)
T487 O_{γ}	2.6 (H-bond)	O33 (O45; R^2)
T487 N	3.2 (H-bond)	O33 (O45; R^2)

[c] As R^1 moiety aztreonam only contains two methyl groups.

9. References

1. Gottfried, R.S., *Black death*. 2010: Simon and Schuster.
2. Adedeji, W., *The treasure called antibiotics.* Annals of Ibadan postgraduate medicine, 2016. **14**(2): p. 56.
3. Pennington, H., *The impact of infectious disease in war time: a look back at WW1*. 2019, Future Medicine.
4. Ligon, B.L. *Penicillin: its discovery and early development*. in *Seminars in pediatric infectious diseases*. 2004. Elsevier.
5. Mulani, M.S., et al., *Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: a review.* Frontiers in microbiology, 2019. **10**: p. 539.
6. Johnston, K.J., et al., *The incremental cost of infections associated with multidrug‐resistant organisms in the inpatient hospital setting—A national estimate.* Health services research, 2019. **54**(4): p. 782-792.
7. Tacconelli, E., et al., *Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis.* The Lancet Infectious Diseases, 2018. **18**(3): p. 318-327.
8. Morris, F.C., et al., *The mechanisms of disease caused by Acinetobacter baumannii.* Frontiers in microbiology, 2019. **10**: p. 1601.
9. Bodey, G.P., et al., *Infections caused by Pseudomonas aeruginosa.* Reviews of infectious diseases, 1983. **5**(2): p. 279-313.
10. Allocati, N., et al., *Escherichia coli in Europe: an overview.* International journal of environmental research and public health, 2013. **10**(12): p. 6235-6254.
11. Sanders Jr, W.E. and C.C. Sanders, *Enterobacter spp.: pathogens poised to flourish at the turn of the century.* Clinical microbiology reviews, 1997. **10**(2): p. 220-241.
12. Lim, S.Y., C.S.J. Teh, and K.L. Thong, *Biofilm-related diseases and omics: global transcriptional profiling of Enterococcus faecium reveals different gene expression patterns in the biofilm and planktonic cells.* Omics: a journal of integrative biology, 2017. **21**(10): p. 592-602.
13. Higuita, N.I.A. and M.M. Huycke, *Enterococcal disease, epidemiology, and implications for treatment.* Enterococci: From commensals to leading causes of drug resistant infection [Internet], 2014.
14. Lowy, F.D., *Staphylococcus aureus infections.* New England journal of medicine, 1998. **339**(8): p. 520-532.
15. Cassini, A., et al., *Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis.* The Lancet infectious diseases, 2019. **19**(1): p. 56-66.
16. O'Neill, J. *Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations.* 2014 [cited 2021 (15. May)]; Available from: https://amr-review.org/Publications.html.
17. Schleifer, K.H., *Classification of Bacteria and Archaea: past, present and future.* Systematic and applied microbiology, 2009. **32**(8): p. 533-542.
18. Hedlund, B.P., J.A. Dodsworth, and J.T. Staley, *The changing landscape of microbial biodiversity exploration and its implications for systematics.* Systematic and applied microbiology, 2015. **38**(4): p. 231-236.
19. Coico, R., *Gram staining.* Current protocols in microbiology, 2006(1): p. A. 3C. 1-A. 3C. 2.
20. Auer, G.K. and D.B. Weibel, *Bacterial cell mechanics.* Biochemistry, 2017. **56**(29): p. 3710-3724.
21. Gould, S.B., *Membranes and evolution.* Current Biology, 2018. **28**(8): p. R381-R385.
22. Sohlenkamp, C. and O. Geiger, *Bacterial membrane lipids: diversity in structures and pathways.* FEMS microbiology reviews, 2016. **40**(1): p. 133-159.

23. Zinser, E., et al., *Phospholipid synthesis and lipid composition of subcellular membranes in the unicellular eukaryote Saccharomyces cerevisiae.* Journal of bacteriology, 1991. **173**(6): p. 2026-2034.
24. Swoboda, J.G., et al., *Wall teichoic acid function, biosynthesis, and inhibition.* Chembiochem: a European journal of chemical biology, 2010. **11**(1): p. 35.
25. D'Elia, M.A., et al., *Wall teichoic acid polymers are dispensable for cell viability in Bacillus subtilis.* Journal of bacteriology, 2006. **188**(23): p. 8313-8316.
