# Targeting farnesyl pyrophosphate synthase of *Trypanosoma cruzi* by fragment-based lead discovery

## Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
dem Fachbereich Pharmazie der Philipps-Universität Marburg
vorgelegt von

**Joy Kristin Petrick**
aus Hamburg

Marburg/Lahn, 2019

| | |
|---|---|
| Erstgutachter | Herr Prof. Dr. Gerhard Klebe |
| Zweitgutachter | Herr Dr. Wolfgang Jahnke |

Tag des Einreichens: 01.08.2019
Tag der mündlichen Prüfung: 12.09.2019
Hochschulkennziffer: 1180

Success is not final,
failure is not fatal:
it is the courage to continue that counts.

Winston Churchill

# Table of contents

# Zusammenfassung

Ziel dieser Arbeit war die Entdeckung von Substanzen, die an *T. cruzi* FPPS (TcFPPS) binden und nicht der Stoffklasse der Bisphosphonate angehören. Zu diesem Zwecke wurde reines und homogenes TcFPPS durch rekombinante Expression in *E. coli* Bakterien und anschließende Aufreinigung mittels IMAC und SEC erhalten (Kapitel **5.1**). Darüber hinaus konnte ein zuverlässiges, reproduzierbares Kristallisationssystem etabliert werden, das Kristalle mit guten Diffraktionseigenschaften liefert. Das System weist ausgezeichnete Eigenschaften für Fragment-basiertes Screening (FBS) auf, da es mit verschiedenen Kristallisationsplatten kompatibel war und Apo-Kristalle lieferte, die bis zu 24 h in 15% DMSO stabil waren und die Aufnahme von Datensätzen mit einer Auflösung von etwa 1,6 Å erlaubten. Die höchste erreichte Auflösung für einen TcFPPS Kristall lag bei 1,28 Å (PDB ID 6R09).

Die allosterische Tasche in TcFPPS wurde mittels Sequenzanalyse und struktureller Überlagerung verschiedener FPPS Homologe untersucht (Kapitel **5.2**). Dabei zeigte sich, dass die allosterische Region in FPPS weniger konserviert ist als das aktive Zentrum. Unterschiede zwischen Aminosäuren an äquivalenten Positionen, die die allosterische Region bilden, wurden festgestellt. Dies ist überraschend, wenn man davon ausgeht, dass dieses Enzym produktinhibiert ist, wie für das humane FPPS (hFPPS) gezeigt werden konnte. Ein interessante Beobachtung war, dass die Aminosäure Phe50 in TcFPPS eine Ausnahme in einer ansonsten hochkonservierten Position ist. Es scheint die Tasche durch sterische Hinderung zu blockieren. Allosterische Inhibitoren von hFPPS wiesen zwar Bindungsaffinität zu TcFPPS auf, aber die beiden erhaltenen Kristallstrukturen zeigten, dass diese an der Proteinoberfläche binden (Bindungsstelle S1 und S2, PDB IDs 6R08 bzw. 6R07).

Die Novartis Haupt- und Fluor-Fragmentbibliotheken (1336 und 482 Verbindungen) wurden auf TcFPPS getestet, was zu 63 bzw. 45 validierten Fragmentbindern führte (Kapitel **5.3**). Die Durchführung des gleichen Screenings mit *T. brucei* FPPS (TbFPPS), dem Erreger der Afrikanischen Schlafkrankheit, und Gegenkontrolle auf hFPPS zeigte, dass einige Verbindungen selektiv an nur eines, oder zwei der Proteine binden. Auffallend war, dass TcFPPS im Allgemeinen mehr Binder hatte als TbFPPS, und auch mehr selektive Binder im Vergleich zu TbFPPS. Nachfolgende Kristallisationsexperimente mit den Bindern der Haupt-Fragmentbibliothek führten zu 3D-Strukturen von zwei TcFPPS-Komplexen. Ein Ligand bindet an die Grenzfläche des Homodimers und der andere im aktiven Zentrum. Letzterer wurde mit Hilfe des Tools Pan-Dataset Density Analysis (PanDDA) identifiziert. FBS mittels Röntgenkristallographie wurden im XChem Labor in Harwell, Großbritannien, und im HTX Labor in Grenoble, Frankreich, durchgeführt (Kapitel **5.4**). Der XChem-Screen identifizierte 35 Fragmentbinder (PDB IDs 5QPD – Z, 5QQ0 – 9, 5QQA – B) in Bindungsstellen, die über das gesamte Protein verteilt waren. Dazu

gehören das aktive Zentrum, die allosterische Bindungsstelle, die Homodimer-Grenzfläche, Bindungsstellen an der Oberfläche und eine neue Tasche in unmittelbarer Nähe des aktiven Zentrums. Erstmals wurden Fragmente identifiziert, die an die allosterische Bindungsstelle von TcFPPS im offenen Zustand binden. Eine Drehung der Phenyl-Seitenkette von Phe50 führte zur Öffnung dieser vorherig geschlossenen Tasche. Der Screen im HTX Labor identifizierte acht weitere Fragmentbinder für die aktive und allosterische Tasche.

Die ersten Optimisierungversuche eines Fragments zu einer Leitstruktur erfolgten mittels virtuellem Screening mit dem webbasierten Tool ANCHOR.QUERY. Sie ging von dem Fragmentbinder **LUY** aus (Kapitel **5.5**) und mittels Eintopf-Mehrkomponentenreaktionen wurden 11 Verbindungen synthetisiert (**MCR-1** – **11**). Allerdings war deren schlechte Löslichkeit in nachfolgenden Tests abträglich, und Kristallisationsexperimente führten nicht zu einem Strukturmodell eines Komplexes. Danach wurde der Ansatz des Fusionierens der Fragmente **AWM**, **LVV**, **LUY**, **LDV** und **AWV** für die chemische Optimierung gewählt (Kapitel **5.6**). Eine Bibliothek von 12 Verbindungen (**MCN-1** – **12**) wurde durch reduktive Aminierung synthetisiert. Kristallstrukturen mit den Verbindungen **MCN-1**, **-4** und **-8** zeigten unerwartete Bindungsmodi. Anstatt an der Bindungsstelle der Ausgangsfragmente, binden die fusionierten Substanzen an die auf der Proteinoberfläche befindliche Bindungsstelle S1 (PDB IDs 6R09, 6R0A, 6R0B).

Die 50 neuen Kristallstrukturen von TcFPPS-Fragment Komplexen, die in dieser Arbeit beschrieben sind, werden neue Impulse für die Medikamentenentwicklung für CD geben. Die große Vielfalt der chemischen Strukturen der Fragmente und die unterschiedlichen Bindungsstellen sind potenzielle Ansatzpunkte für Inhibitoren mit unterschiedlichen physikalisch-chemischen Eigenschaften und einer neuartigen Wirkungsweise, die helfen könnten, die mit den Bisphosphonaten verknüpften Einschränkungen zu überwinden.

# Summary

*Trypanosoma cruzi* (*T. cruzi*) is the causative agent of Chagas disease (CD), which mostly affects underprivileged populations in South and Central America. The current standard of care for this disease are the two empirically discovered drugs benznidazole and nifurtimox. They show low efficacy, difficulties in administration and severe side effects. Moreover, there are *T. cruzi* strains that have formed resistances. Thus, the development of a safe and efficient drug is urgently needed. *T. cruzi* is dependent on isoprenoid biosynthesis as ergosterol and other 24-alkylsterols are essential metabolites that cannot be acquired by other mechanisms. Therefore, it was hypothesised that enzymes along this pathway are promising drug targets. A number of compounds targeting these enzymes were tested and have been shown to inhibit parasite growth. Among those enzymes is farnesyl pyrophosphate synthase (FPPS), a key branch-point enzyme in the isoprenoid pathway, which is in the focus of this work. It catalyses the synthesis of farnesyl pyrophosphate (FPP), a C15 building block in sterol biosynthesis and in protein prenylation of signalling proteins. Bisphosphonates (BPs) are known active site-directed FPPS inhibitors, which exhibit ideal pharmacokinetics to target bone mineral and are used to treat bone diseases. BPs can also combat *T. cruzi* flagellates but are not ideal to treat CD due to their pharmacokinetics. In the search for new chemotypes, several non-BP inhibitors that bind to another pocket were found for human FPPS (hFPPS) by fragment based screening (FBS). Recently, it was shown that the product of FPPS, farnesyl pyrophosphate (FPP), can bind to this pocket and locks the enzyme in an open and inactive state, thus showing the allosteric character of this pocket.

The current work aims at the discovery of non-BP inhibitors of *T. cruzi* FPPS (TcFPPS), which could be starting points for the development of a treatment against CD. Towards this goal, recombinant expression in *E. coli* cells and purification by means of IMAC and SEC yielded pure und homogenous TcFPPS (chapter **5.1**). This includes unlabelled, $^{13}C^{15}N$-labelled and *in vivo* biotinylated avi-tagged TcFPPS. Furthermore, a novel, reliable, highly reproducible, and well-diffracting crystallization system was established. The system exhibits excellent properties for FBS as it was compatible with different types of 96-well plates. Apo crystals were stable for up to 24 h in 15% DMSO and allowed collection of data sets with a diffraction limit of around 1.6 Å. The best achieved diffraction limit was 1.28 Å for a soaked TcFPPS crystal (PDB ID 6R09).

The allosteric region in TcFPPS was investigated by means of sequence analysis and structural superimposition of various orthologous FPPSs (chapter **5.2**). This revealed that the allosteric region is less conserved than the active site. Differences among residues in equivalent positions that form the allosteric site were observed, which is surprising if it is assumed that all FPPSs can be product inhibited as hFPPS. A remarkable finding is that residue Phe50 in TcFPPS is an exception in an otherwise highly conserved position. It causes steric hindrance of the pocket

in TcFPPS. An attempt to reposition established allosteric inhibitors of hFPPS showed binding affinity to TcFPPS but the two obtained crystal structures demonstrated their binding to sites on the protein surface (sites S1 and S2, PDB IDs 6R08 and 6R07, respectively).

The Novartis core and fluorine library (1336 and 482 compounds) were screened on TcFPPS, which resulted in 63 and 45 validated fragment hits, respectively (chapter **5.3**). Performing the same screen with *T. brucei* FPPS (TbFPPS), the causative agent of African sleeping sickness, and counter screening on hFPPS led to unique, pairwise and triple binders demonstrating selectivity at the early stage of FBS. Strikingly, TcFPPS has generally more binders than TbFPPS, and TcFPPS has many unique hits when compared to TbFPPS. Subsequent crystallization experiments with the core library hits resulted in 3D structures of two TcFPPS complexes. One ligand binds to the homodimer interface (site S12) and the other one in the active site. The latter was identified by using the statistical analysis tool Pan-Dataset Density Analysis (PanDDA). FBS by X-ray crystallography at the XChem facility in Harwell, UK, and the HTXlab in Grenoble, France, were conducted (chapter **5.4**). The XChem screen identified 35 fragment binders (PDB IDs 5QPD – Z, 5QQ0 – 9, 5QQA – C) in binding sites that were distributed over the entire protein. This includes the active site, the allosteric site, the homodimer interface, sites on the surface and a new site in close proximity to the active site. Strikingly, the first two fragments binding to the allosteric site of TcFPPS in its open state were identified. Rotation of the phenyl side chain of Phe50 led to opening of the former closed pocket. The HTXlab screen identified additional binders for the active and allosteric site. In total 1244 data sets were collected and analysed. This process was accelerated using PanDDA.

The first fragment-to-lead optimization by means of virtual screening using the web-based platform ANCHOR.QUERY was based on fragment hit **LUY** (chapter **5.5**). Compounds were synthesised using one-pot one-step multi-component reactions. Synthesis of 11 compounds (**MCR-1** – **11**) was successful, but poor solubility was detrimental in subsequent testing on TcFPPS and crystallization experiments did not lead to a structural model of a complex. A second fragment-to-lead optimization using a fragment merging approach for chemical optimization was based on the active site-directed binders **AWM**, **LVV**, **LUY**, **LDV** and **AWV** (chapter **5.6**). A library of 12 compounds (**MCN-1** – **12**) was synthesised by reductive amination. X-ray structures revealed unexpected binding modes for compounds **MCN-1**, **-4** and **-8**. Instead of retaining the binding site of the fragment, the merged compounds bind to the surface-directed binding site S1 (PDB IDs 6R09, 6R0A, 6R0B). Nevertheless, the 50 new crystal structures of TcFPPS-fragment complexes discussed in this work will pave the way for future drug discovery campaigns for CD. The large diversity of the fragments' scaffolds and different binding sites are potential starting points for inhibitors with different physicochemical properties and a novel mode of action that might help to overcome the limitations related to the BP scaffold.

# List of figures

# List of tables

# List of acronyms and abbreviations

| | |
|---|---|
| ALE | alendronate |
| approx. | approximately |
| ATP | adenosine trisphosphate |
| BirA | bifunctional ligase/repressor |
| BisTris | 2,2-Bis(hydroxymethyl)-2,2′,2″-nitrilotriethanol |
| BNZ | benznidazole |
| bp | base pairs |
| BPs | bisphosphonates |
| BSA | bovine serum albumin |
| CD | Chagas disease |
| CPMG | Carr-Purcell-Meiboom-Gill Sequence |
| CYP51 | sterol 14α-demethylase |
| d- | deuterated |
| DCM | dichloromethane |
| dd $H_2O$ | double-distilled water |
| DLS | Diamond Light Source |
| DMAPP | dimethylallyl diphosphate |
| DMSO | dimethylsulfoxid |
| DNDi | Drugs for Neglected Disease *initiative* |
| DSF | differential scanning fluorimetry |
| DSPL | Diamond-SGC poised library |
| DSS | 4,4-dimethyl-4-silapentane-1-sulfonic acid |
| DTT | dithiothreitol |
| DTU(s) | discrete typing unit(s) |
| EDTA | ethylenediaminetetraacetic acid |
| ESI | electrospray ionization |
| ESRF | European Synchrotron Radiation Facility |
| FARM | first aspartate-rich motif |
| FBDD | fragment-based drug discovery |
| FBLD | fragment-based lead discovery |
| FBS | fragment-based screening |
| FCC | flash column chromatography |
| FDA | Food and Drug Administration |
| FPP | farnesyl pyrophosphate |
| FPPS | farnesyl pyrophosphate synthase |
| FT | Fourier transform |
| GGPP | geranylgeranyl pyrophosphate |
| GPP | geranyl pyrophosphate |
| HBA | hydrogen-bond acceptor |
| HBD | hydrogen-bond donor |
| HCS | high content screening |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| hFPPS | human farnesyl pyrophosphate synthase |
| HMGCR | hydroxymethylglutaryl-CoA reductase |
| HMQC | heteronuclear multiple quantum coherence |
| HPLC | high pressure liquid chromatography |
| HRV 3C | cysteine protease from human rhinovirus 3C |
| HSQC | heteronuclear single quantum coherence |
| HTS | high throughput screening |
| $IC_{50}$ | inhibitory concentration achieving 50% inhibition |
| IMAC | immobilized metal affinity chromatography |
| IPP | isopentyl diphosphate |
| IPTG | Isopropyl-β-D-thiogalactopyranosid |
| KanR | kanamycin resistance |
| $K_d$ | dissociation constant |
| lacI | lac repressor |
| lacO | lac operon |
| LB | lysogeny broth |

| | |
|---|---|
| LC | liquid chromatography |
| LE | ligand efficiency |
| logP | partition coefficient |
| MDG | metals, aspartic acid, glucose |
| MCR | multi component reaction |
| MMS | microseed matrix screening |
| modAI | modified autoinducing medium |
| MR | molecular replacement |
| MS | mass spectrometry |
| MVK | mevalonate kinase |
| MW | molecular weight |
| N-BP(s) | nitrogen-containing bisphosphonate(s) |
| ND(s) | neglected disease(s) |
| NFX | nifurtimox |
| NMR | nuclear magnetic resonance |
| NOE | Nuclear Overhauser Effect |
| NTA | nitrilotriacetic acid |
| $OD_{600}$ | optical density, measured at a wavelength of 600 nm |
| P1 | 90 ° pulse, hard pulse |
| PAGE | polyacrylamide gel electrophoresis |
| PAM | pamidronate |
| PanDDA | Pan-Dataset Density Analysis |
| PCR | polymerase chain reaction |
| PDB | protein data bank |
| PEG | polyethylene glycol |
| PFT | protein farnesyltransferase |
| PK(s) | pharmacokinetic(s) |
| PP | pyrophosphate |
| occ. | occupancy |
| RIS | risedronate |
| RMSD | rout mean square deviation |
| Ro3 | rule-of-three |
| Ro5 | rule-of-five |
| SADD | structure aided drug design |
| SAR | structure activity relationship |
| SARM | second aspartate-rich motif |
| SBDD | structure-based drug discovery |
| SDS | sodium dodecyl sulphate |
| SEC | size exclusion chromatography |
| SNPs | single nucleotide polymorphisms |
| SOC | super optimal broth |
| SOFAST | band-selective optimized flip angle short transient |
| SQLE | squalene epoxidase |
| SQS | squalene synthase |
| SPdB12 | pulse for water suppression, soft pulse |
| SPR | surface plasmon resonance |
| STD | saturation transfer difference |
| *T. cruzi* | *Trypanosoma cruzi* |
| TbFPPS | *T. brucei* farnesyl pyrophosphate synthase |
| TCEP | tris(2-carboxyethyl)phosphine) |
| TcFPPS | *T. cruzi* farnesyl pyrophosphate synthase |
| TEA | trimethylamine |
| TE buffer | tris-EDTA buffer |
| TPP | target product profile |
| Tris | tris(hydroxymethyl)aminomethane |
| VDX | vapour diffusion X-ray crystallography |
| waterLOGSY | water-ligand observed via gradient spectroscopy |
| zgesgsp | zero-go excitation sculpting gradient programming |
| ZOL | zoledronate |

Amino acids are abbreviated by their commonly used three letter code.

# 1. Introduction

## 1.1 Chagas disease

Chagas disease (CD) or American trypanosomiasis is a vector-borne disease caused by the parasite *Trypanosoma cruzi* (*T. cruzi*), a parasite that affects mostly underprivileged populations in Southern and Central America[1]. CD is one of 17 neglected diseases listed by the World Health Organisation (WHO)[2]. According to the WHO, six to seven million people are infected all over the world and more than 70 million people are at risk to get infected. Around 10.000 people die every year because of complications linked to this disease[2-3]. In Latin America, CD is a major public health burden[4] associated with the loss of approx. 546.000 disability-adjusted life-years (DALYs)[5]. This results in an estimated economic burden of more than seven billion dollar per year[6]. Countries outside Latin America account for an estimated 4.2% of DALYs and, disproportionately, for 21% of health care costs related to CD[5b]. In the last decades, public health programs significantly reduced the prevalence of CD through vector control programs, improvement of rural housing quality, better screening programs, and access to diagnostics and treatment. Nevertheless, CD remains the most prevalent parasitic disease in the Americas[4, 7].

CD has been present in a sylvatic cycle in America for over 10 million years before the arrival of man[8]. Around 10.000 years ago it became an anthropozoonosis, meaning it primarily affected animals, but was also transmitted to humans in the context of agricultural activity and the domestication of animals[9]. Due to progressive deforestation and a concomitant decrease of wild animal populations, triatomine bugs, which are vectors of *T. cruzi*, lost their main food source. Thus, CD turned into an endemic zoonosis approx. 200 to 300 years ago[9b, 10]. In 1909, Carlos Chagas first described CD in humans and named *T. cruzi* as causative agent and triatomine bugs as its main vector[1b, 1c, 11]. In 1912, Emile Brumpt described the mode of natural transmission of the infection via the feces of the bug[12].

Today, medication is based on two empirically discovered drugs, benznidazole and nifurtimox, which have limitations such as low efficacy in the chronic stage in adults, difficulties in administration, severe side effects and ineffectiveness in resistant *T. cruzi* strains[2, 13]. Thus, an effective drug as a reliable cure is lacking and there is no vaccine for disease prevention either[14]. In consequence, there is a continuing and compelling need for new drugs for a safe and efficacious anti-Chagas treatment[15].

### 1.1.1 *Trypanosoma cruzi*

There are more than 150 species of blood-sucking bugs[16]. The most relevant vectors of CD are the genus Triatoma, known as kissing bug, (*T. infestans*, *T. brasiliensis*), Rhodnius (*R. prolixus*) and Panstrongylus (*P. megistus*). It is mostly assumed that the occurrence of suitable vector species is limited to Southern and Central America[17]. The parasite *T. cruzi* is a homoflagellate protozoan of the order *Kinetoplastida* and family *Trypanosomatidae*[1c, 18]. There are numerous strains that show phenotypic and genetic diversity and are divided into seven discrete typing units (DTUs), TcI to TcVI and Tcbat[19]. Some strains are of higher clinical significance than others, which is due to variations in drug susceptibility[20], virulence strength[21], and the availability to invade host tissues[22]. Strains can be classified by a typing assay identifying key discriminant single nucleotide polymorphisms (SNPs)[23].

*T. cruzi* has the ability to infect and replicate in various tissue types of its host, including cardiac muscle cells, smooth muscle cells, skeletal muscle cells, neurons, macrophages, and dendritic cells[22, 24]. The parasite has a life cycle with four phases that occur in its insect vector and in the mammalian host: (1) Replicative but non-infectious epimastigotes are found in the vector's digestive tract. (2) Epimastigotes differentiate into the metacyclic trypomastigotes form and are subsequently transferred to the mammalian host via contamination of the bite wound with the bug's feces. (3) Trypomastigotes invade host cells and further differentiate into intracellularly replicative amastigotes. (4) Amastigotes differentiate back into trypomastigotes, which invade neighbouring cells after host cell disruption[18, 25]. The cycle is completed when blood-borne trypomastigotes are ingested by a triatomine bug[25b] (**Figure 1**).

*In vitro* studies have shown that infectious trypomastigotes actively attach to and invade mammalian host cells within 5 to 10 min after infection, forming a parasitophorous vacuole[22, 26]. After 1 to 2 h the trypomastigote escapes this vacuole and differentiates into a replicative amastigote in the host cell cytoplasm[26]. After 5 to 6 d and several replication cycles, amastigotes occupy most of the cell volume, transform to trypomastigotes and rupture the host cell[27]. During this process, *T. cruzi* excretes proteins, such as cruzain, P21, phospholipase A and other soluble factors[25a] for protection against the host's immune response and promotion of its own adhesion, recognition and invasion mechanisms by manipulating the host cell signalling pathways[25a, 28].

Complexity and timing of the *T. cruzi* life cycle in mammalian host cells are important factors in cell-based screening experiments where parasite growth is quantified in co-culture with mammalian host cells[24]. Since the amastigote stage is the replicative form in the mammalian host[1a] it is the preferred parasitic target stage in cell-based assays[29]. Zingales *et al.*[30] recommend to validate promising drug candidates for broad activity against each DTU in secondary screens.

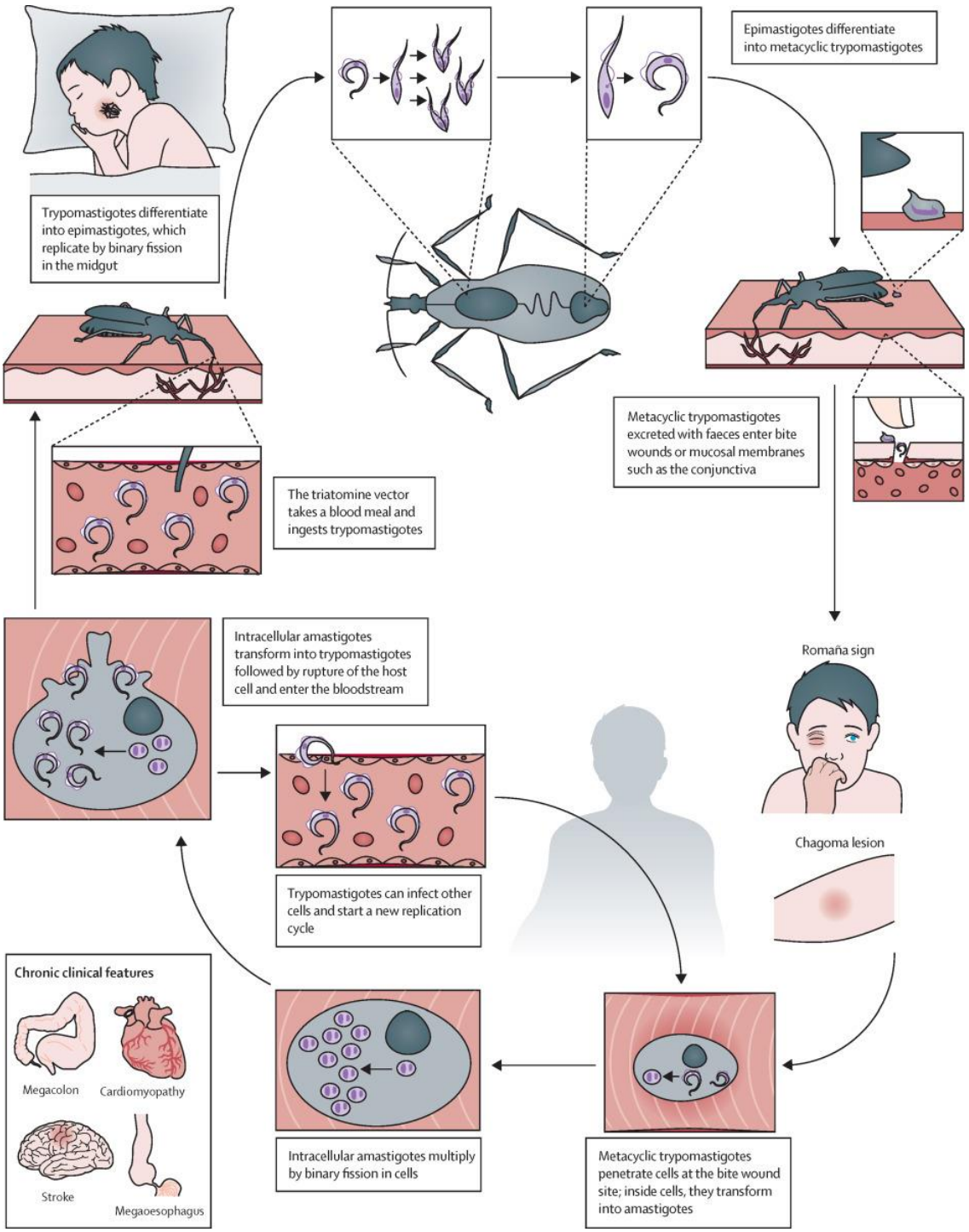



**Figure 1**: Life cycle of *T. cruzi*. Reprinted from Perez-Molina *et al*.[31] with permission from Elsevier.

### 1.1.2 Infection

In endemic regions, mainly in rural areas, natural vectorial transmission of *T. cruzi* via triatomine bugs takes place in the course of the bug's nocturnal blood meal[4]. Infected triatomines

often excrete feces contaminated with parasites next to the bite wound[4]. By unintentional scratching of the itching bite site the parasites enter the wound or near mucosal surfaces[32]. Other infection routes are food born infection[6], congenital transfusion from mother to child[33], transfusion of contaminated blood or transplantation of organs[34] and accidental contact in laboratories[35].

Food born infection occurs by ingestion of food or beverages contaminated with trypomastigotes in sylvatic and rural environments[6, 36]. Food contaminations occur through whole triatomine insects and their feces, or via feces of other vertebrates such as dogs, cats, bats, rats and armadillos[6, 36-37]. Taken together, these vectors still play a crucial role in orally transmitted CD, which often manifests with particularly severe symptoms due to high initial parasite loads[38]. Food preparation techniques such as drying[39] and heating[40] inactivates trypomastigotes, however, refrigeration and freezing[41] show little destructive effects. Vertical transmission from mother to child is becoming a more prominent infection route representing rates of up to one third of new infections[42]. Therefore, screening of pregnant women is critical to prevent disease prevalence[17, 33, 43]. Infected newborns show high parasite loads in their blood, which allows relatively easy diagnosis[17, 44]. Another notable transmission route is the infection after transfusion of contaminated blood or transplantation of organs with persistent parasites[34]. Chemical sterilization of blood samples in endemic regions with gentian violet[34a] prevented transmission, but proved unacceptable due to purplish skin staining of transfusion patients[45]. Therefore, prevention of this route is achieved by better control of donors with serological screening[45].

### 1.1.3 Disease stages

After initial infection and an incubation period of 5 d to 40 d, the disease starts with the acute phase[4, 17]. While mostly asymptomatic and undetected in adults, children and a small subset of adults exhibit fever, headache, decreased appetite, swollen lymph nodes, and show the Romaña sign (swollen eyelid) or a Chagoma (swollen bite wound) (**Figure 1**)[1a, 17]. Around 5% of acutely infected patients, again mostly children, die of acute myocarditis (inflammation of the heart muscle) or meningoencephalitis (inflammation of the brain)[1a, 14].

If the patient is left untreated, the acute phase is followed by an intermediate phase that lasts for 20 to 30 years[1a]. It is an asymptomatic phase, with no physical signs of disease[14]. Despite pathogen persistence, the levels of parasites in the blood are close to the detection limit, therefore making parasitaemia difficult to diagnose. About 70% of intermediate CD patients either clear the infection or just remain asymptomatic for the rest of their lives[1a].

The remaining 30% develop clinical symptoms and become chronic CD patients. They experience irreversible damage to cardiac and gut tissues leading to abnormal heart rate, cardiac arrest, damage of the nervous system, and digestive tract lesions[1a, 14]. It was initially hypothesized

that the organ damage is caused by an autoimmune response[46], however, it was later stated to be a consequence of the inflammatory response triggered by parasite persistence in the patient[47]. Accordingly, *T. cruzi* pathology is related to its presence in muscle tissue during the chronic stage of the disease[1a]. The four most frequent and severe clinical manifestations are Chagastic cardiomyopathy, stroke, and megaoesophagus which are characterized by abnormal enlargement of the heart chambers, the colon and oesophagus, respectively (**Figure 1**)[1a, 15a, 17, 48]. In consequence, heart failure and failure of the gastrointestinal tract function are the most common causes of death[1a]. Despite ongoing efforts, the underlying mechanism that determines which patients develop chronic CD and which patients remain asymptotic are poorly understood[49]. Finally, patients undergoing immunosuppressive therapy or immunocompromised individuals, such as HIV patients, are at higher risk to experience reactivation of *T. cruzi* parasites[36, 50].

## 1.2 Diagnosis

The most appropriate diagnostic strategy depends on the clinical stage of CD[17]. During the acute phase[51], after congenital infection[52], and after transfusion transmission[53], parasite loads in the blood are high and trypomastigotes can be observed in peripheral blood smears under the microscope. The second often applied and much more sensitive method is the polymerase chain reaction (PCR) which assesses the presence of *T. cruzi* DNA in peripheral blood[54]. During the intermediate and chronic phase the levels of parasites in the blood are below the detection limit, therefore making parasitaemia difficult to diagnose. Even PCR can lead to false-negative results[55]. Verification of antibodies against *T. cruzi* in the host's blood by use of trypomastigote excreted-secreted antigens based Western blot analysis (TESA-WB) is an alternative option at this stages of disease[56]. It is recommended to use at least two different serological test methods to confirm a positive diagnosis because the rates of false-positive tests are high[4].

Chemotherapy with benznidazole or nifurtimox reduces the parasite load below the detection limit making it difficult to determine treatment success or to attest cure[1a]. Microscopic quantification of parasitaemia provides a measure of parasite suppression, but is not sufficient to prove parasitological cure, as parasites can circulate at low levels in the blood or remain present in tissues[29, 57]. Parasitological tests are more sensitive but cannot guarantee a cure either. Among them are the aforementioned PCR and xenodiagnosis, in which the feces of previously uninfected bugs is analysed after they had been allowed to take a blood meal, and microscopy after a long term blood culture[29, 58]. Further conventional serological tests, such as enzyme-linked immunosorbent assay (ELISA), indirect immunofluorescence (IIF) and indirect hemagglutination assay (IHA) exist and are available for diagnosis[58-59].

Currently, new tests are under development. Parasitic persistence was assessed by simultaneous profiling of several *T. cruzi* antigens[60] and lately it was shown that the response of the single antibody AB3 is sufficient[61]. Apo lipoprotein A1 and fibronectin fragments were identified as potential markers predictive of cure[62]. In mice the most sensitive measure of cure after a completed drug treatment is obtained by a subsequent immunosuppressive therapy that causes a parasitaemia rebound, which can be detected by microscopy[63], blood culture[64] or PCR[64-65]. Further research on reliable early diagnostic tools and techniques to determine therapeutic responses and evidence of cure are required. The identification of biomarkers to determine parasite clearance versus parasite persistence would allow to dramatically improve the treatment of patients and to evaluate new drugs to fight CD[7c, 37, 62, 66].

### 1.2.1 Medication and vaccines

The ultimate goal of CD chemotherapy is to prevent disease manifestation. Whether this requires complete parasitological cure is unknown. For chronic patients, chemotherapy should prevent disease progression or reverse symptoms[24]. These requirements are partially met by benznidazole (BNZ) (**1**) and nifurtimox (NFX) (**2**) (**Figure 2**), the only available trypanocidal drugs which have been empirically introduced into clinical therapy in the 1970s and 1960s, respectively[67]. BNZ (Abarax®, former Rochagan®) was developed by HOFFMANN-LA ROCHE and is now produced by ELEA[68]. NFX (Lampit®) was developed by Bayer. They provide the drug that can be requested from the WHO[68].

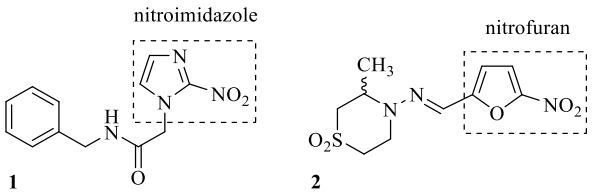



**Figure 2**: Chemical structure of BNZ (**1**) and NFX (**2**). Key scaffolds are highlighted with a box.

Both drugs are activated by type I nitroreductase followed by free-radical formation overwhelming the antioxidant capabilities of *T. cruzi*, as well as by the activity of the formed reduction intermediates which lead to lethal DNA strand breaks[13a, 69]. BNZ has the better safety and efficacy profile and is therefore used as first choice treatment[69a]. Long-term regimes with high dosages are required for an effective treatment[69b]. Treatment regimens suggest 5 to 7 mg · $kg^{-1}$ per day of BNZ divided in two doses for adults for 60 days or 8 to 10 mg · $kg^{-1}$ per day of NFX divided in three doses for 90 days[70]. Multiple doses are needed per day, as both drugs are rapidly metabolized by the cytochrome P450 system[71]. Severe side effects often prompt the

discontinuation of the treatment[69b]. The toxic effect of BNZ and NFX is associated with their chemical features. Both pharmaceuticals belong to the class of nitroaromatic drugs, which are known for chemistry-driven liver damage causing hepatitis (inflammation of the liver)[72]. Other side effects include dermatitis (inflammation of the skin), digestive intolerance (vomiting, anorexia), and peripheral neuropathy (damage to peripheral nerves)[73]. In addition, both compounds exhibit mutagenic properties[73]. These side effects often result in low patient compliance, specifically in intermediate phase patients which are usually symptom free[24, 74].

Nevertheless, treatment in the acute phase shows good efficacy in children, but limited efficacy in adults[68b]. Treatment success of chronic CD ranges from 20 to 50%[74-75]. The effect in advanced chronic patients is low[73a, 76] and efficacy is difficult to assess, since the patient groups vary in age distribution, length of CD manifestation and often suffer from additional diseases[75]. However, there is evidence that chronic patients treated with BNZ benefit from decrease in parasite levels and therefore medication is recommended [73a, 77]. The reasons for failure of treatment have not yet been fully explained, however, different evaluation methods, incomplete treatment, variable virulence among *T. cruzi* strains and differences between host's immune system are contributing factors[68b].

New therapeutic treatments are needed, not only to reduce side effects and toxicity but also because various *T. cruzi* strains show variable susceptibility to BNZ and NFX[68b, 69b]. The Colombian strain for example is highly resistant against both drugs[78]. *T. cruzi* strains with natural resistance against BNZ were shown to overexpress an ABCG-transporter gene that conveys drug resistance[79], but also type I nitroreductase and additional mechanisms play a role in drug-resistance[80].

To date, vaccines against CD are not available, however, preventive and therapeutic vaccines are currently being developed[81]. The recombinant antigens Tc24 and TSA-1 showed promising results in mice[81]. Recently, they have been tested by Villanueva-Lizama *et al*.[14] in a small group of infected humans ($n$ = 20) and healthy volunteers (n = 19). Indeed, both antigens triggered a secondary immune response in Chagastic patients. According to the authors, a therapeutic vaccine aimed at preventing or delaying the development of chronic CD would be an alternative or complement to current drug treatment[14].

### 1.2.2 Control strategies

Public health programs for vector control significantly reduced the prevalence of CD in the last decades[4, 7c]. However, CD control is highly heterogeneous between and within regions and countries and it is not eradicable at all because *T. cruzi* is also present in many different mammals[82]. Chemical vector control is a powerful way to reduce CD prevalence[83]. Spraying rural housings and the surrounding areas with insecticides by professional sprayers led to reduction

of triatomine bugs and thus reduction of transmission[68d, 83]. Among them was *T. infestans*, one of the main vectors in South America[84]. As a consequence, vectorial disease transmission was pushed back[83]. Some species developed insecticide resistance which is of growing concern today. An example is pyrethroid[85] and organophosphate[86] insecticide resistance of *T. infestans* reported for Argentina and Bolivia in the late 1990s. Mougabure-Cueto and Picollo[83] summarized the evolution of many different resistances. They reported on varying resistance profiles and mechanisms between resistant foci, suggesting an independent origin. Due to insecticide resistances, triatomines were observed after spraying with insecticides and the success of spraying campaigns was diminishing[83]. To control resistant foci, other known insecticides can be used for a while, but investigations on new insecticides will also be necessary[83].

One more control tool to mitigate the consequences of pesticide resistances is the improvement of rural housings to minimize colonization by triatomines, and thus minimizing human-triatomine interactions and reducing vector-borne transmission of CD[83]. In this context, initiatives to improve housing of the WHO and Pan American Health Organization (PAHO) have led to significant improvements[7b].

### 1.2.3 Chagas disease in non-endemic countries

Due to increasing migration flows and travelling, CD became a global health threat in non-endemic areas[17, 68d, 82]. This includes Europe, the US, Canada, Asia and Australia[17, 68d, 74, 87]. The relevant mechanisms for transmission in non-endemic countries are congenital transmission from mother to child[52], blood transfusion and organ transplantation[34c, 53].

Around 3.5 million Latin American immigrants live in Europe, mainly in Spain, Italy, France, the UK and Switzerland[17]. Approximating the number of Chagastic patients in Europe is difficult and estimates of CD prevalence vary widely due to different methodological approaches[17]. These problems are further exacerbated by qualitatively poor prevalence data from endemic regions[88]. Although only 4.290 cases have been confirmed in Europe, Bazile *et al.*[89] estimated that approx. 100.000 people are disease carriers.

European countries lack federal screening programs, therefore, tests are rare, by far not exhaustive or even not consistent[17, 90]. According to Requena-Mèndez *et al.*[91], testing Latin-American migrants for CD would be cost-effective and should be supported. The identification of CD infection in pregnant women is a major challenge for the prevention and control of CD in non-endemic countries[90]. Some countries have reference centres, but apart from that access to diagnosis and treatment is often low[17]. That is at least in part because physicians are rarely confronted with CD and lack expertise to accurately diagnose symptoms[17, 88a]. As a first step to improve the patients' situation, physicians need to be trained to recognize and treat

CD[17, 88a]. BNZ and NFX are classified as essential drugs by the WHO but are not registered in Europe and the US due to their severe side effects[1a, 17, 87a].

In the US CD became a major concern not only due to migration, but also due to the spread of triatomine vectors[70, 92]. They were first described in South Texas in the 1930s[93] and in consequence vectorial transmission takes place[94]. Hotez *et al*.[93] name human migration, poverty, climate change, transborder traffic, sea transportation, among others as major external factors driving neglected diseases in Texas.

## 1.3 Drug discovery landscape against Chagas disease

Neglected diseases (NDs), such as CD, account for approx. 11% of the global disease burden[95], however, only 1.3% (21) of the drugs launched between 1975 and 2004 were for their treatment[96]. Thus, the resource investment is disproportionate to the disease burden[15b]. Usually, the discovery of novel therapeutics against NDs is driven by academia and non-profit organizations as the market for such drugs is not of financial interest to pharmaceutical companies[24]. In consequence, the public sector and non-profit organizations finance 90% of resources invested in research on NDs[15b, 97]. Within the last two decades, CD emerged in non-endemic countries, therefore triggering research interest in the US and in European countries[37]. Several public-private partnerships and initiatives, such as Global Health Innovative Technology (GHIT), the Bill & Melinda Gates Foundation and Drugs for Neglected Disease *initiative* (DND*i*) were launched and became a driving force behind drug discovery for CD[7c, 98].

Ongoing efforts in drug research for CD include improvement of current treatments, label extension of drugs in clinical use, drug repositioning, and *de novo* drug discovery applied to phenotypic or target-based screening[7c, 37, 68d, 99]. Drug repositioning, also known as piggy-back or target hopping, in which well-known inhibitors against related targets and thus takes advantage of a former drug development process[15b, 100]. Several computational methods are available for drug repositioning that can either look for potential targets for a known drug or for potential drugs for a specific target[100]. Drug repositioning is inexpensive and saves resources and is thus increasingly used to discover novel drug candidates for NDs[101]. Drug discovery by a phenotypic approach examines the manifestation of parasitic infection without knowledge of the mechanism of action and hence the anti-parasitic activity, membrane permeability and host cell toxicity are directly tested[102]. In contrast, a target approach relies on a validated target, such as an enzyme that is essential in a metabolic pathway[103]. In this approach, differences in pathways, signalling cascades, and protein homologues between the protozoan parasite and the mammalian host are exploited to achieve drug selectivity[67, 104].

A compound for CD chemotherapy first needs to cross the cell membrane of infected mammalian cells and secondly move through the cytoplasm to cross the membrane of amastigotes[104]. Unfortunately, parasite-host interactions, variability of *T. cruzi* strains, and disease progression are not fully understood to date[98]. Nevertheless, a target product profile (TPP) for CD was published by the DND*i*[105] in 2006 and is constantly updated[24, 106]. It guides the efforts towards a curative drug acting by a trypanocidal mechanism[7c, 24, 106b]. The TPP dictates hit and lead criteria for *in vitro* and *in vivo* testing and adherence to the rules of Lipinski[107] and Veber[108], in order to increase the probability of good bioavailability when administered orally. Minimal side effects and low drug-drug interactions are required for better patient compliance. Despite these advances, minimal requirements for *in vitro* and *in vivo* screening strategies are poorly defined, which ultimately leads to poor chances to translate from model systems into clinical trials[7c]. This issue is further exacerbated by the fact that the experimentalists than run clinical trials utilizing diverse experimental models and definitions to rate success of curation. As a result clinical data suffer from poor comparability and require careful evaluation[7c]. Current research advances in drug discovery on CD are described in the next chapters and an overview of review articles is given in **Table 27** in the Appendix.

### 1.3.1 Clinical trials

Currently three new drug candidates are tested for chemotherapy of CD. Two of them are the repositioned anti-fungal azoles, posaconazole (**3**) (Noxafil®, Schering Plough)[109] and the water-soluble prodrug E1224 (**4**) (Eisai, Bristol-Myers Squibb)[110] (**Figure 3**). They are potent inhibitors of sterol 14α-demethylase (CYP51) and block downstream ergosterol biosynthesis, which is essential for the parasite[111]. Posaconazole showed promising results in a patient[111c, 112] but it exhibited lower efficacy in the phase II clinical trials CHAGASAZOL (NCT01162967)[113] and STOP-CHAGAS (NCT01377480)[114] when compared to BNZ controls[113-115]. Unfortunately, similar results were found in the phase II clinical trial of E1224 (NCT01489228)[99a, 116]. Therefore, both azoles are inadequate as monotherapies[116], however, combination therapies of posaconazole or E1224 with BNZ are currently tested[99a]. E1224-BNZ combination showed promising results in mice[117] and the phase II clinical trial BENDITA (NCT03378661)[118] started recently. Based on these preliminary results it seems likely that combination chemotherapy may play a role in future treatment regimens against CD[111a, 119]. The use of additive or synergistic activity of drug combinations may result in higher activity, reduced dosages as well as a decreased incidence of drug resistance[120]. The third candidate, fexinidazole (**5**), is a nitroimidazole with antiprotozoal effect, and currently tested in clinical trials (**Figure 3**). It was initially described five decades ago and the DND*i* successfully rediscovered the substance to treat African sleeping sickness as is supported by phase III studies[121]. Fexinidazole was already tested against *T. cruzi* in 1983[122] and

was shown to affect BNZ-resistant *T. cruzi* strains and to reduce the severity of myocarditis in 2012[123]. Recently, the DND*i* initiated two proof of concept studies to evaluate fexinidazole for the treatment of adult patients with CD (NCT02498782, NCT03587766)[124].

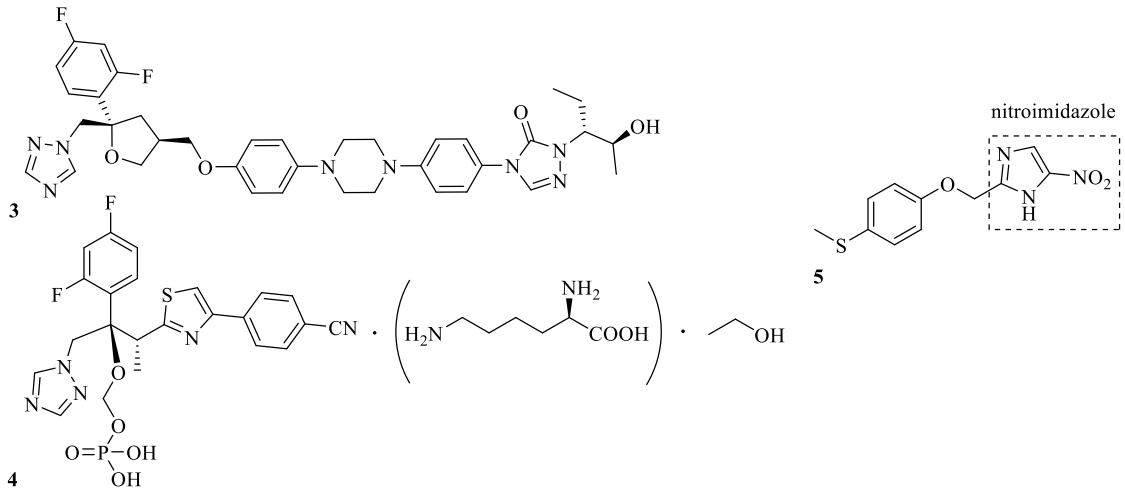



**Figure 3**: Chemical structures of compounds currently investigated in clinical trials. Posaconazole (**3**), prodrug E1224 (**4**) and fexinidazole (**5**) (key scaffolds are highlighted with a box).

Despite these novel approaches, most of the 58 clinical trials on CD investigate optimization of treatment regimens for BNZ and NFX or focus on the treatment of clinical symptoms of chronic CD[99a]. This includes paediatric formulations, new dosage schemes for chronic CD in adults[7c, 99a], diagnostic methodologies[125], and treatment options in Chagastic cardiomyopathy[73a]. The phase III clinical trials of BENEFIT (NCT00123916)[73a, 126] and TRAENA (NCT02386358)[127] showed that BNZ treatment is highly beneficial in chronic CD[77, 128]. The beta-blocker carvedilol (phase IV, NCT01557140)[129] and bisoprolol (phase III, CHARITY, NCT00323973)[130] were successfully tested for the treatment of chronic CD symptoms. Novartis announced to start a clinical trial in 2019 to assess the efficacy and safety of their cardiac drug Entresto® against Chagastic cardiomyopathy[131].

### 1.3.2 Phenotypic approach

The full *T. cruzi* genome was published in 2005[132], which enabled the generation of transgenic *T. cruzi* parasites that express well-established reporter proteins, such as β-galactosidase[133], tandem tomato fluorescence protein[134] or the firefly luciferase protein[29]. By extension the transgenic parasites enzyme activity, is detected by absorption measurements or by imaging after addition of colorimetric and luminescent substrates[75, 135]. Thus, reliable and robust phenotypic *in vitro* assays could be developed[15a, 68d] that are suitable for high-throughput screening

(HTS)[75], as well as high-content screening (HCS)[136]. Although these strains cannot cover the full extent of the *T. cruzi* genetic background[15a], they are highly valuable and behave biologically very similar to their wild-type counterparts[29, 133]. In contrast, the read out of tests that use different life stages of *T. cruzi* showed significant differences that have to be considered[20b]. Often intracellular amastigotes are targeted[29, 133, 137] as they are the replicative form in the mammalian host[1a]. Transgenic parasites also resulted in faster, more accurate, and more animal-friendly *in vivo* assays in mice, the predominant animal model for CD[138]. The severity of mice infected with transgenic parasites can quickly be monitored by detecting light through the skin after injection of luciferin[29, 135, 138b].

In consequence of to the aforementioned innovations and validation of HCS for *T. cruzi* in 2010[139], image-based HTS[137, 140] and HCS[15a, 141] identified a large number of clinically approved drugs that showed activity against *T. cruzi*[102, 140b]. In a subsequent process, which is called target deconvolution, molecular targets and mechanisms of actions were sought by applying target-based screening, genomics, proteomics, metabolomics studies of drug resistant strains and drug affinity responsive target stability (DARTS)[142]. In this context many hits were associated with sterol 14α-demethylase (CYP51) inhibition[67, 143].

Recent HCS campaigns led to a series of xanthines, such as GNF5689 (**6**)[141b] and 5-amino-1,2,3-triazole-4-carboxamide derivatives (**7**)[144] which employ an unknown mode of action (**Figure 4**). HCS, subsequent target identification and optimization revealed highly potent and selective kinetoplastid proteasome inhibitors with a triazolopyrimidine core, such as GNF3849 (**8**)[145], with an $EC_{50}$ of 16 nM (**Figure 4**). Thiazoles, such as compound **9**, have emerged from the scaffold of NFX, which have effects similar to BNZ and were non-mutagenic[146]. The benzothiazole **10** was discovered by drug repurposing, screening the Open Access Malaria Box[147], but was not further developed due to low plasma drug concentrations[99a, 101b]. More promising was compound **11**, a quinoline and derivative of lapatinib, a drug used in lung cancer treatment[99a, 148] (**Figure 4**). Also BNZ derivatives with retained aromatic nitro group, such as indazole **12**[149], 1,2,3-triazole **13**[150], and 1,2,4-tirazole **14**[151] (**Figure 4**) were developed. Silva *et al*.[151] showed that the absence of the nitro group strongly decreases biological activity (compound **15**, **Figure 4**). Ursolic acid (**16**) showed good *in vitro* and *in vivo* results[152] and recently a new formulation, applying nanoemulsion for oral intake, was developed[153] (**Figure 4**). Arylimidamides, such as DB766 (**17**), showed promising results against intracellular parasites and were also successfully tested against *T. cruzi* in 2018 (**Figure 4**)[154], however, some of them were toxic in mice[155].

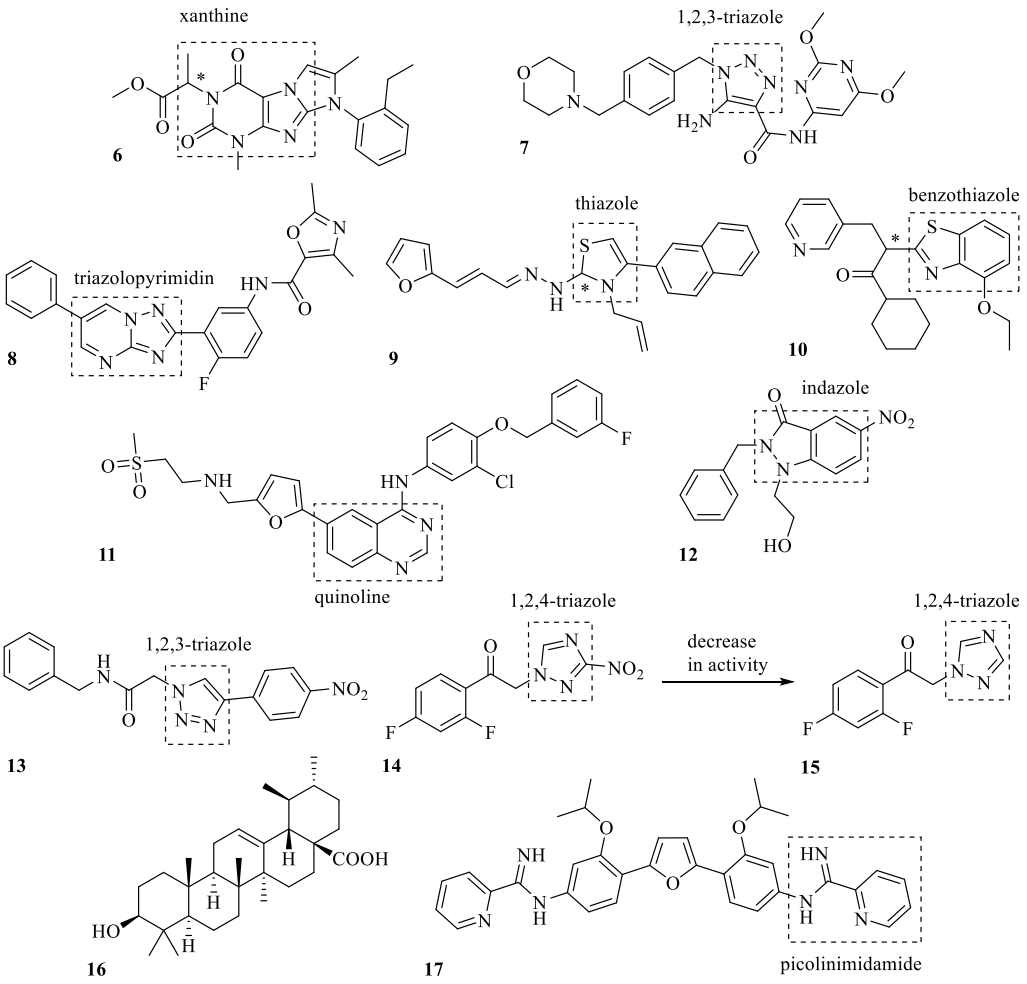



**Figure 4:** Chemical structures of the novel compounds active against *T. cruzi*. Key scaffolds are highlighted with a box. GNF5689 (**6**), 5-amino-1,2,3-triazole-4-carbocamide (**7**), GNF3849 (**8**), thiazole derivative **9**, benzothiazole derivative **10** and quinoline derivative **11**. BNZ derivatives: indazole **12**, 1,2,3-triazole **13**, 1,2,4-tirazole **14**, triazole without nitro group (**15**), ursolic acid (**16**) and arylimidamide DB766 (**17**).

### 1.3.3 Target approach – focus on isoprenoid and sterol biosynthesis

The elucidation of the *T. cruzi* genome sequence[132] enabled target-based drug discovery since it made all potential drug targets accessible for recombinant expression. Currently a large number of targets, for many of which a structure has been deposited in the PDB[100], and inhibitors of various chemotypes are studied for further development of new anti-Chagastic drugs[99a]. One of the pathways under investigation is ergosterol biosynthesis, which includes the mevalonate and isoprenoid pathway[111c, 156]. It is specific in kinetoplastids[156b] and according to genetic profiling[157], it is well understood in *T. cruzi*. Trypanosomes and humans have many isoprenoid and sterol precursors in common, but key steps differ: *T. cruzi* epimastigotes and amastigotes synthesise ergosterol and 24-alkylsterols, respectively, whereas humans produce cholesterol[157-158]. Epimastigotes and amastigotes cannot survive on assimilated cholesterol from their host[111d] and

blocking the pathway leads to depletion and lack of sterols resulting in changes of lipid bilayer integrity and hindrance of proliferation, therefore causing parasite death[156c, 159]. The *in vitro* and *in vivo* susceptibility to ergosterol biosynthesis inhibitors was demonstrated for several steps of the pathway, making these enzymes potential drug targets[13c, 156b] (**Table 1**, **Figure 5**).

**Table 1:** Proteins as potential drug targets in *T. cruzi*..

| Target enzyme | inhibitor | Citation |
|---|---|---|
| hydroxymethylglutaryl-CoA reductase (HMGCR) | statins | [160] |
| mevalonate kinase (MVK) | feedback inhibition by intermediates | [161] |
| farnesyl pyrophosphate synthase (FPPS) | nitrogen-containing bisphosphonates (N-BPs) | [162] |
| squalene synthase (SQS) | quinuclidines | [163] |
| squalene epoxidase (SQLE) | allylamines and hydrazones | [164] |
| lanosterol synthase or oxidosqualene cyclase (OSC) | aminopropylindenes | [165] |
| sterol 14α-demethylase (CYP51) | anti-fungal azoles | [159] |
| sterol 24-methyltransferase (S24MT) | azasterols | [166] |

*T. cruzi* FPPS (TcFPPS), the target enzyme of this work, represents a metabolic branching point and rate limiting step in isoprenoid biosynthesis[156b, 167]. It catalyses the formation of farnesyl pyrophosphate (FPP)[168], an essential building block in biosynthesis of isoprenoids such as sterols, ubiquinones, dolichols and heme A. With over 30,000 known isoprenoids, sterol biosynthesis is quite diverse and its products are ubiquitous and crucial for the survival of the organism[169]. Inhibition of FPPS abrogates all downstream processes of sterol synthesis and other processes relying on FPP due to a lack of starting materials[162, 167, 170].

One of the processes, dependent on FPP, is protein prenylation, a posttranslational modification important for the localization of the signalling proteins Ras, Rho and Rap to membranes and thus for intracellular signal transduction and cell cycle progression[167, 171]. Protein farnesyltransferase (PFT) transfers a farnesyl moiety from FPP to the thiol of a cysteine in a C-terminal CaaX motif (C: cysteine; a: amino acid with aliphatic side-chain; X: variable amino acid)[101d, 172]. Furthermore, FPP is needed for the formation of geranylgeranyl pyrophosphate (GGPP) used in geranylation of proteins catalysed by protein geranylgeranyl-transferase (PGGT)[173]. Besides indirect inhibition such as processes downstream of FPPS, *T. cruzi* PFT and PGGT can also be directly inhibited. Repositioned human PFT inhibitors, used in cancer therapy[101d], as well as monophosphates[172a, 174] and benzophenone derivatives[172a, 174] are active *in vivo* and *in vitro* against *T. cruzi* PFT. N-BPs not only inhibit TcFPPS but also *T. cruzi* PGGT[173]. An overview of the ergosterol pathway and processes depending on FPP are depicted in **Figure 5**.

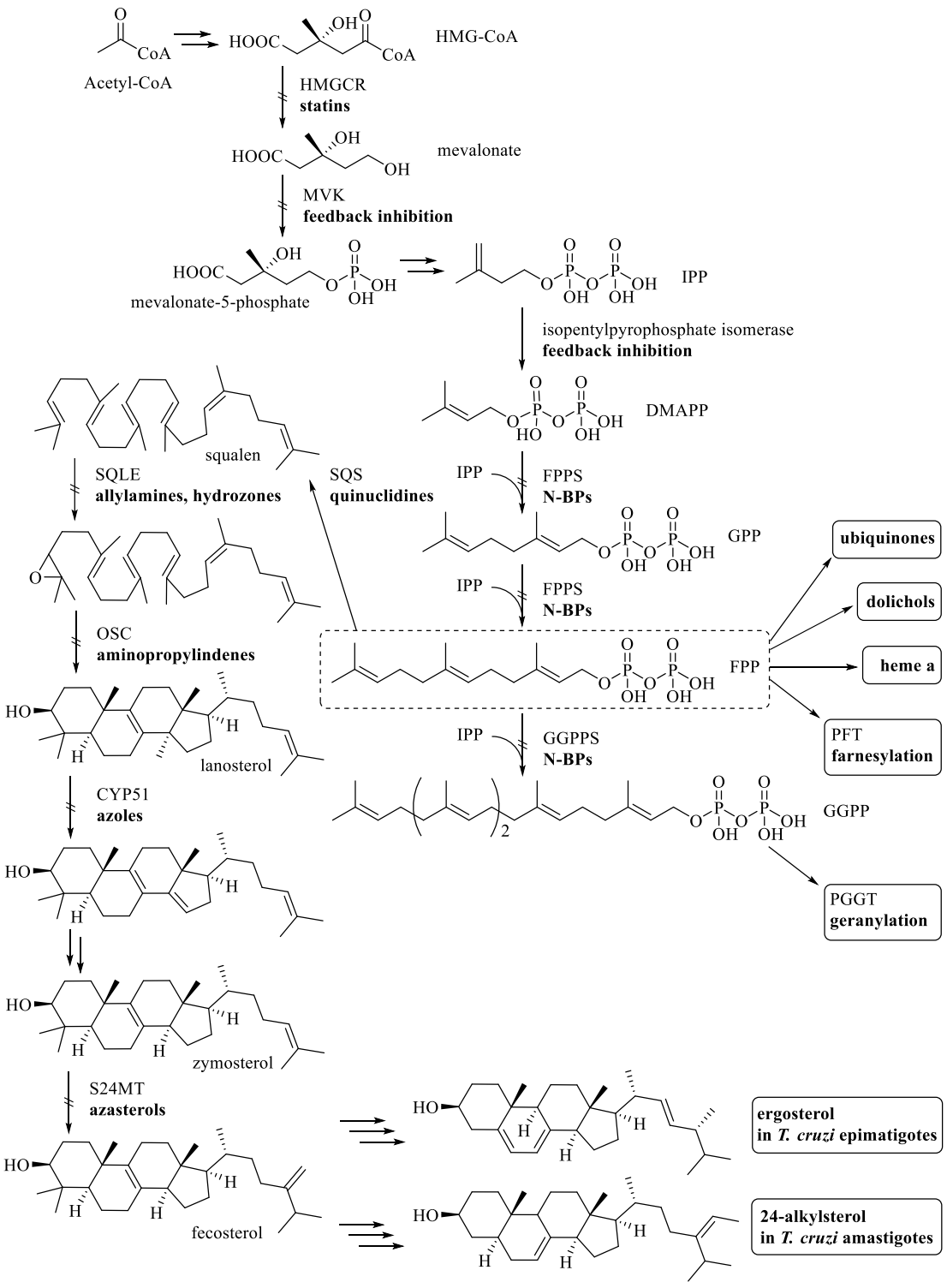



**Figure 5**: Sterol biosynthesis pathway in *T. cruzi*. Metabolic steps start from Acetyl-CoA, lead to mevalonate (mevalonate pathway) and further to FPP (isoprenoid pathway) and give the final products ergosterol and 24-alkylsterols (ergosterol pathway). Inhibitors are written in bold, downstream processes and final product are framed.

In many cases, inhibitors of *T. cruzi* enzymes involved in ergosterol biosynthesis were repositioned from their mammalian homologues. Statins, such as lovastatin (**18**), are well known inhibitors of human HMGCR used in the therapy[175] and were shown to inhibit *T. cruzi* HMGCR[160] (**Figure 6**). Likewise, quinuclidines active against mammalian SQS were found to inhibit *T. cruzi* SQS[111d]. Many compounds active against CYP51, the most studied target enzyme for CD chemotherapy[68d], were derived from phenotypic-based screening[67, 141c]. As mentioned earlier, the outcome of clinical trials with the anti-fungals posaconazole and ravuconazole failed to meet expectations. In 2019, coadministration of the CYP51 inhibitor VFV (**19**) with BNZ showed significantly better results in mice when compared to a monotherapy with BNZ[176] (**Figure 6**).

Another target for anti-Chagastic drug treatment is cruzain, the most abundant cysteine protease in *T. cruzi* essential for intracellular replication, adhesion to host cells and modulation of the host's immune response[25a, 177]. It was validated in mouse models and the vinyl sulfone derivative K777 (**20**) has proven to be a potent inhibitor[177b, 178] (**Figure 6**). Due to tolerability issues in primates, K777 did not proceed into clinical trials[179]. Currently, newly designed benzimidazoles are the most potent inhibitors of cruzain[98].

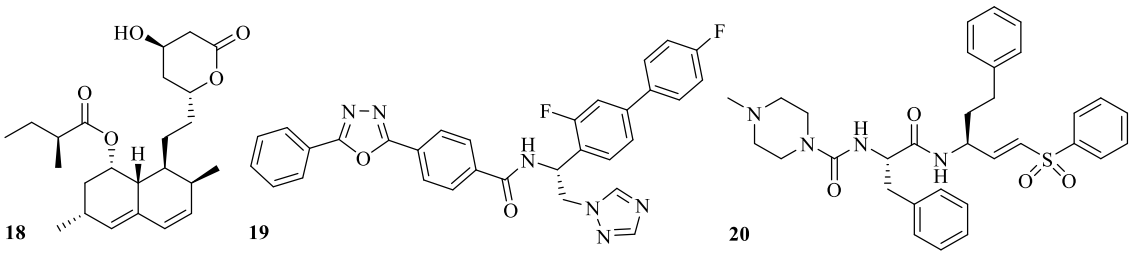



**Figure 6**: Chemical structures of lovastatin (**18**), VFV (**19**) and K777 (**20**).

Other target enzymes are hexokinase[180], triosephosphate isomerase (TIM)[181], and glyceraldehyde 3-phosphate dehydrogenase[182], all of which play important roles in glycolysis. Further targets include topoisomerase, which is involved in DNA supercoiling and entanglement[183], trypanothione reductase[184] and nitroreductase type I[134], which are responsible for cell detoxification, and trans-sialidase, which is important in host cell invasion and immune evasion[185]. Additional approaches to combat flagellate growth are altering tubulin assembly[186] and affecting intracellular calcium homeostasis[187]. The antiarrhythmic drugs amiodarone and dronedarone[187] as well as the antiparkinsonian drug bromocryptine[188] change the mitochondrial electrochemical potential and lead to alkalinisation of acidocalcisomes, vacuole-type storage organelles, rich in pyrophosphate (PP), phosphate, and calcium ions[189].

## 1.4 Farnesyl pyrophosphate synthase (FPPS)

Farnesyl pyrophosphate synthase (FPPS) (EC 2.5.1.10), also known as farnesyl diphosphate synthase (FDPS), was first described in 1959[190]. The *fpps* gene has been cloned to express and purify FPPS of fungi[191], yeast[192], avian[192], algae[193], humans[194], and also *T. cruzi*[195], the parasite of interest. In many organisms FPPS is a cytosolic enzyme, however, in some species FPPS is also localized in other cellular compartments[196]. So far all purified and characterized FPPSs are stable homodimeric enzymes of about 80 kDa size with a catalytic cleft in each monomer[167, 197]. In most reported FPPS crystal structures the two monomers are indistinguishable as they are related by crystallographic symmetry, such as in human FPPS (hFPPS)[168, 198] and avian FPPS[199], the very first solved FPPS crystal structure. In the *E. coli* FPPS crystal structure, the monomers are not related by symmetry, but show only minor differences[200].

FPPS plays an important role as key enzyme and rate limiting step in isoprenoid biosynthesis[167, 201] (chapter **1.3.3**, **Figure 5**) catalysing the formation of the $C_{15}$ building block farnesyl pyrophosphate (FPP) from $C_5$ precursors[168, 202]. FPPS condensates dimethylallyl diphosphate (DMAPP) with its isomer isopentyl diphosphate (IPP) to form intermediate geranyl pyrophosphate (GPP), and consecutively condensates GPP with a second IPP to form FPP[162b, 167, 203]. The reaction runs via a consecutive and stereoselective head-to-tail condensation yielding exclusively (*E*,*E*)-FPP[167] (**Figure 7 (A)**). Despite the availability of crystal structures, it is mechanistically unclear why homodimer formation is required for catalysis, however, it was suggested that the two subunits do not act independently[203].

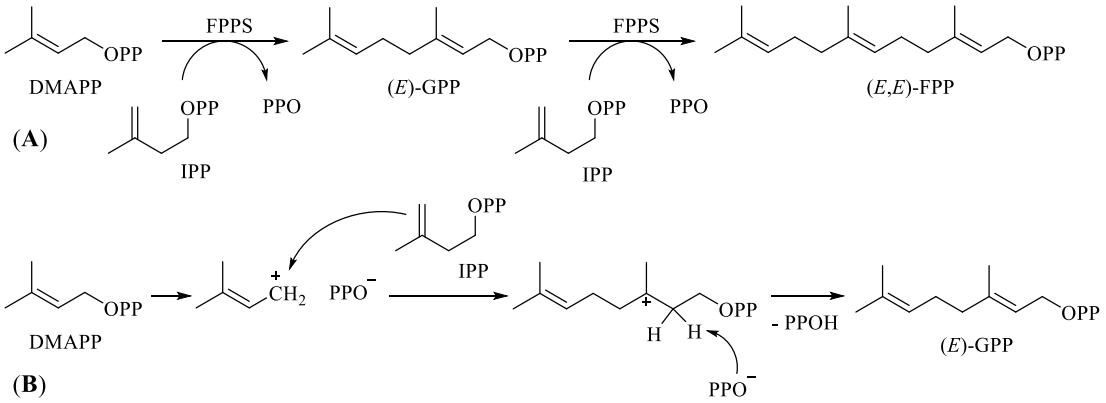



**Figure 7:** Scheme of the condensation reaction catalysed by FPPS. (**A**) Condensation reaction catalysed by FPPS. **(B)** Proposed reaction mechanism via carbocation intermediate.

The comparison of FPPS structures revealed seven conserved regions forming an active site cleft featuring prominent aspartate residues of two highly conserved aspartate-rich motifs

(DDXXD, D: aspartate; X: variable amino acid), which orient their side-chains into this cavity. The latter residues are important for catalytic activity[195, 200, 203-204]. They are called first and second aspartate-rich motif, abbreviated as FARM and SARM, and they are part of the region II and VI, respectively. These are α-helical regions forming the opposing sites of the major cleft, which is approx. 12 Å in diameter in the open-state and approx. 8 Å in the closed-state[167]. The monomer is composed of a two-helix N-terminal hairpin followed by an orthogonal central eight-helix bundle and a bundle of three short helices that protrudes perpendicular from the central bundle[162b, 197, 199-200] (**Figure 8**).

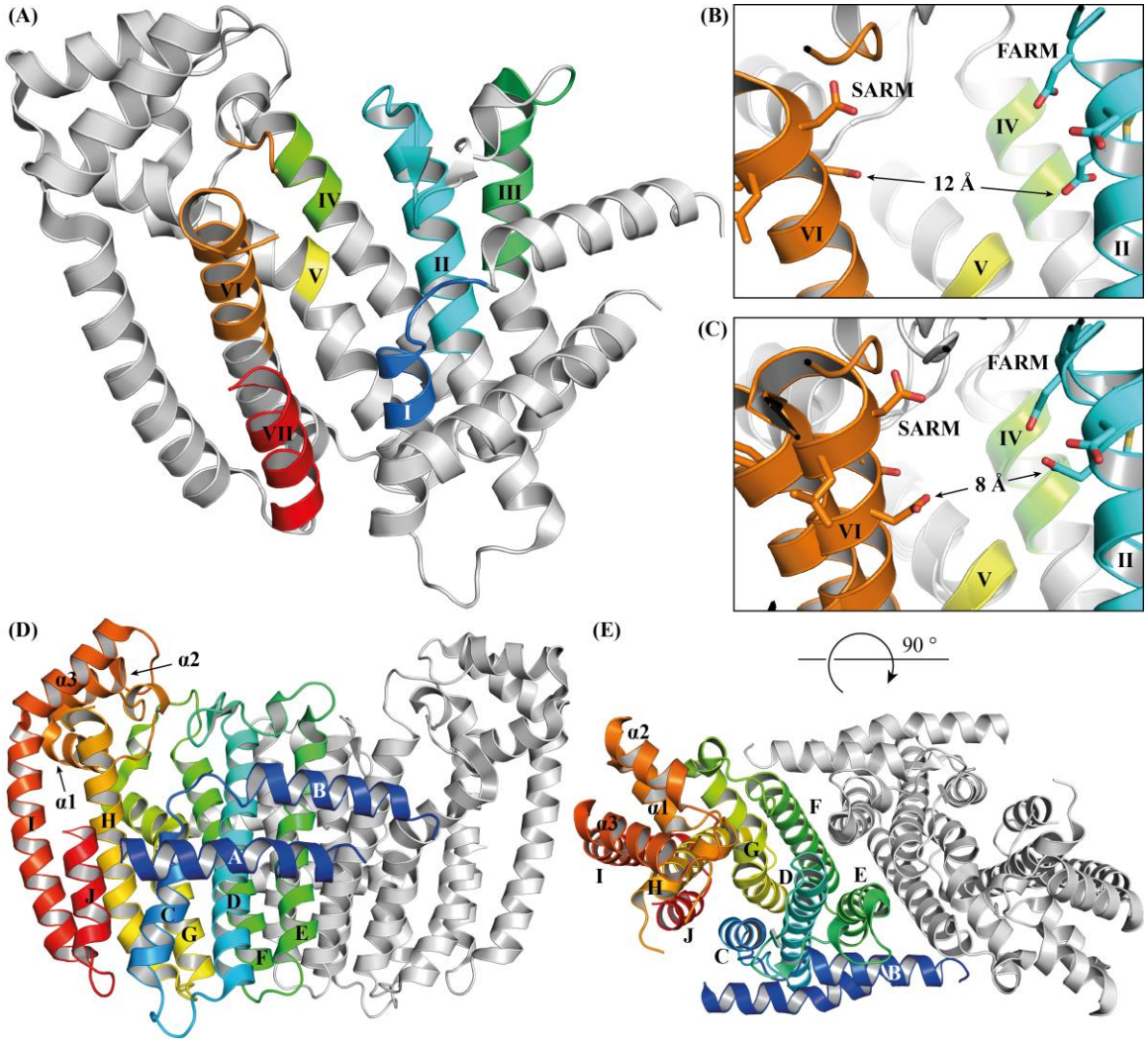



**Figure 8:** Overview of the structure of FPPS. **(A)** Monomeric hFPPS with conserved regions I – VII (PDB ID 5JA0[205]). **(B)** Active cleft in open-state hFPPS (grey cartoon, regions are coloured, residues of the FARM and SARM are shown as sticks, PDB ID 5JA0[205]). **(C)** Active cleft of hFPPS in open-state and closed-state are superimposed (PDB ID 5JA0[205] and 2F8Z[168]). **(D)** hFPPS homodimer with chain A coloured in blue to red gradually moving from the N-terminus to the C-terminus. Helices are labelled accordingly. Chain B is depicted in grey (PDB ID 5JA0[205]). **(E)** A 90° rotation about the horizontal axis of the structure depicted in **(D)** (PDB ID 5JA0[205]).

The canonical substrates DMAPP and GPP, bind via their pyrophosphate moiety to three $Mg^{2+}$ ions which in turn are coordinated by the carboxylate groups of the aspartates of the FARM and SARM. The binding site is therefore referred to as DMAPP or allylic binding site. IPP is binding in close proximity, to the so-called IPP or homoallylic binding site, which features conserved arginine and lysine residues[170a]. Whilst many crystal structures of FPPS in complex with IPP are available[168, 198], the only available structure of FPPS in complex with DMAPP is derived from *Galus galus* (PDB ID 1UBY[206]) (**Figure 9 (A)**).

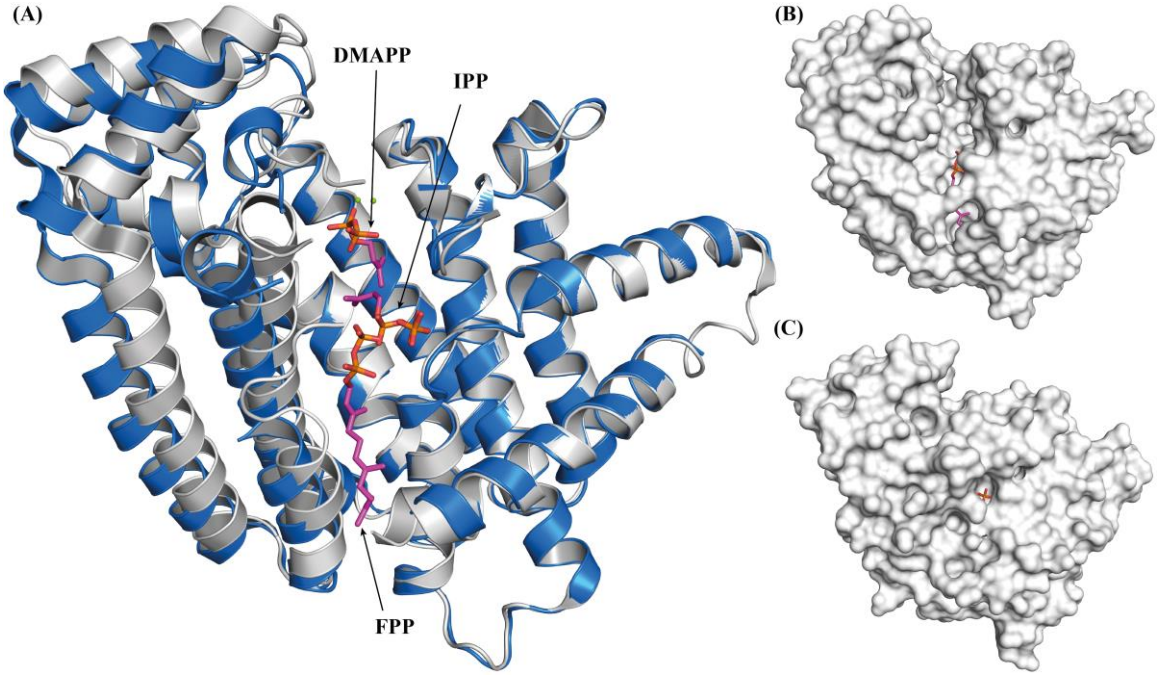



**Figure 9**: Pocket landscape of hFPPS. **(A)** Superimposition of closed-state FPPS (grey cartoon, IPP bound, PDB ID 2F8Z[168]) and open-state hFPPS (blue cartoon, FPP bound, PDB ID 5JA0[205]). Additionally, DMAPP and Mg2+ ions (green spheres) are superimposed (PDB ID 1UBY[206]), backbone not shown). **(B)** Surface representation of open-state FPPS with FPP bound (PDB ID 5JA0[205]). **(C)** Surface representation of closed-state FPPS (PDB ID 2F8Z[168]). In **(A)** – **(C)** all ligands are represented by sticks. Carbon, oxygen, and sulphate atoms are coloured in pink, red, and orange, respectively.

The way of substrate binding and conformational changes during the course of catalysis were elucidated by superimposition of crystal structures of unliganded FPPS and FPPS in complexes with its substrates and N-BPs[168, 198, 200]. The conformational changes from an open inactive hFPPS conformation into a close active conformation can be described as a two-step rigid body motion of the last 130 C-terminal residues[168]. Upon initial occupancy of the allylic site by DMAPP or GPP, the active site undergoes conformational rearrangement and the IPP binding site is fully formed by tightening of the FARM and SARM motifs. This intermediate state represents the partially-closed conformation. Subsequently, binding of IPP to the homoallylic site induces further rearrangement of the highly basic four-residue C-terminal tail, thereby closing the

homoallylic site and shielding the active site from the solvent exposure. This represents the fully-closed state[168, 205]. The mechanism of catalysis was proposed to involve dissociation of DMAPP into a carbocation that reacts with the double bond of IPP, and a pyrophosphate leaving group that subsequently subtracts a proton at the C-2 position of the former IPP moiety yielding GPP as first intermediate (**Figure 7 (B)**). The enzyme reverts back to the open-state, releases the pyrophosphate, translocates GPP, returns back to the fully-closed state and after a subsequent cycle of catalysis it releases FPP[162b, 205]. The dimer interface constitutes a rigid core and is not affected by the conformational switch[168]. For activity the bivalent metal ions $Mg^{2+}$ or $Mn^{2+}$ are required as they enable binding of DMAPP and GPP[207]. The 4$^{th}$ and 5$^{th}$ amino acids upstream of FARM were shown to play a role in product chain length determination as their aromatic side-chains form a hydrophobic floor of the pocket[208].

In addition to the aforementioned active site, an allosteric site adjacent to the IPP binding site and close to the C-terminal tail was described in some FPPS structures[209]. Recently, the binding of FPP to the allosteric pocket in hFPPS was reported, thus suggesting feedback inhibition of FPPS by its own product[205] (**Figure 9**).

### 1.4.1 *T. cruzi* FPPS

TcFPPS is a physiological homodimer[170a] in which each monomer has a length of 362 residues and a molecular weight of 41.2 kDa[202b]. TcFPPS is localized in the cytosol of the parasite[196]. The optimum catalytic activity of TcFPPS was observed at a $Mg^{2+}$ concentration of 1 mM to 5 mM and pH 8.5[195].

The proteins tertiary structure of a monomeric unit can be described as a two-helix N-terminal hairpin (helices A and B) followed by an orthogonal central eight-helix bundle (helices C to J) that is connected by loops with two exceptions[162b, 210]. Between helices F and G an 11-residue insertion loop is formed by the residues Lys179-Thr189 with a reverse turn at Pro182[162b, 210]. This insertion is unique to trypanosomal FPPS[162b, 210]. Between helices H and I are three short helices, named α1 to α3, which protrude perpendicular and orthogonal to the central eight-helix bundle. Helix α1 and helix α2 form an antiparallel hairpin and α3 is connecting back to the eight-helix bundle[162b] (**Figure 10**). The homodimer interface is composed of the N-terminal hairpin and helices E, F, G and D, which form together an interface of approx. 6028 Å[162b]. The insertion loops contribute to the dimer interface as the loop of monomer B is located above the hairpin of monomer A and vice versa[162b]. To date, their function is unknown[204]. Taken together, with the exception of the 11-residue insertion, the TcFPPS tertiary and quaternary structure correspond to those of other FPPSs (chapter **1.4**).

A BLAST search of the protein data base showed an identity of 35% to 39% and a similarity of 48% to 55% for the amino acid sequence of TcFPPS with other representative FPPSs

(mammalian, plant and yeast)[195]. Sequence alignment of TcFPPS and TbFPPS with hFPPS, avian FPPS and others showed that the residues involved in catalysis are conserved in the trypanosomal FPPS[162b, 195, 202b, 204]. In *T. cruzi* FPPS the FARM and SARM, which interact with the pyrophosphate moiety of DMAPP or GGP via bivalent ions, are residues Asp98-Asp99-Ile100-Met101-Asp102 in helix D and Asp250-Asp251-Val252-Met253-Asp254 in helix H, respectively[162b]. While FARM is perfectly conserved in TcFPPS, TbFPPS and hFPPS, the residues of the SARM in trypanosomal FPPS differ from the ones in human and avian FPPS (Asp-Asp-Tyr-Leu-Asp). Furthermore, in trypanosomal FPPS the 4$^{th}$ and 5$^{th}$ residue upstream of FARM are histidine (His93 in TcFPPS) and tyrosine (Tyr94 in TcFPPS), but in human and avian FPPS phenylalanine residues are at these positions[170a, 204]. IPP is bound to the enzyme by interacting directly with the arginine residues Arg51, Arg108 and Arg360 and to the lysine residues Lys48 and Lys362[162b, 170a]. Whilst the afore enumerated residues are conserved, the four-residue C-terminal tail slightly differs between TcFPPS (Lys359-Arg360-Lys361-Lys362) and hFPPS (Lys350-Arg351-Arg352-Lys353)[162b].