26. Schirner, K., et al., *Distinct and essential morphogenic functions for wall‐and lipo‐teichoic acids in Bacillus subtilis.* The EMBO journal, 2009. **28**(7): p. 830-842.
27. Santa Maria, J.P., et al., *Compound-gene interaction mapping reveals distinct roles for Staphylococcus aureus teichoic acids.* Proceedings of the national academy of sciences, 2014. **111**(34): p. 12510-12515.
28. Vollmer, W. and U. Bertsche, *Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2008. **1778**(9): p. 1714-1734.
29. Linkevicius, M., et al., *Fitness of Escherichia coli mutants with reduced susceptibility to tigecycline.* Journal of Antimicrobial Chemotherapy, 2016. **71**(5): p. 1307-1313.
30. Egan, A.J., et al., *Outer-membrane lipoprotein LpoB spans the periplasm to stimulate the peptidoglycan synthase PBP1B.* Proceedings of the National Academy of Sciences, 2014. **111**(22): p. 8197-8202.
31. Fraipont, C., et al., *The integral membrane FtsW protein and peptidoglycan synthase PBP3 form a subcomplex in Escherichia coli.* Microbiology, 2011. **157**(1): p. 251-259.
32. Sauvage, E., et al., *The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis.* FEMS microbiology reviews, 2008. **32**(2): p. 234-258.
33. Breukink, E. and B. de Kruijff, *Lipid II as a target for antibiotics.* Nature reviews Drug discovery, 2006. **5**(4): p. 321-323.
34. Leclercq, S., et al., *Interplay between penicillin-binding proteins and SEDS proteins promotes bacterial cell wall synthesis.* Scientific reports, 2017. **7**(1): p. 1-13.
35. Sauvage, E. and M. Terrak, *Glycosyltransferases and transpeptidases/penicillin-binding proteins: valuable targets for new antibacterials.* Antibiotics, 2016. **5**(1): p. 12.
36. Goffin, C. and J.-M. Ghuysen, *Biochemistry and comparative genomics of SxxK superfamily acyltransferases offer a clue to the mycobacterial paradox: presence of penicillin-susceptible target proteins versus lack of efficiency of penicillin as therapeutic agent.* Microbiology and molecular biology reviews, 2002. **66**(4): p. 702-738.
37. Piette, A., et al., *Structural determinants required to target penicillin-binding protein 3 to the septum of Escherichia coli.* Journal of bacteriology, 2004. **186**(18): p. 6110-6117.
38. Vigouroux, A., et al., *Class-A penicillin binding proteins do not contribute to cell shape but repair cell-wall defects.* Elife, 2020. **9**: p. e51998.
39. Lupoli, T.J., et al., *Lipoprotein activators stimulate Escherichia coli penicillin-binding proteins by different mechanisms.* Journal of the American Chemical Society, 2014. **136**(1): p. 52-55.
40. Typas, A., et al., *Regulation of peptidoglycan synthesis by outer-membrane proteins.* Cell, 2010. **143**(7): p. 1097-1109.
41. Den Blaauwen, T., et al., *Penicillin‐binding protein PBP2 of Escherichia coli localizes preferentially in the lateral wall and at mid‐cell in comparison with the old cell pole.* Molecular microbiology, 2003. **47**(2): p. 539-547.
42. Denome, S.A., et al., *Escherichia coli mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis.* Journal of bacteriology, 1999. **181**(13): p. 3981-3993.
43. Baquero, M.-R., et al., *dacD, an Escherichia coli gene encoding a novel penicillin-binding protein (PBP6b) with DD-carboxypeptidase activity.* Journal of Bacteriology, 1996. **178**(24): p. 7106-7111.
44. Vollmer, W., et al., *Bacterial peptidoglycan (murein) hydrolases.* FEMS microbiology reviews, 2008. **32**(2): p. 259-286.

45. González-Leiza, S.M., M.A. de Pedro, and J.A. Ayala, *AmpH, a bifunctional DD-endopeptidase and DD-carboxypeptidase of Escherichia coli.* Journal of bacteriology, 2011. **193**(24): p. 6887-6894.
46. Jacoby, G.A., *AmpC β-lactamases.* Clinical microbiology reviews, 2009. **22**(1): p. 161-182.
47. Sauvage, E., et al., *Crystal structure of penicillin-binding protein 3 (PBP3) from Escherichia coli.* Public Library of Science one, 2014. **9**(5): p. e98042.