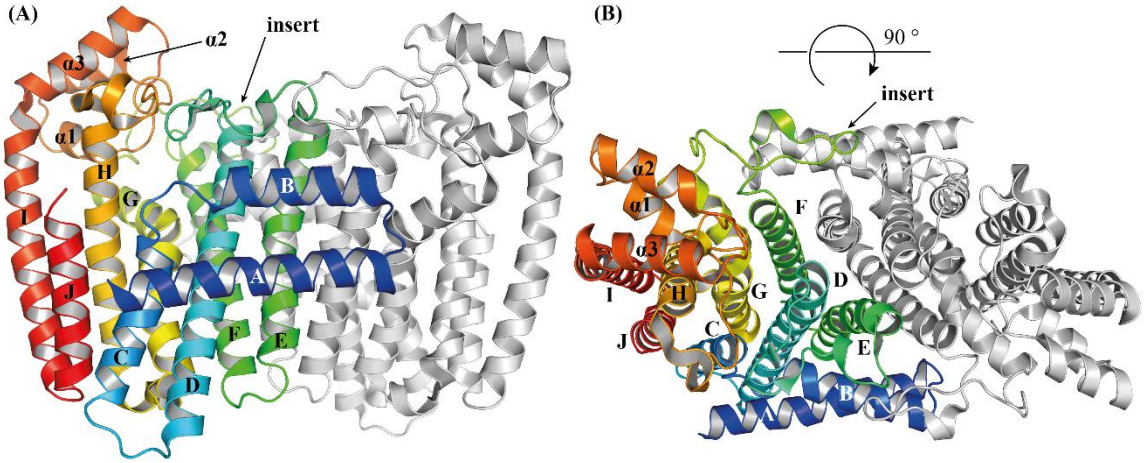



**Figure 10**: Crystal structure of TcFPPS. (**A**) Cartoon representation of TcFPPS homodimer (PDB ID 1YKL[162b], chain A coloured in blue to red gradually moving from the N-terminus to the C-terminus and labelled helices A – J, α1 – α3 and insertion loop (insert), chain B is depicted in grey). (**B**) A 90° rotation of the structure depicted in (**A**) about the horizontal axis.

To summarize, comparative studies of trypanosomal FPPS and human FPPS revealed that most of their active site residues are conserved and in consequence the enzyme-substrate and enzyme-N-BP interactions are very similar[162b, 195, 202b, 204, 211]. Huang *et al*.[211] stated that the high level of conservation leads to difficulties when designing parasite-specific drugs. However, the slight differences, such as between the 4$^{th}$ and 5$^{th}$ residue upstream of FARM offer some options for the development of TcFPPS-specific inhibitors[204] and inhibition assays performed on the same inhibitors using various FPPSs already showed different potencies[212].

### 1.4.2 Human FPPS identified as target enzyme of active site-directed N-BPs

BPs (**21**) had already been used in the clinic for decades to treat bone diseases which are related to osteoclast-mediated bone loss, such as osteoporosis, post-menopausal osteoporosis, osteitis, Paget's disease of bone, hypercalcemia and tumour metastases in bone mineral[213]. In 1966 and 1969 studies confirmed the positive effect of simple BPs, such as etidronate (**22**) and clodronate (**23**), on calcification of bone mineral[214] (**Figure 11**). Later they were found to function via formation of non-hydrolysable ATP analogues, which trigger osteoclast apoptosis[215]. Further development led to several FDA-approved nitrogen-containing bisphosphonates (N-BPs), which are orders of magnitude more potent than first generation BPs[216]. Their mode of action, inhibition of FPPS and blockade of carotenoid biosynthesis, was first described in a patent in 1998 for their use as bleaching herbicides[217].

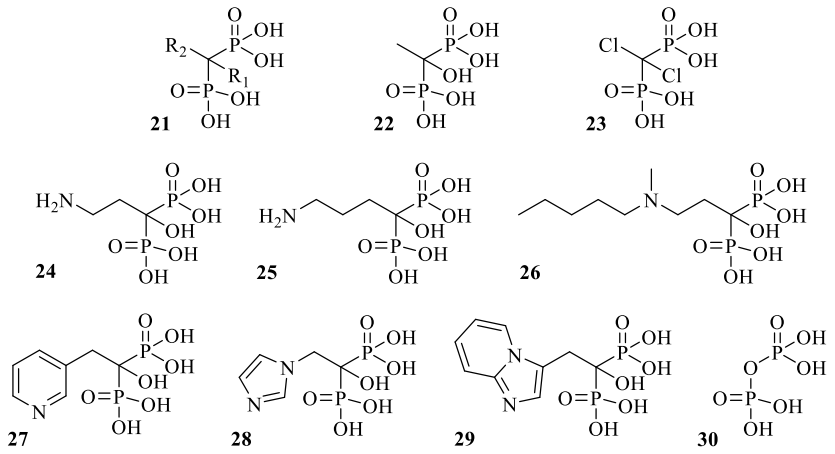



**Figure 11**: Chemical structure of bisphosphonates. Generic BP scaffold (**21**), etidronate (**22**), clodronate (**23**), pamidronate (PAM) (**24**), alendronate (ALE) (**25**), ibandronate (**26**), risedronate (RIS) (**27**), zoledronate (ZOL) (**28**), minodronate (**29**) and pyrophosphate (PP) (**30**).

One year later, tests identified that also recombinant hFPPS[218] is targeted by N-BPs[215a, 219]. Among these drugs are the primary amines pamidronate (PAM) (**24**) (Aredia®, Novartis)[220] and alendronate (ALE) (**25**) (Fosamax®, Merck)[171c, 221], the tertiary amine ibandronate (**26**) (Boniva®, Roche)[222] and nitrogen atoms localized within an aromatic ring, such as risedronate (RIS) (**27**) (Actonel®, Merck)[223], zoledronate (ZOL) (**28**) (Zometa®, Novartis)[220], and minodronate (**29**) (Onobis®, Ono pharmaceuticals and Astellas Pharma)[224] (**Figure 11**). Structure-activity relationships for inhibition of farnesyl diphosphate synthase *in vitro* match inhibition of bone resorption *in vivo* by N-BPs[168, 213c, 225]. N-BPs have a high affinity to the bone mineral hydroxyapatite and accumulate accordingly in bone mineral[226], where they are taken up by osteoclasts via fluid-phase endocytosis[227]. The inhibition of FPPS in osteoclasts results in a lack of FPP which blocks downstream processes, disrupts sterol biosynthesis and disables

prenylation of cell signalling proteins[216]. Finally, intracellular accumulation of IPP and subsequent formation of the cytotoxic ATP analogue ApppI, an ester formed of IPP and AMP, triggers osteoclast apoptosis thus preventing bone resorption[167, 228].

N-BPs are pyrophosphate (PP) analogues and therefore inhibit hFPPS by competing with its natural PP substrates[168, 198]. Replacement of the oxygen bridge in the PP backbone (P-O-P) (**30**) with a carbon (P-C-P) resulted in BPs (**21**) which are metabolically stable due to their non-hydrolysable backbone (**Figure 11**). The carbon backbone allowed various substituents, which are referred to as $R_1$ and $R_2$ from hereon[214c, 229]. N-BPs are characterized by a hydroxyl-group as substituent $R_1$, mimicking the p$K_a$ value of the pyrophosphoric acid and various $R_2$ side chains that contain nitrogen atom(s) (**24 – 30**) (**Figure 11**). Crystal structures revealed that N-BPs bind to the active site. They span the site usually occupied by DMAPP[168, 198], as their phosphate backbone mimics the major interactions of PPs. These are electrostatic interactions with three $Mg^{2+}$ ions, which are coordinated by the residues of the FARM and SARM, and interactions with the three basic side-chains Arg112, Lys200 and Lys257[168, 230] (**Figure 12**). The $R_2$ side-chain binds to the hydrophobic cleft that normally accommodates the growing isoprenoid chain. N-BPs bind to FPPS with protonated $R_2$ side-chain, i.e. as pyridinium, imidazolium, alkylammonium or amidinium-containing species[231], thus mimicking the carbocation transition state, specifically when the nitrogen atom is at C-4 position[232].

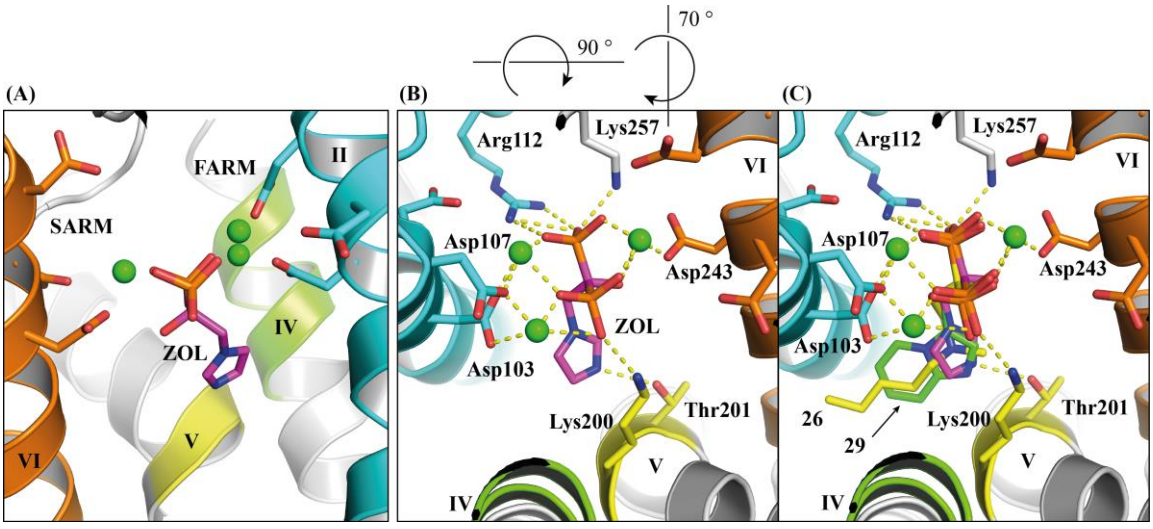



**Figure 12:** Active site of FPPS. **(A)** Binding of ZOL (**28**) in the active site (PDB ID 2F8Z[168], grey cartoon with coloured conserved regions, ZOL in stick representation, carbon, oxygen, nitrogen and phosphorous atoms coloured in pink, red, blue and orange, respectively, $Mg^{2+}$ green spheres, residues of the FARM and SARM are also shown in stick representation). **(B)** Representation as in (A) rotated by 180 ° and tilted by 70 ° (yellow dashes indicate interactions, residues forming main interactions in stick representation) **(C)** Representation of **(B)** superimposed with ibandronate (**26**) (PDB ID 2F94[168], stick representation, C in yellow,) and minodronate (**29**) (PDB ID 3B7L, stick representation, C in green).

N-BPs are classified as slow, tight-binding inhibitors[198, 233]. Their time-dependent inhibition is caused by the conformational change of the enzyme upon inhibitor binding[228b]. RIS

and ZOL showed initial $IC_{50}$ values of roughly 1.0 µM and 0.5 µM, respectively, however, after incubation they showed final $IC_{50}$ values of 57 nM and 41 nM, respectively[225b]. Rondeau *et al*.[168] described that IPP binds to the hFPPS-N-BP complex with a stabilizing effect due to full pocket closure. Binding of chemically stable N-BPs locks the enzyme in the closed-state, keeping FPPS from cycling through its three conformational-states needed throughout catalysis and therefore, N-BP binding is considered to be near irreversible[234]. Hence, inhibition by N-BPs is contrary to competitive inhibition where displacement occurs with increasing substrate concentrations[168].

### 1.4.3 *T. cruzi* FPPS inhibition by N-BPs

In 1999, Urbina *et al*.[189] published that *T. cruzi* contains large amounts of PP, of which 23% is stored in acidocalcisomes. In addition, they showed *in vitro* and *in vivo* inhibition of amastigote proliferation by the N-BPs PAM, ALE and RIS without toxicity to host cells[189]. With the uncovering of FPPS as target enzyme of N-BPs in the same year (chapter **1.4.2**), a phase of extensive testing of BPs as potential drug repositioning candidates on TcFPPS started in the hope that they could be used to treat CD[189, 195, 212, 235]. Montalvetti *et al*.[195] expressed recombinant TcFPPS and successfully demonstrated its inhibition by PAM, ALE and RIS, confirming that N-BPs target TcFPPS, too. The $IC_{50}$ values against *T. cruzi* amastigotes of PAM, ALE and RIS were determined to 60 µM, 147 µM and 123 µM, respectively[212]. *In vivo* testing of RIS showed 90% reduction of parasite loads in the blood of infected mice and significantly increased animal survival, suggesting trypanocidal activity of the compound[212, 236]. PAM showed inhibition of intracellular replication of amastigotes in *in vitro* assays and also reduced parasitaemia in mice[13c]. Hence, FPPS function is essential for *T. cruzi* viability in animal models of infection[170b, 212, 237]. Similar to the findings in human osteoclasts, the analysis of sterols in treated parasites showed that TcFPPS inhibition disrupts sterol biosynthesis and blocks downstream processes such as prenylation[195, 212, 238], which was shown to directly affect *T. cruzi* cell growth[174, 239]. Non-nitrogenous BPs, such as clodronate and etidronate, did not affect parasite proliferation[212].

Further insights were provided by crystal structures. The first predicted 3D model of TcFPPS was based on avian FPPS[195] and was followed by an X-ray structure of unliganded TcFPPS (PDB ID 1YHK) published by Gabelli *et al*.[162b] in 2006. They also crystallized protein-ligand complexes of TcFPPS with its natural substrates IPP and $Mg^{2+}$ or DMAPP and $Mg^{2+}$ alongside with ALE (PDB ID 1YHM) and RIS (PDB ID 1YHL), respectively[162b]. As shown for hFPPS, ALE was active site-directed in TcFPPS, mimicking the major interactions of the allylic substrate, interacting with $Mg^{2+}$ ions coordinated by the FARM and SARM (Asp98, Asp102 and Asp250)[162b]. In addition, the 3D structure revealed a conformational change in form of a hinge-like closure of the FPPS binding site when bound to these substrates[162b]. Again, these findings are very similar to the findings for hFPPS. Binding of N-BPs to TcFPPS with long

side-chains was shown to be enabled by the movement of residues Tyr94 and Gln167[162b, 170a, 202b, 211].

Huang *et al*.[211] did further co-crystallization experiments with ZOL (PDB ID 3IBA) and minodronate (PDB ID 3ICK) in the presence of IPP, confirming the binding mode of N-BPs[211]. However, the complexes showed partial asymmetry within the homodimer depending on the crystal structure. The authors related this finding to the binding event of the N-BPs[211]. Nevertheless, it is not present all crystal structures. As of 2019, 14 TcFPPS structures[162b, 170a, 211] have been deposited in the protein data bank[240] (Appendix, **Table 28**). They were all refined in the same hexagonal space group ($P6_122$) but correspond two different sizes of unit cells. In the first case, the cell dimensions a = b = 58 Å and c = 390 Å with one monomer in the asymmetric unit resulting in a homodimer generated by crystallographic symmetry. The second has the cell dimensions a = b = 103 Å and c = 390 Å and three monomers in the asymmetric unit. Crystallographic symmetry produces one symmetry constrained homodimer and two unconstrained homodimers with no imposed symmetry conditions allowing for slightly different subunits, as described by Huang *et al*.[211]

The activity of N-BPs against *T. cruzi* was repeatedly explained by accumulation in the parasites' acidocalcisomes which were reported to behave equivalent to human bone mineral, hence facilitating their antiparasitic action[163b, 200, 206, 229, 232, 243]. Acidocalcisomes have an average diameter of around 200 nm and their number and location per cell varies[241]. In *T. cruzi* epimastigotes more than 40 vacuoles of varying size were observed[241]. In amastigotes the vacuoles are arranged in rows near the cell periphery and in trypomastigotes they are located close to the flagellum[242].

### 1.4.4 BPs and treatment of non-bone diseases

BPs became the most transformative drugs of the last 25 years[243] due to their importance in treatment of bone diseases and rare occurrence of side effects according to their highly selective binding to bone mineral representing ideal pharmacokinetics (PKs) to treat this type of diseases[244]. Furthermore, synthesis of BPs is straightforward and cost-efficient[197, 245]. The high degree of evolutionary conservation of the active site of FPPS explains why N-BPs inhibit FPPS from various sources[198]. Acidocalcisomes, needed for selectivity of the treatment[170b, 189, 246], also exist in other Trypanosoma (*T. brucei*[247]), and also in Leishmania (*L. donovani*[248], *L. major*[249]), Toxoplasma (*T. gondii*[250]) and Plasmodium (*P. vivax*[251]). In consequence, N-BPs showed good inhibitory activity against many parasitic species *in vitro* and *in vivo*, e.g. RIS showed *in vivo* activity against *T. brucei* in mice[204], PAM, ALE and RIS were active in *in vivo* experiments with *L. donovani* and *T. gondii*[252] and various BPs showed growth inhibition of *P. vivax*[253]. Thus, BPs have been regarded as good drug candidates to treat tropical and neglected diseases[170b].

However, BPs are charged hydrophilic compounds that accumulate strong and rapid binding to bone mineral. They exhibit poor membrane permeability and bioavailability, undergo rapid renal clearance and therefore exhibit very low concentrations in serum and non-skeletal tissues[254]. In consequence, BPs are inappropriate to treat diseases which are not related to bone mineral[168, 230]. Thus, new and non-bisphosphonate FPPS inhibitors, which should be characterized by low affinity to bone mineral and less rapid removal from systematic circulation, are of interest as they may be better suited for non-skeletal related indications, such as CD[209a]. When used as antimicrobial, such FPPS inhibitors would also need to be selective as broad inhibition of hFPPS may result in toxicity[255]. Novel inhibitors would also be beneficial in tumour treatment, where inhibition of hFPPS was shown to trigger $\gamma\delta$ T cells activation, thus providing immuno-surveillance against tumours[256]. Another indication could be neurodegenerative diseases, such as Alzheimer's disease, which was previously linked to high levels of FPPS[257]. Even for the treatment of bone diseases such inhibitors could be beneficial, as they could distribute more evenly in bone mineral and minimise adverse effects associated with N-BP treatment[254d], such as osteonecrosis of the jaw and atypical femoral fractures[244, 258].

Approaches to optimize PK properties and to reduce bone affinity by changing the BP moiety or reducing polarity by introducing side-chains with increasing lipophilicity were made, but remained ultimately unsuccessful[168, 259]. Attempts to move away from the BP scaffold were unsuccessful because the resulting compounds did not mimic the natural substrate comparably well as BPs[259]. Recently reported non-BP inhibitors, binding to an allosteric site previously addressed in hFPPS, could lead to a breakthrough as they are assumed to have the potential to treat infectious diseases and soft-tissue cancer[209a, 209b, 230, 260].

### 1.4.5 Research on *T. cruzi* FPPS inhibitors

Rational modification of BPs to improve affinity to TcFPPS or to change their PK properties, while maintaining high ligand affinity started early on. Among them were BPs derived from fatty acids, such as alkyl-1,1-bisphosphonates **31** and **34**, 1-hydroxy-1,1-bisphosphonates **32** and **35** and 1-amino-1,1-bisphosphonates **33** and **36** to **40**[235, 238c, 261] (**Figure 13**). The latter showed growth inhibition of amastigotes but had no effect against epimastigotes[235, 238c, 261]. Compound **39** inhibited TcFPPS at nanomolar level and was thus more potent than previously tested 1-hydroxy-1,1-bisphosphonates[235, 238c, 261] (**Figure 13**). The structure activity relationship (SAR) of $R_1$ substituents attached at C-1 positions showed that, 1-hydroxy (**32**) and 1-amino (**33**) have higher binding affinities than 1-alkyl (**31**)[235, 238c, 261]. Linear α-fluoro-1,1-bisphosphonates **41** to **49** were found to neither inhibit TcFPPS nor being efficient *in vitro*[262]. These findings showed that the substituent at C-1 plays an important role, and although it is not actively involved in binding of

$Mg^{2+}$ ions it influences the p$K_a$ of the geminal phosphonate moieties and thus changes its capacity for coordination[197].

Linear 2-alkylaminoethyl-1,1-bisphosphonate derivatives (**50**-**54**) inhibit growth of amastigotes with $IC_{50}$ values between 38 nM and 1.84 µM[170a, 263] and were later described as potent SQS inhibitors, too[264] (**Figure 13**).

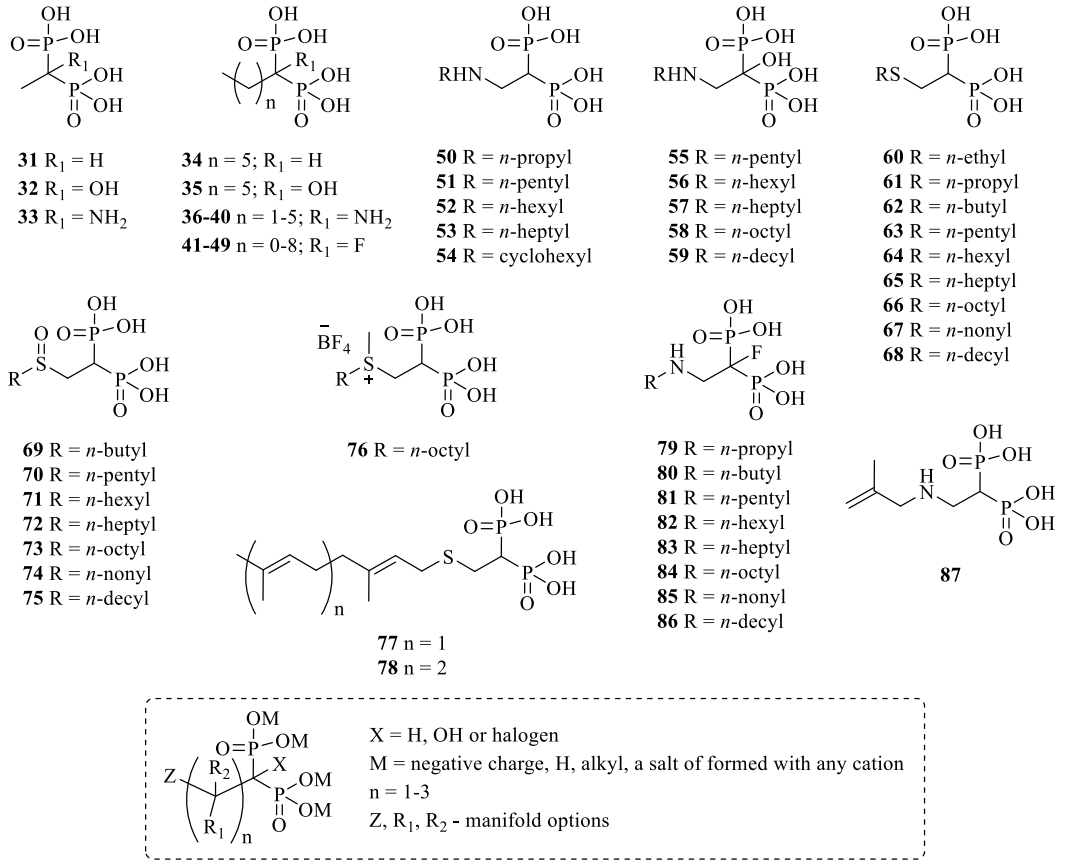



**Figure 13**: Chemical structures of BPs **31** to **87**. Compounds were tested on TcFPPS and/or *in vitro* against amastigotes and/or trypomastigotes and structure of generic BPs (box) patented as anti-parasitic agents.

Complexes of these compounds together with $Mg^{2+}$ and IPP were crystallized by Aripiralla *et al*.[170a] (PDB IDs 4DWB, 4DXY, 4DWG, 4EIE, 4DZW). The obtained 3D structures were used in molecular modelling but the resulting 2-alkylaminoehtyl-1-hydroxy-1,1-bisphosphonic acids **55** to **59**, representing 1-hydroxy analogues of the potent series **50** to **54**, were neither active against TcFPPS nor the parasite, however, several of them were active against *T. gondii* FPPS and *T. gondii* tachyzoites [162a] (**Figure 13**). Long chain length sulphur-containing BPs, such as thioethers **60** to **68**, sulfoxides **69** to **75** and a methyl-sulfonium derivative **76** were tested. Compounds **60** to **62**, **66** to **68** and **77** were active against TcFPPS but less potent than RIS. Compounds **61** to **68** were more potent against *T. gondii* FPPS. Compounds **72** to **74** and all sulfoxides **69** to **75** were only successfully tested against *T. gondii*[265] (**Figure 13**). Further

sulphur-containing BPs **77** and **78** and 2-alkylaminoethyl-1-fluoro-1,1-bisphosphonates **79** to **86**, tested by Galaka *et al*.[266], were inactive. In addition, they synthesised and tested compound **87**, which had been proposed by Aripiralla *et al*. [170a] but has proven to be inactive *in vitro* and had negligible affinity for the enzyme (**Figure 13**). Some of the aforementioned BPs inhibiting TcFPPS are protected by a patented generic BP that confirms BPs as anti-trypanosomal agents with known mode of action[267] (**Figure 13**).

Despite all previously described investigations and efforts of patent protection, the disadvantage of poor oral bioavailability of BPs[268] remained. To address this problem, BP metal complexes were formed and tested for TcFPPS inhibition in 2010[269]. The concept to use synergisms of metals and drugs was introduced in the development of anti-Chagastic chemotherapy in 1993 for the anti-fungal azol-derivative clotrimazol[270]. Indeed, the metal complexes of RIS, $[Ni^{II}(RIS)_2(H_2O)_2] \cdot H_2O$ and $[M^{II}(RIS)_2] \cdot 4\,H_2O$, with M = Cu, Co or Mn, showed enhanced anti-proliferative effects against amastigotes and dramatically improved $IC_{50}$ values of 2.7 nM and 2.9 nM, respectively[269a]. PAM and ALE complexes showed to be more potent, too[269a]. A promising finding was the interaction of the metal complexes with albumin in the blood, which functioned as transport vehicle to tissues[269a]. In addition, binding to plasma proteins was shown to prolong the plasma half-life which can be beneficial in drug therapy[271]. The latest results for ibandroante metal complexes were inconclusive, as the inhibitory effect on the protein under *in vitro* conditions was disproportionally small when compared to the effect in the cell under *in vivo* conditions. This observation indicates that additional targets might be affected once the compound is exposed to whole cells[272].

There are few efforts that go beyond BPs. Recent computational drug repositioning for TcFPPS showed that the anti-viral foscarnet (**88**) showed a good overlap with the binding position of ALE making it a top candidate for further investigation[100] (**Figure 14**).

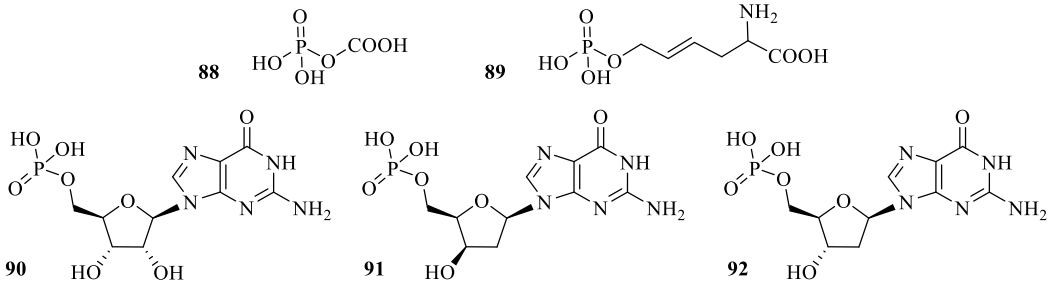



**Figure 14**: Chemical structure of foscarnet (**88**) and compounds from the ZINC database: monophosphate derivative, ZINC2139872 (**89**), guanosine monophosphate, ZINC1532555 (**90**), ZINC12296728 (**91**), ZINC01730395 (**92**).

The latest virtual screening campaign on TcFPPS used a pharmacophore model based on the best known TcFPPS inhibitors and identified four natural products in the ZINC database[273]:

The monophosphate derivative **89**, guanosine monophosphate (**90**) and two derivatives **91** and **92** thereof await in-depth validation[274] (**Figure 14**). In summary, most compounds evaluated for TcFPPS are BPs and recent approaches underscored the potential of monophosphate derivatives for clinical applications. To date, N-BPs are the substance class with the strongest reduction of parasitaemia[238b]. Structural data of TcFPPS and TcFPPS in complexes with N-BP inhibitors are available and pave the way for further development[104].

### 1.4.6 Allosteric site binding of novel scaffold inhibitors

In 2010, Jahnke *et al*.[209a] addressed the issue of BPs being too specific for bone minerals, in order to use them effectively in cancer treatment, however, they found inhibitors of new compound classes inhibiting hFPPS activity. This inhibitors are the benzothiophenes **93** to **95** and the indoline **96** which were identified in a fragment-based lead discovery approach and showed only weak interactions with FPPS characterized by $IC_{50}$ values >500 µM (**Figure 15**). Studying them by a series of NMR experiments uncovered their binding to an additional pocket of the protein. In contrast to the allylic binding site, binding to the allosteric site in hFPPS is independent of $Mg^{2+}$ ions[209a]. 3D structures of the protein-fragment complexes obtained by crystallography (PDB IDs 3N1V, 3N1W, 3N3L, 3N45) showed that this pocket is adjacent to the IPP binding site and close to the C-terminus[209a]. It is formed by helices C, G, H and J as well as the B-C and H-I loop. The hydrophobic floor and back faces of the pocket are shaped by residues Tyr10, Phe206, Phe239, Leu344, and Ile348. The front site is defined by the positively charged residues Lys57, Arg60 and Lys347 and the polar Asn59 and Thr63[209a, 275]. The described distribution of amino acids leads to an amphipathic pocket (**Figure 16**).

By merging useful features of the SAR studies of their fragment series, Jahnke *et al*.[209a] designed and synthesised benzindole derivatives **97** (PDB ID 3N6K) and **98** that lack affinity to bone mineral and exhibit $IC_{50}$ values of 200 nM and 80 nM, respectively (**Figure 15**)[209a]. The mechanism of action for hFPPS inhibition by these allosteric binders was hypothesized to be as follows: The negatively charged benzindole inhibitors prevent IPP from binding via repulsive electrostatic interactions as the binding sites are in close spatial proximity. When a benzindole-type inhibitor is bound, the C-terminal tail (Lys350-Arg351-Arg352-Lys353) remains disordered and hence cannot reinforce full pocket closure which is required for catalysis. Furthermore, the mechanism that drives conformational rearrangement from an open to a closed-state and vice versa is likely to be disrupted upon binding of such an inhibitor[209a]. Taken together, by exploiting a previously uncharacterised binding pocket of hFPPS, compounds with a novel scaffold were discovered that entail new and desired properties, thereby overcoming previous limitations of BPs[275-276].

Jahnke *et al.*[209a] suggested the term allosteric binding pocket for the site adjacent to the active site, because they hypothesized that it might play a regulatory role in a kind of feedback inhibition, a process that was first described by Changeux in 1961[277]. Recently Park *et al.*[205] showed that hFPPS is indeed product inhibited by binding of its own condensation product FPP to this pocket and locking of the enzyme in the open, inactive conformation. The crystal structure of hFPPS in complex with FPP (PDB ID 5JA0) revealed an induced-fit conformational change accompanied by FPP binding, allowing accommodation of its tail[205]. Tyr10 swings out from helix A, thereby generating space for the FPP tail and induces a tilt in helix A that lead to some more side-chain rearrangements. Thus, this binding site will be referred to as allosteric site of FPPS. All other sites described in this work will be termend additional binding sites although they might also function like an allosteric pocket.

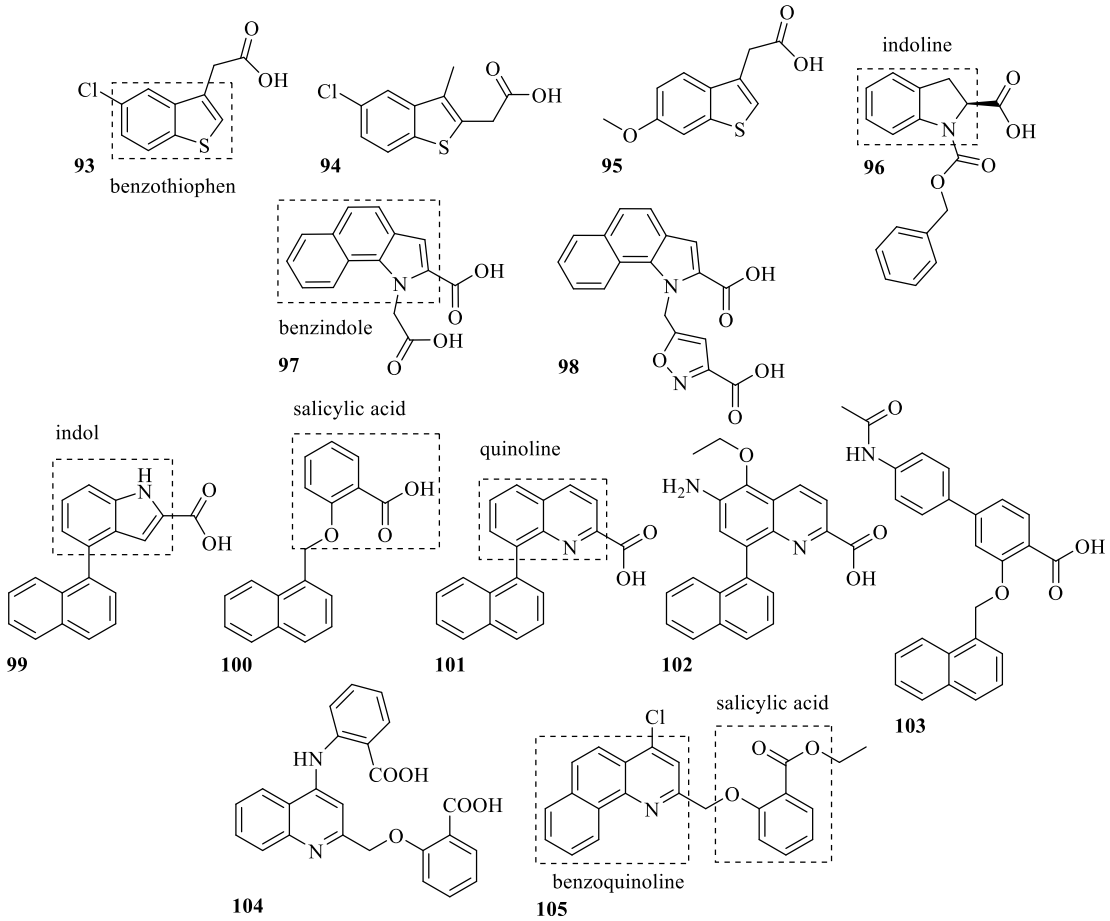



**Figure 15**: Chemical structure of hFPPS inhibitors with a new scaffold. Fragments found by FBS and optimization: benzothiophenes **93** to **95**, indoline **96**; benzindoles **97** and **98**. Integrated lead finding lead to indole **99**, salicylic acid derivative **100**, quinolines **101** and **102**, lead compound **103** and quinolines **104** and **105** (key scaffolds are highlighted with a box).

In 2015, Marzinzik *et al.*[230] published two more chemical lead series binding to the allosteric binding pocket of hFPPS: quinolines and salicylic acids which were found by integrated lead discovery. The most potent inhibitors they identified were an indole derivative **99** (PDB ID 5DJP) and a salicylic acid derivative **100** (PDB ID 5DIQ), exhibiting $IC_{50}$ values of 7.1 µM and 6.8 µM, respectively (**Figure 15**, **Figure 16**). The indole **99** was evolved by iterative medicinal chemistry efforts into a quinoline **101** (PDB ID 5DGN) with an improved $IC_{50}$ of 1.2 µM and was further developed into **102** with an $IC_{50}$ of 24 nM, ultimately being 300-fold more potent than the starting compound **37** (**Figure 15**, **Figure 16**). A library synthesis of salicylic acid derivatives resulted in **103** with an $IC_{50}$ value of 17 nM thus being 400-fold more potent than the starting compound **100** (**Figure 15**). Also Liu *et al.*[260] synthesised a series of quinolines. The most potent substances among them were **104** and **105**, both with an $IC_{50}$ of 3.5 µM (**Figure 15**). Neither the quinoline nor the salicylic acid series identified for hFPPS induced such significant conformational change[209a, 230] than observed for FPP[205].

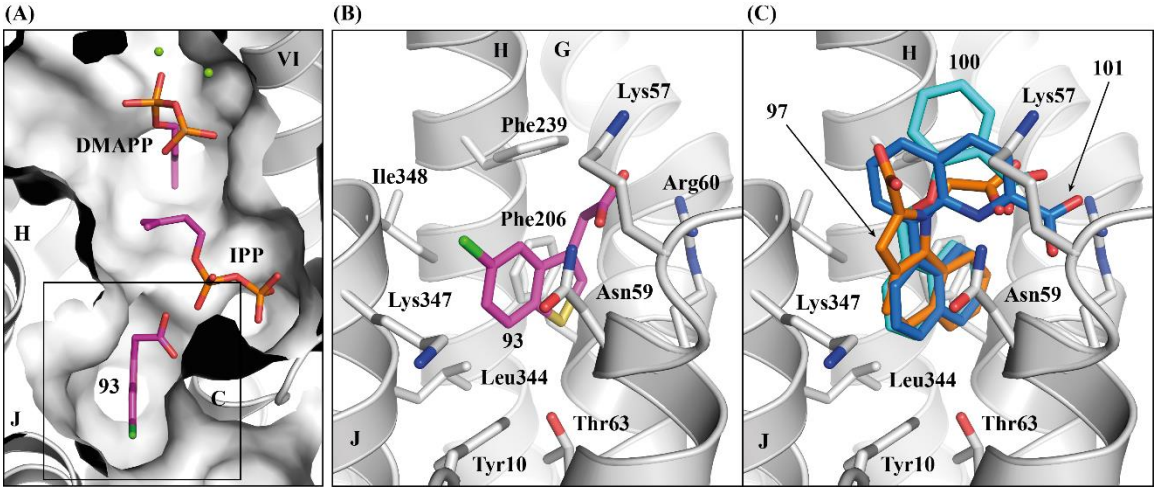



**Figure 16:** FPPS in complex with allosteric inhibitors. **(A)** Open-state FPPS with allosteric inhibitor **93** (grey cartoon and surface representation, PDB ID 3N1W[209a]). Superimposition with DMAPP and IPP (backbones not shown, $Mg^{2+}$ green spheres, PDB IDs: 1UBY[206], 2F8Z[206]). **(B)** Allosteric pocket with inhibitor **93** (grey cartoon, residues forming the pocket in stick representation, PDB ID 3N1W[209a]). In **(A)** and **(B)** C, O, S and Cl atoms are coloured in pink, red, orange and green, respectively.**(C)** Allosteric pocket with inhibitors **97**, **100** and **101** (ligands in stick representation; carbon atoms in orange, blue and cyan, respectively, backbones not shown, PDB IDs 3N6K[209a], 5DIQ[230] and 5DGN[230]).

As crystal structure analysis showed, the described scaffolds are bound to the previously novel allosteric pocket in the open state of the enzyme. The compounds have no affinity for bone mineral and were claimed to represent a potential lead series for the treatment of non-bone diseases[230]. However, both compound series showed low cellular permeability, resulting from the carboxylic acid functionality and replacement by a tetrazole had limited success[230]. Replacement by a phosphate moiety increased the bone affinity and reduced ligand affinity, but the compounds maintained binding to the allosteric site[254d]. Patents for a whole range of quinoline derivatives[278]

and salicylic acid derivatives[279] were filed in 2009 and 2010, respectively. All other patented FPPS inhibitors in the period from 2006 to 2010 were BPs[280].

Efforts to exploit the allosteric pocket of FPPS for therapeutic purposes are ongoing[205]. The pocket is dubbed the Achilles' heel of FPPS[230] and binders are expected to have a wide range of applications, e.g. in cancer treatment[230], as cholesterol-lowering agents, and whenever excessive lipid production causes disease [205, 230], as well as in neurodegenerative diseases[276], and as anti-parasitic agents, notably for CD, leishmaniasis and malaria[230]. In fact, ligands of this pocket in FPPS of the gram-negative bacterium *Pseudomonas aeruginosa* were described in 2015 by Schmidberger *et al.*[209b]. They claimed that the allosteric site is significantly less conserved than the active site between human and bacterial FPPSs, allowing the development of a selective inhibitor for each enzyme[209b]. In 2017 the allosteric pocket and potential inhibitors were described for FPPS of *Plasmodium falciparum* the causative agent of Malaria[209c]. To date, an allosteric pocket or inhibitors binding to sites other than the active site have not been described for trypanosomal FPPS.

### 1.4.7 Further approaches in FPPS inhibition

Gao *et al.*[281] pioneered a new class of BP-based derivatives replacing a hydroxyl group of a phosphorous moiety by a geranyl moiety. Compounds **106** and **107** showed $IC_{50}$ values of 7.8 nM and 13 nM for rat FPPS, respectively (**Figure 17**). The authors surmise that these types of compounds might interact with the allylic and homoallylic binding site. The same group envisioned fluorescent BP derivatives (**108**, **109**) as active site inhibitors of multiple enzymes involved in early steps of isoprenoid biosynthesis. Both were successfully tested on the rat enzymes mevalonate kinase (MVK), phosphomevalonate kinase (PMK), mevalonate 5-disphosphate decarboxylase (MDD) and FPPS with increasing potency starting with $IC_{50}$ values in the single digit µM range for MVK to $IC_{50}$ values in the double digit nM range for FPPS[282] (**Figure 17**).

Docking studies based on allosteric inhibitors of hFPPS[209a] resulted in the discovery of bisamidines[275]. The most potent was bisamidine **110** with an $IC_{50}$ of 1.8 µM[275]. It was co-crystallized later by Liu *et al.*[283] and found to bind to at the protein surface (PDB ID 4RXA)[197, 283]. Thienopyrimidine bisphosphonates, such as compound **111**, were described as inhibitors with several binding positions either binding to the DMAPP binding site in the presence of $Mg^{2+}$ ions (PDB ID 4JVJ) or the IPP binding site in the absence of $Mg^{2+}$ ions (PDB ID 4LPG)[284]. Although not of biological relevance this showed that BPs with hydropathic cores can be directed to the allosteric site[276]. Thienopyrimidine monophosphates, such as compound (**112**) exclusively bound to the allosteric pocket (PDB ID 4LPH) and showed an *in vitro* potency similar to the benzindole derivative **97**[276] (**Figure 17**). Previous attempts to remove one phosphate moiety, turning a BP into a mono-phosphate, led to complete loss of potency[225b].