48. Freischem, S., et al., *Interaction Mode of the Novel Monobactam AIC499 Targeting Penicillin Binding Protein 3 of Gram-Negative Bacteria.* Biomolecules, 2021. **11**(7): p. 1057.
49. Bellini, D., et al., *Novel and Improved Crystal Structures of H. influenzae, E. coli and P. aeruginosa Penicillin-Binding Protein 3 (PBP3) and N. gonorrhoeae PBP2: Toward a Better Understanding of β-Lactam Target-Mediated Resistance.* Journal of molecular biology, 2019. **431**(18): p. 3501-3519.
50. Sainsbury, S., et al., *Crystal structures of penicillin-binding protein 3 from Pseudomonas aeruginosa: comparison of native and antibiotic-bound forms.* Journal of molecular biology, 2011. **405**(1): p. 173-184.
51. Han, S., et al., *Structural basis for effectiveness of siderophore-conjugated monocarbams against clinically relevant strains of Pseudomonas aeruginosa.* Proceedings of the National Academy of Sciences, 2010. **107**(51): p. 22002-22007.
52. Bermingham, A. and J.P. Derrick, *The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery.* Bioessays, 2002. **24**(7): p. 637-648.
53. Bush, K. and P.A. Bradford, *β-Lactams and β-lactamase inhibitors: an overview.* Cold Spring Harbor perspectives in medicine, 2016. **6**(8): p. a025247.
54. Zervosen, A., et al., *Development of new drugs for an old target—the penicillin binding proteins.* Molecules, 2012. **17**(11): p. 12478-12505.
55. Fuda, C., et al., *The basis for resistance to β-lactam antibiotics by penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus.* Journal of Biological Chemistry, 2004. **279**(39): p. 40802-40806.
56. Lima, L.M., et al., *β-lactam antibiotics: An overview from a medicinal chemistry perspective.* European Journal of Medicinal Chemistry, 2020: p. 112829.
57. Fleming, A., *On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae.* British journal of experimental pathology, 1929. **10**(3): p. 226.
58. Enright, M.C., et al., *The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA).* Proceedings of the National Academy of Sciences, 2002. **99**(11): p. 7687-7692.
59. Schwartz, M., et al., *Stability of methicillin.* Journal of pharmaceutical sciences, 1965. **54**(1): p. 149-150.
60. Turner, N.A., et al., *Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research.* Nature Reviews Microbiology, 2019. **17**(4): p. 203-218.
61. Raynor, B.D., *Penicillin and ampicillin.* Primary Care Update for OB/GYNS, 1997. **4**(4): p. 147-152.
62. Knudsen, E., G.N. Rolinson, and R. Sutherland, *Carbenicillin: a new semisynthetic penicillin active against Pseudomonas pyocyanea.* British medical journal, 1967. **3**(5557): p. 75.
63. Clissold, S.P., P.A. Todd, and D.M. Campoli-Richards, *Imipenem/cilastatin.* Drugs, 1987. **33**(3): p. 183-241.
64. Birnbaum, J., et al., *Carbapenems, a new class of beta-lactam antibiotics: discovery and development of imipenem/cilastatin.* The American journal of medicine, 1985. **78**(6): p. 3-21.
65. Kropp, H., et al., *Antibacterial activity of imipenem: the first thienamycin antibiotic.* Reviews of infectious diseases, 1985. **7**(Supplement_3): p. S389-S410.
66. Rodloff, A., E. Goldstein, and A. Torres, *Two decades of imipenem therapy.* Journal of Antimicrobial Chemotherapy, 2006. **58**(5): p. 916-929.

67. Queenan, A.M. and K. Bush, *Carbapenemases: the versatile β-lactamases.* Clinical microbiology reviews, 2007. **20**(3): p. 440-458.
68. Sykes, R., et al., *Monocyclic β-lactam antibiotics produced by bacteria.* Nature, 1981. **291**(5815): p. 489-491.
69. Sykes, R. and D. Bonner, *Aztreonam: the first monobactam.* The American journal of medicine, 1985. **78**(2): p. 2-10.
70. Westley-Horton, E. and J.A. Koestner, *Aztreonam: a review of the first monobactam.* The American journal of the medical sciences, 1991. **302**(1): p. 46-49.
71. McPherson, C.J., et al., *Clinically relevant Gram-negative resistance mechanisms have no effect on the efficacy of MC-1, a novel siderophore-conjugated monocarbam.* Antimicrobial agents and chemotherapy, 2012. **56**(12): p. 6334-6342.
72. Flanagan, M.E., et al., *Preparation, Gram-negative antibacterial activity, and hydrolytic stability of novel siderophore-conjugated monocarbam diols.* American Chemical Society medicinal chemistry letters, 2011. **2**(5): p. 385-390.
73. Blais, J., et al., *In vitro activity of LYS228, a novel monobactam antibiotic, against multidrug-resistant Enterobacteriaceae.* Antimicrobial agents and chemotherapy, 2018. **62**(10): p. e00552-18.
74. Reck, F., et al., *Optimization of novel monobactams with activity against carbapenem-resistant Enterobacteriaceae–identification of LYS228.* Bioorganic & medicinal chemistry letters, 2018. **28**(4): p. 748-755.
75. Schurek, K.N., et al., *Faropenem: review of a new oral penem.* Expert review of anti-infective therapy, 2007. **5**(2): p. 185-198.
76. Replidyne, I.a.F.L., Inc. *Orapem FDA Approval Status*. 2006 [cited 2021 (25. August)]; Available from: https://www.drugs.com/nda/orapem_061023.html.
77. Gandra, S., et al., *Faropenem consumption is increasing in India.* Clinical Infectious Diseases, 2016. **62**(8): p. 1050-1052.
78. GlobeNewswire, I. *Iterum Therapeutics Announces U.S. FDA Filing Acceptance of New Drug Application for Oral Sulopenem*. 2021 [cited 2021 (25. August)]; Available from: https://www.drugs.com/nda/sulopenem_210125.html.
79. Reading, C. and M. Cole, *Clavulanic acid: a beta-lactamase-inhibiting beta-lactam from Streptomyces clavuligerus.* Antimicrobial agents and chemotherapy, 1977. **11**(5): p. 852-857.
80. Ehmann, D.E., et al., *Kinetics of avibactam inhibition against class A, C, and D β-lactamases.* Journal of Biological Chemistry, 2013. **288**(39): p. 27960-27971.
81. Ehmann, D.E., et al., *Avibactam is a covalent, reversible, non–β-lactam β-lactamase inhibitor.* Proceedings of the National Academy of Sciences, 2012. **109**(29): p. 11663-11668.
82. Pemberton, O.A., et al., *Mechanism of proton transfer in class A β-lactamase catalysis and inhibition by avibactam.* Proceedings of the National Academy of Sciences, 2020. **117**(11): p. 5818-5825.
83. Klenke, B., et al., *Amidine substituted beta-lactam compounds, their preparation and use as antibacterial agents* 2013.
84. Woestenhemke, K., *AiCuris Initiates Clinical Development of AIC499, a Novel Resistance-Breaking Antibiotic against a Broad Range of MDR Gram-Negative Bacteria*. 2017.
85. López-Pérez, A., et al., *Discovery of Pyrrolidine-2, 3-diones as Novel Inhibitors of P. aeruginosa PBP3.* Antibiotics, 2021. **10**(5): p. 529.
86. Murray, C.W. and D.C. Rees, *The rise of fragment-based drug discovery.* Nature chemistry, 2009. **1**(3): p. 187-192.
87. Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Advanced drug delivery reviews, 1997. **23**(1-3): p. 3-25.
88. Hubbard, R.E., *Fragment approaches in structure-based drug discovery.* Journal of synchrotron radiation, 2008. **15**(3): p. 227-230.
89. Erlanson, D.A., et al., *Site-directed ligand discovery.* Proceedings of the National Academy of Sciences, 2000. **97**(17): p. 9367-9372.

90. Kenakin, T., *Pharmacology in drug discovery and development: Understanding drug response*. 2016: Academic Press.
91. Curatolo, W., *Physical chemical properties of oral drug candidates in the discovery and exploratory development settings.* Pharmaceutical Science & Technology Today, 1998. **1**(9): p. 387-393.
92. Clark, R.W., et al., *Description of the torcetrapib series of cholesteryl ester transfer protein inhibitors, including mechanism of actions.* Journal of lipid research, 2006. **47**(3): p. 537-552.
93. Hay, M., et al., *Clinical development success rates for investigational drugs.* Nature biotechnology, 2014. **32**(1): p. 40-51.