Substituted indoles and azabenzimidazoles, such as compound **113** and **114**, respectively, were shown to bind to all three binding sites in hFPPS and were thus named chameleon inhibitors[285] (**Figure 17**). Multiple binding modes and events were previously reported; IPP binding with a 2:1 stoichiometry to hFPPS[198], binding of two homo-risedronate molecules to *T. brucei* FPPS (PDB ID 4RXC)[283] and DMAPP binding to the IPP binding site in TcFPPS (PDB ID 1YHL)[162b]. Binding to the IPP site in hFPPS was demonstrated for the non-BPs taxodione (**115**) (PDB ID 4P0V) and arenarone (**116**) (PDB ID 4P0W), compounds with known anti-bacterial and anti-tumour activity[286] (**Figure 17**). Furthermore, hFPPS was identified as a target of *N*6-iso-pentenyladenosine (**117**) and NMR and docking studies suggested active site-directed binding for this non-BP scaffold[287] (**Figure 17**). In a recent virtual screening, guanosine monophosphate and derivatives thereof were identified as potential binders of TcFPPS[274].

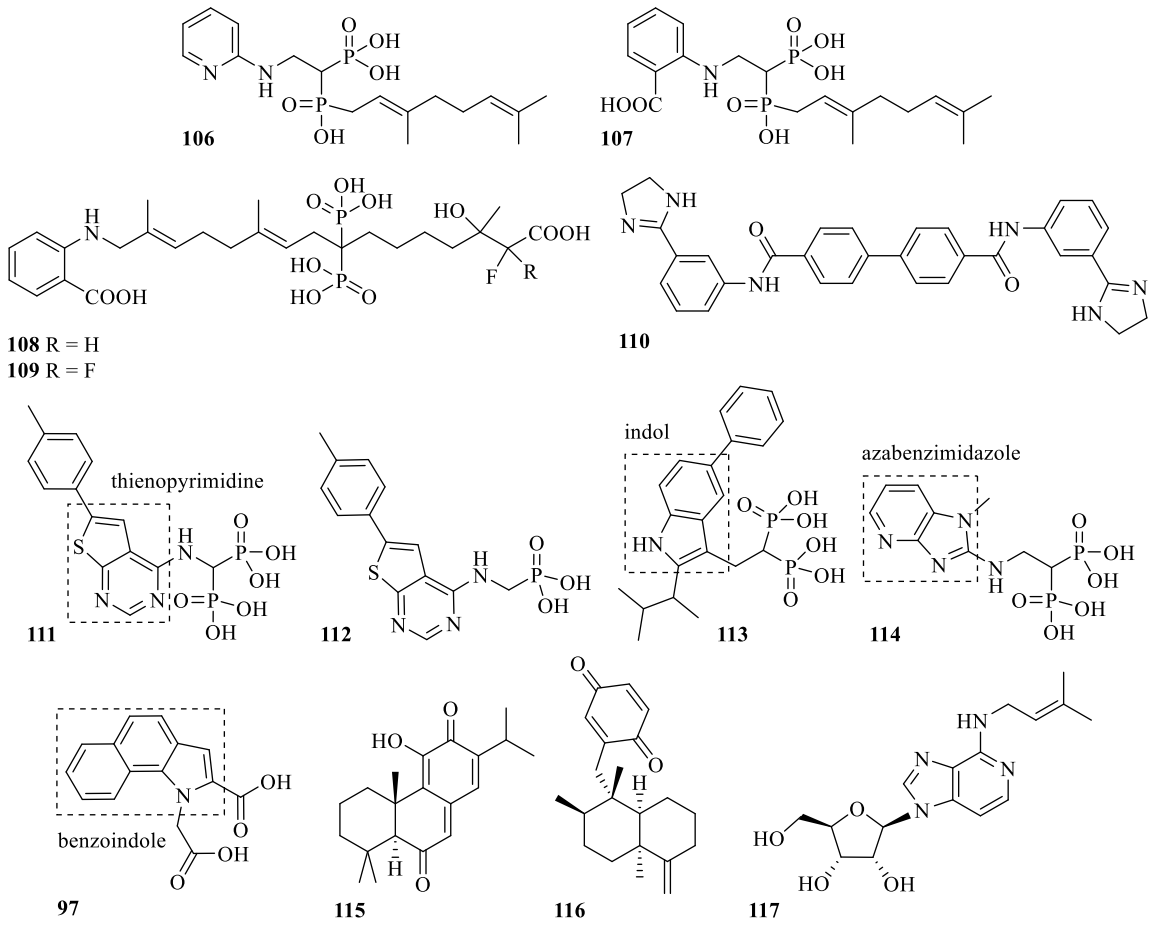



**Figure 17**: Chemical structure of BP-based and other inhibitors. BP-based inhibitors **106** and **107**, fluorescent BP derivatives **108** and **109**, bisamidine **110**, thienopyrimidine bisphosphonate **111**, thienopyrimidine monophosphate **112**, benzindole **97** (showed for comparison), substituted indole **113** and azabenzimidiazole **114**, taxodione (**115**), arenarone (**116**) and *N*6-isopentenyladenosine (**117**).

## 1.5 Fragment-based lead discovery (FBLD)

For several years, biochemical and cell-based HTS was the dominant approach in lead discovery, which led to many initial starting points[288]. However, issues with false positives constantly occurred[289] and in some cases only a few hits were observed[290], although libraries for HTS contain up to $10^6$ lead- or drug-like compounds[291]. Instead of further increasing the number of tested compounds, fragment-based screening (FBS) used libraries that contained around $10^2$ to $10^4$ smaller compounds, so-called fragments that are screened against a target protein[292]. Fragments are characterized by a MW ranging from 110 to 300 Da with less than 20 heavy (non-hydrogen) atoms[293]. A retrospective analysis of 145 fragment-to-lead campaigns showed that the starting fragments and the final lead compounds had an average number of 15 and 28 heavy atoms, respectively[294]. As chemical space increases by approx. 8-fold with each added heavy atom[295], 1000 fragments with a MW of 190 Da cover chemical space as effectively as $10^8$ compounds of 280 Da or $10^{18}$ compounds of 440 Da[292]. When fragments are screened instead of larger compounds, it is more likely to identify motifs that match the protein's requirements[293a], because fragments can only form few interactions due to their limitation in size, while larger compounds can form many more interactions, and for some of which a higher chance is given to be counterproductive[293a].

Usually, fragments bind to hot spots of binding in the protein[296], and in most cases all binding sites of a given protein can be mapped with fragments[297]. Furthermore, fragment screens can also reveal unknown and secondary binding sites of proteins[298], which was lately shown for FPPS[209a, 209b] and K-Ras[299]. Hence, FBS can stimulate new interest in known targets[292]. However, fragment screening hits are characterized by low affinity with an equilibrium dissociation constant ($K_d$), ranging from 0.1 mM to 10 mM and thus, sensitive biophysical methods are needed for their detection[292, 300]. The properties of the fragment hits are decisive for the success of the subsequent optimization process into drug-like molecules[291]. Lead compounds, which were derived by optimization of fragments, were shown to have more-drug-like properties, e.g. lower MW and/or lower lipophilicity, when compared to starting compounds of higher complexity derived from HTS campaigns. The latter can only be adjusted slightly because their size leaves little room for improvement[291]. In short, the advantages of FBS compared to HTS are a smaller screening library, higher hit rates and the smaller and less lipophilic fragments hits that have proven to be good starting points for chemical optimization[293b, 301]. Nevertheless, a potent lead can fail long after the incipient screening phase. Potential roadblocks are poor ADME (absorption, distribution, metabolism, excretion) properties, poor cell permeability, unexpected toxicity and/or the lack of the desired biological response[102].

One of the first fragment-based lead discovery (FBLD) campaigns was conducted by Shuker *et al.*[302] in 1996. In a so called structure-activity relationship (SAR) study by NMR, they

identified ligands that were bound to different sites and their chemical linkage resulted in potent inhibitors[302-303]. Since then a whole range of tools and procedures for FBLD were developed[293b], some of them by companies, such as Abbott[304], Astex[305], Sunesis[306] and Vernalis[293c]. To date, after 23 years of investment, FBLD is frequently used in academia and pharmaceutical companies and has resulted in many fragment-to-lead success stories[292, 293b, 307]. Currently, there are approx. 30 compounds in clinical trials, and two of them were approved as drugs[293b, 308].

Basically, all FBLD campaigns include four steps: assembly of a suitable fragment library, identification of fragment hits by biophysical screening, biophysical characterization and structure determination and finally fragment-to-lead optimization by chemical optimization[292] (**Figure 18**). Biophysical methods to screen the fragment library against the target protein need to be robust and sensitive to identify the low affinity binders. Methods matching these criteria are nuclear magnetic resonance (NMR) spectroscopy[309], surface plasmon resonance (SPR)[310], differential scanning fluorimetry (DSF)[311], X-ray crystallography[312] and mass spectrometry (MS)[313]. The methods vary in their affinity detection range and thus cover the range of binding affinities within a fragment-to-lead optimization process[293c, 314]. NMR and X-ray crystallography require larger amounts of protein (10s of mg) than all other methods (<1 mg)[292] (**Figure 18**).

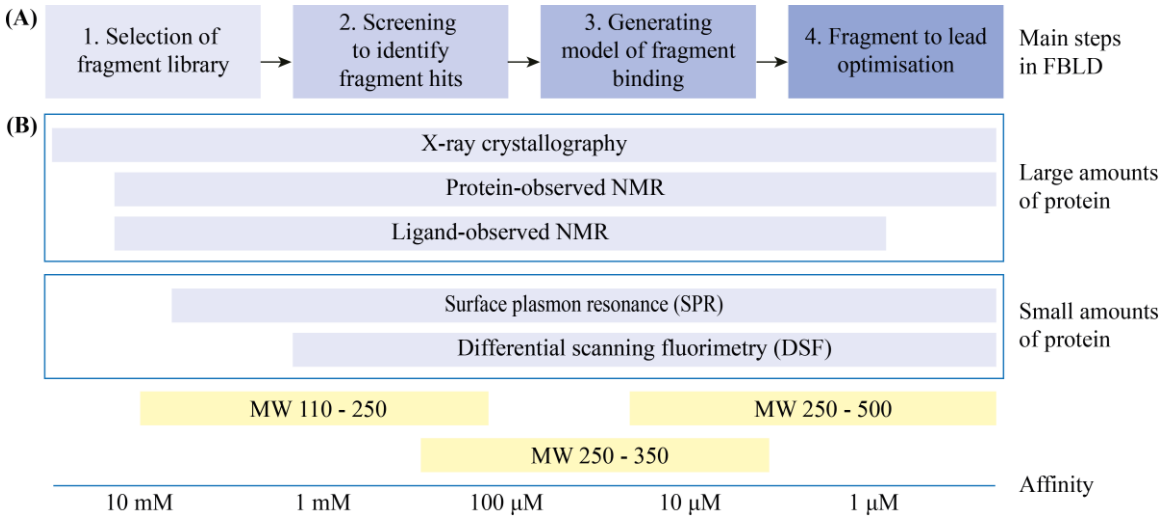



**Figure 18**: Overview of FBLD and methods used for fragment screening. **(A)** Four major steps made in FBLD. **(B)** Affinity ranges of fragment screening methods for the detection of compound binding to macromolecular targets. Binding affinities for compounds of different size and levels of FBLD are indicated. Figure **(B)** was adapted from Hubbard *et al.*[314]

For a screening campaign the optimal combination or sequential application of the previously enumerated techniques has to be identified. They can either be applied independently of each other or consecutively in a screening cascade or funnel format, with the highest sensitivity and throughput run first[300b, 312]. In the first case, the resulting hit lists are analysed with attention to similarities and deviations[315]. In both cases, the number of fragment hits can be significantly

reduced, even if many compounds have been tested. Hit lists of different methods can differ significantly and result in quite little overlap[315-316], and the outcome of a screening cascade is limited by the least sensitive and robust method[292, 293b]. Throughout the whole FBLD campaign, target enabling plays an important role, as it includes production of pure, functional and homogenous protein, setup of robust assays and methods for the generation of binding models[292].

### 1.5.1 Fragment libraries

Fragment libraries with high fragment diversity and good coverage of chemical space ensure holistic testing and increase the chance to identify fragments that are binders[291-292, 317]. Although many libraries have been compiled in the last two decades, it is unknown how diverse these are, because compound lists are usually not disclosed as they are considered a competitive advantage or confidential for company know-how[291]. Nevertheless, criteria for library assembly have been published in several publications[291, 295b, 297a, 318].

A set of physicochemical properties frame the selection of fragment-like compounds to guarantee high solubility (>2 mM in aqueous solution) and optimal interactions with the target protein[292, 295a, 314]. Fragment complexity can be limited according to the rule-of-three (Ro3), which includes an MW ≤300 Da, a logarithm of the partition coefficient (logP) ≤3, a number of H-bond donors (HBD) ≤3, and a number of H-bond acceptors (HBA) ≤6[305]. The Ro3 originated from Lipinski's rule-of-five (Ro5)[107], where the numerical values are equal or an even multiple of five. It is assumed that $10^{60}$ possible organic compounds comply with the Ro5[107, 319]. Apart from the Ro3 and Ro5, further criteria considered in library assembly are the Heavy Atom Count (HACnt) ranging from 5 to 18, the topological polar surface area (PSA) ≤90 Å$^2$, the number of rotatable bonds ≤3, and the aromatic ring count ≤3[320]. Some libraries contain a number of special fragments, such as compounds with a higher amount of $sp^3$-hypbridized carbon atoms for shape diversity[321], fragments derived from natural products[322], or compounds with $CF_n$ groups[323]. Recently, libraries of ligands with even lower molecular-weight were reported[324]. When the desired boundaries are settled, surface plasmon resonance (SPR) and ligand-observed NMR can help to identify sticky compounds or self-aggregators[293b, 325]. In addition, compounds are assessed for reactive or toxic groups, so called structural alerts[326], likely to produce pan-assay interference (PAINS)[327]. They are discarded if necessary.

When a new library should fill gaps in chemical space, additional filtering tools to access novelty, such as extended connectivity fingerprints (ECFP4) are applied or data analysis tools, such as principal component analysis (PCA) and principal moments of inertia (PMI) are applied, to compare the new compound set with already known libraries[291]. Some scientist claim that increasing the diversity and novelty of the compound collection is a constant concern[291, 328]. However, libraries are often assembled from commercially available substances and synthesised

compounds, which share common parent scaffolds[291, 329]. By 2012 drug space showed 351 ring systems and 1197 frameworks[330], whereby only a small number of new ring systems entered drug space each year[331]. The SHAPES strategy even fosters known scaffolds, as it looks for fragments that match known therapeutic agents[332]. According to Schuffenhauer *et al*.[333], the number of commercially available reagents is still much larger than any compound collection and when proprietary fragments are included there is potential for novelty.

For medicinal chemistry efforts, fragments should contain a suitable functional group, such as carboxylic acids, amines and alcohols, and masked linker groups[333]. Functional groups can also serve as linker in subsequent optimization but when they form the key interaction to the protein, their chemical modification will likely results in an affinity loss[332, 334]. The use of building block-like fragments with masked linking groups proved to be a highly useful strategy[302, 326c, 335]. In addition, it enables screening of scaffolds that are highly reactive in their unmasked form, such as isocyanates, aldehydes and acid chlorides[333]. The concept of fragment pairs considers masked screening fragments and the corresponding building blocks, and hence, the linking strategies, as well as the masking or capping reactions have to be specified before the library is assembled[333]. In chemically-poised libraries, fragments can be deconstructed in two scaffolds that are based on the most commonly used chemical reactions, and thus allow rapid parallel synthesis of fragment analogues[336].

Finally, high compound quality, with regards to purity, stability and solubility, as well as suitable plating formats are key requirements for the successful deployment of a compound library[291]. Compound quality controls, careful handling and appropriate storage are essential to guarantee and maintain their quality, because impurities can significantly increase the number of false-positives[291], when screened at high compound concentrations, e.g. at 1 mM in protein-observed NMR experiments[309b].

### 1.5.2 Fragment-based screening by NMR

The finding of the physical basis and key developments of NMR spectroscopy were associated by several Nobel Prizes. Stern and Rabi, as well as Bloch and Purcell were awarded with the Nobel Prize in physics in 1943, 1944 and 1952, respectively[337]. Whilst Stern discovered the magnetic moments of protons[338], Rabi was the first to record the magnetic properties of atomic nuclei[339] and Bloch and Purcell were the first, who independently of each other conducted an NMR experiment[340]. NMR spectroscopy is based on the fact that atomic nuclei have a magnetic moment and a nuclear spin. They align to an external magnetic field in few, defined orientations with a specific energy, according to the laws of quantum mechanics. Exposure of the sample to radio waves of certain frequencies make the nuclear spins to invert the population between energy levels. If the frequency matches the characteristic frequency of the nuclei (resonance or Lamor

frequency), an electromagnetic resonance can be detected and can be plotted as function of frequency. The nuclear resonance frequencies depend not only on the type of atomic nuclei but also on their chemical environment[341]. Among others, Purcell mentioned the possibility to use NMR as a tool for chemical analysis[342]. Modern NMR experiments are based on the work of Ernst and Wüthrich, who were awarded with the Nobel Prize in Chemistry in 1991 and 2002, respectively[337a]. Ernst found that applying short, intense pulses of radio waves instead of slow, sweeping radio waves increased the sensitivity of NMR spectra dramatically. The signal was recorded as function of time and many pulses summed up, before they are transferred into a frequency-resolved NMR spectrum using Fourier transformation (FT)[343]. He also developed many pulse sequences for 2D NMR experiments[344]. Wüthrich showed that NMR spectroscopy can also be used for the elucidation of 3D structures of proteins in solution[345].

NMR spectroscopy was the first method used for fragment screening[302] and to date, NMR-based screening is a well-established and commonly used technology in drug discovery[293b, 302]. NMR experiments are sensitive, detect interactions in solution[309b] and can be divided in ligand-observed and protein-observed experiments[346]. Thus, NMR experiments can be applied sequentially in order to do a comprehensive FBS campaign by NMR spectroscopy including screening of fragments and validation of primary hits[309b]. Gossert and Jahnke developed a concept, named validation cross, to easily keep track of the validation and integrity status of a fragment hit throughout such a cascade of experiments[309b]. For primary screening, ligand-observed experiments are favoured, as they require less protein than protein-observed experiments. They are label free, have no upper size limit in molecular weight for the target protein and can identify binders from mixtures, if each fragment has at least one distinct peak in the overlay of all recorded 1D spectra[292, 309b, 346b, 347]. Typical one-dimensional ligand-observed NMR experiments are saturation transfer difference (STD), T1ρ relaxation, water-ligand observed via gradient spectroscopy (waterLOGSY) and $^{19}$F T2 experiments[293b, 348].

STD experiments are based on the transfer of magnetization from the protein to the bound fragment[349]. The experiment is based on the Nuclear Overhauser Effect (NOE)[350], which enables the transfer of magnetisation trough space instead of chemical bonds. The peaks of binders show a decrease in signal intensity and can be easily identified, when compared to compound blanks. Although STD NMR is commonly used, false positives can result from saturation of fragment methyl protons and thus STD experiments should be used in combination with other experiments[351]. T1ρ relaxation experiments are based on the fact that spin-spin relaxation time (T2) differs for ligand nuclei that are free in solution and for ligand nuclei that are bound in a complex[352]. T2 is slow for free ligands, as they tumble very fast in solution and faster for ligands, which formed a complex and thus tumble as slow as the protein. To detect a change in T2, the spins are recorded after allowing relaxation for a short and for a long time (spin lock time). The comparison of the two spectra shows a significant broadening in linewidth and a loss of signal

intensity for binders. The reduction of the signal intensity varies, as it is dependent on the mass difference between binder and protein. Measuring compound blanks is an easy way to prevent false positives that may occur for example due to aggregation of the compound[309b]. WaterLOGSY is based on the transfer of magnetisation from bulk water through space to the ligand[353], thus also using the NOE[350]. Magnetisation is transferred to the ligand via its hydration shell in solution and from water molecules located at the ligand-protein interface. Due to phase shifting, the two cases differ in their algebraic sign. Chemical shifts of compounds forming a protein-ligand complex have a positive sign, whereas chemical shifts of free ligands have a negative sign[353b]. Including an internal standard helps phasing the spectra correctly and measuring compound blanks is an easy way to prevent false positives which can easily occur as exchangeable protons give positive signals, too. Gossert and co-workers developed polarization optimized PO-waterLOGSY[346b], which reduced the measurement time by a factor of five[346b]. Such time savings are a huge advantage when measuring large numbers of samples in screening mode. $^{19}$F NMR has become the tool of choice for fluorine containing fragments in drug discovery[354]. The wide chemical shift range of $^{19}$F and the strong chemical shift anisotropy allow mixtures of >30 $CF_3$-fragments without overlapping of signals[309a, 323]. For $^{19}$F T2 experiments, the Carr-Purcell-Meiboom-Gill Sequence (CPMG)[355] is commonly used to measure T2 relaxation times. The readout is similar to T1ρ relaxation experiments, as also T2 relaxation times differ for fragments free in solution or when binding to a protein[292]. Binders are detected due to their increase of linewidth, which can even lead to disappearance of the signal in the background of the spectrum. Unfortunately, 1D experiments do not give information on the binding mode but when applied as an reporter screen[356] they can make use of known ligands and identify completive binding to known binding sites[351].

Chemical shift changes in protein-observed NMR spectra are considered as gold standard for fragment screening[293b, 309b]. Frequently used methods are HSQC experiments[302] or SOFAST-HMQC experiments[357], which correlate the chemical shifts of a nitrogen or carbon atom to an attached or neighbouring proton, respectively. However, these two-dimensional methods require isotope-labelled protein, as the natural abundance of the isotopes $^{15}$N and $^{13}$C is 0.37% and 1.1%, respectively[358]. Protein-observed NMR is better suited for proteins with a MW <30 kDa for $^{15}$N labelling and <100 kDa for $^{13}$C labelling[309b]. Resonance assignment can reveal the ligand binding site[359] and in some cases measuring the chemical shift differences in a dose response series can provide the $K_d$[360].

A big advantage of NMR over other screening techniques is the large number of different experiments that make the method very versatile[361]. 1D $^1$H spectra for quality control can be measured from the same sample tube by applying solvent suppression and the excitation sculpting principle (zgesgp pulse program)[362]. If the compound stability and solubility are monitored,

misinterpretation of results due to erroneous assumptions is avoided and the number of false positives and false negatives is decreased[292, 309b].

### 1.5.3 Fragment-based screening by X-ray crystallography

Crystals of haemoglobin were first described by Hunefeld in 1840[363]. Throughout the late 19th and early 20th century, crystallization became a powerful purification tool for naturally occurring peptides and proteins from supersaturated extracts[364]. Many Nobel Prize winners are associated with crystallography[365], among them are Sumner[366], Northrop[367] and Stanley[368], who were awarded in 1946 with the Nobel Prize in chemistry for their work on purification and crystallization of urease, pepsin and the tobacco mosaic virus, respectively[369]. The basis for the development of X-ray crystallography was the discovery of X-rays by Röntgen[370] in 1895, the observation of diffraction patterns of crystals by von Laue in 1912[371] and the formulation of Bragg's law in 1913[372]. When a well-ordered single crystal interacts with electromagnetic radiation[373], a part of the X-rays are scattered in all directions whereas others diffract, according to Bragg's law[374]. It describes 3D diffraction as a reflection of an incipient X-ray beam by imaginary planes in the crystal lattice which occurs if the interference is constructive. For a planar interspacing $d$ and an incident angle $\theta$, this is true if the path difference between waves with the wavelength $\lambda$ is equal to an integer number $n$ (**Figure 19**). The minimum distance $d_{min}$ that can be resolved corresponds to the maximum angle $\theta_{max}$ and is called resolution of the diffraction pattern. High resolution is directly related to higher crystalline order[364].

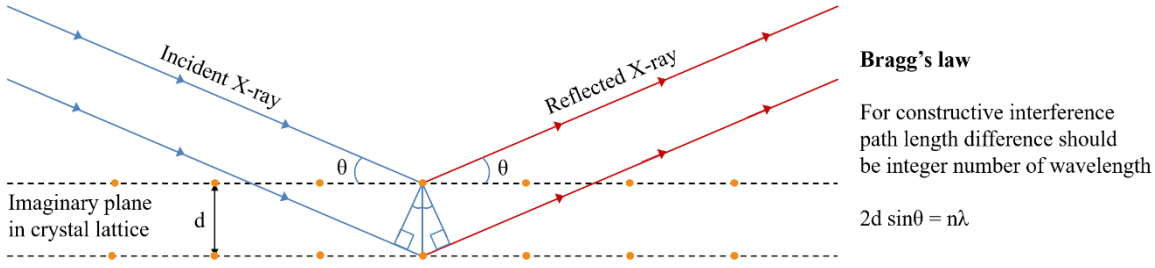



**Figure 19:** Bragg's law. Reflection of X-rays by imaginary planes in a crystal lattice.

The first protein structure elucidated by X-ray crystallography was the structure of myoglobin, which was published by Kendrew[375] in 1958 and was awarded with the Nobel Prize in chemistry in 1962[376]. At this time, the diffraction spots (reflections) were detected on photographic film and analysed to elucidate the protein crystal structure[376]. Then and now, the coordinates of the reflections represent a pattern which provides information about the crystals' space group. The measured intensities contain information about the molecules forming the lattice[377]. Each reflection contains information from all atoms in the crystal as it results from

interactions of all atoms in the crystal at the same angle. The phases of the reflections cannot be measured, which is known as phase problem in crystallography and therefore the electron density distribution ρ throughout the unit cell cannot be calculated by applying a Fourier transformation from the recorded reflection data[373, 378]. Perutz and co-workers[379] solved the phase problem in 1954 by using multiple isomorphous replacement through heavy atom soaks and hence decisively assisted Kendrew's work on myoglobin. Molecular replacement (MR) uses the phases of an unravelled structure of a homologue to phase a new protein structure. Owing to the large number of available crystal structures it became the major procedure used today[380].

In the last decades, methods in and around crystallography developed tremendously and turned X-ray crystallography into a cornerstone of structural biology[372]. It is the most powerful and common method to elucidate the three-dimensional structures of macromolecules such as proteins, nucleic acids, protein or nucleic acid complexes and their complexes with low-molecular weight ligands [373, 381]. The development began when many proteins with low abundance in natural systems and membrane proteins became available through genetic engineering, recombinant expression and better handling, using special reagents[364, 372]. Protein crystals are usually soft, contain a large amount of water, are fragile, limited in size, and are sensitive to temperature changes. They can disintegrate upon dehydration, in consequence they show poor optical properties and poorly diffract X-ray radiation. Therefore, key steps in X-ray crystallography were the introduction of cryo-crystallography[382], in which crystals are measured at 100 K to minimize radiation damage, which allowed the usage of high intensity synchrotron radiation[383].

Recent developments turned the method into a high throughput technique. Therefore and because X-ray crystallography is particularly sensitive and simultaneously allows high fragment concentrations, it became a hit identification tool in FBS[384]. This was first demonstrated by Nienaber *et al*.[304] and Hartshorn *et al*.[385] in 2002 and 2005, respectively. The developments included miniaturisation and automation of crystallization trials[386], sophisticated crystal handling with robots[387], stable, brighter and tuneable radiation sources[388] and high speed detectors with high resolution[389]. Today's beamlines are software driven and run mostly in automation, which includes strategy routines for optimal collection of diffraction images[388, 390] and automatic data processing pipelines[391] that replace the labour-intensive and time-consuming manual processing[392]. The subsequent manual refinement work of the crystallographer towards a final structural model is accelerated by software tools, which can identify and/or place ligands[391a, 393] or ions[394]. The quality of the obtained 3D structure results from a combination of the diffraction quality of the crystal, the optimized data collection strategy, and the quality of data processing and refinement[395].

Despite all advances, identification of fragment hits remained difficult when compared to more potent ligands, because their low affinity can result in partial occupancy, causing weak and ambiguous electron density that can be misleading[394, 396]. High-quality diffraction data with a

resolution of at least 2.5 Å are needed to evaluate such weak density[397] and when an automatic processing pipeline is set up to identify fragment hits. Special attention is required, e.g. when placing water molecules[384a]. Recently, Pearce *et al*.[393a] developed a multi-crystal approach, called Pan-Dataset Density Analysis (PanDDA). The software computes a background electron density estimate as a mean of ground-state measurements from apo structures and in a second step, significant changes from the mean are identified by a weighted subtraction of the background from each electron density map. The resulting partial-difference map is termed event map and corresponds to the bound fraction in the crystal, i.e. ligand binding event. Thus, PanDDA allows sensitive detection of binding sites as it reveals regions of an individual data set that represent a statistical outlier[393, 398]. This approach was shown to be ideal for data analysis of high-throughput fragment screening campaigns by X-ray crystallography[399]. In Europe, such campaigns can be conducted at the XChem lab[400] at the Diamond Light Source in Harwell, UK, at the HTX lab[387b] at the EMBL in Grenoble, France, and at the BESSY II MX-beamlines of the Helmholtz-Zentrum in Berlin, Germany[401]. Both sites offer state of the art equipment, and access to fragment libraries. However, the presence of a well-established crystallization system yielding high quality crystals of the apo protein of interest that are suited for soaking experiments remains an absolute requirement[385].

To establish such crystallization system, a pure and homogenous protein formulation is brought to supersaturation under various conditions. In most cases this leads to precipitation, but sometimes it leads to nuclei formation and subsequent crystal growth until the equilibrium is re-established[364]. There is no comprehensive theory to guide crystallization efforts and in consequence, protein crystallography is empirical and trial and error is the main method to succeed[364]. The most common technique to achieve supersaturation is vapour diffusion in hanging drops or sitting-drops[364]. Others are free interface diffusion, dialysis, and batch crystallization[364]. To find and optimize crystallization conditions, either commercial screening matrices are used in a shotgun approach, or all parameters are varied as systematically as possible in several test campaigns[364, 402]. A combination of both strategies is applied, when crystal quality or size obtained from primary conditions need further optimization[364]. Commercial screening matrices cover a wide range of conditions and are done on micro-scale in plastic multi-chambered trays. Thus, they explore a large range of crystallization conditions while using little material and became the method of choice when facing a new crystallization problem[364]. The most important variables in a crystallization system include the precipitant of the macromolecule, such as salts, polymers and organic solvents, pH and temperature[364]. These variables can be correlated to each other resulting in a non-linear problem[364, 403]. When the formation of stable nuclei is an obstacle, seeding is used to directly induce crystal growth, by adding crystalline material from various sources[364]. Seeds are added by pipetting or they are introduced by using whiskers or horse hair, which results in a so-called streak seeding[404]. An effective method, used during crystallization screening, is

microseed matrix screening (MMS)[405]. In this method, a seed stock is prepared from protein crystals with the seed bead method[406] and used as an additive in a matrix condition screen of the same protein.

With a well-established crystallization system for the apo protein, it can still be difficult to obtain structures of protein-ligand complexes by soaking or by co-crystallization[378]. For soaking, apo crystals of known structure and good quality are prepared and incubated with the ligands of interest[378]. The symmetrically arranged protein chains in a protein crystal are loosely packed and interact only by a few contacts with their neighbouring chains[364]. Large solvent channels traverse the protein crystal and account for 30 to 80% of its volume[407]. Fragments can penetrate the preformed crystal through these channels by diffusion and bind to the protein, what makes soaking a simple method, which can achieve high throughput and good reproducibility[373, 378]. The success of soaking experiments depends on the accessibility of the desired binding site through channels in the crystal lattice, the channel size, which typically varies from 20 to 100 Å[408], the channel configuration, e.g. shape, surface charge distribution, viscosity of the bulk solvent, solubility of the ligand in the mother liquor and ligand affinity[364]. Conformational changes required for ligand binding might not be tolerated by the crystal packing[409]. An alternative to soaking is co-crystallization, where protein and ligand are mixed to form the complex in solution which is then crystallized. Each new complex can differ from the apo protein as well as from other complexes and therefore the system does not necessarily crystallize under the known conditions of the apo protein. If every complex leads to a new crystallization problem, co-crystallization becomes a demanding and time consuming effort[378].

### 1.5.4 Fragment-to-lead optimisation

After the identification of low affinity fragment hits, the fragment hit-to-lead optimization starts, which is also known as lead generation[292, 378, 410]. Usually, so called SAR-by-archive and SAR-by-catalogue studies are made directly after finishing the fragment screen and without any knowledge of the binding pose of the fragment hit, to test substructures and similar compounds[411]. As a second step, medicinal chemisty is applied in iterative cycles of design and testing to develop the potency, selectivity, activity and pharmacokinetic properties of the fragment hit[67, 99b, 136c, 295b, 410, 412]. Fragment optimization relies on the fact that each interacting part of the molecule contributes to the free energy of binding[413]. Initially, fragments form few interactions, which results in low affinity. Optimization leads to additional binding interactions and a gain in binding affinity[413-414]. In most cases, three-dimensional structures of protein-ligand complexes are accessible, often by X-ray crystallography[378], and thus the iterative optimization is driven by rational compound design[24]. This process is called structure-based lead design (SBLD), structure-based drug design (SBDD) or structure-aided drug design (SADD) and represents the

decisive advantage of target-based approaches over cell-based approaches[378, 415]. Several case studies demonstrated its success[416]. Prominent examples are inhibitors of HIV-1 protease[417], HIV-1 integrase[418], and influenza neuraminidase[419]. The three main strategies in SBDD when associated with FBLD are fragment linking, fragment growing and fragment merging (**Figure 20**)[333, 412b]. In fragment linking, fragments that are located in close proximity to each other, are linked by a suitable spacer[292]. Finding a spacer with the right length and geometry is important, to avoid negative effects on binding[420]. This was successfully done in the first FBLD campaign[302], the development of Venetoclax[421], and others[422].

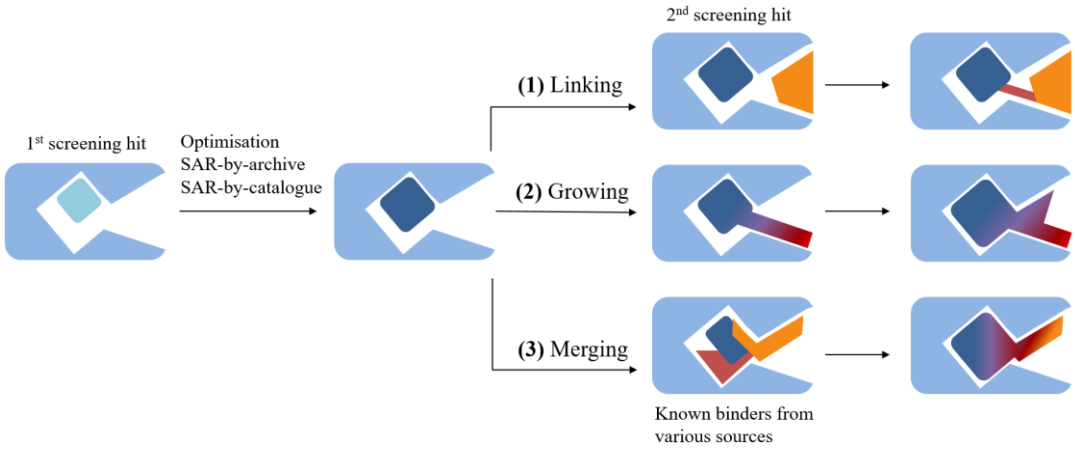



**Figure 20**: Main strategies in SBDD. After the fragment screening, SAR-by-archive and/or SAR-by-catalogue studies are applied, followed by SBDD, either using linking, growing or merging. Figure was adapted from Lamoree *et al.*[292]

Fragment growing is the most commonly used strategy in SBDD[292, 293b]. Here, a new series of compounds is extended step by step using a single group to identify further interactions and to be able to distinguish between beneficial and detrimental effects[397]. This approach led to the development of Vemurafenib[423], as well as to compound AT9283, an inhibitor of the Aurora kinase activity[424], and compound AUY922, an inhibitor of chaperone Hsp90[425]. Fragment merging combines scaffolds that originate from multiple crystal structures and also takes results of other screening campaigns, such as docking studies and literature searches into account[292, 426]. Examples are an inhibitor of PDPK1 kinase[426] and the compound BEP800, another inhibitor of the chaperone Hsp90[425, 427]. A lead compound inhibiting mitochondrial branched-chain aminotransferase (BCATm), resulted from merging a hit derived from FBS with a hit derived from HTS[428].

All in all, SBDD contains a significant number of options for chemical optimization. Thus, it requires a large amount of decision-making[292], which is guided by synthetic accessibility, the best design opportunities offered by the binding pocket, and the resulting binding affinities[412b]. Characterization of ligand binding and determination of $K_d$ by SPR[310b], isothermal titration

calorimetry (ITC)[429] or NMR[360a] plays an important role in the optimization process. For prioritizing FBS hits and optimized compounds, the ligand efficiency (LE) is a good indicator, as the value of the whole ligand and of added atoms to the molecule can be estimated by the LE. It is defined as the binding free energy per non-hydrogen atom[430]. If the ligand binding event does not cause conformational changes, *in silico* docking studies, such as dynamic un-docking (DUCK)[431], or virtual screening of compound libraries, such as ANCHOR.QUERY[432], can deliver additional knowledge and guidance for the optimization process[292]. Further *in silico* approaches are FTMAP[433] and Molecular Dynamics simulations with mixed solvents (MDmix)[434], which map the protein surface for binding of functional groups. Using structural information from docking studies, NMR[345], or covalent fragments (e.g. tethering[306]) can help to optimize fragments in the absence of crystal structures[411].

# 2. Aim of the thesis

Millions of people, mostly underprivileged populations in South and Central America, suffer from Chagas disease (CD). They have an urgent need for novel drugs exhibiting reduced adverse effects and increased efficacy. A novel mechanism of action is preferred to circumvent emerging resistance against benznidazole and nifurtimox, the current standard of cure. *T. cruzi* is dependent on isoprenoid biosynthesis as ergosterol and other 24-alkylsterols are essential metabolites that cannot be acquired by other mechanisms. Therefore, it was hypothesised that enzymes along this pathway are promising drug targets. A number of compounds targeting these enzymes were tested and have been shown to inhibit parasite growth. Among those enzymes is farnesyl pyrophosphate synthase (FPPS), which is in the focus of this work. It catalyses the synthesis of farnesyl pyrophosphate (FPP), a C15 building block in sterol biosynthesis and in protein prenylation of signalling proteins. Therefore, it is a key branch-point enzyme in the isoprenoid pathway.

Nitrogen-containing bisphosphonates (N-BPs) are blockbuster drugs for bone diseases that inhibit FPPS by mimicking its allylic substrate. The N-BPs pamidronate and risedronate showed efficacy in mice infected with *T. cruzi*. Unfortunately, N-BPs have inappropriate pharmacokinetics to treat CD as they avidly bind to bone mineral and show poor bioavailability. Nevertheless, BPs and BP analogues are the only and well-studied lead series. Thus, finding binders of *T. cruzi* FPPS (TcFPPS) based on novel scaffolds with suitable drug properties is desirable and the aim of this thesis. A recent success story was the identification of non-bisphosphonate inhibitors of human FPPS (hFPPS) that bind to an allosteric binding site. Early fragment hits were developed by medicinal chemistry into three lead series with up to low nanomolar affinity. The distinct mode of inhibition and different physicochemical properties of these inhibitors overcome the limitations related to the N-BP scaffold.

Encouraged by the findings for hFPPS and owing to the paucity of lead series available for TcFPPS, this work focuses on the early phase of the drug discovery process. The goal is to identify TcFPPS binders of a novel scaffold, to explore potential binding sites in TcFPPS and to start structure-based lead discovery. For this purpose, fragment-based screening will be applied, using different biophysical methods, such as NMR and X-ray crystallography. Regions other than the active site are less conserved and are thought to have higher potential for specific inhibitors over the human homologue. Thus, revealing further binding sites in TcFPPS would give rise to new options to develop inhibitors of a novel scaffold specific for TcFPPS and would give new impulses for the drug discovery for CD. The overarching goal, which exceeds the scope of this thesis, is the development of a tool compound to prove the concept of allosteric inhibition of TcFPPS.