94. Matthews, H., J. Hanison, and N. Nirmalan, *"Omics"-informed drug and biomarker discovery: opportunities, challenges and future perspectives.* Proteomes, 2016. **4**(3): p. 28.
95. Norton, R.S., et al., *Applications of 19F-NMR in fragment-based drug discovery.* Molecules, 2016. **21**(7): p. 860.
96. Congreve, M., et al., *A'rule of three'for fragment-based lead discovery?* Drug discovery today, 2003. **8**(19): p. 876-877.
97. Clark, D.E. and S.D. Pickett, *Computational methods for the prediction of 'drug-likeness'.* Drug discovery today, 2000. **5**(2): p. 49-58.
98. Gossert, A.D. and W. Jahnke, *NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules.* Progress in nuclear magnetic resonance spectroscopy, 2016. **97**: p. 82-125.
99. Li, Q. and C. Kang, *A practical perspective on the roles of solution NMR spectroscopy in drug discovery.* Molecules, 2020. **25**(13): p. 2974.
100. Lamoree, B. and R.E. Hubbard, *Current perspectives in fragment-based lead discovery (FBLD).* Essays in biochemistry, 2017. **61**(5): p. 453-464.
101. Boehm, H.-J., et al., *Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening.* Journal of medicinal chemistry, 2000. **43**(14): p. 2664-2674.
102. Hajduk, P.J., et al., *Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis.* Journal of medicinal chemistry, 1999. **42**(19): p. 3852-3859.
103. Fejzo, J., C. Lepre, and X. Xie, *Application of NMR screening in drug discovery.* Current topics in medicinal chemistry, 2003. **3**(1): p. 81-97.
104. Shuker, S.B., et al., *Discovering high-affinity ligands for proteins: SAR by NMR.* Science, 1996. **274**(5292): p. 1531-1534.
105. Hajduk, P., et al., *Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR.* Journal of the American Chemical Society, 1997. **119**(25): p. 5818-5827.
106. Oltersdorf, T., et al., *An inhibitor of Bcl-2 family proteins induces regression of solid tumours.* Nature, 2005. **435**(7042): p. 677-681.
107. Erlanson, D.A., R.S. McDowell, and T. O'Brien, *Fragment-based drug discovery.* Journal of medicinal chemistry, 2004. **47**(14): p. 3463-3482.
108. Baell, J. and M.A. Walters, *Chemistry: Chemical con artists foil drug discovery.* Nature News, 2014. **513**(7519): p. 481.
109. Baell, J.B. and G.A. Holloway, *New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays.* Journal of medicinal chemistry, 2010. **53**(7): p. 2719-2740.
110. McGovern, S.L., et al., *A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening.* Journal of medicinal chemistry, 2002. **45**(8): p. 1712-1722.
111. Cramer, J., et al., *A False‐Positive Screening Hit in Fragment‐Based Lead Discovery: Watch out for the Red Herring.* Angewandte Chemie International Edition, 2017. **56**(7): p. 1908-1913.
112. Dalvit, C. and A. Vulpetti, *Ligand-based fluorine NMR screening: principles and applications in drug discovery projects.* Journal of medicinal chemistry, 2018. **62**(5): p. 2218-2244.

113. Chessari, G., et al., *Fragment-based drug discovery targeting inhibitor of apoptosis proteins: discovery of a non-alanine lead series with dual activity against cIAP1 and XIAP.* Journal of medicinal chemistry, 2015. **58**(16): p. 6574-6588.
114. Walpole, S., et al., *STD NMR as a technique for ligand screening and structural studies.* Methods in enzymology, 2019. **615**: p. 423-451.
115. Dalvit, C., et al., *WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability.* Journal of biomolecular NMR, 2001. **21**(4): p. 349-359.
116. Shi, L. and N. Zhang, *Applications of Solution NMR in Drug Discovery.* Molecules, 2021. **26**(3): p. 576.
117. Sugiki, T., et al., *Current NMR techniques for structure-based drug discovery.* Molecules, 2018. **23**(1): p. 148.
118. Vulpetti, A. and C. Dalvit, *Fluorine local environment: from screening to drug design.* Drug discovery today, 2012. **17**(15-16): p. 890-897.
119. Dolbier Jr, W.R., *Guide to fluorine NMR for organic chemists*. 2016: John Wiley & Sons.
120. Dalvit, C. and A. Vulpetti, *Fluorine–protein interactions and 19F NMR isotropic chemical shifts: an empirical correlation with implications for drug design.* ChemMedChem, 2011. **6**(1): p. 104-114.