# 3. Materials

## 3.1 Chemicals

**Table 2:** List of used chemicals. All solvents listed had the purity grade *pro analysi* (p. a.).

| Chemical | Manufacturer |
|---|---|
| Acetic acid | SIGMA |
| Acetonitrile, LC-MS CHROMASOLV® | FLUKA |
| Aspartic acid | FLUKA |
| Biotin | SIGMA-ALDRICH |
| BisTris | FLUKA |
| 2,5-Dichlorobenzo[d]thiazole, 95% | ABCR |
| $CaCl_2$ | MERCK |
| Chloramphenicol | APPLICHEM |
| 2-Chlorobenzothiazole, 99% | ALDRICH |
| 7-Chloro-1H-indole-3-carbaldehyde | MATRIX SCIENTIFIC |
| 2-Chloro-5-(trifluoromethyl)benzo[d]thiazole, 95% | ENAMINE |
| $CoCl_2 \cdot 6H_2O$ | RIEDEL-DE HAËN |
| cOmplete™, protease inhibitors, EDTA-free | ROCHE DIAGNOSTICS |
| $CuCl_2 \cdot 2H_2O$ | MERCK |
| $d_6$-DMSO | EURISOTOP |
| $D_2O$ | EURISOTOP |
| DCM | BRENNTAG SCHWEIZERHALL AG |
| DCM, 99.8% anhydrous | SIGMA |
| DMSO | SIGMA-ALDRICH |
| DSS | CAMBRIDGE ISOTOPE LABORATORIES INC. |
| DTT, 1.0 M in $H_2O$ | SIGMA |
| Ethylacetate, 99 – 100% | BRENNTAG SCHWEIZERHALL AG |
| $FeCl_3 \cdot 6\ H_2O$ | SIGMA-ALDRICH |
| Folic acid (vitamin $B_9$) | SIGMA-ALDRICH |
| D-(+)-Glucose | FLUKA |
| $^{13}$C-D-(+)-Glucose | ALDRICH |
| Glycerol | SIGMA-ALDRICH |
| Guanidine · HCl | SIGMA |
| Guanidine · HCl solution, 8.0 M in $H_2O$ | SIGMA |
| $H_3BO_3$ | FLUKA |
| HCl in dioxan, 4.0 M | SIGMA-ALDRICH |
| Heptane, mixture of isomers | BRENNTAG SCHWEIZERHALL AG |
| 5-Hydroxy-1H-indole-3-carbaldehyde | J&W PHARMLAB |
| Imidazole | SIGMA-ALDRICH |
| IPTG | SIGMA LIFE SCIENCE |
| Kanamycin sulphate | SIGMA LIFE SCIENCE |
| $KH_2PO_4$ | FLUKA |
| α-Lactose · $H_2O$ | SIGMA |
| $MgSO_4$ | SIGMA |
| 7-Methoxy-1H-indole-3-carbaldehyde | BIOFINE INTERNATIONAL |
| $MnCl_2 \cdot 4\ H_2O$ | MERCK |

| | |
|---|---|
| $Na_2HPO_4$ | MERCK |
| $Na_2MoO_4 \cdot 2\ H_2O$ | SIGMA-ALDRICH |
| $Na_2SeO_3 \cdot 5\ H_2O$ | SIGMA-ALDRICH |
| $Na_2SO_4$, anhydrous | SIGMA LIFE SCIENCE |
| NaCl | SIGMA-ALDRICH |
| $NH_4Cl$ | SIGMA LIFE SCIENCE |
| $^{15}NH_4Cl$ | CAMBRIDGE ISOTOPE LABORATORIES INC. |
| $NiCl_2 \cdot 6\ H_2O$ | RIEDEL-DE HAËN |
| Nicotinamide | FLUKA |
| D-Pantothenic acid (vitamin $B_5$) | SIGMA-ALDRICH |
| 2-(Piperazin-1-yl)benzo[d]thiazole, 95% | CHEMBRIDGE CORPORATION |
| PPG antifoam, polypropylene glycol 2000 | VWR CHEMICALS |
| Protein standard, Precision Plus Protein Kaleidoscope | BIORAD |
| Pyridoxal hydrochloride | SIGMA |
| Riboflavin (vitamin $B_2$) | SIGMA |
| SOC medium | FLUKA |
| Sodium triacetoxyborohydride | SIGMA-ALDRICH |
| SYPRO® Orange, 5 mM (5000x) in DMSO | SIGMA |
| TCEP · HCl | SIGMA |
| $d_{16}$-TCEP·HCl | CAMBRIDGE ISOTOPE LABORATORIES INC. |
| TEA, anhydrous | SIGMA |
| Tert-butyl piperazine-1-carboxylate | COMBI-BLOCKS |
| Tert-butyl 3-formyl-1H-indole-1-carboxylate, 95% | ABCR |
| Thiamine hydrochloride (vitamin $B_1$) | FLUKA |
| TRIS TRIZMA® Base | SIGMA LIFE SCIENCE |
| TRIS · HCl, TRIZMA® hydrochloride | SIGMA LIFE SCIENCE |
| $d_{11}$-TRIS | CAMBRIDGE ISOTOPE LABORATORIES INC. |
| Tryptone | SIGMA-ALDRICH |
| Water, CHROMASOLV® Plus, for HPLC | SIGMA-ALDRICH |
| Yeast extract | SIGMA-ALDRICH |
| $ZnSO_4 \cdot 7\ H_2O$ | SIGMA |

## 3.2 Plasmids and *E. coli* strains

1. Plasmid encoding for TcFPPS, Uniprot ID Q8WS26, ec_opt, Met64-Lys425 in pACE0, T7 promotor, lac operon, 5596 bp, $Kan^r$
   N – $His_6$-tag – HRV 3C cleavage site – $TcFPPS_{64\text{-}425}$ – C
2. Plasmid encoding for Avi-tagged TcFPPS, Uniprot ID Q8WS26, ec_opt, Met64-Lys425, in pACE-GP9, T7 promotor, lac operon, 5683 bp, $Kan^r$
   N – $His_6$-tag – spacer (3xGGGS) – HRV 3C cleavage site – Avi-tag – $TcFPPS_{64\text{-}425}$ – C
3. Plasmid encoding for hFPPS, Uniprot ID P14324, ec_opt, Met67-Lys419, in pACE0, T7 promotor, lac operon, 5569 bp, $Kan^r$
   N – $His_6$-tag – HRV 3C cleavage site – $hFPPS_{67\text{-}419}$ – C
4. Plasmid encoding for Avi-tagged hFPPS, Uniprot ID P14324, ec_opt, Met67-Lys419, in pACE-GP9, T7 promotor, lac operon, 5656 bp, $Kan^r$
   N – $His_6$-tag – spacer (3xGGGS) – HRV 3C cleavage site – Avi-tag – $hFPPS_{67\text{-}419}$ – C

5. Plasmid encoding for human rhinovirus 3C protease (HRV 3C), Uniprot ID P03303, Gly1-Gln182, T7 promotor, lac operon, 6722 bp, $Kan^r$
   N – MBP – Thrombin cleavage site – HRV $3C_{1538-1719}$ – $His_6$-tag – C
6. Plasmid encoding for *E. coli* bifunctional ligase/repressor (BirA), Uniprot ID P06709, Met1-Lys321, in pACYC184, araBAD promoter, araC operon, $Cam^r$.
   N – $His_6$-tag – $BirA_{1-321}$ – C

All previously listed constructs were designed by Dr. Felix Freuler, NIBR, Novartis Pharma AG, Basel, Switzerland. Plasmids 1. and 3. were purchased from INVITROGEN GENEART and were obtained as lyophilized plasmids that were dissolved at 0.1 µg · $\mu L^{-1}$ in TE buffer and stored at -80 °C. Plasmids 2. and 4. were cloned by Simon Haenni and Lena Muenzker, NIBR, Novartis Pharma AG, Basel, Switzerland. These plasmids were dissolved at 0.1 µg · $\mu L^{-1}$ in elution buffer (MARCHEREY-NAGEL kit, no 740615.50) and stored at -80 °C.

For transformation competent *E. coli* BL21 (DE3) (genotype: *fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*, T1 phage resistance, 0.05 mL per tube, NEW ENGLAND BIOLABS) were used. For expression of *in vivo* biotinylated avi-tagged protein, the above listed *E. coli* strain was transformed with a plasmid encoding for bifunctional ligase/repressor (BirA). They were kindly provided as competent cells by Cecile Delmas, NIBR, Novartis Pharma AG, Basel, Switzerland.

## 3.3 Proteins

All enzymes listed in **Table 3** have been overexpressed in *E. coli* and purified as part of this work. The hFPPS was expressed and purified in collaboration with Lena Muenzker, NIBR, Novartis Pharma AG, Basel, Switzerland. In some measurements TbFPPS (Uniprot ID Q86C09) was included. It was expressed and purified as part of the doctoral thesis of Lena Muenzker. In addition, Pierce™ bovine serum albumin (BSA) standard at 1.5 mg · $mL^{-1}$ and 2.0 mg · $mL^{-1}$, THERMO FISHER SCIENTIFIC, were used.

**Table 3:** List of expressed and purified enzymes.

| Enzyme | MW (Da) | Correction factor A280 EC 1 mg · mL$^{-1}$ |
|---|---|---|
| TcFPPS | 41313 | 1.412 |
| $^{13}C^{15}N$-labelled TcFPPS | 43657 | 1.412 |
| Biotinylated avi-tagged TcFPPS | 43352 | 1.430 |
| hFPPS | 40686 | 1.346 |
| $^{13}C^{15}N$-labelled hFPPS | 43007 | 1.346 |
| Biotinylated avi-tagged hFPPS | 42724 | 1.330 |
| HRV 3C | 62853 | 1.110 |

## 3.4 Chromatography resins

For protein purification affinity chromatography columns (Ni-NTA Superflow 5 mL, QIAGEN), desalting columns (HiPrep™ 26/10, GE HEALTHCARE) and a size exclusion column (HiLoad™ Superdex™ 16/60 S200, GE HEALTHCARE) were used. For buffer exchange pre-packed desalting columns (PD-10, Sephadex™ B-25 Medium, GE HEALTHCARE) were used. For purification of chemical reaction mixtures two types of pre-packed normal phase silica flash columns (RediSep® Rf, 12 g, TELEDYNE ISCO and FlashPure12 g, silica 40 µm irregular, BÜCHI) were used. Thin layer chromatography was performed on precoated silica gel plates (0.2 mm, particle size 25 µM, FLUKA).

## 3.5 Buffers and solutions

All buffers and solutions were prepared with ultrapure water (Millipore) and sterile filtered (0.22 µm). The pH was adjusted using HCl or NaOH.

**Table 4**: List of buffers, solutions and media for protein expression and purification.

| Solution / medium | Composition |
|---|---|
| TE buffer | 10 mM TRIS·HCl, pH 8.0, 0.1 mM EDTA |
| SOC medium, FLUKA | 20 g · L$^{-1}$ tryptone, 5 g · L$^{-1}$ yeast extract, 4.8 g · L$^{-1}$ $MgSO_4$, 3.603 g · L$^{-1}$ dextrose, 0.5 g · L$^{-1}$ NaCl, 0.186 g · L$^{-1}$ KCl |
| LB medium | 10 g · L$^{-1}$ tryptone, 5 g · L$^{-1}$ yeast extract, 10 g · L$^{-1}$ NaCl |
| MDG medium | 25 mM $Na_2HPO_4$, 25 mM $KH_2PO_4$, 50 mM $NH_4Cl$, 5 mM $Na_2SO_4$, 2 mM $MgSO_4$, 0.2 × metals, 0.5% (w/v) D-(+)-glucose, 0.25% (w/v) aspartic acid |
| 1000x metals | 50 mM $FeCl_3 \cdot 6\ H_2O$, 20 mM $CaCl_2$, 10 mM $MnCl_2 \cdot 4\ H_2O$, 10 mM $ZnSO_4 \cdot 7\ H_2O$, 2 mM $CoCl_2 \cdot 6\ H_2O$, 2 mM $CuCl_2 \cdot 2\ H_2O$, 2 mM $NiCl_2 \cdot 6\ H_2O$, 2 mM $Na_2MoO_4 \cdot 2\ H_2O$, 2 mM $Na_2SeO_3 \cdot 5\ H_2O$, 2 mM $H_3BO_3$ |

| | |
|---|---|
| modAI medium | 25 mM $Na_2HPO_4$, 25 mM $KH_2PO_4$, 50 mM $NH_4Cl$, 5 mM $Na_2SO_4$, 2 mM $MgSO_4$, 1 × metals, 2.5% (w/v) tryptone, 5% (w/v) yeast extract, 1.0% (w/v) glycerol, 0.1% (w/v) D-(+)-glucose, 0.4% (w/v) α-lactose |
| 100x BME vitamin mix | 50 g · $L^{-1}$ thiamine hydrochloride, 10 g · $L^{-1}$ D-pantothenic acid, 10 g · $L^{-1}$ biotin, 10 g · $L^{-1}$ pyridoxal hydrochloride, 10 g · $L^{-1}$ folic acid, 10 g · $L^{-1}$ nicotinamide, 1 g · $L^{-1}$ riboflavin |
| $^{13}C^{15}N$-uniform labelling medium | 50 mM $Na_2HPO_4$, 50 mM $KH_2PO_4$, 5 mM $Na_2SO_4$, 2 mM $MgSO_4$, 0.4% (w/v) $^{13}C$-D-(+)-glucose, 0.25 % (w/v) $^{15}NH_4Cl$, 1 × BME vitamin mix, 0.2 × metals |
| Lysis buffer | 50 mM TRIS, pH 8.0, 300 mM NaCl, 5 mM imidazole, 2 mM TCEP · HCl, 10% (v/v) glycerol |
| Elution buffer | 50 mM TRIS, pH 8.0, 150 mM NaCl, 200 mM imidazole, 2 mM TCEP · HCl, 10% (v/v) glycerol |
| SEC buffer | 50 mM TRIS, pH 8.0, 200 mM NaCl and 2 mM TCEP · HCl |

**Table 5**: List of buffers for NMR spectroscopy.

| Buffer | Composition |
|---|---|
| TRIS buffer | 10 mM $d_{11}$-TRIS, pH 8.0, 150 mM NaCl, 1 mM $d_{16}$-TCEP · HCl |
| BisTris buffer | 25 mM BisTris, pH 6.5, 50 mM NaCl, 2 mM TCEP · HCl |
| Hepes buffer | 25 mM $d_{18}$-Hepes, pH 7.4, 150 mM NaCl, 1 mM TCEP · HCl, 2 mM $MgCl_2$, 0.01% (v/v) Tween, 10% (v/v) $D_2O$, 150 µM DSS |

Seven commercial crystallization screens were purchased in deep well block format: $AmSO_4$ Suite, Cryos Suite, JCSG+ Suite and MBClass II Suite from QIAGEN and Index HT, PegRx HT and SaltRx HT from HAMPTON RESEARCH. For final crystallization conditions, the buffers and reservoir solutions listed in **Table 6** were used. Some of the screened conditions were prepared for further investigations. Crystallization plates were prepared with a pipetting robot (Formulator) from stock solutions (**Table 7**).

**Table 6**: List of buffers and solutions for crystallization.

| Buffer | Composition |
|---|---|
| Low salt protein buffer | 10 mM TRIS, pH 7.4, 25 mM NaCl, 2 mM TCEP · HCl |
| Seed stock buffer | 80 mM NaOAc, pH 5.0, 160 mM $(NH_4)_2SO_4$, 20% (w/v) PEG 4000, 20% (v/v) glycerol |
| Reservoir 24-well plate | 80 mM MES, pH 6.5, 8.5 mM $ZnSO_4$, 19.42% (v/v) PEG MME 550, 15% (v/v) glycerol |
| Reservoir 96-well plate | 80 mM MES, pH 6.5, 4 mM $ZnSO_4$, 12.36% (v/v) PEG MME 550, 11.57% (v/v) glycerol |

**Table 7**: List of stock solutions used for crystal plate preparation with the formulator.

| Stock | Vendor / preparation |
|---|---|
| 2.5 M ammonium citrate dibasic, pH 4.7 – 4.8 | HAMPTON RESEARCH |
| 75% (v/v) glycerol | Prepared from 100% glycerol, SIGMA-ALDRICH |
| 1.0 M MES, pH 5.2 | HAMPTON RESEARCH |
| 1.0 M MES, pH 7.1 | HAMPTON RESEARCH |
| 1.0 M NaOAc · 3 $H_2O$, pH 3.6 | HAMPTON RESEARCH |
| 1.0 M NaOAc · 3 $H_2O$, pH 5.6 | HAMPTON RESEARCH |
| 3.5 M $(NH_4)_2SO_4$ | HAMPTON RESEARCH |
| 8.0 M $NH_4OAc$, pH 4.0 – 5.4 | HAMPTON RESEARCH |
| 50% (w/v) PEG 3350 | HAMPTON RESEARCH |
| 50% (w/v) PEG 4000 | HAMPTON RESEARCH |
| 75% (v/v) PEG MME 550 | Prepared from 100% PEG MME 550, FLUKA |
| 50% (w/v) PEG MME 2000 | HAMPTON RESEARCH |
| 1 M TRIS, pH 7.0 | HAMPTON RESEARCH |
| 1 M TRIS, pH 9.0 | HAMPTON RESEARCH |
| 2 M $ZnSO_4$ · 7 $H_2O$ | HAMPTON RESEARCH |

All other used buffers and solutions that were used in various experiments are listed in the following table.

**Table 8**: List of other buffers and solutions.

| Buffer | Composition |
|---|---|
| SDS-PAGE running buffer | 25 mM TRIS, pH 8.3 192 mM glycine, 0.1% (w/v) SDS |
| SDS sample buffer | NuPage® LDS Sample buffer 4x, NOVEX LIFE TECHNOLOGIES |
| SDS-PAGE staining solution | Instant Blue™, EXPEDEON |
| DSF buffer | 25 mM BisTris, pH 6.5, 50 mM NaCl, 2 mM TCEP · HCl |
| Solubility and stability screen | Solubility and Stability Screen II, HAMPTON RESEARCH |
| SPR buffer | 50 mM Hepes, pH 7.4, 150 mM NaCl, 2 mM TCEP · HCl, 2 mM $MgCl_2$, 0.01% (v/v) Tween |
| SPR buffer with DMSO | 50 mM Hepes, pH 7.4, 150 mM NaCl, 1 mM TCEP · HCl, 2 mM $MgCl_2$, 0.01% (v/v) Tween, 0.9% (v/v) DMSO |
| LC-MS cleaning solution | 7.62 M guanidine hydrochloride, 48 mM DTT |
| LC-MS eluate A | 0.05% TFA |
| LC-MS eluate B | 0.04% TFA in acetonitrile |
| HPLC cleaning solution | 20 mM TRIS, pH 8.0, 6 M guanidine, 100 mM NaCl, 10 mM DTT |
| HPLC eluate A | 90% (v/v) acetonitrile, 0.1% (v/v) TFA |
| HPLC eluate B | 0.1% (v/v) TFA |

## 3.6 Fragment libraries

The **4th generation Novartis core fragment library** contained 1408 fragments, which were available in 176 mixtures of eight compounds as 50 mM $d_6$-DMSO stock solutions (6.25 mM per compound) that were divided on two 96-well plates. In some mixtures compounds had been replaced by an equal amount of $d_6$-DMSO and at the time of the screen 1336 compounds were available. All compounds were also available as singles as 50 mM $d_6$-DMSO stock solution for follow up experiments. The fragments fulfilled the following criteria: MW ≤300 Da, clogP <3, 1 – 3 aromatic rings, maximal linker length 1 – 3 bonds, rotatable bonds <3, HBD <3 and HBA <5, solubility >200 µM in aqueous solution.

The **1st generation Novartis fluorine library** contained 540 $CF_3$-compounds in 18 mixtures of 30 compounds as 50 mM $d_6$-DMSO stock solutions (1.66 mM per compound). In some mixtures compounds were replaced by an equal amount of $d_6$-DMSO leading to 470 $CF_3$-compounds that were screened. Additionally, 1 mixture with 12 $CF_2$-compounds was screened.

The **1st generation Diamond-SGC poised library** (DSPL)[336] contained 406 fragments as singles (380 used) as 500 mM DMSO compound stocks on 384-well Echo-compatible source plates.

The **Edelris keymical fragments™**[435] contained 279 fragments as singles as 250 mM $d_6$-DMSO compound stocks on a 384-well Echo-compatible source plate. The library was enriched in 3D fragments and compliant to the Ro3.

The **Enamine Golden fragment library**[436] contained 500 fragments as singles as 100 mM DMSO compounds stocks on 96-well source plates. All fragments fulfilled the following criteria: MW = 140 – 300 Da, HBD <3 and HBA <3, rotatable bonds <3, HACnt = 10 – 21, 1 – 4 rings, ≤2 aromatic rings, ≤3 fused rings and clogP = -0.5 – 3. All compounds are described with 1176 different Bemis-Murcko loose frameworks[437].

## 3.7 Equipment and devices

**Table 9:** List of used devices and tools.

| Equipment / Device, trade name | Manufacturer |
|---|---|
| Acoustic liquid handling system, ECHO 550 | LABCYTE |
| Batch disperser, Polytron® PT 1200 E | POLYTRON |
| Benchtop centrifuge, centrifuge 5810 R | EPPENDORF |
| Centrifugation bottles, Nalgene™, 500 mL | THERMO SCIENTIFIC |
| Centrifuge, Sorvall RC 3BP | THERMO SCIENTIFIC |
| Centrifuge, Avanti J30 I | BECKMAN COULTER |

| | |
|---|---|
| Chromatography system, automated purification, Äktaxpress™ | Amershan Bioscience |
| Chromatography system, SEC, Äkta Avant 25 | GE Healthcare |
| Cryoloops, CrystalCap™ SPINE HT Cryoloops, various sizes | Hampton |
| Cryoloops, Dual-thickness MicroLoops LD™ | MiTeGen, LLC |
| Crystal harvester, Crystal Shifter | Oxford Lab Technologies |
| Crystallization plate, VDX micro plate, 24-well, 18 mm, greased | Hampton Research |
| Crystallization plates, 96-well, 2 drop and 3 drop, SwissCi/MRC | Hampton Research |
| Crystallization plate, 96-well CrystalDirect™ | MiTeGen, LLC |
| Crystallization plate storage and imager, Rock Imager® system | Formulatrix Inc |
| Cuvettes, Cuvettes PS semi-micro | VWR |
| Deepwell block, Masterblock 2 mL sterile 96-well v-shape | Greiner bio-one |
| Dynamic light scattering, DynaPro Plate Reader | Wyatt Technology |
| Electrophoresis chamber, Mini-Protean Tetra-System with PowerPAC™ Basic | BioRad |
| Fermentor, Labfors 5 with fermentor cleaning system LabCIP | Infors HT |
| Spin filtration, Amicon Ultra-15, MWCO 30 kDa | Merck Millipore |
| Filter, Glass fibre prefilter, non-sterile | Merck |
| Filter units, Millipore® Stericup™, 0.10 µm and 0.45 µm, PVDF, 1000 mL | Merck Millipore |
| Filter units, inlet for 2 mL tube 0.22 µM and 0.45 µm | Millipore |
| Flash column chromatography system, CombiFlash® Rf200 | Teledyne Isco |
| Flash column chromatography system, Reveleris® X2 | Grace Materials Technologies |
| French press, EmulsiFlex-C50 | Avestin |
| Gel imager, Gel Doc™ EZ Imager | BioRad |
| HPLC, 1290 Infinity II LC System | Agilent Technologies |
| Incubator, INCU-line | VWR |
| Micro centrifuge, Centrifuge 5415 R | Eppendorf |
| Microwave, Microwave 1700 | Koenig |
| Microwave synthesiser, Initiator™ | Biotage |
| UV-Vis spectrometer, microvolume, NanoDrop™ OneC | Thermo Scientific |
| NMR spectrometer, DPX 401 MHz | Bruker |
| NMR spectrometer, AVANCE™ 500 MHz | Bruker |
| NMR spectrometer, AVANCE™ III HD 600 MHz, with SampleJet™ | Bruker |
| NMR spectrometer, AVANCE™ 600 MHz, quadruple cryoprobe, SampleJet™ | Bruker |
| NMR spectrometer, AVANCE™ 800 MHz, with SampleJet™ | Bruker |
| NMR tubes, SampleJet™ Rack, 96-format for 3 mm tubes | Bruker BioSpin AG |
| Normalizing solution for SPR, BIAnormalizing solution | GE Healthcare |
| PCR detection system, CFX384 real-time PCR detection system | BioRad |
| Photometer, BioPhotometer UV/VIS | Eppendorf |
| Pipettes, Multichannel, 10 µL, 50 µL , 200 µL | Thermo Scientific |
| Pipettes, Research plus®, 2.5 µl, 10 µl, 100 µl, 200 µl and 1000 µl | Eppendorf |
| Pipetting robot for crystallization, Cartesian PixSys 4200 | Genomic Solutions |
| Pipetting robot for crystallization, Mosquito with humidity chamber | TTP Labtech |
| Pipetting robot for reservoir solutions, Formulator | Formulatrix Inc. |
| Pipetting robot, automated, CyBi-well simultaneous pipettor | CyBio |
| Pipetting robot for NMR sample preparation, Freedom evo | Tecan |
| Pipetting system, 96 manual, Liquidator™ | Mettler-Toledo Inc. |

| | |
|---|---|
| Plate sealant crystallization, Adhesive PCR Sealing Foil sheets, aluminium | THERMO SCIENTIFIC |
| Plate sealant crystallization, Crystal clear sealing film | HAMPTON RESEARCH |
| Plate sealant used in DSF, Microseal® B Adhesive sealing film | BIORAD |
| Plate shaker, MixMate® | EPPENDORF |
| Plate, DSF, 384-well, Hard-shell 384 microplate | BIORAD |
| Plate, SPR/NMR, 96-well Greiner plate | GREINER |
| Plate, DLS, 384-well, Corning® 384-well microplate | SIGMA-ALDRICH |
| Preparative LC, AutoPurification™ mass-directed HPLC system | WATERS |
| Sample bags, Whirl-Pak® Stand-up bag | NASCO |
| SDS-PAGE gels, Mini-PROTEAN® TGX™ Precast gels, 4-20%, 15-well | BIORAD |
| Seed bead tube, Seed Bead™ | HAMPTON RESEARCH |
| Sensor ship, Series S Sensor chip SA | GE HEALTHCARE |
| Shaking incubator, Thermomixer comfort 2 mL | EPPENDORF |
| Shaking incubator, Shaker X | KUHNER |
| SPR machine, Biacore T200 | GE HEALTHCARE |
| SPR vials, 4 mm and 7 mm with rubber cap type 3 and 5, respectively | GE HEALTHCARE |
| Supercritical fluid chromatography-MS, Investigator Semi-prep 15 SFC-MS | WATERS |
| UPLC-ESI-Q-TOF-MS, Xevo-G2-S QTof, Zspray™ source, ESI, modular Lockspray™ interface, Acquity™ UPLC system | WATERS |
| Water bath, TW12 | JULABO |

## 3.8 Software

**Table 10:** List of used software.

| Name and version | Source / Reference |
|---|---|
| ChemBioDraw® Ultra, 14.0 | PERKINELMER |
| UNICORN™, version 5.31 | GE HEALTHCARE LIFE SCIENCE |
| Glide, Release 2018-1 | SCHRÖDINGER, LLC |
| IconNMR | BRUKER BIOSPIN |
| MS Office 2016 | MICROSOFT |
| PyMOL, up to version 2.2.3 | SCHROEDINGER, LLC |
| TopSpin, up to version 3.2 | BRUKER |
| TopSpin FBS tool, test version | BRUKER |
| RockMaker, up to version 3.12.4.1 | FORMULATRIX INC. |
| Mnova | MESTRELAB RESEARCH |
| PoseView | University of Hamburg, Accessed via ProteinsPlus, http://proteins.plus/ |
| ProtParam tool | ExPASy web server[438], https://www.expasy.org/ |
| fitKD | NOVARTIS, in-house script by Armin Widmer |
| gedit | Python text editor |
| AnchorQuery™ | University of Pittsburgh, http://anchorquery.csb.pitt.edu/ |

# 4. Methods

## 4.1 Recombinant protein expression and purification

### 4.1.1 Transformation of *E. coli*

For LB agar plates 200 mL autoclaved LB medium were heated for 2 min in a microwave, mixed, cooled down to 60 °C to add 50 µg · $mL^{-1}$ kanamycin and poured into 100 × 15 mm plates. Plates were stapled to reduce condensation while the agar solidified and then, plates were stored at 4 °C and used within two weeks. For enzyme production by overexpression in *E. coli* BL21(DE3), competent bacteria were transformed with the corresponding plasmid. 1 µL of plasmid preparation was added to one aliquot of cells (0.05 mL), gently mixed by tapping, incubated on ice for 30 min, heat shocked for 1 min at 42 °C in a water bath and again incubated on ice for 5 min. 400 µL of SOC medium were added and the cells incubated in a shaking incubator for 1 h at 37 °C at 500 rpm. 150 µL of the solution were plated on an LB agar plate and incubated at 37 °C, overnight. The plate was visually inspected for single colonies, which were either directly picked for a cell preculture or the plate was stored at 4 °C and colonies were picked within two weeks.

### 4.1.2 Expression and purification of FPPS

For the cell culture in a fermentor two precultures were prepared. For preculture I two colonies from LB agar plates were picked, transferred to 2 mL of LB medium containing 50 µg · $mL^{-1}$ kanamycin and were incubated in a shaking incubator at 37 °C for 6 h at 200 rpm. For preculture II, preculture I was transferred to 40 mL MDG medium containing 50 µg · $mL^{-1}$ kanamycin and was further incubated in a shaking incubator overnight at 37 °C and 200 rpm. The next morning, 1.5 L of modified auto induction (modAI) medium were pre-heated in a fermentor to 37 °C, inoculated with preculture II (final $OD_{600}$ around 8) and 0.1% (v/v) PPG antifoam were added. The cell culture was stirred at 37 °C, pH 7.0, $pO_2$ 80% and 900 rpm. The $OD_{600}$ was frequently checked and at an $OD_{600}$ of 10, the temperature was reduced to 18 °C for overnight growth. The next morning, the cell culture (final $OD_{600}$ around 70) was drained to sample bags and cells were harvested by centrifugation at 3566 × *g* for 30 min at 4 °C. Cell pellets were stored at -80 °C until purification.

For purification a cell pellet was thawed and resuspended in 10 mL lysis buffer per 1 g of cell pellet. One EDTA free protease inhibitor tablet (cOmplete™) was added per 100 mL of

solution and the mixture homogenized with a batch disperser. After disruption of the cells by passing them four times through a French press, the cell debris was removed by centrifugation at 15000 × *g* for 60 min at 4 °C and the supernatant filtered through a 0.45 µm filter unit equipped with a glass fibre prefilter.

The filtrate was subjected to immobilized metal affinity chromatography (IMAC) at 4 °C using a chromatography system for an automated multi-step purification processes (ÄKTAxpress™ running on UNICORN™). The filtrate was loaded on a Ni-NTA column (Ni-NTA Superflow 5 mL), eluted with elution buffer and automatically loaded onto a desalting column (HighPrep™ 26/10), which was equilibrated in lysis buffer. The eluate was collected in a falcon tube and incubated overnight for $His_6$-tag cleavage with 0.25 mg HRV 3C. Successful cleavage was confirmed by LC-MS. A reverse IMAC purification step on a second Ni-NTA column (Ni-NTA Superflow 5 mL) yielded the tag-free protein. Fractions were combined according to their purity determined by SDS-PAGE and LC-MS and the protein concentration determined by measurement of the absorbance at 280 nm. The protein solution was concentrated at 2300 × *g* in a spin filtration device (Amicon Ultra-15, MWCO 30 kDa) at 4 °C up to concentration of 30 mg · $mL^{-1}$ and subjected to further purification.

Size exclusion chromatography **(**SEC) was the last polishing purification and buffer exchange step. Depending on the amount of protein to be purified, several runs were made, the fractions analysed by SDS-PAGE and combined accordingly to get one final and homogeneous protein batch. Per run 2 mL to 3 mL concentrated protein solution was injected with a maximum protein amount of 60 mg. The column (HiLoad™ Superdex™ 16/60 S200) was run in SEC buffer at 4 °C with a flow rate of 1 mL · $min^{-1}$. The concentration of the final protein batch was determined by measuring the absorbance at 280 nm, concentrated if necessary by centrifugation in a spin filtration device (Amicon Ultra-15, MWCO 30 kDa) and analysed by SDS-PAGE, HPLC and LC-MS. The solution was flash frozen in small aliquots in liquid nitrogen and stored at -80 °C until further use for up to three years.

### 4.1.3 Expression and purification of $^{13}C^{15}N$-labelled FPPS

For the expression and purification of $^{13}C^{15}N$-labelled protein the same methods and devices have been used as described in chapter **4.1.2**, with the following deviations: For the cell culture 1 L of $^{13}C$,$^{15}N$-uniform labelling medium with 50 µg · $mL^{-1}$ kanamycin was prepared. While sterile filtering through a 0.22 µm filter unit, 10 ml 100 × BME vitamin mix and 0.2 ml 1000 × metals were added. For preculture II 100 mL of the medium were inoculated with preculture I and incubated in a shaking incubator overnight at 37 °C and 200 rpm (final $OD_{600}$ around 6). The remaining 900 mL of medium were stored at 4 °C overnight and were pre-heated in the fermentor to 37 °C the next day and inoculated with preculture II. At an $OD_{600}$ of around 3

the temperature was reduced to 18 °C. For further growth 6 g $^{13}$C-D-(+)-glucose were solved in water and added resulting in a final amount of 1.0% (w/v) $^{13}$C-D-(+)-glucose. The expression was induced 5 min later by 0.5 mM IPTG and the culture stirred for overnight growth (final $OD_{600}$ around 15). The SEC column was either run in SEC buffer or in a BisTris buffer suitable for later NMR measurements. The labelling rate was calculated based on the mass difference of the measured mass, determined by LC-MS, and the theoretically expected mass for 100% deuterated protein. Atomic numbers were calculated by uploading the protein sequence to the web-based ProtParam tool from ExPASy[438].

### 4.1.4 Expression and purification of *in vivo* biotinylated Avi-tagged FPPS

For the expression and purification of *in vivo* biotinylated protein the same methods and devices have been used as described in chapter **4.1.2**, with the following deviations: In addition to 50 µg · mL$^{-1}$ kanamycin, 36 µg · mL$^{-1}$ chloramphenicol were added to all media. The $OD_{600}$ of the cell culture was frequently checked and at an $OD_{600}$ around 3 the expression of BirA was induced with 4 g · L$^{-1}$ L-arabinose and shortly after 200 µM biotin were added. At an $OD_{600}$ of 10 the temperature was reduced to 18 °C for overnight growth.

### 4.1.5 Expression and purification of HRV 3C

For the expression and purification of HRV 3C the same methods and devices have been used as described in chapter **4.1.2**, with the following deviations: Instead of a full purification procedure only one IMAC step was conducted and followed by desalting to lysis buffer. Consequently, the protein was used for $His_6$-tag cleavage with N-terminal MBP-tag and C-terminal $His_6$-tag enabling the separation of this cleavage enzyme from the enzyme of interest by means of reverse IMAC.

### 4.1.6 Protein characterization by mass spectrometry

Mass spectrometry (MS) was performed on protein samples to analyse the status of tag-cleavage, the final protein batch, the $^{13}$C$^{15}$N-labelling or biotinylation rate and protein degradation. For the measurements an UPLC-ESI-Q-TOF-MS system from WATERS was used, composed of a Xevo-G2-S QTof with a Zspray™ source, positive-ion electrospray ionization (ESI) and modular Lockspray™ interface, coupled to an Acquity™ UPLC system. A 10 min standard method designed for proteins was used. Separation in UPLC was done on a reversed-phase column (Acquity UPLC BEH C4, 2.1 mm × 100 mm column, 1,7 µm) running an acetonitrile gradient of

5% to 60% at a flow rate of 0.5 mL · $min^{-1}$ and at a column temperature of 80 °C. A sample volume of 1 µL to 5 µL, containing 0.5 µg to 10 µg protein, were injected. Before and after sample application the system was washed by injecting 10 µL of a cleaning solution. Mass spectra over a mass range from 700 m/z to 3000 m/z were acquired with a deconvolution range from 10 kDa to 150 kDa using maximum entropy (MaxEnt).

### 4.1.7 Protein characterization by SDS-PAGE

SDS-PAGE was used to estimate purity and amount of protein after purification by IMAC and SEC. Samples were mixed with SDS sample buffer, heated to 95 °C for 5 min at 500 rpm in a shaking incubator and loaded to 4% to 20% Mini-PROTEAN TGX Precast Protein Gels (BIORAD). Sample volumes between 2 µL and 8 µL were loaded. For size comparison, 4 µL Precision Plus Protein Kaleidoscope Protein Standard was used. Gels were run for 35 min at 200 V in SDS running buffer in a Mini Protein Tetra system and were afterwards stained overnight in a Coomassie staining solution (Instant Blue™) with subsequent destaining in deionized water (2 × for 1 h). For documentation gels were imaged in a gel imager.

### 4.1.8 Determination of protein concentration

Protein concentration was determined by measuring the absorbance at 280 nm. This was done throughout the purification process for monitoring purposes, for final sample analysis and during buffer exchange and sample concentration steps. The absorbance at 280 nm was measured as 1 Abs = 1 mg · $mL^{-1}$. 2 µL of sample were pipetted onto the measurement pedestal of a microvolume UV-Vis spectrometer (NanoDrop™ OneC). The corresponding sample buffer was used as blank. The protein concentration of the sample was determined by correcting the measured concentration with the corresponding extinction correction factor, as listed in **Table 3**. Triplicates were measured and the mean calculated to minimize the error.

In some cases protein concentration and purity were additionally determined by high pressure liquid chromatography (HPLC). The measurements were performed on a 1290 Infinity II LC System from AGILENT TECHNOLOGIES employing a 100 × 2 mm column packed with POROS R1, 10 µm (DR. MAISCH GMBH) calibrated with BSA. The column ran in 80:20 (v/v) of eluate A and eluate B at a flow rate of 0.8 mL · $min^{-1}$ with maximum pressure of 400 bar. Injections of 30 µL sample were done in duplicates or triplicates. To prevent overloading of the column, the sample was diluted to 100 µg · $mL^{-1}$ in water, based on previous absorbance experiments at 280 nm, resulting in a maximum amount of 3 µg of protein. Before and after sample application the system was washed by injecting 10 µL of a cleaning solution. Data were collected and

processed automatically using MassHunter Walkup Software and ChemStation Rev. B.04.03, AGILENT TECHNOLOGIES. Chromatograms were manually reintegrated if necessary.

## 4.2 Nuclear magnetic resonance spectroscopy

### 4.2.1 General procedures

Proteins where either already stored in NMR buffer or the buffer was exchanged prior to NMR sample preparation either using PD-10 desalting columns according to the manufacturer's gravity protocol or spin filtration devices (Amicon Ultra-15, MWCO 30 kDa). The latter were washed with water and the desired buffer before washing the protein five times with the approx. 10-fold of the own sample volume. Compounds were stored in 90% $d_6$-DMSO and 10% $D_2O$ (v/v) as 100 mM stock. In case of poor solubility 50 mM or 25 mM stocks were prepared. Titration series were also done in 90% $d_6$-DMSO and 10% $D_2O$ (v/v). An amount of 10% $D_2O$ (v/v) was added to all NMR samples. To allow standardization and quality control (chemical shift, signal intensity, line width) of samples and ensure comparability of different samples, DSS was added as an internal standard. For multiple samples a master mix was prepared and dispersed to single samples to minimize differences between the individual samples. Additional to sample tubes containing protein and compound, a protein blank (protein in sample buffer) and a $d_6$-DMSO blank (protein in sample buffer and $d_6$-DMSO equivalent to the amount in sample tubes) were measured. If not stated differently, samples with a sample volume of 170 µL were prepared and measured in 3 mm NMR spine tubes. A pipetting robot (Freedom evo) was used when large numbers of samples were prepared.

Experiments were performed on a BRUKER AVANCE™ III HD 600 MHz spectrometer and a BRUKER AVANCE™ 800 MHz spectrometer, both equipped with a 5 mm triple resonance inverse cryoprobe $^{1}H/^{13}C/^{15}N$ with deuterium lock and z-gradient, operating at a an $^{1}H$ resonance frequency of 600.23 MHz and 800.19 MHz, respectively. $^{19}F$ spectra were collected on a BRUKER AVANCE™ 600 MHz spectrometer, equipped with a 5 mm quadruple cryoprobe $^{1}H/^{19}F/^{13}C/^{15}N$ at an $^{1}H$ resonance frequency of 600.13 MHz. All spectrometers were equipped with a sample changer cooled to 4 °C (SampleJet™) to store samples until acquisition. NMR spectra were acquired at a temperature of 296 K, if not otherwise stated. Prior to every measurement the lock was set to $D_2O$. Then the impedance matching and coil tuning to the sample was done and the magnetic field shimmed. The 90° pulse (p1) was calibrated and when necessary, also the soft pulse for water suppression was calibrated. Usually, 1D $^{1}H$ NMR experiments (zgesgp, 128 scans) were recorded for each sample before and after the actual experiment to monitor sample quality. To determine compound solubility for subsequent experiments, such as SPR, samples with 1 mM

compound were prepared in the corresponding buffer system and 180 µM DSS were added as internal standard. 1D $^{1}$H NMR experiments (zgesgp, 128 scans) were recorded and for solubility estimation, an aromatic signal with known number of protons was integrated and normalized to the DSS peak. Experiments were set up in IconNMR and spectra were analysed in Topspin.

### 4.2.2 Ligand-observed NMR

In a fragment-based screen the Novartis core fragment library was screened at 10 µM protein and 200 µM compound (ratio 1:20) to detect interactions with TcFPPS using waterLOGSY and T1ρ experiments[348, 360a]. The needed amount of compound mixtures for 18.5 µM compound in 180 µL were ordered from the Novartis compound management. For sample preparation, protein was quickly thawed in the hand balm and buffer exchanged to $d_{11}$-TRIS buffer using PD-10 desalting columns. A master solution containing 10 µM protein, 10% $D_2O$ and 150 µM DSS was prepared. With a pipetting robot the master mix was added to the compound mixtures, the samples were mixed and 175 µL transferred to 3 mm NMR spine tubes. For each sample a zgesgp (128 scans), a T1ρ 10 ms, a T1ρ 200 ms (128 scans) and a waterLOGSY (256 scans) experiment were recorded on a BRUKER AVANCE™ III HD 600 MHz spectrometer. Acquisition time were approx. 4 min, 6 min, 7 min and 18 min, respectively, leading to an overall acquisition time of 4.3 d. Spectra of blanks of the compound mixtures in the same buffer system were already available. For data analysis association files of the recorded spectra and the corresponding compound blanks were created in TopSpin and visually inspected. First, the aromatic signals were considered and, if necessary, also the aliphatic signals. For T1ρ experiments the strength of linewidth broadening and loss of intensity was determined as difference of the signal intensities in T1ρ 10 ms and T1ρ 200 ms spectra. A compound was considered as primary fragment hit, when the effect of signal broadening was >20% and the readout in waterLOGSY was also positive. To exclude false-positives, which occurred due to effects in the mixture, the same series of experiments were repeated for all identified fragment hits as singletons at 1 mM compound concentration. Confirmed hits were further employed to protein-observed NMR spectroscopy for validation.

In a second fragment-based screen the Novartis fluorine library was screened at 3.7 µM protein and 18.5 µM compound (ratio 1:5). Interactions with TcFPPS were detected using $^{19}$F CPMG NMR experiments. The needed amount of compound mixtures for 18.5 µM compound in 180 µL were ordered from the Novartis compound management. For compound mixture blanks, a master mix of BisTris buffer with 10% $D_2O$ and 100 µM DSS was prepared and added to the compound mixtures with a pipetting robot. The samples were mixed and 175 µL transferred to NMR tubes. A zgesgp (128 scans), $^{19}$F CPMG 80 ms (512 scans) and a $^{19}$F CPMG 400 ms (512 scans) were recorded of each sample on a BRUKER AVANCE™ 600 MHz spectrometer,

equipped with a quadruple cryoprobe. Acquisition times were approx. 4.8 min, 18.3 min and 21.2 min, respectively, leading to an overall acquisition time of 15 h. For the fragment screen, protein was quickly thawed in the hand balm, filtered (0.45 µm, 4 °C) and the concentration determined by measuring the absorbance at 280 nm. Protein was added to the compound mixture blanks with the pipetting robot (dilution by 2.5%), mixed, and the same NMR experiments were recorded again. For data analysis a test version of the Topspin FBS tool was used. The strength of linewidth broadening and loss of intensity was determined as difference of the signal intensities in $^{19}$F CPMG 80 ms and $^{19}$F CPMG 400 ms spectra. Primary hits with an effect of signal reduction ≥40% were further employed to protein-observed NMR spectroscopy for validation.