121. Abraham, R.J. and M. Mobli, *Modelling 1H NMR spectra of organic compounds: theory, applications and NMR prediction software*. 2008: John Wiley & Sons.
122. Schaefer, T., et al., *Spin–spin coupling constants between side-chain and ring fluorine nuclei in some benzotrifluoride, benzal fluoride, and benzyl fluoride derivatives: coupling mechanisms.* Canadian Journal of Chemistry, 1979. **57**(7): p. 807-812.
123. Richards, R. and T. Schaefer, *Nuclear resonance spectra of some fluorine derivatives of toluene and benzotrifluoride.* Transactions of the Faraday Society, 1958. **54**: p. 1447-1453.
124. Brey, W., L. Jaques, and H. Jakobsen, *A $^{13}C-\{^{1}H\}$ double resonance study of the signs of $^{1}H-^{19}F$ and $^{13}C-^{19}F$ spin coupling constants in fluorobenzenes and 2-fluoropyridine.* Organic Magnetic Resonance, 1979. **12**(4): p. 243-246.
125. Guo, W. and T.C. Wong, *Study of $^{13}C-^{19}F$ and $^{1}H-^{19}F$ couplings in some fluorinated aromatic compounds using two-dimensional $^{13}C-^{1}H$ chemical shift correlation spectroscopy with proton homonuclear decoupling.* Magnetic resonance in chemistry, 1986. **24**(1): p. 75-79.
126. Dalvit, C., *Ligand-and substrate-based ^{19}F NMR screening: Principles and applications to drug discovery.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2007. **4**(51): p. 243-271.
127. Dalvit, C. and A. Vulpetti, *Weak intermolecular hydrogen bonds with fluorine: detection and implications for enzymatic/chemical reactions, chemical properties, and ligand/protein fluorine NMR screening.* Chemistry–A European Journal, 2016. **22**(22): p. 7592-7601.
128. Dalvit, C., et al., *Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability.* Journal of the American Chemical Society, 2003. **125**(25): p. 7696-7703.
129. Rüdisser, S.H., et al., *Efficient affinity ranking of fluorinated ligands by ^{19}F NMR: CSAR and FastCSAR.* Journal of Biomolecular NMR, 2020. **74**(10): p. 579-594.
130. Stadmiller, S.S., et al., *Rapid quantification of protein-ligand binding via ^{19}F NMR lineshape analysis.* Biophysical journal, 2020. **118**(10): p. 2537-2548.
131. Purslow, J.A., et al., *NMR methods for structural characterization of protein-protein complexes.* Frontiers in molecular biosciences, 2020. **7**: p. 9.
132. Fielding, L., *NMR methods for the determination of protein–ligand dissociation constants.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2007. **51**(4): p. 219-242.
133. Frueh, D.P., *Practical aspects of NMR signal assignment in larger and challenging proteins.* Progress in nuclear magnetic resonance spectroscopy, 2014. **78**: p. 47-75.

134. Xu, Y., et al., *A new strategy for structure determination of large proteins in solution without deuteration.* Nature methods, 2006. **3**(11): p. 931-937.
135. Kainosho, M., et al., *Optimal isotope labelling for NMR protein structure determinations.* Nature, 2006. **440**(7080): p. 52-57.
136. Ulrich, E.L., et al., *BioMagResBank.* Nucleic acids research, 2007. **36**(suppl_1): p. D402-D408.
137. Sattler, M. and S.W. Fesik, *Use of deuterium labeling in NMR: overcoming a sizeable problem.* Structure, 1996. **4**(11): p. 1245-1249.
138. Pervushin, K., *Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology.* Quarterly reviews of biophysics, 2000. **33**(2): p. 161-197.
139. Pervushin, K., et al., *Attenuated T 2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution*, in *NMR WITH BIOLOGICAL MACROMOLECULES IN SOLUTION: A Selection of Papers published from 1996 to 2020 by Kurt Wüthrich*. 2021, World Scientific. p. 61-66.
140. Takeuchi, K., et al., *Nitrogen detected TROSY at high field yields high resolution and sensitivity for protein NMR.* Journal of biomolecular NMR, 2015. **63**(4): p. 323-331.
141. Facey, G. *TROSY*. 2015 [cited 2021 (22. September)]; Available from: http://u-of-o-nmr-facility.blogspot.com/2015/05/trosy.html.