### 4.2.3 Protein-observed NMR

Primary fragment hits identified by ligand observed NMR were validated in an orthogonal method by screening uniform $^{13}C^{15}N$-labelled TcFPPS for interactions with these fragments with a 2D protein-observed NMR technique. The tested primary hits derived from previously described fragment screens by ligand-observed NMR of the 6$^{th}$ Novartis core library and the Novartis fluorine library. Compounds were ordered as powder from the Novartis compound archive and stocks prepared according to the general procedure. Validation tests were conducted at a protein concentration of 30 µM and a compound concentration of 1 mM (Novartis core library) and 700 µM (Novartis fluorine library), respectively, recording a zgesgp (512 scans, 10 min) and a [$^{13}$C, $^{1}$H]-SOFAST-HMQC[357a, 357b] (32 scans, 42 min) at 31.85 °C on a Bruker AVANCE™ 800 MHz spectrometer. Prior to sample preparation protein was quickly thawed in the hand balm, filtered (0.45 µm, 4 °C) and the buffer exchanged to BisTris buffer in a spin filtration device (Amicon Ultra-15, MWCO 30 kDa). A master mix containing protein, 10% $D_2O$ and 150 µM DSS was made and samples of 170 µL were prepared by adding the corresponding amount of compound. If chemical shift differences occurred between a sample and the DMSO blank, a primary fragment hit was successfully validated. Compounds were categorized into weak, medium and strong binder according to the number and strength of chemical shift differences.

Protein-observed NMR experiments were also used to test compounds from various sources for their interaction with TcFPPS. This includes: allosteric inhibitors of hFPPS (1), compound analogues (2), primary fragment hits by X-ray crystallography (3) and compounds synthesised in medicinal chemistry campaigns (4). In cases (1) and (2) compounds were ordered from the Novartis compound archive. In case (3) they were purchased from Enamine or abcr GmBH. Stocks were prepared according to the general procedures. For experimental set up, the methods and devices, which have been previously described, were used with the following specifications: Inhibitors to bind to the allosteric pocket of hFPPS (1) were tested at 1 mM. Compound analogues (2) were measured around the $K_d$ concentration of the parental compound.

If necessary the compound stocks were diluted to reach a manually pipettable volume between 1 µL and 2 µL. In addition to a protein blank and a DMSO blank, the starting compound was measured at the selected concentration for better comparison. Primary fragment hits by X-ray crystallography (3) were tested at 700 µM. The majority of compounds synthesised in the medicinal chemistry campaign (4) showed poor solubility in SPR buffer (>1 mM in aqueous buffer). Nevertheless, 700 µM compound were added to the NMR samples to measure at saturation level.

### 4.2.4 $K_d$ determination

To determine the $K_d$ of some compounds that were positively tested in protein-observed NMR, the same experiments, a zgesgp (512 scans, 10 min) and a [$^{13}$C, $^{1}$H]-SOFAST-HMQC (32 scans, 42 min), were performed for a series of samples with constant protein concentration and increasing compound concentrations. A dilution series of the compound stock was prepared and equal volumes of the dilutions were added to protein samples that were derived from a master mix, to ensure comparability by keeping the $d_6$-DMSO and protein concentration on a constant level. Additionally, a protein blank and a $d_6$-DMSO blank were measured to exclude chemical shifts of protein resonances caused by $d_6$-DMSO. Signal shifts were analysed with fitKD. To generate an overlay of spectra in fitKD, an input file with the file locations of the spectra, $d_6$-DMSO blank and the corresponding compound concentrations was generated with gedit. Curve generation by plotting the chemical shift versus the ligand concentration, curve fit and $K_d$ calculations were done automatically by fitKD, based on a series of chemical shifts that were manually selected. The $K_d$ was determined for the $^{1}$H dimension of several signals to check if it is in the same range.

## 4.3 Crystallization at Novartis laboratories

### 4.3.1 General procedures

For crystallization trials at the Novartis laboratories 96-well SwissCi/MCR plates (2-drop) and 24-well VDX micro plates (18 mm, greased) were used, employing the sitting drop vapour diffusion technique and the hanging drop vapour diffusion technique, respectively. Experiments on both plate types were designed in RockMaker and the reservoir solutions were pipetted from stock solutions directly into the plates with a Formulator pipetting robot and mixed manually. A manual pipetting system (Liquidator™) was used to add the reservoir solution to 96-well plates when commercial crystallization screens or other deep well blocks were used.

In general, sitting drops were set up in 96-well plates at nanoliter scale with a Mosquito pipetting robot, which was equipped with a humidity chamber (60% - 70% humidity). Drops of

300 nL protein formulation were pipetted in multi-dispersion mode. Then 100 nL – 200 nL reservoir solution were added in batch mode and when applicable 100 nL – 200 nL seed dilution were added in multi-dispersion mode. Drops were equilibrated against 80 µL reservoir solution. Promising conditions were selected for transfer and optimization in 24-well plates or optimization in 96-well plates. A direct transfer of parameters of the initial hits were tested and parameters, such as precipitant amount, salt concentration, and pH, were also changed in small increments to meet conditions for optimal crystal growth. Only one parameter was changed within a row or column at a time. In general, hanging drops were pipetted manually into the centre of a round cover slide (18 mm, siliconized). Drops of 1.0 µL – 1.2 µL protein formulation were pipetted, 0.5 µL – 1.8 µL reservoir solution were added and if applicable another 0.4 µL seed dilution were added successively to the drop. Drops were equilibrated against 500 µL reservoir solution. Sitting drops were set up as described above. Crystallization plates were set up and incubated at 20 °C. Plates were imaged at regular intervals for a week or in case of screening experiments over a period of 90 d in a Rock Imager system. Images were visually inspected in RockMaker.

Two different types of protein formulations were used: 6.81 mg · $mL^{-1}$ TcFPPS in SEC buffer (50 mM TRIS, pH 8.0, 200 mM NaCl, 2 mM TCEP · HCl), which is referred to as protein formulation I from hereon, or 12.20 mg · $mL^{-1}$ – 12.70 mg · $mL^{-1}$ TcFPPS in low-salt buffer (10 mM TRIS, pH 7.4, 25 mM NaCl, 2 mM TCEP · HCl), which is referred to as protein formulation II. Both protein formulations were stored at -80 °, were quickly thawed in the hand palm and kept on ice until pipetting. The formulation in low-salt buffer was obtained by buffer exchange from TcFPPS stored in SEC buffer. Samples were quickly thawed, filtered (0.45 µm, 4 °C), transferred to a spin filtration device (Amicon Ultra-15, MWCO 30 kDa), which was previously washed with water and low salt buffer, and washed 5 times with the approx. 10-fold of the new sample buffer and finally concentrated. This protein solution was either directly used for plate set up or aliquoted, flash frozen in liquid nitrogen and stored at -80 °C for later usage. Seed dilutions were either used after storage at 4 °C or were quickly thawed in the hand palm when stored at -80 °C, kept on ice and vigorously shaken before pipetting.

### 4.3.2 Screening for crystallization conditions and optimization I

In order to find new crystallization conditions seven commercially available crystallization screens (AmSO4 Suite, Cryos Suite, JCSG+ Suite, MBClass II Suite, Index HT, PegRx HT and SaltRX HT) were tested. Drops of 300 nL protein formulation I and 200 nL reservoir solution (3:2 (v/v)) were set up. In a first round of optimization on 24-well plates (**Table 11**), drops of 1.2 µL protein formulation I and 0.8 µL reservoir (3:2 (v/v)) were pipetted. First crystals appeared after 1 d – 3 d and reached full size after 3 d – 5 d. In a 2$^{nd}$ round of optimization (**Table 11**), drops of 1.0 µL protein formulation I and 0.5 µL reservoir (2:1 (v/v)) were pipetted. In a 3$^{rd}$ round

(**Table 11**), the most promising variations of round 1 of reservoir condition H1 of the Cryos Suite were further optimized. Drops of 1.0 µL protein formulation I and 0.5 µL reservoir (2:1 (v/v)) were pipetted.

**Table 11:** List of conditions transferred and optimized in 24-well plates.

| **Round 1** | | | |
|---|---|---|---|
| Screen | well | Variation along row | pH variation along column |
| Cryos Suite | G11 | ± 5% PEG MME 2000 | 4.4 to 5.0 |
| Cryos Suite | H1 | ± 5% PEG 4000 | 4.4 to 5.0 |
| Index HT | D5 | ± 5% PEG 3350 | 4.4 to 5.0 |
| Index HT | G9 | 20-30 PEG 3350 | 8.1 to 8.7 |
| SaltRX HT | B4 | ± 200 mM ammonium citrate dibasic | 4.4 to 5.0 |
| **Round 2** | | | |
| Screen | well | Variation along row | pH variation along column |
| Cryos Suite | G11 | 10% - 25% PEG MME 2000 | ± 5% glycerol |
| Cryos Suite | H1 | 5% - 20% PEG 4000 | 10% - 20% glycerol |
| SaltRX HT | B4 | none | pH at 5.0, 5.4 and 5.6 |
| **Round 3** | | | |
| Screen | well | composition | |
| Cryos Suite | H1 | 80 mM NaOAc · 3 H2O, pH 4.6, 160 mM $(NH_4)_2SO_4$, 20% PEG 4000, 20% glycerol | |
| Cryos Suite | VarA6 | 80 mM NaOAc · 3 H2O, pH 4.4, 160 mM $(NH_4)_2SO_4$, 25% PEG 4000, 20% glycerol | |
| Cryos Suite | VarD5 | 80 mM NaOAc · 3 $H_2O$, pH 5.0, 160 mM $(NH_4)_2SO_4$, 20% PEG 4000, 20% glycerol | |

### 4.3.3 Seed crystals

Seed crystals were grown on 24-well plates. Drops of 1.0 µL of protein formulation I and 0.5 µL reservoir (160 mM $(NH_4)_2SO_4$, pH 5.0, 80 mM NaOAc, 20% (w/v) PEG 4000 and 20% (v/v) glycerol) (2:1, (v/v)) were pipetted. First crystals appeared after 1 d to 2 d and reached full size after 3 d to 4 d. For seed stock preparation the seed bead method[406] was used. Fresh crystals from two wells were crushed mechanically with a small metal spatula. Obtained seed crystals were added to 100 µL reservoir in a seed bead tube and further crushed by vigorous vortexing. A dilution series of 1:10, 1:100, 1:1000 and 1:10 000 (v/v) (1 mL each) was prepared in seed crystal reservoir. Solutions were stored at 4 °C for several months or aliquoted, flash frozen in liquid nitrogen and stored at -80 °C. Residual reservoir solution was collected and stored at -80 °C to allow further dilutions. When a new seed stock was prepared, it was tested once which dilution led to a high number of wells with 5 to 10 crystals per plate. In most cases dilutions of 1:100 or 1:1000 were used. When a larger amount of seed stock was prepared, crystals from multiple wells were crushed and added to 200 µL reservoir. A seed stock was used until it was used

up or no more crystals grew. TcFPPS seed crystals older than 6 d could not be used for seed stock preparation because they did not deliver a high quality seed stock.

### 4.3.4 Screening for crystallization conditions and optimization II

In a second round of crystallization condition screening in 96-well plates, four commercially available screens (AmSO4 Suite, Cryos Suite, Index HT and SaltRX HT) were rescreened. In the first well, drops of 300 nL protein formulation II and 200 nL reservoir solution (3:2 (v/v)) were set up. In the second well, MMS was applied. Drops of 300 nL of protein formulation II, 200 nL reservoir solution and 100 nL seed stock (first seed stock, chapter **4.3.3**) (3:2:1 (v/v)) were set up.

The reservoir condition 80 mM MES, pH 6.5, 8.5 mM $ZnSO_4$, 19.42% (v/v) PEG MME 550, 15% (v/v) glycerol (well G7, Cryos Suite) was transferred to 24-well plates (Round 1). Drops of 1.2 µL protein formulation II, 0.8 µL reservoir solution and 0.4 µL seed dilution (3:2:1 (v/v)) (chapter **4.3.3**) were set up. Crystals appeared after 1 d and reached full size after 3 d. At a later stage, optimization experiments were started in 96-well plates (Round 2) to overcome problems with occurring precipitate on 24-well plates and to enable fragment screening by X-ray crystallography. The reservoir solution was changed, while all other parameters were left unchanged (300 nL protein formulation II, 200 nL reservoir solution, 100 nL seed dilution, 3:2:1 (v/v)), (**Table 12**, **Optimization**).

**Table 12:** Optimization of reservoir condition in 96-well plates.

| **Optimization** | | |
|---|---|---|
| Variation along row | Variation along column | Additional deviations from initial condition |
| 8% - 20% (v/v) PEG MME 550 | 4 mM - 11 mM $ZnSO_4$ | none |
| 8% - 20% (v/v) PEG MME 550 | 2 mM - 9 mM $ZnSO_4$ | none |
| 8% - 20% (v/v) PEG MME 550 | 7% - 15% (v/v) glycerol | 4 mM $ZnSO_4$ |
| pH 5.8 - 6.8 | 40 mM - 110 mM MES | 4 mM $ZnSO_4$, 13.86% (v/v) glycerol 17.85% (v/v) PEG MME 550 |

| **Selection** | |
|---|---|
| Well | Composition |
| E5 | 80 mM MES, pH 6.5, 4 mM $ZnSO_4$, 12.36% (v/v) PEG MME 550, 11.57% (v/v) glycerol |
| F6 | 80 mM MES, pH 6.5, 4 mM $ZnSO_4$, 13.45% (v/v) PEG MME 550, 12.71% (v/v) glycerol |
| F9 | 80 mM MES, pH 6.5, 4 mM $ZnSO_4$, 16.73% (v/v) PEG MME 550, 12.71% (v/v) glycerol |
| G1 | 80 mM MES, pH 6.5, 4 mM $ZnSO_4$, 17.85% (v/v) PEG MME 550, 13.86% (v/v) glycerol |

Promising conditions (**Table 12**, **Selection**) were reproduced in 96-well plate to investigate reproducibility, the number of wells with crystals per plate and precipitate formation in the crystallization drops. To ensure comparability, all experiments were performed with the same seed dilution. Finally, apo crystals in 96-well plates were grown with a drop ratio of (3:2:1) or (3:1:2). First crystals appeared after 1 d – 3 d and reached full size within 2 d – 3 d. The percentage of wells with crystals per plate ranged from 40% to 95%.

### 4.3.5 Soaking

Validated hits derived from the Novartis core library were subjected to soaking experiments. Fresh and fully grown TcFPPS apo crystals, which were set up in 24-well plates (chapter **4.3.4**, Round 1) were used for soaking. In general, 2 – 3 crystals were transferred to a 10 µL drop of reservoir solution or a mix of protein buffer, reservoir solution and seed buffer (3:2:1, (v/v)) that contained 5 mM to 75 mM compound and 4.5% to 13.5% (v/v) DMSO. Crystals were soaked for 2 min to 24 h. At an early stage 100 mM compound stocks in 90% $d_6$-DMSO and 10% $D_2O$ (v/v) were used to prepare the soaking solutions. At a later stage 250 mM or 500 mM compound stocks were used to increase concentration and/or minimize the amount of DMSO.

Compounds from the medicinal chemistry campaign at the University of Groningen and compounds from the medicinal chemistry campaign at Novartis were also subjected to soaking experiments, using fresh and fully grown TcFPPS apo crystals from 96-well plates (chapter **4.3.4**, Round 2 and 3). Drops of 105 nL of 100 mM or 500 mM compound stock were added by manual pipetting to 600 nL crystallization drops, resulting in 15 mM or 75 mM compound and 13.4% (v/v) DMSO. The stock solution was pipetted to the edge of the drop to minimize the osmotic shock for the crystals. Crystals were soaked for approx. 24 h and in most cases, a back-up crystal was only soaked for approx. 4 h. Some compounds were purified as trifluoracetates and had to be neutralized. Equal amounts of the stock and 90% $d_6$-DMSO and 10% 5 M NaOH (v/v) were mixed and added to the crystallization drop, resulting in a final compound concentration of 37.5 mM.

### 4.3.6 Co-crystallization

Validated hits derived from the Novartis core library were subjected to co-crystallization. This includes hits that (1) were distinct for TcFPPS, (2) formed the intersection of TcFPPS, TbFPPS and hFPPS or (3) formed the intersection with hFPPS. Crystallization drops were set up in 24-well plates as described in chapter **4.3.4** with the following deviations: Instead of 0.8 µL reservoir solution, 0.8 µL of a mix of 100 mM compound stock and reservoir was added. For (1) co-crystallization was conducted at a compound concentration of 2.564 mM (17x compound excess, dilution of 100 mM stock in the reservoir 1:13 (v/v)), 2.5% (v/v) DMSO). For (2) and (3)

experiments were conducted at 5.138 mM (34x compound excess, dilution of the 100 mM stock in the reservoir 2:13 (v/v)), 5.0% (v/v) DMSO). For each compound four crystallization drops were set up.

### 4.3.7 Data collection at the Swiss Light Source

Crystals were manually mounted in cryoloops (CrystalCap™ SPINE HT Cryoloops), flash frozen and stored in liquid nitrogen for data collection. If the reservoir did not contain any cryoprotectant, a backup crystal was mounted after incubation with 2 M $(NH_4)_2SO_4$ and 18% (v/v) glycerol or 2.5 M $LiSO_4$ for 30 s. Diffraction data were collected at 100 K on a Pilatus 6M detector (25 Hz, DECTRIS)[439] at beamline PXII (X10SA) of the Swiss Light Source (SLS), Paul Scherrer Institut, Villigen, Switzerland. The beamline provided monochromatic radiation at a wavelength of 0.99995 Å to 1.00000 Å. For a full dataset, 720 images at 0.25 °, with an exposure time of 0.25 s per image were recorded.

### 4.3.8 Data processing, structure determination and refinement

Diffraction data were indexed and integrated with XDS[391c], release 20180226, and symmetry-related reflections were scaled in AIMLESS[440], release 0.7.2. Further diffraction data processing was carried out using Global Phasing Pipedream automatic pipeline[391a], which used autoPROC[441], version 1.1.7, and Phaser[442], version 2.8.2, for MR. PDB ID 4DWG[170a] was used as search model in MR to solve the first apo TcFPPS structure. For this purpose ligand coordinates were removed from the file. From then on, various in-house models of apo TcFPPS were used as search model, thus indirectly making use of the phases of PDB ID 4DWG. In addition to visual inspection of density maps, statistical data analysis by PanDDA[393b] was used to identify data sets with binding events. For this purpose, PanDDA was run on autoPROC input files and using DIMPLE[443]. Stepwise manual model correction was done in Coot[444], release 0.8.9.1, and the structure was refined using BUSTER[445], version 2.11.7. Ligands were manually fitted into difference electron density and occupancy refinement was done with the help of the geometry module GELLY within BUSTER. $R_{free}$ values[446] were generated from randomly selected 5 % of unique reflections excluded from the refinement. All TcFPPS crystals were in space group $P6_122$ with a monomer per asymmetric unit. X-ray data collection and refinement statistics are summarized in **Table 29** in the Appendix.

### 4.3.9 Data deposition and accession codes

Crystal structures were deposited in the PDB with the status on hold for publication. The file upload was prepared with pdb_extract, version 3.24. The apo protein structure of TcFPPS was deposited under PDB ID 6R04. The structures of TcFPPS in complex with **CS-18** (**JNE**), **CS-33** (**JMN**), **93** (**3N2**), **119** (**GO1**), **MCN-1** (**JMK**), **MCN-4** (**JMT**) and **MCN-4** (**JMW**) were deposited under PDB IDs 6R05, 6R06, 6R07, 6R08, 6R09, 6R0A and 6R0B, respectively.

## 4.4 Crystallization at XChem laboratories

A fragment screen by X-ray crystallography was conducted in collaboration with the PhD student Elliot Nelson in Frank von Delft´s lab at the SGC, Oxford, UK. The fragment screen was performed at beamline I04-1 and associated laboratories of the Diamond Light Source, Harwell, UK, in a one-week on-site stay in October 2017. A workflow for the set-up of high-throughput X-ray screening experiments was implemented on site, called XChem facility[400].

### 4.4.1 Crystallization experiments and fragment screen

All materials needed for crystallization were shipped to the UK either at 4 °C or on dry ice. Shipped material included 12.21 mg · $mL^{-1}$ TcFPPS in low salt buffer (10 mM TRIS, pH 7.4, 25 mM NaCl, 2 mM TCEP · HCl), TcFPPS seed crystal dilution (160 mM $(NH_4)_2SO_4$, pH 5.0, 80 mM NaOAc, 20% (w/v) PEG 4000, 20% (v/v) glycerol) and reservoir buffer (4 mM $ZnSO_4$, 80 mM MES, pH 6.5, 12.36% (v/v) PEG MME 550, 11.57% (v/v) glycerol). The plate set up was conducted at Novartis laboratories (chapter **4.3.1**). In brief: Protein and seed stock dilution were quickly thawed in the hand palm and kept on ice until pipetting. The sitting drop vapour diffusion technique was employed in 2-drop and 3-drop 96-well SwissCi/MRC plates filled with 80 µL and 20 µL reservoir solution, respectively. Drops were set up on seven 2-drop SwissCi/MRC plates and seven 3-drop SwissCi/MRC plates at 20 °C by mixing 300 nL protein solution, 200 nL reservoir solution and 100 nL seed stock dilution using a Mosquito pipetting robot (without humidity chamber). The crystallization plates were incubated at 20 °C and imaged in a Rock Imager system. Crystals appeared after 3 d – 4 d on both plate types with rates of wells with crystals of approx. 40%. Methodology details of the XChem fragment screening platform can be found on its webpage[400], in the literature[447], and is briefly described here: An Echo acoustic liquid handling system was used to transfer individual fragments as multiple 2.5 nL acoustic droplets to crystal drops[448]. To generate the transfer scheme, images of the crystallization plates were visually inspected in TeXRank[449]. Crystallization drops were ranked according to the presence and quality

of crystals and in selected drops a position for the compound transfer, which was as far away from the crystal as possible in order to minimize the osmotic shock when adding the compound by acoustic dispensing[448], was chosen. To access crystal stability to DMSO, soaks with 2.5%, 5.0%, 7.5%, 10%, 15% and 20% (v/v) DMSO for incubation times of approx. 1 h 20 min and 3 h 30 min were conducted and crystal diffraction was tested. Crystals showed unchanged diffraction power for the maximum DMSO amount and soaking time. Fresh crystals were soaked with fragments from the Diamond-SGC Poised library[336] (DSPL) and the Keymical fragments library (KFL) by EDELRIS. For soaking 74.5 mM of the DSPL fragments were delivered to the crystallization drops from 500 mM 100% DMSO stock solution and 37.25 mM of the KFL fragments were delivered from 250 mM 100% DMSO stock solutions (15% DMSO). In many cases the actual compound concentration in the crystallization drops was lower, due to lower compound solubility in the aqueous buffer system. The soaking time ranged from approx. 3 h to 4 h.

### 4.4.2 Data collection at the Diamond Light Source

Without additional cryoprotectant, crystals were mounted in Dual-thickness MicroLoops LD™ (MiTeGen, LLC), which matched the crystal size. Mounting was done in semi-automation at a speed of approx. 60 crystals per hour by using the Crystal Shifter (Oxford Lab Technologies). The crystals were flash frozen and stored in liquid nitrogen for data collection. X-ray diffraction data were collected at 100 K in automated and unattended loop centring mode[447] on a Pilatus 6M-F (25 Hz, Dectris) at beamline I04-1[387a] at the Diamond Light Source, Harwell, UK, running at a fixed wavelength of 0.92 Å.

### 4.4.3 Data processing, structure determination and refinement

During data collection, diffraction data were immediately processed with the Diamond autoprocessing pipeline, which uses xia2,[450] DIALS,[391e] XDS[391c], POINTLESS,[451], DIMPLE[443], REFMAC5[452] and CCP4[453]. Finally, diffraction data were indexed and integrated with XDS[391c], and symmetry-related reflections were scaled in AIMLESS[440], release 0.5.32. Results were displayed in the ISPyB data management system[454]. For MR the structural model of unliganded TcFPPS (in-house) was used (MR with PDB ID 4DWG[170a]). Data were further processed by PanDDA[393b] in XChemExplorer[392a]. In pandda.inspect visual inspection of all events was done in Coot[444a] and 85 ligands were manually modelled into the bound-state models. Ensemble models were generated using the *pandda.export* function. Iterative refinement and manual model building was performed using REFMAC5[452], version 5.8.0189, or Phenix[455], version 1.13_2998, and Coot[444a], respectively. Ligand restraints were generated with AceDRG[456],

Grade[457] and Phenix.elbow[458]. X-ray data collection and refinement statistics are summarized in **Table 35** in the Appendix.

### 4.4.4 Data deposition and accession codes

The coordinates of the bound-state models of 35 structures of TcFPPS in complex with **LT7**, **AWG**, **AWM**, **AWV**, **LUS**, **GQM**, **JGJ**, **LUY**, **M0J**, **LV1**, **LDV**, **GQP**, **LV4**, **LV7**, **LVD**, **LVP**, **LVV**, **JHS**, **LWA**, **JH7**, **LWD**, **JH1**, **AYV**, **LWV**, **LX4**, **MJ4**, **LXA**, **LX7**, **JJM**, **LXJ**, **LXM**, **LXS**, **M0D**, **LZV** and **LZY** have been deposited as a group in the PDB with the status on hold for publication under PDB IDs 5QPD – Z, 5QP0 – 9, 5QPA and 5QPB, respectively. Additionally, a ground state model was deposited under PDB ID 5QPC. In addition, files that document the PanDDA analysis have been made publicly available on Zenodo under DOI 10.5281/zenodo.2649077. For each processed dataset a model of the unbound state, structure factors, an average map for the corresponding resolution bin, a PanDDA Z-map and as many PanDDA event maps as existing. For datasets with a fragment bound, additionally the refined ground state model and bound state model as separate PDB files, restraint files for Phenix and Refmac used for ensemble refinement as well as ligand restraints.

## 4.5 Crystallization at EMBL laboratories

A proposal for remote access to the facility of the High Throughput Crystallization laboratory (HTXlab), EMBL Outstation, Grenoble, France, and beam time at the European Synchrotron Radiation Facility (ESRF) was granted by the iNEXT framework (European Union's framework programme for research and innovation Horizon 2020, grant agreement ID 653706, project number 2847). A fragment screen by X-ray was performed with the help of the web-based Crystallization Information Management System (CRIMS v.4). The author thanks the staff from the HTXlab and the scientist from the ESRF for set up of the crystallization plates and data collection.

### 4.5.1 Crystallization experiments and fragment screen

The crystallization protocol for 2-drop 96-well SwissCi/MRC plates (chapter **4.3.4**) was successfully transferred to CrystalDirect™ plates at Novartis laboratories and was shared with the HTXlab staff. All needed materials were shipped on dry ice to the HTX facility, including 12.21 mg · $mL^{-1}$ TcFPPS in low salt buffer (aliquoted and flash frozen in liquid nitrogen), TcFPPS seed stock and seed dilutions (aliquoted and flash frozen in liquid nitrogen), seed stock buffer and

reservoir solution (4 mM $ZnSO_4$, 80 mM MES, pH 6.5, 12.36% (v/v) PEG MME 550 and 11.57% (v/v) glycerol in a deep well block, 96x 1.8 mL, prepared with a Formulator pipetting robot). The sitting-drop vapour diffusion technique was employed by setting up drops of 150 nL protein solution, 50 nL reservoir solution and 100 nL seed stock dilution in CrystalDirect™ plates using a Cartesian PixSys 4200 crystallization robot. The crystallization plates were incubated at 20 °C in a Rock Imager system and the images were remotely accessed for visual inspection. After 2 d – 3 d crystals grew to full size with 50% - 75% wells per plate that showed crystals. Wells suitable for soaking experiments were selected.

Fresh crystals were soaked with fragments of the Enamine Golden Fragment Library through diffusion, by adding 53 nL 100 mM compound stock solution to a 300 nL crystallization drop using a Cartesian PixSys 4200 pipetting robot. This lead to a final concentration of 10 mM compound and 15% (v/v) DMSO in the drop. Crystals were incubated for 20 h to 24 h.

### 4.5.2 Data collection at the European Synchrotron Radiation Facility

Automated high-throughput crystal harvesting and cryo-cooling was performed with the CrystalDirect™ technology as described elsewhere[386b, 387b] and is briefly described here: After inserting a crystal plate in the system, crystals were located on the plate (scanner, SUNNY TECHNOLOGY), crystallization liquid was aspirated through a small hole in the foil in two intervals à 500 ms and a crystallization pin was glued next to it. A laser (Satsuma femtosecond laser, AMPLITUDE SYSTEMS) excised the film around the crystal and the crystallization pin tip. Finally, the resulting crystal pins were transferred by a robotic arm from the plate to liquid nitrogen for flash freezing and storage until data collection. Diffraction data were collected at 100 K using X-ray centering[390b] mode on a Pilatus3 2M or Pilatus 6M (DECTRIS) at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, at the fully automated MASSIF-1 beamline (ID30A-1)[388b, 390b] or at ID30B[390a], respectively. For a full data set 3600 diffraction images in a 360 ° rotation range with 0.1 ° oscillation steps were collected.

### 4.5.3 Data processing, structure determination and refinement

Diffraction data were processed with the ESRF autoprocessing pipeline, which uses EDNA[459], GrenADES[460], autoPROC[441], XDSAPP[391d], xia2_DIALS[391e, 450]. Results were displayed in the ISPyB data management system[454] and fed into CRIMS, version 4.0. Data were processed by the Global Phasing Pipedream automatic pipeline[391a], which uses autoPROC[441], Phaser[442] for MR and ligand placement with RHOFIT[461]. For MR the structural model of unliganded TcFPPS (in-house) was used (MR with PDB ID 4DWG[170a]). Results were displayed in CRIMS. Data were transferred to Novartis laboratories and were reprocessed as described in

chapter **4.3.8**. X-ray data collection and refinement statistics are summarized in **Table 37** in the Appendix.

### 4.5.4 Data deposition and accession codes

Crystal structures were deposited in the PDB with the status on hold for publication. The file upload was prepared with pdb_extract, version 3.25. The structures of TcFPPS in complex with **HTX-1** (**LEQ**) and **HTX-8** (**LDW**) were deposited under PDB IDs 6SI5 and 6SHV, respectively. Structural models of **HTX-2** to **HTX-7** are described in this work but have not been deposited in the PDB.

## 4.6 Structure aided lead design

### 4.6.1 Virtual screening

For virtual screening the interactive web-based application ANCHOR.QUERY (http://anchorquery.csb.pitt.edu and Koes *et al*.[462]) was used for rational SBLD. Crystal structures of TcFPPS complexes, which resulted from the XChem screen (chapter **4.4**), were used for the query construction. The protein backbone and the ligand were separately loaded and an anchor mimic in the ligand was recognized by the software (ANCHOR.QUERY supports seven residues as starting point for the query: Trp, Tyr, Phe, Val, Leu, Asp and Glu). A pharmacophore query was created by adding additional features of the ligand, e.g. ions, hydrogen donors/acceptors or hydrophobic rings. A rapid pharmacophore search was conducted to screen the library for matching compounds. A root mean square deviation (RMSD) alignment was applied and the aligned poses further refined by minimisation. Additional filters were applied, such as selecting certain reaction types and limiting the MW to 450 Da. The output coordinate files were visually inspected in PyMol with attention to quality of the binding poses and possible interactions of the compounds with adjacent binding site residues.

### 4.6.2 Docking

Virtual inspection and superimposition of various crystal structures of TcFPPS complexes in PyMOL led to a number of candidates that were proposed for synthesis. To evaluate their biding position, they were employed to *in silico* docking.

Compound series **MCN-1** to **MCN-8**: Ligands to be docked were prepared with Schrödinger LigPrep and the protein structure with thiazole **AWM** was prepared with the Protein

Preparation wizard[463] at standard settings. Ligands were then docked with Glide[464] (release 2018-1) using standard precision (SP) docking and default settings. Docking studies were conducted by Rainer Wilcken, NIBR, Novartis Pharma AG, Basel, Switzerland.

Compound **DNDi-1**: The compound was washed, and a single low-energy conformer was generated with RDKit (v2018.09.1), using the MMFF94 force field, according to a previously described procedure[465]. **DNDi-1** was docked with MOE (v2016.08.02) to PDB ID 1YHL[162b]. The protein structure was imported into MOE, and protonated using the LigX tool and Protonate3D, adding explicit hydrogens and performing in situ rigid minimisation, to a gradient of 0.1 kCal/Mol/Angstrom. For docking, 30 placements were performed with triangle-matching, followed by ten rounds of minimisation under Amber10:EHT[466]. The pose with best S-score (-15.99) was chosen for further consideration. Docking studies were conducted by Ryan Byrne, who is a PhD student on the AEGIS project in the group of Prof. Gisbert Schneider at the ETH Zürich, Switzerland.

## 4.7 Medicinal chemistry at the University of Groningen

### 4.7.1 General procedures

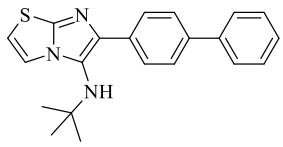
All chemicals and solvents purchased were used without further purification. All isocyanides were kindly provided by Markella Konstantinidou and other group members from the lab of Prof. Alexander Dömling, University of Groningen, the Netherlands. There they were made in-house by either performing the Hoffman or Ugi procedure. All microwave radiation reactions were carried out in a BIOTAGE Initiator™ Microwave Synthesiser. The solvent was removed *in vacuo* and the crude reaction mixture was purified by flash column chromatography (FCC) using a Reveleris® X2 flash chromatography system by GRACE MATERIALS TECHNOLOGIES. Samples were dry loaded on normal phase FlashPure columns (12 g, silica 40 µm irregular, BÜCHI). Thin layer chromatography was performed on FLUKA precoated silica gel plates (0.2 mm thick, particle size 25 µm). According to thin layer chromatography, pure fractions were pooled and the solvent removed *in vacuo*. The final product was washed with TCM and dried under high-vacuum. $^{1}$H and $^{13}$C NMR spectra were recorded on a BRUKER AVANCE™ 500 MHz spectrometer. Chemical shifts δ were reported in parts per million (ppm) and coupling constants *J* in Hertz (Hz). Spin multiplicity was designated as follows: s, singlet; brs, broad singlet; d, doublet; dd, doublet of doublets; ddd, doublet of doublet of doublets; t, triplet; dt, doublet of triplets; dq, doublet of quartets; td triplet of doublets, and m, multiplet. ESI-MS was performed on a WATERS Investigator Semi-prep 15 SFC-MS instrument. All data are consistent with the assigned structures.

### 4.7.2 Synthetic procedures

Synthetic procedure A: Groebke-Blackburn-Bienaymé products; Reactions were carried out at 1 mmol scale. A microwave reaction vial was filled with 1 mL of acetonitrile. Aldehyde (1 mmol, 1 eq), amidine (1 mmol, 1 eq), catalyst (0.2 mmol, 0.2 eq) and isocyanide (1 mmol, 1 eq) were added under stirring. The reaction mixture was subjected to microwave radiation for 1 h at 120 °C. Synthetic procedure B: Tetrazoles by Ugi-4CR; Reactions were carried out at 1 mmol scale. A 2 mL screwcap glass vial, equipped with magnetic stirrer, was filled with 1 mL MeOH. Aldehyde (1 mmol), amine (1 mmol) and isocyanide (1 mmol) were added in this order under stirring at RT. When everything was dissolved, finally trimethylsilyl azide (1 mmol) was added. The vial was closed tightly and the mixture further stirred overnight at RT. Synthetic procedure C: β-lactams; Reactions were carried out at 1 mmol scale. A microwave reaction vial was filled with 1 mL of 2,2,2-trifluoroethanol (TFE). β-amino acid (1 mmol, 1 eq), aldehyde (1 mmol, 1 eq) and isocyanide (1 mmol, 1 eq) were added under stirring. The reaction mixture was subjected to microwave radiation for 1 h at 100 °C.

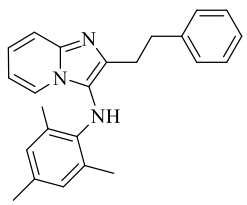
### 4.7.3 Experimental procedures and characterization data

6-([1,1'-biphenyl]-4-yl)-*N*-(*tert*-butyl)imidazo[2,1-*b*]thiazol-5-amine (**MCR-1**)



Synthesised according to synthetic procedure A using [1,1'-biphenyl]-4-carbaldehyde (182 mg, 1.0 mmol), thiazol-2-amine (100 mg, 1.0 mmol), $Sc[OTf]_3$ (98 mg, 0.2 mmol) and 2-isocyano-2-methylpropane (113 µl, 1.0 mmol). FCC in PE:EA (0% – 100% EA in PE). Product eluates at 34% EA. 191 mg of **MCR-1** obtained as off-yellow solid. Yield 55%. $^{1}$H NMR (500 MHz, $CDCl_3$-*d*) δ 8.06 (d, *J* = 8.2 Hz, 2H), 7.67 (t, *J* = 8.1 Hz, 4H), 7.46 (t, *J* = 7.7 Hz, 2H), 7.37 (d, *J* = 4.8 Hz, 2H), 6.71 (d, *J* = 4.5 Hz, 1H), 1.12 (s, 9H). $^{13}$C NMR (126 MHz, $CDCl_3$-*d*) δ 145.67, 140.87, 139.7, 139.23, 134.32, 128.80, 127.50, 127.19, 126.94, 126.82, 125.68, 117.90, 111.52, 55.91, 30.37. Calc. exact mass for $C_{21}H_{21}N_3S$ $[M]^+$ 347.15, SFC found $[M+H]^+$ 348.21.

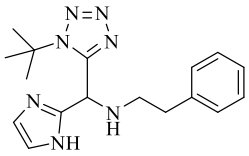
*N*-mesityl-2-phenethylimidazo[1,2-*a*]pyridin-3-amine (**MCR-2**)



Synthesised according to synthetic procedure A using 3-phenylpropanal (131 µl, 1 mmol), pyridin-2-amine (110 mg, 1.0 mmol), $Sc[OTf]_3$ (98 mg, 0.2 mmol) and 2-isocyano-1,3,5-trimethylbenzene (145 mg, 1.0 mmol). FCCin DCM:MeOH (0% – 10% MeOH in DCM). Product eluates at 5% MeOH. 285 mg of **MCR-2** obtained as a brown syrup. Yield 79%. $^{1}$H NMR

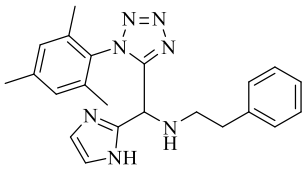
(500 MHz, $CDCl_3$-*d*) δ 7.92 – 7.89 (m, 1H), 7.55 – 7.51 (m, 1H), 7.23 – 7.18 (m, 2H), 7.17 – 7.09 (m, 2H), 6.97 – 6.91 (m, 2H), 6.82 (s, 2H), 6.75 (td, *J* = 6.6 Hz, 0.8 Hz, 1H), 4.71 (s, 1H), 2.83 – 2.78 (m, 2H), 2.77 – 2.72 (m, 2H), 2.27 (s, 3H), 1.91 (s, 6H). $^{13}$C NMR (126 MHz, $CDCl_3$-*d*) δ 142.08, 141.09, 139.04, 138.87, 130.98, 130.02, 128.58, 128.21, 127.13, 125.80, 123.43, 122.18, 122.05, 116.84, 111.84, 35.36, 29.83, 20.50, 18.20 Calc. exact mass for $C_{24}H_{25}N_3$ $[M]^+$ 355.20, SFC found $[M+H]^+$ 356.27.

*N*-((1-(*tert*-butyl)-1*H*-tetrazol-5-yl)(1*H*-imidazol-2-yl)methyl)-2-phenylethan-1-amine (**MCR-3**)



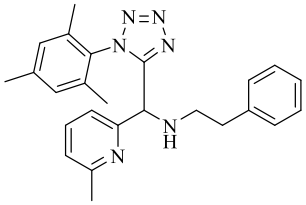
Synthesised according to synthetic procedure B using 1*H*-imidazole-2-carbaldehyde (96 mg, 1.0 mmol), 2-phenylethan-1-amine (126 µl, 1.0 mmol), 2-isocyano-2-methylpropane (113 µl, 1.0 mmol) and TMS azide (131 µl, 1.0 mmol). FCC in DCM:MeOH (0% – 10% MeOH in DCM). Product eluates at 5% MeOH. 169 mg of **MCR-3** obtained as off-yellow solid. Yield 52%. $^1$H NMR (500 MHz, $CDCl_3$-*d*) δ 7.26 – 7.21 (m, 2H), 7.20 – 7.15 (m, 1H), 7.13 – 7.09 (m, 2H), 7.01 (s, 2H), 5.72 (d, *J* = 1.5 Hz, 1H), 2.86 (td, *J* = 7.1 Hz, 6.7 Hz, 2.8 Hz, 2H), 2.74 (td, *J* = 6.9 Hz, 3.3 Hz, 2H), 1.74 (s, 9H). $^{13}$C NMR (126 MHz, $CDCl_3$-*d*) δ 154.14, 144.76, 139.38, 128.72, 128.52, 126.36, 62.44, 52.89, 48.45, 36.20, 29.95 Calc. exact mass for $C_{17}H_{23}N_7$ $[M]^+$ 325.20, SFC found $[M+H]^+$ 326.30.

*N*-((1*H*-imidazol-2-yl)(1-mesityl-1*H*-tetrazol-5-yl)methyl)-2-phenylethan-1-amine (**MCR-4**)



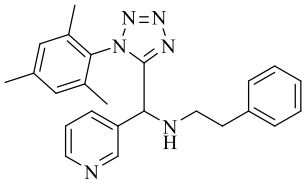
Synthesised according to synthetic procedure B using 1*H*-imidazole-2-carbaldehyde (96 mg, 1.0 mmol), 2-phenylethan-1-amine (126 µl, 1.0 mmol), 2-isocyano-1,3,5-trimethylbenzene (145 mg, 1.0 mmol) and TMS azide (131 µl, 1.0 mmol). FCC in PE:EA (0% – 100% EA in PE). Product eluates at 60% EA. 91 mg of **MCR-4** obtained as brown solid. Yield 24%. $^1$H NMR (500 MHz, $CDCl_3$-*d*) δ 7.24 – 7.19 (m, 2H), 7.17 (d, *J* = 7.1 Hz, 1H), 7.05 (d, *J* = 6.9 Hz, 2H), 6.99 (t, *J* = 6.9 Hz, 4H), 5.20 (s, 1H), 2.88 – 2.74 (m, 2H), 2.66 (td, *J* = 7.0 Hz, 4.0 Hz, 2H), 2.36 (s, 3H), 1.87 (s, 3H), 1.64 (s, 3H). $^{13}$C NMR (126 MHz, $CDCl_3$) δ 155.56, 143.85, 141.39, 139.25, 135.82, 135.48, 129.64, 129.50, 128.73, 128.62, 128.45, 126.27, 51.08, 48.44, 35.87, 21.25, 17.30, 16.97. Calc. exact mass for $C_{22}H_{25}N_7$ $[M]^+$ 387.22, SFC found $[M+H]^+$ 388.28.

*N*-((1-mesityl-1*H*-tetrazol-5-yl)(6-methylpyridin-2-yl)methyl)-2-phenylethan-1-amine (**MCR-5**)



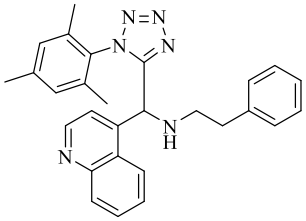
Synthesised according to synthetic procedure B using 6-methyl-pyridine-2-carboxaldehyde (121 mg, 1.0 mmol), 2-phenyl-ethan-1-amine (126 µl, 1.0 mmol), 2-isocyano-1,3,5-trimethyl-benzene (145 mg, 1.0 mmol) and TMS azide (131 µl, 1.0 mmol). FCC in PE:EA (0% – 10% EA in PE). Product eluates at 35% EA. 366 mg of **MCR-5** obtained as off-orange syrup. Yield 89%. $^{1}$H NMR (500 MHz, $CDCl_3$-*d*) δ 7.35 (t, *J* = 7.7 Hz, 1H), 7.27 (s, 1H), 7.17 – 7.13 (m, 2H), 7.10 – 7.06 (m, 1H), 7.05 – 7.02 (m, 2H), 6.90 (dd, *J* = 7.7 Hz, 3.1 Hz, 3H), 6.80 (s, 1H), 4.84 (s, 1H), 2.85 – 2.76 (m, 1H), 2.76 – 2.65 (m, 3H), 2.59 (s, 1H), 2.28 (s, 3H), 2.26 (s, 3H), 1.74 (s, 3H), 1.40 (s, 3H). $^{13}$C NMR (126 MHz, $CDCl_3$) δ 157.94, 156.32, 155.84, 140.58, 139.19, 136.69, 135.71, 134.82, 129.03, 129.00, 128.98, 128.32, 128.08, 125.86, 122.35, 118.78, 58.75, 48.81, 36.04, 23.85, 20.88, 16.91, 16.67. Calc. exact mass for $C_{25}H_{28}N_6$ $[M]^+$ 412.24, SFC found $[M+H]^+$ 423.32.

*N*-((1-mesityl-1*H*-tetrazol-5-yl)(pyridin-3-yl)methyl)-2-phenylethan-1-amine (**MRC-6**)



Synthesised according to synthetic procedure B using pyridine-3-aldehyde (93 µl, 1.0 mmol), 2-phenylethan-1-amine (126 µl, 1.0 mmol), 2-isocyano-1,3,5-trimethylbenzene (145 mg, 1.0 mmol) and TMS azide (131 µl, 1.0 mmol). FCC in PE:EA (0% – 100% EA in PE). Product eluates at 70% EA. 329 mg of **MRC-6** obtained as yellow syrup. Yield 83%. $^{1}$H NMR (500 MHz, $CDCl_3$-*d*) δ 8.38 (dd, *J* = 4.8 Hz, 1.7 Hz, 1H), 8.01 (d, *J* = 2.3 Hz, 1H), 7.47 (dt, *J* = 8.0 Hz, 1.9 Hz, 1H), 7.16 – 7.09 (m, 2H), 7.09 – 7.03 (m, 2H), 7.02 – 6.97 (m, 2H), 6.92 (s, 1H), 6.77 (s, 1H), 4.63 (s, 1H), 2.79 – 2.72 (m, 1H), 2.71 – 2.63 (m, 3H), 2.24 (s, 3H), 1.75 (s, 3H), 1.11 (s, 3H). $^{13}$C NMR (126 MHz, $CDCl_3$-*d*) δ 156.07, 149.67, 148.87, 141.14, 138.94, 135.40, 134.93, 134.56 , 133.09, 129.40, 129.30, 128.36, 128.30, 128.20, 126.02, 123.51, 55.00, 48.47, 35.88, 20.91, 16.93, 16.35. Calc. exact mass for $C_{24}H_{26}N_6$ $[M]^+$ 398.22, SFC found $[M+H]^+$ 399.27.

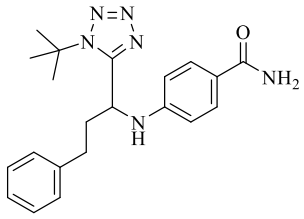
*N*-((1-mesityl-1*H*-tetrazol-5-yl)(quinolin-4-yl)methyl)-2-phenylethan-1-amine (**MCR-7**)



Synthesised according to synthetic procedure B using quinoline-4-carbaldehyde (157 mg, 1 mmol), 2-phenylethan-1-amine (126 µl, 1.0 mmol), 2-isocyano-1,3,5-trimethylbenzene (145 mg, 1.0 mmol) and TMS azide (131 µl, 1.0 mmol). FCC in PE:EA (0% – 100% EA in PE). Product eluates at 50% EA. 280 mg of **MCR-7** obtained as brownish solid. Yield 62%. $^{1}$H NMR (500 MHz, $CDCl_3$-*d*) δ 8.52 (s, 1H), 7.97 (d, *J* = 8.4 Hz, 1H),

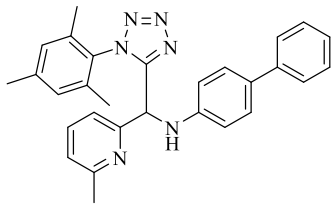
7.72 (s, 1H), 7.60 (t, *J* = 7.7 Hz, 1H), 7.54 (d, *J* = 8.2 Hz, 1H), 7.43 (t, *J* = 7.6 Hz, 1H), 7.17 – 7.05 (m, 4H), 7.02 (d, *J* = 7.1 Hz, 2H), 6.89 (s, 1H), 6.69 (s, 1H), 4.89 (s, 1H), 2.88 – 2.79 (m, 1H), 2.77 – 2.68 (m, 3H), 2.24 (s, 3H), 1.75 (s, 3H), 1.08 (s, 3H). $^{13}$C NMR (126 MHz, $CDCl_3$-*d*) δ 156.07, 149.71, 147.71, 141.11, 139.00, 135.38, 134.68, 134.63, 130.02, 129.81, 129.36, 129.31, 128.91, 128.53, 128.36, 128.24, 127.63, 127.12, 126.90, 126.07, 55.30, 48.57, 35.96, 20.92, 16.97, 16.47. Calc. exact mass for $C_{28}H_{28}N_6$ $[M]^+$ 448.24, SFC found $[M+H]^+$ 449.29.

4-((1-(1-(*tert*-butyl)-1*H*-tetrazol-5-yl)-3-phenylpropyl)amino)benzamide (**MCR-8**)



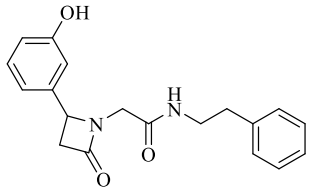
Synthesised according to synthetic procedure B using 3-phenylpropanal (131 µl, 1.0 mmol), 4-aminobenzamide (136 mg, 1.0 mmol), 2-isocyano-2-methylpropane (113 µl, 1.0 mmol) and TMS azide (131 µl, 1.0 mmol). FCC in DCM:MeOH (0% – 100% MeOH in DCM). Product eluates at 6% MeOH. 194.1 mg of **MCR-8** obtained as fluffy white powder. Yield 51%. $^1$H NMR (500 MHz, Chloroform-*d*) δ 7.63 (d, *J* = 8.8 Hz, 2H), 7.29 – 7.24 (m, 2H), 7.20 (t, *J* = 7.3 Hz, 1H), 7.11 (d, *J* = 6.9 Hz, 2H), 6.52 (d, *J* = 8.7 Hz, 2H), 6.33 (d, *J* = 5.4 Hz, 2H), 5.47 – 5.35 (m, 1H), 4.93 (dt, *J* = 10.1 Hz, 6.9 Hz, 1H), 2.84 (dt, *J* = 14.5 Hz, 7.5 Hz, 1H), 2.76 (dt, *J* = 13.9 Hz, 6.8 Hz, 1H), 2.35 (q, *J* = 7.1 Hz, 2H), 1.57 (s, 9H). $^{13}$C NMR (126 MHz, $CDCl_3$-*d*) δ 169.50, 155.75, 149.44, 140.17, 129.51, 128.73, 128.66, 126.61, 123.07, 112.80, 61.95, 48.12, 36.47, 31.92, 30.04. Calc. exact mass for $C_{21}H_{26}N_6O$ $[M]^+$ 378.22, SFC found $[M+H]^+$ 379.27.

*N*-((1-mesityl-1*H*-tetrazol-5-yl)(6-methylpyridin-2-yl)methyl)-[1,1'-biphenyl]-4-amine (**MRC-9**)



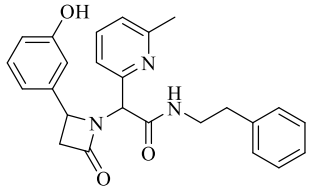
Synthesised according to synthetic procedure B using 6-methylpyridine-2-carboxaldehyde (121 mg, 1 mmol), [1,1' biphenyl]-4-amine (169 mg, 1.0 mmol), 2-isocyano-1,3,5-trimethylbenzene (145 mg, 1.0 mmol) and TMS azide (131 µl, 1.0 mmol). FCC in PE:EA (0% – 100% EA in PE). 453 mg of **MCR-9** obtained as brown syrup. Yield 98%. $^1$H NMR (500 MHz, Chloroform-*d*) δ 7.53 – 7.43 (m, 3H), 7.44 – 7.38 (m, 3H), 7.37 (t, *J* = 7.7 Hz, 2H), 7.30 (d, *J* = 8.0 Hz, 1H), 7.25 – 7.22 (m, 1H), 7.00 (d, *J* = 7.6 Hz, 1H), 6.98 (s, 1H), 6.91 (s, 1H), 6.75 (d, *J* = 8.6 Hz, 2H), 5.95 (d, *J* = 7.0 Hz, 1H), 5.80 (d, *J* = 7.0 Hz, 1H), 2.39 (s, 3H), 2.37 (s, 3H), 1.68 (s, 3H), 1.56 (s, 3H). $^{13}$C NMR (126 MHz, $CDCl_3$) δ 157.92, 156.38, 154.90, 145.15, 140.82, 140.73, 137.30, 136.41, 135.14, 131.29, 129.26, 128.96, 128.57, 127.87, 126.20, 126.15, 122.73, 118.79, 113.70, 53.98, 23.98, 21.14, 17.24, 16.66. Calc. exact mass for $C_{29}H_{28}N_6$ $[M]^+$ 460.24, SFC found $[M+H]^+$ 461.34.

2-(2-(3-hydroxyphenyl)-4-oxoazetidin-1-yl)-*N*-phenethylacetamide (**MRC-10**)



Synthesised according to synthetic procedure C using 3-amino-3-(3-hydroxyphenyl)propanoic acid (181 mg, 1.0 mmol), polyoxymethylene (30 mg, 1.0 mmol) and (2-isocyanoethyl) benzene (138 µl, 1.0 mmol). FCC in DCM:MeOH (0% – 100% MeOH in DCM). Product eluates at 5% MeOH. 303.5 mg of **MRC-10** obtained as orange to brownish syrup. Yield 93%. $^{1}$H NMR (500 MHz, Chloroform-*d*) δ 7.24 – 7.17 (m, 2H), 1.41 – 1.19 (m, 0H), 8.57 (s, 1H), 7.28 (t, *J* = 7.3 Hz, 3H), 7.14 (dd, *J* = 7.0 Hz, 1.7 Hz, 2H), 6.86 (dd, *J* = 8.0 Hz, 2.4 Hz, 1H), 6.79 (t, *J* = 2.1 Hz, 1H), 6.76 – 6.69 (m, 2H), 4.62 (dd, *J* = 5.1 Hz, 2.3 Hz, 1H), 4.10 – 4.03 (m, 1H), 3.50 – 3.38 (m, 3H), 3.34 (dd, *J* = 15.0 HZ, 5.2 Hz, 1H), 2.84 (dd, *J* = 14.9 Hz, 2.4 Hz, 1H), 2.75 (t, *J* = 7.2 Hz, 2H). $^{13}$C NMR (126 MHz, $CDCl_3$-*d*) δ 168.99, 167.67, 157.58, 138.51, 138.44, 130.39, 128.74, 128.66, 126.63, 118.04, 116.21, 113.11, 55.43, 46.53, 44.32, 40.84, 35.25. Calc. exact mass for $C_{19}H_{20}N_2O_3$ $[M]^+$ 324.15, SFC found $[M+H]^+$ 325.24.

2-(2-(3-hydroxyphenyl)-4-oxoazetidin-1-yl)-2-(6-methylpyridin-2-yl)-*N*-phenethylacet amide (**MCR-11**)



Synthesised according to the synthetic procedure C using 3-amino-3-(3-hydroxyphenyl)propanoic acid (181 mg, 1.0 mmol), 6-methyl-picolinaldehyde (121 mg, 1.0 mmol) and (2-isocyano ethyl)benzene (138 µl, 1.0 mmol). FCC in DCM:MeOH (0% – 100% MeOH in DCM). Product eluates at 5% MeOH. 197 mg of racemic **MCR-11** obtained as orange to brownish syrup. Yield 47%. $^{1}$H NMR (500 MHz, $CDCl_3$-*d*, shifts for both enantiomers are given) δ 8.67 (brs, 2H), 7.70 (q, *J* = 6.2 Hz, 2H), 7.54 (td, *J* = 7.7, 1.3 Hz, 1H), 7.37 (td, *J* = 7.8 Hz, 1.8 Hz, 1H), 7.25 – 7.10 (m, 8H), 7.08 (s, 1H), 7.07 (s, 1H), 7.05 (d, *J* = 7.6 Hz, 1H), 7.04 (d, *J* = 7.6 Hz, 1H), 7.02 – 6.95 (m, 3H), 6.93 (brs, 1H) 6.91 (brs, 1H), 6.82 (dd, *J* = 7.9 HZ, 2.4 Hz, 1H), 6.74 (d, *J* = 7.6 Hz, 1H), 6.72 – 6.68 (m, 2H), 6.56 (d, *J* = 7.6 Hz, 1H), 5.24 (s, 1H), 5.19 (s, 1H), 4.74 (dd, *J* = 5.4 Hz, 2.5 Hz, 1H), 4.71 (dd, *J* = 5.4 Hz, 2.6 Hz, 1H), 3.55 (dq, *J* = 13.2 Hz, 6.8 Hz, 1H), 3.47 (dq, 1H), 3.39 – 3.25 (m, 3H), 3.25 – 3.16 (m, 1H), 2.89 (dd, *J* = 14.9 Hz, 2.6 Hz, 1H), 2.81 (dd, *J* = 15.0 Hz, 2.4 Hz, 1H), 2.76 (t, *J* = 7.0 Hz, 2H), 2.53 (td, *J* = 7.0 Hz, 2.9 Hz, 2H), 2.34 (s, 3H), 2.32 (s, 3H).$^{13}$C NMR (126 MHz, $CDCl_3$-*d*, shifts for higher abundant enantiomer are given) *δ* 169.08 (s, 1C), 167.70, 158.10, 157.15, 153.53, 139.55, 138.65, 137.39, 129.60, 128.70, 128.46, 126.36, 122.9, 120.46, 117.90, 116.03, 113.18, 62.15, 55.38, 46.06, 40.96, 35.19, 24.10. Calc. exact mass for $C_{25}H_{25}N_3O_3$ 415.19, SFC found $[M+H]^+$ 416.26.

# 4.8 Medicinal chemistry at Novartis

## 4.8.1 General procedures

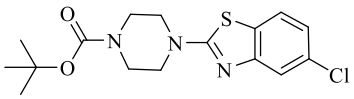
All chemicals and solvents purchased were used without further purification. FCC was performed on a CombiFlash® Rf 200 by TELEDYNE ISCO. Samples were injected directly onto prepacked 12 g RediSep® Rf normal phase silica flash columns. Preparative LC was done on a AutoPurification™ mass-directed HPLC system by WATERS, applying Method A: TFA gradient, Sunfire 30 × 150, C18, 50 mL · min$^{-1}$, $H_2O$+TFA/acetonitrile, 1.0 min 95/5, 11.0 min 20/80, 11.1 min 0/100, 14.0 min 0/100 or Method B: $NH_4HCO_3$ gradient, XBridge 30 × 150, C18, 50 mL · min$^{-1}$, $H_2O$+$NH_4HCO_3$/acetonitrile, 1.0 min 95/5, 11.0 min 20/80, 11.1 min 0/100, 14.0 min 0/100. Evidence of the chemical structure of synthesised compounds was provided by analytical data including NMR spectra and high resolution MS. $^1$H NMR spectra of precursors and intermediates were recorded on a Bruker DPX 401 MHz. $^1$H and $^{13}$C NMR spectra of final products were recorded on a BRUKER AVANCE™ 600 MHz spectrometer equipped with a 5 mm TXI probe ($^1$H/$^{13}$C/$^{15}$N) with deuterium lock and triple-axis. Chemical shifts δ were reported in ppm. Multiplicity was designated as followed: s, singlet; brs, broad singlet; d, doublet; dd, doublet of doublets; ddd, doublet of doublet of doublets; t, triplet; dt, doublet of triplet; td triplet of doublets; and m, multiplet. Purity analysis and mass spectra were performed on a WATERS Acquity UPLC/SQD MS (ESI +/-). High resolution mass spectrometry of final products was performed on an Ultimate 3000 UHPLC by coupled to a Q Exactive Plus mass spectrometer by THERMO SCIENTIFIC using electrospray ionisation in positive ion modus. The high mass accuracy below 1.5 ppm was obtained by using a lock mass. The elemental composition was derived from the mass spectra acquired at the high resolution of about 35’000. All data were consistent with the assigned structures.

## 4.8.2 General synthetic procedure for amination of 2-chlorobenzothiazoles

For the amination of 2-chlorobenzothiazoles a sustainable chemistry approach described by Kumar *et al.*[467] was used. Unlike classical reaction conditions, 2-chlorobenzthiazoles and amines are taken in water and stirred at RT or up to 100 °C.

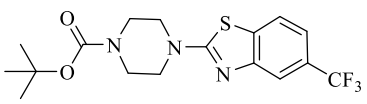
### 4.8.3 Experimental procedures and characterization data

tert-butyl 4-(5-(trifluoromethyl)benzo[d]thiazol-2-yl)piperazine-1-carboxylate (**MCN-S1**)



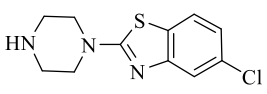
2,5-dichlorobenzo[d]thiazole (400 mg, 1.96 mmol, 1.0 Eq) and tert-butyl piperazine-1-carboxylate (730 mg, 3.92 mmol, 2.0 Eq) were taken in water (7.5 mL) and stirred at RT overnight. Product formation was monitored by LC-MS. After additional stirring for 6 h at 80 °C, the reaction mixture was worked up with EA (2x 10 mL) and 0.1 M HCl. Combined organic phases were dried over anhydrous $Na_2SO_4$, concentrated and purified by FCC (Heptane:EA, product eluates at 20% EA). Concentrating the pure fractions yielded 538 mg of the product as white solid with 92% purity (yield 71%). $^1$H NMR (400 MHz, DMSO-$d_6$) δ 7.81 (d, $J$ = 8.4 Hz, 1H), 7.49 (d, $J$ = 2.0 Hz, 1H), 7.11 (dd, $J$ = 8.4 Hz, 2.1 Hz, 1H), 3.58 (dd, $J$ = 6.6 Hz, 3.7 Hz, 4H), 3.49 (dd, $J$ = 6.5 Hz, 3.7 Hz, 4H), 1.43 (s, 9H). Calc. exact mass for $C_{16}H_{20}ClN_3O_2S$ 353.10, LC-MS found $[M+H]^+$ 354.2.

tert-butyl 4-(5-(trifluoromethyl)benzo[d]thiazol-2-yl)piperazine-1-carboxylate (**MCN-S2**)



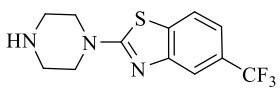
2-chloro-5-(trifluoromethyl)benzo[d]thiazole (500 mg, 2.10 mmol, 1.0 Eq) and tert-butyl piperazine-1-carboxylate (784 mg, 4.20 mmol, 2.0 Eq) were taken in water (7.5 mL) and stirred at RT overnight. Product formation was monitored by LC-MS. After additional stirring for 6 h at 80 °C, the reaction mixture was worked up with EA (2x 10 mL) and 0.1 M HCl. Combined organic phases were dried over anhydrous $Na_2SO_4$ and concentrated. 628.4 mg of the product were obtained as white solid with 99% purity (yield 76%). $^1$H NMR (400 MHz, DMSO-$d_6$) δ 8.03 (d, $J$ = 8.2 Hz, 1H), 7.73 (s, 1H), 7.39 (d, $J$ = 8.3 Hz, 1H), 3.62 (dd, $J$ = 6.3 Hz, 4.0 Hz, 4H), 3.58 – 3.46 (m, 4H), 1.44 (s, 9H). Calc. exact mass for $C_{17}H_{20}F_3N_3O_2S$ 387.12, LC-MS found $[M+H]^+$ 388.3.

5-chloro-2-(piperazin-1-yl)benzo[d]thiazole (**MCN-S3**)



tert-butyl 4-(5-chlorobenzo[d]thiazol-2-yl)piperazine-1-carboxylate (**MCN-S1**), 534 mg, 1.51 mmol) was taken in 8 mL 4.0 M HCl in dioxan and stirred overnight at RT. Concentrating the reaction mixture yielded 445 mg of the HCl salt of the product as white solid with 97% purity (yield 99%). $^1$H NMR (400 MHz, DMSO-$d_6$, as HCl salt) δ 9.39 (brs, 2H), 7.85 (d, $J$ = 8.4 Hz, 1H), 7.53 (d, $J$ = 2.1 Hz, 1H), 7.16 (dd, $J$ = 8.4 Hz, 2.1 Hz, 1H), 3.82 (t, $J$ = 5.4 Hz, 4zH), 3.30 – 3.22 (m, 4H). Calc. exact mass for $C_{11}H_{12}ClN_3S$ 253.75, LC-MS found $[M+H]^+$ 254.2.

2-(piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole (**MCN-S4**)



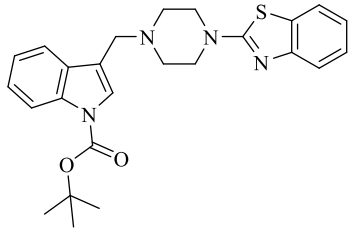
tert-butyl 4-(5-(trifluoromethyl)benzo[d]thiazol-2-yl)piperazine-1-carboxylate (**MCN-S2**), 629.4 mg, 1.62 mmol) was taken in 8 mL 4.0 M HCl in dioxan and stirred at RT for 4 h. Concentrating the reaction mixture yielded 534 mg of the HCl salt of **MCN-S4** as white solid with 97% purity (yield 99%). $^1$H NMR (400 MHz, DMSO-$d_6$, as HCl salt) δ 9.60 (brs, 2H), 8.09 (d, $J$ = 8.2 Hz, 1H), 7.78 (s, 1H), 7.44 (dd, $J$ = 8.3 Hz, $J$ = 1.1 Hz, 1H), 3.93 – 3.84 (m, 4H), 3.29 – 3.23 (m, 4H). Calc. exact mass for $C_{12}H_{12}F_3N_3S$ 287.07, LC-MS found $[M+H]^+$ 288.2.

### 4.8.4 General synthetic procedure for the reductive amination of aldehydes

The reductive amination of aldehydes with sodium triacetoxyborohydride used here are similar to standard conditions[468]. Reaction times were chosen in accordance with Jeankumar *et al*.[469] To a solution of 2-piperazinebenzothiazole (1.0 mmol) and the corresponding indoleacetaldehyde (1.1 mmol) in dry DCM (2 mL) under argon atmosphere catalytic amounts of acetic acid were added. The reaction mixture was stirred at RT for 6 h and cooled to 0 °C. Sodium triacetoxyborohydride (1.5 mmol) was added and the stirring continued at RT overnight. The reaction mixture was extracted by further dilution with DMC (5 mL) and water (5 mL). The aqueous phase was back-extracted with DCM (2x 10 mL). If needed brine solution was added. Combined organic phases were dried over anhydrous $Na_2SO_4$, filtered, concentrated *in vacuo*, resolved in acetonitrile:$H_2O$ (9:1) and purified by preparative LC-MS using either method A or method B, yielding the TFA salt or free base, respectively. When using the HCl salts of the 2-piperazinebenzothiazole (1.0 mmol) deprotonation was done in-situ by adding TEA (1.5 mmol). After stirring for 30 min at RT, the corresponding indoleacetaldehyde (1.1 mmol) and acetic acid were added (7.0 mmol). Subsequent steps were done as previously described.

### 4.8.5 Experimental procedures and characterization data of benzothiazole series

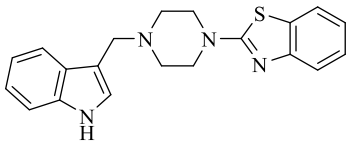
tert-butyl 3-((4-(benzo[d]thiazol-2-yl)piperazin-1-yl)methyl)-1H-indole-1-carboxylate (**MCN-S5**)



The compound was synthesised according to the above general procedure using tert-butyl 3-formyl-1H-indole-1-carboxylate (50 mg, 0.204 mmol, 1.0 Eq), 2-(piperazin-1-yl)benzo[d]thiazole (49.2 mg, 0.224 mmol, 1.1 Eq) and sodium triacetoxyborohydride (64.8 mg, 0.306 mmol, 1.5 Eq). The combined, dried, filtered and *in vacuo* concentrated organic fractions yielded 92 mg of the TFA salt of **MCN-S5** as brownish

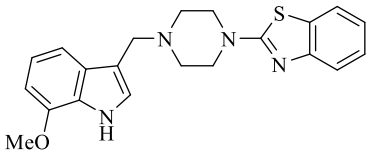
coloured syrup with a purity of 92% (yield 93%). Calc. exact mass for $C_{25}H_{28}N_4O_2S$ 448.19, LC-MS found $[M+H]^+$ 449.4. No further analytics done.

2-(4-((1H-indol-3-yl)methyl)piperazin-1-yl)benzo[d]thiazole (**MCN-1**)



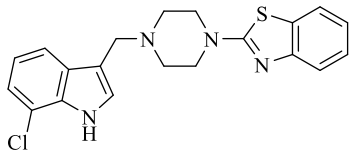
tert-butyl 3-((4-(benzo[d]thiazol-2-yl)piperazin-1-yl) methyl)-1H-indole-1-carboxylate (**MCN-S5**), as TFA salt, 92 mg, 0.204 mmol) was solved in a mixture of 1 mL DCM and 1 mL TFA and stirred at RT for 1h. The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 66.5 mg of the TFA salt of **MCN-1** as a red solid with 89% purity (yield 63%). $^1$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.57 (d, *J* = 2.6 Hz, 1H), 10.11 (s, 1H), 7.88 – 7.72 (m, 2H), 7.59 (d, *J* = 2.6 Hz, 1H), 7.51 (dd, *J* = 8.1 Hz, 1.1 Hz, 1H), 7.46 (d, *J* = 8.1 Hz, 1H), 7.32 (ddd, *J* = 8.3 Hz, 7.7 Hz, 1.3 Hz, 1H), 7.18 (ddd, *J* = 8.1 Hz, 6.9 Hz, 1.2 Hz, 1H), 7.15 – 7.10 (m, 2H), 4.57 (s, 2H), 4.21 (d, *J* = 14.1 Hz, 2H), 3.55 (d, *J* = 12.4 Hz, 2H), 3.46 (t, *J* = 13.7 Hz, 2H), 3.34 – 3.17 (m, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 167.56, 151.98, 136.02, 130.86, 128.88, 127.38, 126.20, 121.94, 121.85, 121.47, 119.72, 119.04, 118.59, 111.97, 102.31, 50.67, 49.29, 45.06. Calc. exact mass for $C_{20}H_{20}N_4S$ 348.15, FTMS found $[M+H]^+$ 349.15.

2-(4-((7-methoxy-1H-indol-3-yl)methyl)piperazin-1-yl)benzo[d]thiazole (**MCN-2**)



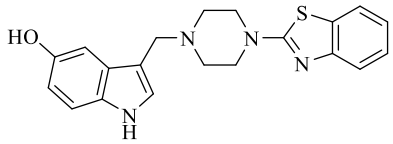
The compound was synthesised according to the above general procedure using 7-methoxy-1H-indole-3-carb-aldehyde (50 mg, 0.285 mmol, 1.0 Eq), 2-(piperazin-1-yl)benzo[d]thiazole (68.9 mg, 0.314 mmol, 1.1 Eq) and sodium triacetoxyborohydride (91 mg, 0.428 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified twice by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 41.4 mg of the TFA salt of **MCN-2** as an off-white solid with 100% purity (yield 30%). $^1$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.68 (d, *J* = 2.8 Hz, 1H), 10.15 (brs, 1H), 7.82 (dd, *J* = 7.9 Hz, 1.2 Hz, 1H), 7.55 – 7.45 (m, 2H), 7.37 (d, *J* = 8.0 Hz, 1H), 7.32 (ddd, *J* = 8.3 Hz, 7.3 Hz, 1.3 Hz, 1H), 7.13 (td, *J* = 7.6 Hz, 1.2 Hz, 1H), 7.05 (t, *J* = 7.8 Hz, 1H), 6.74 (d, *J* = 7.7 Hz, 1H), 4.54 (s, 2H), 4.20 (d, *J* = 14.0 Hz, 2H), 3.93 (s, 3H), 3.64 – 3.50 (m, 2H), 3.45 (t, *J* = 12.8 Hz, 2H) , 3.27 – 3.24 (m, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 167.55, 151.98, 146.35, 130.85, 128.96, 128.34, 126.19, 126.14, 121.93, 121.46, 120.46, 119.04, 111.22, 102.85, 102.24, 55.24, 50.74, 49.26, 45.05. Calc. exact mass for $C_{21}H_{22}N_4OS$ 378.15, FTMS found $[M+H]^+$ 379.16.

2-(4-((7-chloro-1H-indol-3-yl)methyl)piperazin-1-yl)benzo[d]thiazole (**MCN-3**)



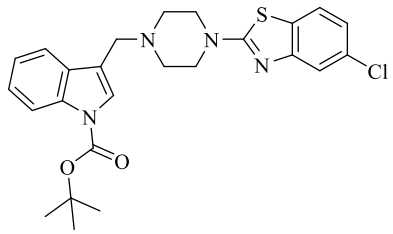
The compound was synthesised according to the above general procedure using 7-chloro-1H-indole-3-carbaldehyde (50 mg, 0.278 mmol, 1.0 Eq), 2-(piperazin-1-yl)benzo[d]thiazole (67.2 mg, 0.306 mmol, 1.1 Eq) and sodium triacetoxyborohydride (89 mg, 0.418 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 101.8 mg of the TFA salt of **MCN-3** as an off-white solid with 96% purity (yield 71%). $^{1}$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.96 (d, *J* = 2.8 Hz, 1H), 10.31 (brs, 1H), 7.87 – 7.78 (m, 2H), 7.66 (d, *J* = 2.7 Hz, 1H), 7.51 (dd, *J* = 8.1 Hz, 1.1 Hz, 1H), 7.32 (ddd, *J* = 8.3 Hz, 7.3 Hz, 1.3 Hz, 1H), 7.27 (dd, *J* = 7.5 Hz, 0.8 Hz, 1H), 7.14 (q, *J* = 7.9 Hz, 2H), 4.58 (s, 2H), 4.20 (d, *J* = 14.0 Hz, 2H), 3.54 (d, *J* = 12.5 Hz, 2H), 3.49 (s, 2H), 3.33 – 3.15 (m, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 167.54, 151.99, 132.86, 130.86, 130.12, 129.39, 126.19, 121.93, 121.46, 121.39, 120.79, 119.03, 117.84, 116.30, 103.84, 50.37, 49.34, 45.07. Calc. exact mass for $C_{20}H_{19}N_4ClS$ 382.10, FTMS found $[M+H]^+$ 383.11.

3-((4-(benzo[d]thiazol-2-yl)piperazin-1-yl)methyl)-1H-indol-5-ol (**MCN-4**)



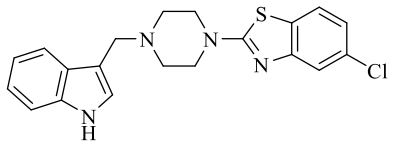
The compound was synthesised according to the above general procedure using 5-hydroxy-1H-indole-3-carbaldehyde (64.7 mg, 0.401 mmol, 1.0 Eq), 2-(piperazin-1-yl)benzo[d] thiazole (80 mg, 0.365 mmol, 1.0 Eq) and sodium triacetoxy-borohydride (116.0 mg, 0.547 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified twice by preparative LC-MS (method B). Freeze-drying of the pure fractions yielded 32 mg of **MCN-4** as a white solid with 92% purity (yield 22%). $^{1}$H NMR (600 MHz, DMSO-$d_6$) δ 10.67 – 10.62 (m, 1H), 8.60 (brs, 1H), 7.76 – 7.69 (m, 1H), 7.44 (d, *J* = 7.7 Hz, 1H), 7.29 – 7.23 (m, 1H), 7.14 (d, *J* = 8.7 Hz, 2H), 7.09 – 7.02 (m, 1H), 6.97 (d, *J* = 2.3 Hz, 1H), 6.60 (dd, *J* = 8.6 Hz, 2.3 Hz, 1H), 3.60 (s, 2H), 3.56 – 3.49 (m, 4H) 3.34 (brs, 2H), 2.53 – 2.50 (m, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$) δ 168.05, 152.48, 150.30, 130.90, 130.37, 128.29, 125.95, 125.27, 121.18, 121.14, 118.53, 111.65, 111.35, 109.44, 103.03, 53.32, 51.80, 48.18. Calc. exact mass for $C_{20}H_{20}ON_4S$ 364.14, FTMS found $[M+1]^+$ 365.14.

tert-butyl 3-((4-(5-chlorobenzo[d]thiazol-2-yl)piperazin-1-yl)methyl)-1H-indole-1-carboxylate (**MCN-S6**)



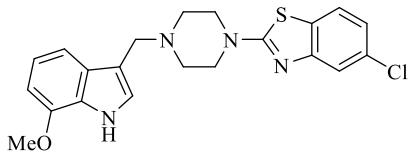
The compound was synthesised according to the above general procedure using 5-chloro-2-(piperazin-1-yl)benzo[d]thiazole as HCl salt (**MCN-S3**) (70.0 mg, 0.241 mmol, 1.00 Eq), TEA (50 µl, 0.362 mmol, 1.50 Eq), tert-butyl 3-formyl-1H-indole-1-car-boxylate (65.1 mg, 0.265 mmol, 1.10 Eq), acetic acid (100 µl, 1.747 mmol, 7.24 Eq) and sodium triacetoxyborohydride (77.0 mg, 0.362 mmol, 1.50 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 86.9 mg of the TFA salt of **MCN-S6** as a white solid with 92% purity (yield 56%). $^{1}$H NMR (400 MHz, DMSO-$d_6$, as TFA salt) δ 10.17 (s, 1H), 8.11 (d, $J$ = 8.2 Hz, 1H), 7.96 (s, 1H), 7.87 (dd, $J$ = 10.6 Hz, 8.0 Hz, 2H), 7.54 (d, $J$ = 2.1 Hz, 1H), 7.51 – 7.25 (m, 2H), 7.17 (dd, $J$ = 8.4 Hz, 2.1 Hz, 1H), 4.57 (s, 2H), 4.32 – 4-13 (m, 4H), 3.64 – 3.39 (m, 4H), 1.65 (s, 9H). Calc. exact mass for $C_{25}H_{27}ClN_4O_2S$ 482.59, LC-MS found $[M+H]^+$ 483.3.

2-(4-((1H-indol-3-yl)methyl)piperazin-1-yl)-5-chlorobenzo[d]thiazole (**MCN-5**)



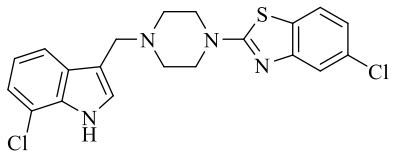
tert-butyl 3-((4-(5-chlorobenzo[d]thiazol-2-yl)piperazin-1-yl) methyl)-1H-indole-1-carboxylate (**MCN-S6**), as TFA salt, (92 mg, 0.204 mmol) was solved in a mixture of 1 mL DCM and 1 mL TFA and stirred at RT for 1h. The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 51.6 mg of the TFA salt of **MCN--5** as a white solid with 99% purity (yield 61%). $^{1}$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.57 (d, $J$ = 2.7 Hz, 1H), 10.16 (s, 1H), 7.86 (d, $J$ = 8.4 Hz, 1H), 7.80 (d, $J$ = 7.8 Hz, 1H), 7.59 (d, $J$ = 2.6 Hz, 1H), 7.54 (d, $J$ = 2.1 Hz, 1H), 7.46 (dd, $J$ = 8.1 Hz, 0.9 Hz, 1H), 7.21 – 7.15 (m, 2H), 7.12 (ddd, $J$ = 8.0 Hz, 6.9 Hz, 1.1 Hz, 1H), 4.57 (s, 2H), 4.25 – 4.16 (m, 2H), 3.55 (d, $J$ = 12.4 Hz, 2H), 3.48 (t, $J$ = 13.4 Hz, 2H), 3.27 (brs, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 169.09, 153.24, 136.02, 130.87, 129.58, 128.88, 127.37, 122.84, 121.85, 121.65, 119.72, 118.58, 118.37, 111.97, 102.28, 50.66, 49.25, 44.99. Calc. exact mass for $C_{20}H_{19}ClN_4S$ 382.10, FTMS found $[M+1]^+$ 383.11.

5-chloro-2-(4-((7-methoxy-1H-indol-3-yl)methyl)piperazin-1-yl)benzo[d]thiazole (**MCN-6**)



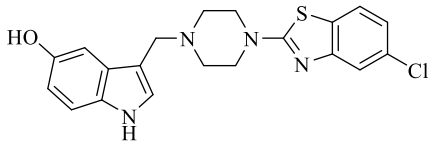
The compound was synthesised according to the above general procedure using 7-methoxy-1H-indole-3-carbaldehyde (85 mg, 0.482 mmol, 2.0 Eq), 5-chloro-2-(piperazin-1-yl)benzo[d]thiazole (**MCN-S3**), HCl salt, 70 mg, 0.241 mmol, 1.0 Eq) and sodium triacetoxyborohydride (128 mg, 0.603 mmol, 2.5 Eq). The mixture was stirred for 36 h. The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 72.3 mg of the TFA salt of **MCN-6** as a yellow solid with 85% purity (yield 48%). $^{1}$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.68 (d, *J* = 2.7 Hz, 1H), 10.13 (brs, 1H), 7.85 (d, *J* = 8.4 Hz, 1H), 7.53 (d, *J* = 2.1 Hz, 1H), 7.48 (d, *J* = 2.7 Hz, 1H), 7.37 (d, *J* = 8.0 Hz, 1H), 7.16 (dd, *J* = 8.5 Hz, 2.1 Hz, 1H), 7.05 (t, *J* = 7.8 Hz, 1H), 6.74 (d, *J* = 7.7 Hz, 1H), 4.54 (s, 2H), 4.20 (d, *J* = 14.1 Hz, 2H), 3.93 (s, 3H), 3.58 – 3.42 (m, 4H), 3.25 (brs, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 169.08, 153.24, 146.35, 130.87, 129.58, 128.95, 128.36, 126.14, 122.84, 121.65, 120.47, 118.37, 111.22, 102.82, 102.24, 55.24, 50.75, 49.24, 44.98. Calc. exact mass for $C_{21}H_{21}ClN_4OS$ 412.11, FTMS found $[M+1]^+$ 413.12.