142. Hyberts, S.G., et al., *Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling.* Journal of biomolecular NMR, 2012. **52**(4): p. 315-327.
143. Hyberts, S.G., et al., *FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation.* Journal of biomolecular NMR, 2009. **45**(3): p. 283-294.
144. Kazimierczuk, K. and V. Orekhov, *Non‐uniform sampling: post‐Fourier era of NMR data collection and processing.* Magnetic Resonance in Chemistry, 2015. **53**(11): p. 921-926.
145. Hyberts, S.G., K. Takeuchi, and G. Wagner, *Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein NMR data.* Journal of the American Chemical Society, 2010. **132**(7): p. 2145-2147.
146. Berman, H.M., et al., *The protein data bank.* Nucleic acids research, 2000. **28**(1): p. 235-242.
147. D'Arcy, A., et al., *Microseed matrix screening for optimization in protein crystallization: what have we learned?* Acta Crystallographica Section F: Structural Biology Communications, 2014. **70**(9): p. 1117-1126.
148. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of biomolecular NMR, 1995. **6**(3): p. 277-293.
149. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: development of a software pipeline.* Proteins: structure, function, and bioinformatics, 2005. **59**(4): p. 687-696.
150. Kabsch, W., *Xds.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(2): p. 125-132.
151. Tickle, I., et al., *Staraniso.* Cambridge, United Kingdom: Global Phasing Ltd, 2018.
152. Vagin, A. and A. Teplyakov, *Molecular replacement with MOLREP.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(1): p. 22-25.
153. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallographica Section D: Structural Biology, 2019. **75**(10): p. 861-877.
154. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 486-501.
155. Schrodinger, L. *The PyMOL Molecular Graphics System, Version 2.4.0*. [cited 2021 (27. August)]; Available from: https://pymol.org/2/support.html?#citing.
156. Laskowski, R.A. and M.B. Swindells, *LigPlot+: multiple ligand–protein interaction diagrams for drug discovery*. 2011, American Chemical Society Publications.

157. Shen, Y. and A. Bax, *Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks.* Journal of biomolecular NMR, 2013. **56**(3): p. 227-241.
158. Marion, D., *Timing and related artifacts in multidimensional NMR.* Concepts in Magnetic Resonance Part A, 2012. **40**(6): p. 326-340.
159. Shi, P., et al., *In situ 19F NMR studies of an E. coli membrane protein.* Protein Science, 2012. **21**(4): p. 596-600.
160. Rosenau, C.P., et al., *Exposing the origins of irreproducibility in fluorine NMR spectroscopy.* Angewandte Chemie International Edition, 2018. **57**(30): p. 9528-9533.
161. Panwalkar, V., et al., *The Nedd4–1 WW Domain Recognizes the PY Motif Peptide through Coupled Folding and Binding Equilibria.* Biochemistry, 2016. **55**(4): p. 659-674.
162. Vulpetti, A., et al., *Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine.* Journal of the American Chemical Society, 2009. **131**(36): p. 12949-12959.
163. Mitton-Fry, R.M., et al., *Conserved structure for single-stranded telomeric DNA recognition.* Science, 2002. **296**(5565): p. 145-147.
164. Sticht, J., et al., *A peptide inhibitor of HIV-1 assembly in vitro.* Nature structural & molecular biology, 2005. **12**(8): p. 671-677.
165. Findeisen, M., T. Brand, and S. Berger, *A 1H‐NMR thermometer suitable for cryoprobes.* Magnetic Resonance in Chemistry, 2007. **45**(2): p. 175-178.
166. Markley, J.L., et al., *Recommendations for the presentation of NMR structures of proteins and nucleic acids (IUPAC Recommendations 1998).* Pure and applied chemistry, 1998. **70**(1): p. 117-142.
167. Wishart, D.S., et al., *1 H, 13 C and 15 N chemical shift referencing in biomolecular NMR.* Journal of biomolecular NMR, 1995. **6**(2): p. 135-140.
168. Sattler, M., J. Schleucher, and C. Griesinger, *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution.* Progress in nuclear magnetic resonance spectroscopy, 1999. **34**: p. 93-158.
169. Barkhuijsen, H., et al., *Retrieval of frequencies, amplitudes, damping factors, and phases from time-domain signals using a linear least-squares procedure.* Journal of Magnetic Resonance (1969), 1985. **61**(3): p. 465-481.
170. Bauman, J.D., J.J.E. Harrison, and E. Arnold, *Rapid experimental SAD phasing and hot-spot identification with halogenated fragments.* IUCrJ, 2016. **3**(1): p. 51-60.
171. Kurutz, J. *Background-free ^{19}F Experiments*. 2020 [cited 2021 (22. September)]; Available from: https://voices.uchicago.edu/chemnmr/2020/05/21/background-free-19f-experiments/.
172. Shi, P., et al., *Site‐specific 19F NMR chemical shift and side chain relaxation analysis of a membrane protein labeled with an unnatural amino acid.* Protein Science, 2011. **20**(1): p. 224-228.
173. Jordan, J.B., et al., *Fragment based drug discovery: practical implementation based on ^{19}F NMR spectroscopy.* Journal of medicinal chemistry, 2012. **55**(2): p. 678-687.
174. Nagatoishi, S., et al., *A combination of ^{19}F NMR and surface plasmon resonance for site-specific hit selection and validation of fragment molecules that bind to the ATP-binding site of a kinase.* Bioorganic & medicinal chemistry, 2018. **26**(8): p. 1929-1938.
175. Dalvit, C. and S. Knapp, *^{19}F NMR isotropic chemical shift for efficient screening of fluorinated fragments which are racemates and/or display multiple conformers.* Magnetic Resonance in Chemistry, 2017. **55**(12): p. 1091-1095.
176. Vulpetti, A., N. Schiering, and C. Dalvit, *Combined use of computational chemistry, NMR screening, and X‐ray crystallography for identification and characterization of fluorophilic protein environments.* Proteins: Structure, Function, and Bioinformatics, 2010. **78**(16): p. 3281-3291.
177. Schumann, F.H., et al., *Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions.* Journal of biomolecular NMR, 2007. **39**(4): p. 275-289.
178. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Progress in nuclear magnetic resonance spectroscopy, 2013. **73**: p. 1-16.

179. Morrison, J., et al., *Solution NMR study of the interaction between NTF2 and nucleoporin FxFG repeats.* Journal of molecular biology, 2003. **333**(3): p. 587-603.
180. Rajavel, M., et al., *Structural Characterization of Diazabicyclooctane β-Lactam "Enhancers" in Complex with Penicillin-Binding Proteins PBP2 and PBP3 of Pseudomonas aeruginosa.* mBio, 2021. **12**(1): p. e03058-20.
181. Markin, C.J. and L. Spyracopoulos, *Increased precision for analysis of protein–ligand dissociation constants determined from chemical shift titrations.* Journal of biomolecular NMR, 2012. **53**(2): p. 125-138.
182. Strelow, J., et al., *Mechanism of action assays for enzymes.* Assay Guidance Manual [Internet], 2012.
183. Rohs, R., C. Etchebest, and R. Lavery, *Unraveling proteins: a molecular mechanics study.* Biophysical journal, 1999. **76**(5): p. 2760-2768.
184. Buehler, M.J., *Tu (r) ning weakness to strength.* Nano Today, 2010. **5**(5): p. 379-383.
185. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* Journal of molecular biology, 2007. **372**(3): p. 774-797.
186. Kozlowski, L.P., *Proteome-pI: proteome isoelectric point database.* Nucleic acids research, 2017. **45**(D1): p. D1112-D1116.
187. Tugarinov, V. and L.E. Kay, *Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods.* Journal of the American Chemical Society, 2003. **125**(45): p. 13868-13878.
188. Xun, Y., et al., *Cell-free synthesis and combinatorial selective 15N-labeling of the cytotoxic protein amoebapore A from Entamoeba histolytica.* Protein expression and purification, 2009. **68**(1): p. 22-27.
189. Michel, E. and F.H.-T. Allain, *Selective amino acid segmental labeling of multi-domain proteins.* Methods in enzymology, 2015. **565**: p. 389-422.
190. Yang, D. and L.E. Kay, *TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein.* Journal of the American Chemical Society, 1999. **121**(11): p. 2571-2575.
191. Jung, Y.-S. and M. Zweckstetter, *Mars-robust automatic backbone assignment of proteins.* Journal of biomolecular NMR, 2004. **30**(1): p. 11-23.
192. Musafia, B., V. Buchner, and D. Arad, *Complex salt bridges in proteins: statistical analysis of structure and function.* Journal of molecular biology, 1995. **254**(4): p. 761-770.