5-chloro-2-(4-((7-chloro-1H-indol-3-yl)methyl)piperazin-1-yl)benzo[d]thiazole (**MCN-7**)



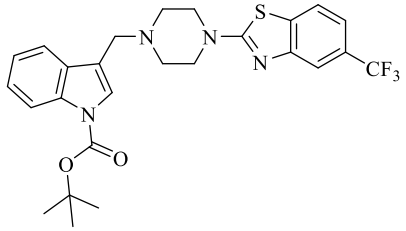
The compound was synthesised according to the above general procedure using 5-chloro-2-(piperazin-1-yl)benzo[d]thiazole as HCl salt (**MCN-S3**) (70 mg, 0.241 mmol, 1.0 Eq), TEA (67 µl, 0.482 mmol, 2.0 Eq), 7-chloro-1H-indole-3-carbaldehyde (47.7 mg, 0.265 mmol, 1.1 Eq), acetic acid (97 µl, 1.688 mmol, 7.0 Eq) and sodium triacetoxy-borohydride (77.0 mg, 0.362 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 57.9 mg of the TFA salt of **MCN-7** as an off-white solid with 98% purity (yield 44%). $^{1}$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.95 (s, 1H), 10.31 (brs, 1H), 7.86 (d, *J* = 8.4 Hz, 1H), 7.79 (d, *J* = 8.0 Hz, 1H), 7.66 (d, *J* = 2.7 Hz, 1H), 7.54 (d, *J* = 2.1 Hz, 1H), 7.27 (d, *J* = 7.5 Hz, 1H), 7.19 – 7.11 (m, 2H), 4.57 (s, 2H), 4.27 – 4.14 (m, 2H), 3.65 – 3.33 (m, 4H), 3.26 (brs, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 169.08, 153.25, 132.87, 130.86, 130.12, 129.57, 129.38, 122.83, 121.64, 121.39, 120.78, 118.35, 117.84, 116.29, 103.8, 50.39, 49.32, 45.00. Calc. exact mass for $C_{20}H_{18}Cl_2N_4S$ 416.06, FTMS found $[M+1]^+$ 417.07.

3-((4-(5-chlorobenzo[d]thiazol-2-yl)piperazin-1-yl)methyl)-1H-indol-5-ol (**MCN-8**)



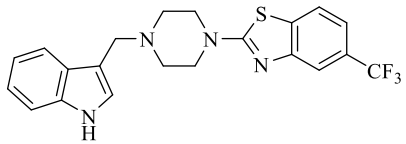
The compound was synthesised according to the above general procedure using 5-chloro-2-(piperazin-1-yl)benzo [d]thiazole as HCl salt (**MCN-S3**), 70.0 mg, 0.241 mmol, 1 Eq), TEA (67 µl, 0.482 mmol, 2.0 Eq), 5-hydroxy-1H-indole-3-carbaldehyde (42.8 mg, 0.265 mmol, 1.1 Eq), acetic acid (97 µl, 1.688 mmol, 7.0 Eq) and sodium triacetoxyborohydride (128.0 mg, 0.603 mmol, 2.5 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method B). Freeze-drying of the pure fractions yielded 2.37 mg of **MCN-8** as an off-white solid with 88% purity (yield 2%). Poor solubility and difficult phase separation during extraction caused a very low yield. Therefore no NMR spectra have been recorded. Calc. exact mass for $C_{20}H_{19}ClN_4OS$ 398.10, LC-MS found $[M+1]^+$ 399.2.

tert-butyl 3-((4-(5-(trifluoromethyl)benzo[d]thiazol-2-yl)piperazin-1-yl)methyl)-1H-indole-1-carboxylate (**MCN-S7**)



The compound was synthesised according to the above general procedure using 2-(piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole as HCl salt (**MCN-S4**), 100 mg, 0.309 mmol, 1.0 Eq), TEA (86 µl, 0.618 mmol, 2.0 Eq), tert-butyl 3-formyl-1H-indole-1-carboxylate (83 mg, 0.34 mmol, 1.1 Eq), acetic acid (159 µl, 2.78 mmol, 9.0 Eq) and sodium triacetoxyborohydride (98.0 mg, 0.463 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 95.2 mg of the TFA salt of **MCN-S7** as a white solid with 92% purity (yield 45%). $^1$H NMR (400 MHz, DMSO-$d_6$, as TFA salt) δ 10.08 (s, 1H), 8.10 (dd, $J$ = 11.4, 8.4 Hz, 2H), 7.96 (brs, 1H), 7.89 (d, $J$ = 7.4 Hz, 1H), 7.78 (s, 1H), 7.52 – 7.33 (m, 3H), 4.58 (brs, 2H), 4.24 (brs, 2H), 1.66 (s, 9H). Calc. exact mass for $C_{26}H_{27}F_3N_4O_2S$ 516.18, LC-MS found $[M+1]^+$ 517.30.

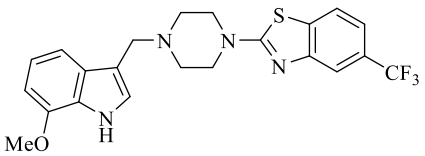
2-(4-((1H-indol-3-yl)methyl)piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole (**MCN-9**)



tert-butyl 3-((4-(5-(trifluoromethyl)benzo[d]thiazol-2-yl)piperazin-1-yl)methyl)-1H-indole-1-carboxylate (**MCN-S7**), as TFA salt, 93.2 mg, 0.148 mmol) was solved in a mixture of 1 mL DCM and 1 mL TFA and stirred at RT for 1h. The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 95.2 mg of the TFA salt of **MCN-9** as a white solid with 97% purity

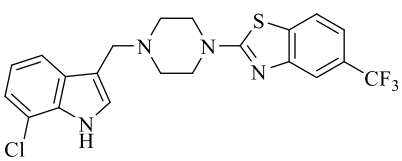
(yield 63.7%). $^{1}$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.57 (d, $J$ = 2.7 Hz, 1H), 10.17 (brs, 1H), 8.08 (d, $J$ = 8.2 Hz, 1H), 7.82 – 7.75 (m, 2H), 7.59 (d, $J$ = 2.6 Hz, 1H), 7.50 – 7.42 (m, 2H), 7.18 (ddd, $J$ = 8.1 Hz, 6.9 Hz, 1.2 Hz, 1H), 7.13 (ddd, $J$ = 7.5 Hz, 6.9 Hz, 1.0 Hz, 1H), 4.57 (s, 2H), 4.25 (d, $J$ = 14.1 Hz, 2H), 3.60 – 3.39 (m, 4H), 3.29 (brs, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 169.13, z152.11, 136.03, 135.27, 128.89, 127.38, 127.14 (q, $J$ = 31.5 Hz), 124.47 (q, $J$ = 270.10 Hz), 122.64, 121.86, 119.73, 118.58, 117.93, 115.06 (d, $J$ = 3.5 Hz, 1C), 111.97, 102.27, 50.68, 49.24, 45.06. Calc. exact mass for $C_{21}H_{19}F_3N_4S$ 416.13, FTMS found $[M+1]^+$ 417.14.

2-(4-((7-methoxy-1H-indol-3-yl)methyl)piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole (**MCN-10**)



The compound was synthesised according to the above general procedure using 2-(piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole as HCl salt (**MCN-S4**) (78 mg, 0.241 mmol, 1.0 Eq), TEA (67 µl, 0.482 mmol, 2.0 Eq), 7-methoxy-1H-indole-3-carbaldehyde (46.4 mg, 0.265 mmol, 1.1 Eq), acetic acid (124 µl, 2.168 mmol, 9.0 Eq) and sodium triacetoxyborohydride (77.0 mg, 0.361 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 60.5 mg of the TFA salt of **MCN-10** as a white solid with 94% purity (yield 42.1%). $^{1}$H NMR (600 MHz, DMSO-$d_6$, as TFA salt) δ 11.68 (s, 1H), 10.21 (s, 1H), 8.08 (d, $J$ = 8.2 Hz, 1H), 7.77 (d, $J$ = 1.8 Hz, 1H), 7.49 (d, $J$ = 2.7 Hz, 1H), 7.44 (dd, $J$ = 8.4 Hz, 1.8 Hz, 1H), 7.37 (d, $J$ = 8.0 Hz, 1H), 7.05 (t, $J$ = 7.9 Hz, 1H), 6.74 (d, $J$ = 7.7 Hz, 1H), 4.54 (s, 2H), 4.24 (d, $J$ = 14.0 Hz, 2H), 3.93 (s, 3H), 3.60 – 3.45 (m, 4H), 3.26 (brs, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 169.12, 152.11, 146.35, 135.26, 128.96, 128.34, 127.13 (q, $J$ = 31.6 Hz), 126.15, 124.47 (q, $J$ = 272.3 Hz), 122.63, 120.46, 117.92, 115.04 (d, $J$ = 3.3 Hz), 111.22, 102.81, 102.23, 55.24, 50.76, 49.22, 45.06. Calc. exact mass for $C_{22}H_{21}F_3N_4OS$ 446.14, FTMS found $[M+1]^+$ 417.15.

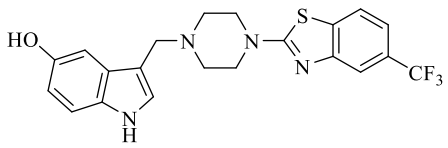
2-(4-((7-chloro-1H-indol-3-yl)methyl)piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole (**MCN-11**)



The compound was synthesised according to the above general procedure using 2-(piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole as HCl salt (**MCN-S4**), (78.0 mg, 0.241 mmol, 1.0 Eq), TEA (67 µl, 0.482 mmol, 2.0 Eq), 7-chloro-1H-indole-3-carbaldehyde (47.6 mg, 0.265 mmol, 1.1 Eq), acetic acid (124 µl,

2.168 mmol, 9.0 Eq) and sodium triacetoxyborohydride (77.0 mg, 0.361 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method A). Freeze-drying of the pure fractions yielded 60.5 mg of the TFA salt of **MCN-11** as a white solid with 94% purity (yield 42%). $^{1}$H NMR (600 MHz, DMSO-$d_6$, TFA salt) δ 11.95 (d, $J$ = 2.7 Hz, 1H), 10.31 (brs, 1H), 8.08 (d, $J$ = 8.2 Hz, 1H), 7.86 – 7.76 (m, 2H), 7.66 (d, $J$ = 2.7 Hz, 1H), 7.44 (dd, $J$ = 8.4 Hz, 1.8 Hz, 1H), 7.28 (d, $J$ = 7.5 Hz, 1H), 7.15 (t, $J$ = 7.8 Hz, 1H), 4.58 (s, 2H), 4.32 – 4.16 (m, 2H), 3.64 – 6.44 (m, 4H), 3.28 (brs, 2H). $^{13}$C NMR (151 MHz, DMSO-$d_6$, as TFA salt, TFA signal not included) δ 169.11, 152.11, 135.26, 132.87, 130.14, 129.38, 127.14 (q, $J$ = 31.7 Hz), 124.47 (q, $J$ = 271.4 Hz), 122.63, 121.40, 120.80, 117.92, 117.84, 116.30, 115.05 (d, $J$ = 3.2 Hz), 103.79, 50.39, 49.29, 45.06. Calc. exact mass for $C_{21}H_{18}ClF_3N_4S$ 450.09, FTMS found $[M+1]^+$ 451.10.

3-((4-(5-(trifluoromethyl)benzo[d]thiazol-2-yl)piperazin-1-yl)methyl)-1H-indol-5-ol (**MCN-12**)




The compound was synthesised according to the above general procedure using 2-(piperazin-1-yl)-5-(trifluoromethyl)benzo[d]thiazole as HCl salt (**MCN-S4**), 100 mg, 0.309 mmol, 1.0 Eq), TEA (86 µl, 0.618 mmol, 2.0 Eq), 5-hydroxy-1H-indole-3-carbaldehyde (54.8 mg, 0.34 mmol, 1.1 Eq), acetic acid (124 µl, 2.162 mmol, 7.0 Eq) and sodium triacetoxyborohydride (98 mg, 0.463 mmol, 1.5 Eq). The *in vacuo* concentrated sample was purified by preparative LC-MS (method B). Freeze-drying of the pure fractions yielded 29.2 mg of **MCN-12** as a white solid with 88% purity (yield 19%). $^{1}$H NMR (600 MHz, DMSO-$d_6$) δ 10.65 (d, $J$ = 2.5 Hz, 1H), 8.60 (s, 1H), 7.98 (d, $J$ = 8.2 Hz, 1H), 7.70 (d, $J$ = 1.7 Hz, 1H), 7.36 (dd, $J$ = 8.3 Hz, 1.8 Hz, 1H), 7.17 – 7.11 (m, 2H), 6.97 (d, $J$ = 2.3 Hz, 1H), 6.60 (dd, $J$ = 8.6 Hz, 2.4 Hz, 1H), 3.65 – 3.53 (m, 6H), 2.52 (t, $J$ = 5.1 Hz, 4H). $^{13}$C NMR (151 MHz, DMSO-$d_6$) δ 169.37, 152.60, 150.29, 134.78, 130.88, 128.27 (q, $J$ = 31.1 Hz), 125.26, 124.52 (q, $J$ = 271.3 Hz), 122.20, 117.15 (d, $J$ = 3.7 Hz), 117.14, 114.60 – 114.35, 111.63, 111.34, 109.35, 103.00, 53.23, 51.70, 48.22. Calc. exact mass for $C_{21}H_{19}F_3N_4OS$ 432.12, FTMS found $[M+1]^+$ 433.13.

## 4.9 Surface plasmon resonance

Surface plasmon resonance (SPR) experiments were established for orthogonal fragment screening and $K_d$ determination of fragment hits identified by NMR spectroscopy or X-ray crystallography. In addition, compounds derived from medicinal chemistry efforts were examined. Experiments were conducted at 22 °C on a Biacore T200 using the Biacore T200 Control Software. Biotinylated avi-tagged FPPS was immobilized on a Series S Sensor chip SA carrying a

carboxymethylated dextran matrix pre-immobilized with streptavidin. While one flow channel functioned as reference channel, biotinylated avi-tagged TcFPPS, TbFPPS and hFPPS were immobilized on the three remaining channels, thus allowing parallel testing on the three homologues. Experiments were run in collaboration with Lena Muenzker, NIBR, Novartis Pharma AG, Basel, Switzerland.

For an experiment 2.0 L of SPR buffer were prepared and to 1.3 L thereof 0.9% (v/v) DMSO were added. The device was primed with dd $H_2O$, the chip docked and pre-conditioned by three injections of a solution containing 50 mM NaOH and 1 M NaCl, followed by two injections of dd $H_2O$, all at a flow rate of 30 µL · $min^{-1}$. Then the device was primed with SPR buffer, the chip normalized with normalizing solution (70% glycerol) and then rinsed with SPR buffer at 30 µL · $min^{-1}$ to obtain a stable base line. Biotinylated avi-tagged FPPSs were thawed, filtered (0.45 µm) and diluted to 50 µg · $mL^{-1}$ (approx. 1:50) in SPR buffer. Proteins were immobilized in intervals, starting with 1 min, at a flow rate of 10 µg · $mL^{-1}$. Injections were repeated and time spans adjusted if necessary to achieve a final load of 2500 RUs to 3500 RUs. Protein immobilization was followed by several injections of SPR buffer without DMSO to check for baseline drifting. After loading was completed, the device was primed twice with SPR buffer with DMSO and the flow channels rinsed with SPR buffer with DMSO at a flow rate of 50 µL · $min^{-1}$ with injections for 20 min. For excluded volume correction (EVC) calibration an eight point dilution series from 0.4% to 1.7% (v/v) DMSO in SPR buffer was pipetted. The preparation of compound dilution varied depending on experiment design (fragment screening or $K_d$ determination). In both cases, compound solubility in SPR buffer was previously tested in an NMR experiment (chapter **4.2.1**). Compounds were prediluted in a solution of 90% $d_6$-DMSO and 10% $D_2O$ (v/v) to a concentration 100x higher than the final sample on the SPR source plate. In screening mode compounds were tested at the highest possible concentration, but at a maximum of 500 µM (pre-dilution 50 mM). For $K_d$ determination twofold dilution series with 12 dilution points (up to the highest possible concentration but to a maximum of 500 µM) were pipetted into a 96-well plate. Finally, an SPR source plate (96-well Greiner bio-one PP-microplate) was prepared by further dilute the predilutions or dilution series 1:100 in SPR buffer. The plate was covered with a microplate foil, mixed at 600 rpm for 2 min at RT and centrifuged for 1 min at 200 × *g*. Additionally, compound control samples with a compound concentration around the $K_d$ value were prepared. Since no potent binder was available for all three proteins, two controls were used: Compound 50 µM **CS-18** for TcFPPS and 200 µM compound **97** at for TbFPPS and hFPPS. Further, 2 mL of SPR buffer with DMSO, referred to as start-up solution, and a fresh pipetted mix of SPR buffer with 0.9% DMSO, referred to as buffer, were prepared.

A run started with 20 injections of the start-up solution, which were followed by 24 injections of samples. A sample was injected for 30 s at a flow rate of 30 µL · $min^{-1}$ and was followed by a dissociation period of 60 s or 180 s. The device ran in automation for approx. 18 h

per run. Analysis, curve fitting and $K_d$ calculations were done in the Biacore T200 Evaluation Software. Base line drift and compound behaviour on the control channel were evaluated. Curves were fitted assuming a 1:1 stoichiometry using affinity analysis because the observed interactions mainly had very fast on and off rates.

# 5. Results

## 5.1 Target enabling

A prerequisite for biochemical and structural work is the production of pure, homogenous and monodisperse protein. Whilst expression and purification of TcFPPS as well as medium resolution crystal structures were described in the literature[162b, 170a, 211], the expression and purification of isotope-labelled TcFPPS, biotinylated TcFPPS, and high resolution structures have not been reported to date. This thesis describes the development of a robust, reproducible and highly ordered i.e. well diffracting crystal system that enables FBS by X-ray crystallography. In the following chapter the exploration for such a crystallization system and the development of a soaking protocol is described. Crystal structures of ligand-protein complexes that were obtained in the soaking experiments described here are discussed in detail in later chapters.

### 5.1.1 Recombinant protein expression and purification

*E. coli* BL21(DE3) cells were used as expression system for all proteins expressed and purified in this work. In order to improve protein-expression, the plasmids used were codon optimized for *E. coli* to increase expression rates (in-house plasmid design, Felix Freuler, Novartis Pharma AG, Basel). Protocols used for protein expression in high-density shaking cultures using auto-inducing medium were similar to the procedures described in literature[470]. TcFPPS and hFPPS were successfully expressed with a cleavable N-terminal $His_6$-tag. After overnight growth in a fermentor $OD_{600}$ values of around 70 were measured accounting for approx. $5.6 \cdot 10^{10}$ cells per mL of medium. Harvesting yielded cell pellets of approx. 120 g (wet weight) that were purified in batch. In brief, mechanical cell lysis in a French Press was followed by protein purification applying IMAC, cleavage of the $His_6$-tag with HRV 3C protease, reverse IMAC, and SEC. Purification yielded ≥95% pure, homogenous and monodisperse protein as indicated by SDS-PAGE gels and LC-MS (**Figure 21**).

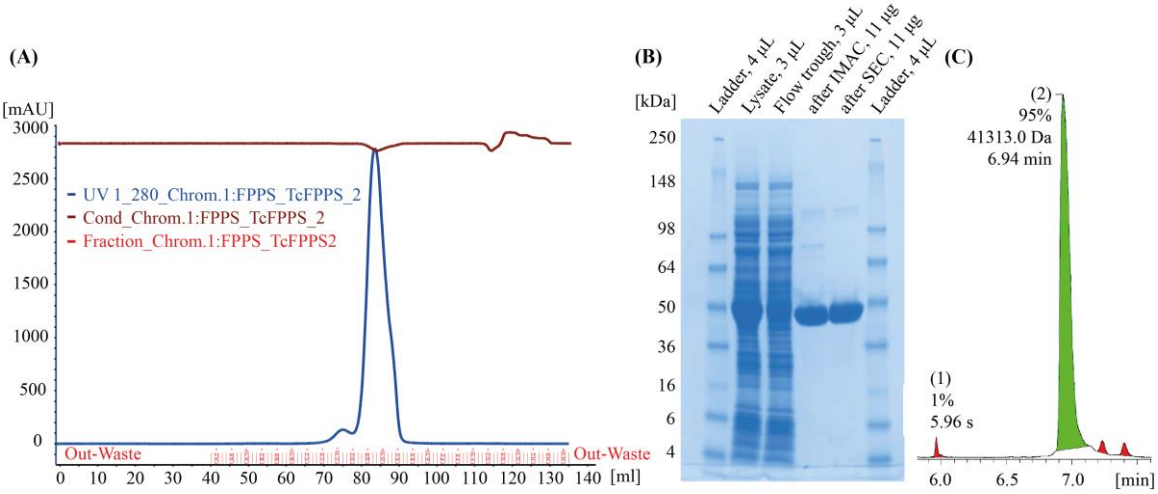



**Figure 21**: Purification of TcFPPS. **(A)** SEC profile of approx. 60 mg TcFPPS (HiLoad™ Superdex™ 16/60 S200, 1 mL · min$^{-1}$, 50 mM TRIS, pH 8.0, 200 mM NaCl, 2 mM TCEP · HCl). **(B)** Overview of purification steps on a SDS-PAGE gel (Coomassie blue stained): Lysate obtained after cell lysis, flow-through of IMAC, protein after purification by IMAC-reverse-IMAC, and after purification by SEC (final TcFPPS sample) were loaded. **(C)** Cut-out from the LC spectrum of the LC-MS run of the final TcFPPS sample.

TcFPPS purification yielded 1.75 mg protein per gram of cell pellet corresponding to 130 mg per L of medium. For hFPPS the yield of the purification was even higher. The expression of $^{13}C^{15}N$-labelled TcFPPS and hFPPS in minimal medium led to a decrease in cell densities. After overnight cell growth, the cell density showed $OD_{600}$ values of 13 accounting for approx. $1.04 \cdot 10^{10}$ cells per mL of medium. The harvested cell pellets were approx. 10 times less in weight (wet weight) when compared to expression in auto-inducing medium. Nevertheless, the protein yields of the purifications per g of cell pellet were higher. All proteins were obtained at a purity ≥95%. LC-MS studies revealed high labelling rates of TcFPPS and hFPPS with $^{13}C$ and $^{15}N$. Avi-tagged TcFPPS and hFPPS got completely biotinylated by *in vivo* biotinylation using *E. coli* cells that were expressing BirA (**Table 13**). In summary, pure protein was obtained in sufficient amounts for structural experiments and fragment screening campaigns.

**Table 13**: List of purified proteins.

| Enzyme | Yield per g pellet[a] (wet weight) (mg) | Purity[b] (%) | MW calc. (Da) | MW obs. $[M+H]^+$ (Da) | Labelling / biotinylation rate (%) |
|---|---|---|---|---|---|
| TcFPPS | 1.8 | ≥95 | 41313.21 | 41314.0 | - |
| $^{13}C^{15}N$-labelled TcFPPS | 3.8 | 95 | 43657.61 | 43580.0 | 97.31 |
| Biotinylated avi-tagged TcFPPS | 1.8 | 96 | 43350.48 | 43352.4 | 100 |
| hFPPS | 2.2 | ≥97 | 40686.56 | 40686.7 | - |
| $^{13}C^{15}N$-labelled hFPPS | 2.9 | 99 | 43007.50 | 43288.6 | 96.92 |
| Biotinylated avi-tagged hFPPS | 2.4 | 99 | 42723.34 | 42725.7 | 100 |

[a] Yields varied slightly between batches. The values given are examples.
[b] Purity is given according to LC.

### 5.1.2 High resolution crystals of *T. cruzi* FPPS – The power of MMS

To find appropriate crystallization conditions for TcFPPS that result in high resolution crystal structures, commercial screening matrices were used in a shotgun approach. In the first round, seven screens were tested on 2-drop 96-well SwissCi/MRC plates, employing the sitting-drop vapour diffusion technique. Drops of a 3:2 (v/v) mixture of 6.81 mg · $mL^{-1}$ TcFPPS in SEC buffer (50 mM TRIS, pH 8.0, 200 mM NaCl, 2 mM TCEP · HCl) and screening solution were pipetted. This resulted in 4.09 mg · $mL^{-1}$ TcFPPS in the crystallization drop. After incubation of the plates at 20 °C for a period of 90 d, most crystallization drops showed precipitate and only a few conditions with crystals were identified (**Table 14**, Var A). These were found in wells G11 and H1 from the Cryos Suite by QIAGEN, wells D5 and G9 of the Index HT by HAMPTON RESEARCH and well B4 from the SaltRX HT also by HAMPTON RESEARCH (**Table 15 (A)**). These conditions were selected for optimisation on 24-well VDX plates, employing the hanging-drop vapour diffusion technique. A direct transfer of parameters of the initial hits was tested and variables, such as precipitant concentration, salt concentration, and pH, were changed in small increments/decrements to identify conditions for optimal crystal growth. Whilst condition Index_D5 yielded salt crystals, condition Index_G9 was not reproducible. However, the other three conditions were further optimized. Crystals that grew in variations of condition SaltRX_B4 remained small in size and all tested variations of condition Cryos Suite_G11 showed precipitation. Crystals from two variations of Cryos Suite_H1 looked promising, but were not monocrystalline and hence, they were selected for a third round of optimisation. This time a drop ratio of protein formulation to reservoir of 2:1 (v/v) was used. Thus, the initial protein concentration was increased to 4.54 mg · $mL^{-1}$. This condition reliably yielded crystals that were first detected after 1 d – 2 d and were fully grown after 3 d – 4 d, however, the crystals grew as agglomerates ranging from 200 µm – 500 µm. These agglomerates were subsequently used to prepare seed stocks for microseeding.

A second round of condition screening was conducted to increase the scope of starting conditions. Four of the previously tested commercial screens were rescreened, using a protein formulation at approx. twice the original concentration to enhance protein concentration in the drop. In addition, a protein formulation in a low salt buffer was prepared using a spin filtration column to make use of the conditions at low ionic strength. This modification showed only slight improvements when compared to the first screening round (**Table 14**, Var B), and no additional conditions were found for further optimisation. In a third screening round, microseed matrix screening (MMS)[405b] was applied to overcome poor nucleation performance[364]. Crystallization drops were set up of protein formulation, reservoir solution and seed stock in a ratio of 3:2:1 (v/v) resulting in 6.31 mg · $mL^{-1}$ TcFPPS. Applying MMS revealed many more wells with crystals when compared to the non-seeded trials. This finding is consistent with the observation of D'Arcy and

co-workers[471]. The number of wells with large amounts of TcFPPS microcrystals was even higher than the ones with crystals (**Table 14**, Var C), which shows that it is difficult to find the correct concentration of nuclei, when seeding is applied for the first time[364].

**Table 14**: Screening for crystallization conditions of TcFPPS.

A)

| | Precipitation (%) | | | Clear drop (%) | | | Microcrystal (%) | | | Crystal (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Var / No | A | B | C | A | B | C | A | B | C | A | B | C |
| 1 | 77 | 79 | 46 | 19 | 16 | 18 | 0 | 1 | 24 | 0 | 0 | 8 |
| 2 | 46 | 43 | 26 | 46 | 32 | 20 | 0 | 4 | 28 | 4 | 7 | 16 |
| 3 | 52 | - | - | 42 | - | - | 1 | - | - | 1 | - | - |
| 4 | 27 | - | - | 68 | - | - | 0 | - | - | 1 | - | - |
| 5 | 35 | 62 | 37 | 60 | 31 | 45 | 1 | 2 | 6 | 1 | 1 | 8 |
| 6 | 38 | - | - | 58 | - | - | 0 | - | - | 0 | - | - |
| 7 | 47 | 59 | 22 | 47 | 36 | 55 | 1 | 1 | 13 | 1 | 0 | 2 |

B)

| No | Commercial condition screen, 96 conditions | Var | |
|---|---|---|---|
| 1 | AmSO4 Suite, QIAGEN | A | protein in SEC buffer<br>6.81 mg · mL$^{-1}$, mixed 3:2 with reservoir |
| 2 | Cryos Suite, QIAGEN | | |
| 3 | JCSG+ Suite, QIAGEN | B | protein in low salt buffer<br>12.62 mg · mL$^{-1}$, mix 3:2 with reservoir |
| 4 | MBClass II Suite, QIAGEN | | |
| 5 | Index HT, HAMPTON RESEARCH | C | protein in low salt buffer<br>12.62 mg · mL$^{-1}$, mix 3:2:1 with reservoir and seed stock |
| 6 | PegRX HT, HAMPTON RESEARCH | | |
| 7 | SaltRx HT, HAMPTON RESEARCH | | |

A) Results of visual inspection of the crystallization plates by visual inspection over a period of 90 d.
B) No 1 to 7 are commercial condition screens, variations A to C conditions used for screening.

Some crystals from the third screening round were chosen and their diffraction properties were tested at beamline X10SA of the Swiss Light Source (SLS), Villigen, Switzerland. X-ray data were collected at 100 K and diffraction patterns with diffraction limits ranging from 1.8 Å to 4.5 Å were obtained (**Table 15 (B)**). The best crystal grew in condition G7 of the Cryos Suite by QIAGEN, which was composed of 80 mM MES, pH 6.5, 8.5 mM $ZnSO_4$, 19.42% (v/v) PEG MME 550, 15% (v/v) glycerol (**Figure 22 (B)**). The condition was successfully transferred to 24-well plates applying the hanging drop vapour diffusion technique. Whilst the volumes were adjusted to the new set up, all other variables, such as buffer composition and ratios were kept constant. Hexagonal protein crystals appeared after 1 d – 2 d and grew to full size of approx. 150 µm × 50 µm × 50 µm after 3 d – 4 d (**Figure 22 (C)**). These crystals diffracted up to a resolution of 1.5 Å and therefore, this condition was selected for soaking experiments. Notably, crystals that were older than one week did not diffract and therefore, fresh apo crystals were used in all experiments. Thus, seed stocks and seed dilutions were regularly prepared. Seed stocks could be reproduced well and they

could be stored at 4 °C for up to three months or at -80 °C for an even longer period (**Table 16**, **Figure 22 (A)**).The number of wells per plate that showed crystals particularly of a reasonable size were strongly dependent on the quality and concentration of the used seed dilution. With an increase in the number of crystals per drop the average size of the crystals decreased. Strikingly, crystal size did not influence diffraction quality. An apo TcFPPS structure with a diffraction limit of 1.47 Å and with good data collection and refinement statistics was deposited under PDB ID 6R04. The TcFPPS apo crystal belonged to the hexagonal space group $P6_122$ and had the unit-cell parameters of $a = b = 57.65$ Å, $c = 397.59$ Å and $\alpha = \beta = 90$ ° and $\gamma = 120$ °. Assuming the presence of one protein chain per asymmetric unit, the specific volume $V_M$, also known as Matthews coefficient was 2.30 $Å^3 \cdot Da^{-1}$ and accordingly the solvent content was estimated to 47% (Appendix, **Table 29**).

**Table 15**: Reservoir conditions that yielded TcFPPS crystals.

| Screen[a] | Well | Reservoir composition | |
|---|---|---|---|
| Cryos Suite | G11 | 85 mM NaOAc · 3 $H_2O$, pH 4.6, 170 mM $(NH_4)_2SO_4$, 15% (v/v) glycerol, 25.5% (w/v) PEG MME 2000 | |
| Cryos Suite | H1 | 80 mM NaOAc · 3 $H_2O$, pH 4.6, 160 mM $(NH_4)_2SO_4$, 20% (v/v) glycerol, 20% (w/v) PEG 4000 | |
| Index HT | D5 | 100 mM NaOAc · 3 $H_2O$, pH 4.5, 25% (w/v) PEG 3350 | |
| Index HT | G9 | 100 mM TRIS, pH 8.5, 200 mM $NH_4OAc$, 25% (w/v) PEG 3350 | |
| SaltRX HT | B4 | 100 mM NaOAc · 3 $H_2O$, pH 4.6, 180 mM ammonium citrate dibasic | |
| Screen[b] | Well | Reservoir composition | Diffraction limit (Å), comment |
| $AmSO_4$ Suite | E7 | 0.1 M citric acid, pH 4.0, 1.6 M $(NH_4)_2SO_4$ | 3.5, 4.0, ice rings |
| $AmSO_4$ Suite | G3 | 0.1 M NaOAc · 3 $H_2O$, pH 4.6, 1.0 M $(NH_4)_2SO_4$ | no diffraction |
| $AmSO_4$ Suite | H6 | 2.2 M $(NH_4)_2SO_4$, 20% (w/v) glycerol | 2.0 |
| Cryos Suite | G11 | 85 mM NaOAc · 3 $H_2O$, pH 4.6, 170 mM $(NH_4)_2SO_4$, 15% (v/v) glycerol, 25.5% (w/v) PEG MME 2000 | 2.2 |
| Cryos Suite | F8 | 0.08 M sodium cacodylate, pH 6.5, 0.16 mM $Mg(OAc)_2 \cdot 4\ H_2O$, 20% (v/v) glycerol, 16.0% (w/v) PEG 8000 | 2.2, 2.7 |
| Cryos Suite | G2 | 0.095 M HEPES sodium salt, pH 7.5, 0.19 M $CaCl_2 \cdot 2\ H_2O$, 26.6% (v/v) PEG 400, 5% (v/v) glycerol | salt |
| Cryos Suite | G7 | 0.085 M MES, pH 6.5, 0.0085 M $ZnSO_4$, 19.42% (v/v) PEG 550 MME, 15.0% (v/v) glycerol | 1.8 |
| Index HT | G3 | 0.1 M Bis-Tris, pH 6.5, 0.2 M $Li_2SO_4 \cdot H_2O$, 25% (w/v) PEG 3350 | 2.9 |
| Index HT | G4 | 0.1 M HEPES, pH 7.5, 0.2 M $Li_2SO_4 \cdot H_2O$, 25% (w/v) PEG 3350 | 3.0, 4.5, anisotropic |

[a] Conditions of the 1st round of screening (variation A, see **Table 14**).
[b] Conditions of the 3rd round of screening (variation C, see **Table 14**).

Soaking experiments were conducted with apo crystals grown in 24-well plates. Crystal stability in DMSO containing conditions was highly variable and in many cases diffraction quality

decreased when the DMSO concentration and/or soaking times were increased. Crystals started to show cracks perpendicular to the long axis when incubated with the compound. Thus, several soaking experiments were conducted and several data sets were collected to determine the highest tolerated DMSO concentration and the longest soaking time that did not compromise crystal diffraction quality. For most compounds a data set of a soaked crystal could be collected with a diffraction limit ranging from 1.5 Å – 3.4 Å. Among these the compound concentrations and soaking times differed widely from 5 mM – 75 mM and 5 min to overnight, respectively. To overcome the high variability in tolerance to DMSO and the resulting experimental error, apo crystals were grown in drops that contained 4.5% DMSO in the crystallization drop, therefore effectively priming the crystals for subsequent DMSO exposure. In the subsequent soaking experiment the same amount of DMSO was used (PDB ID 6R06, results described in chapter **5.3.2**).

Notwithstanding first successful soaking experiments that were set up manually, the variability in DMSO tolerance and the 24-well plate format were not suited for high-throughput crystallization experiments. Therefore, further optimization experiments were conducted on 2-drop 96-well SwissCi/MRC plates applying the sitting-drop vapour diffusion technique. The concentration of buffer, salt and precipitant were changed in small increments/decrements to meet conditions for optimal crystal growth. The finally optimized reservoir solution contained 53% less $ZnSO_4$, 36% less PEG MME 550 and 23% less glycerol (80 mM MES, pH 6.5, 4 mM ZnSO4, 12.36% (v/v) PEG MME 550, 11.57% (v/v) glycerol). Experiments were conducted at a ratio of protein formulation, reservoir and seed dilution of 3:2:1 (v/v) in the drop (**Figure 22 (D)**). Later, also a drop ratio of 3:1:2 (v/v) was used. In both cases, crystals obtained could support soaks with up to 15% DMSO for up to 24 h, which is exceptionally high.

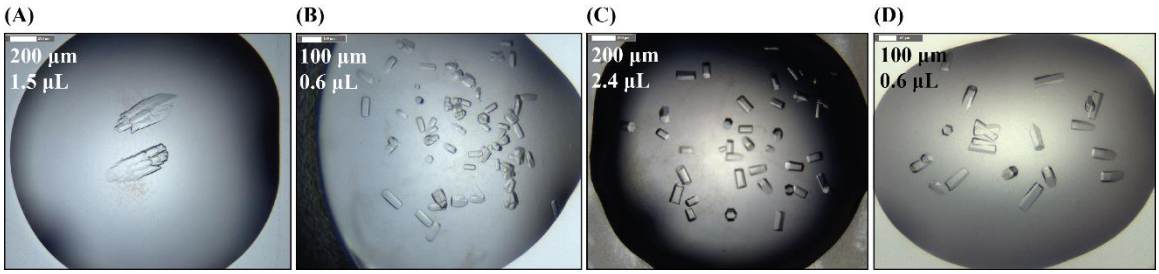



**Figure 22**: TcFPPS crystals. **(A)** Seed crystals. **(B)** Hexagonal crystals with the best diffraction in the third round of condition screening, applying MMS (well G7, Cryos Suite, QIAGEN). **(C)** Apo crystals on 24-well plates applying hanging drop vapour diffusion. **(D)** Apo crystals on 2-drop 96-well plates (drop ratio 3:2:1 (v/v)) applying sitting drop vapour diffusion. Scale and drop size are given in each picture.

After this crystallization system was established at the Novartis laboratories, crystallization experiments were also transferred to other laboratories to conduct FBS by X-ray crystallography. At laboratories of beamline I04-1 at the Diamond Light Source in Harwell, UK, TcFPPS crystals

were grown in 3-drop 96-well SwissCi/MRC plates. Instead of 80 µL reservoir, which was used on 2-drop plates, drops were equilibrated against 30 µL reservoir. At laboratories of the HTX lab in Grenoble, France, CrystalDirect™ plates were used. In addition, drop volumes were downsized to 300 nL and a ratio of protein formulation to reservoir to seed dilution of 3:1:2 (v/v) was used in the drop setup. This resulted in 80% of wells with crystals per plate, which was a prerequisite to pass an evaluation phase and enter the screening phase. Consider **Table 16** for final buffer conditions and plate setups. A comprehensive overview of all crystallization experiments conducted with TcFPPS are given in **Table 30** in the Appendix.

**Table 16:** Crystallization conditions of TcFPPS.

| Formulation/Buffer[a] | Composition |
|---|---|
| TcFPPS in SEC buffer (Formulation I) | 6.81 mg · mL$^{-1}$ TcFPPS in 50 mM TRIS, pH 8.0, 200 mM NaCl, 2 mM TCEP · HCl |
| TcFPPS in low salt buffer (Formulation II) | 12.20 mg · mL$^{-1}$ – 12.70 mg · mL$^{-1}$ TcFPPS in 10 mM TRIS, pH 7.4, 25 mM NaCl, 2 mM TCEP · HCl |
| Reservoir 24-well seeds | 80 mM NaOAc, pH 5.0, 160 mM $(NH_4)_2SO_4$, 20% (w/v) PEG 4000, 20% (v/v) glycerol |
| Reservoir 24-well | 80 mM MES, pH 6.5, 8.5 mM $ZnSO_4$, 19.42% (v/v) PEG MME 550, 15% (v/v) glycerol |
| Reservoir 96-well | 80 mM MES, pH 6.5, 4 mM $ZnSO_4$, 12.36% (v/v) PEG MME 550, 11.57% (v/v) glycerol |
| Seed buffer | 80 mM NaOAc, pH 5.0, 160 mM $(NH_4)_2SO_4$, 20% (v/v) PEG 4000, 20% (v/v) glycerol |

| Usage[b] | Well | Plate | Drop (µL) | Ratio (v/v) | Components |
|---|---|---|---|---|---|
| Seed crystals | 24 | VDX 18 mm | 1.5 | 2:1 | Formulation I : reservoir 24-well seeds |
| Apo crystals | 24 | VDX 18 mm | 2.4 | 3:2:1 | Formulation II : reservoir 24-well : seed dilution |
| Apo crystals | 96 | 2-, 3-drop SwissCi/MRC | 0.6 | 3:2:1 | Formulation II : reservoir 96-well : seed dilution |
| Apo crystals | 96 | 2-drop SwissCi/MRC | 0.6 | 3:1:2 | Formulation II : reservoir 96-well : seed dilution |
| Apo crystals | 96 | CrystalDirect™ plates | 0.3 | 3:1:2 | Formulation II : reservoir 96-well : seed dilution |

[a] Protein formulations and buffers used in crystallization trials.
[b] Set up of crystallization plates for different purposes and in different formats.

### 5.1.3 Discussion

When fragment binding is investigated by X-ray crystallography, a diffraction limit of at least 2.5 Å is highly desirable[397] as problems resulting from the weak diffraction power of small fragments and the often experienced partial ligand occupancies are exacerbated at low and medium resolution [392b]. Conversely, high resolution data were shown to make the identification of bound fragments easier and more reliable[384, 397, 440]. To date, 14 crystal structures of TcFPPS with an average diffraction limit of 2.36 Å were deposited in the PDB [162b, 170a, 211] (Appendix, **Table 28**). While this resolution is technically feasible for FBS, high-throughput screenings would highly benefit from a diffraction limit below 2.0 Å.

The developed crystallization system yielded TcFPPS crystals with high-quality diffraction. The reservoir buffer contains cryoprotectant that reliably prevents formation of ice rings and made treatment with an additional cryoprotectant prior to flash-freezing superfluous. Other benefits are easy, quick and gentle crystal handling, because crystals can be easily picked from the mother liquor. Furthermore, cryoprotectants dilute the mother liquor, which is disadvantageous for soaking experiments. While early soaking experiments followed the method of trial and error, further optimization in 2-drop 96-well SwissCi/MRC plates resulted in conditions that allowed soaking in up to 15% DMSO for up to 24 h. A further strength of the crystallization conditions found for TcFPPS was demonstrated by the use of a wide variety of plate formats and successful transfer to other laboratories. A weakness is the aging of the TcFPPS crystals, which resulted in a loss of diffraction power. This observation was already reported in the literature for TcFPPS crystals and was related to the decrease in reducing agent in the crystallization drop over time [162b].

All TcFPPS crystals measured as part of this work belonged to the hexagonal space group $P6_122$ (No. 178, International Tables for Crystallography[377b]) and showed unit-cell parameters of approx. a = b = 58 Å, c = 397 Å and $\alpha = \beta = 90$ ° and $\gamma = 120$ °. The 14 previously published TcFPPS structures also belong to this space group and show similar cell dimensions [162b, 170a, 211] (Appendix, **Table 28**). The length of the unit cell axis is inversely-proportional to the distance between Bragg reflections. Hence, reflections along the c-axis are very close to each other. However, with the advent of Pilatus detectors (DECTRIS)[389, 439] that enable data collection at extremely fine oscillation angles[472], reflections were successfully resolved spatially and data processing and refinement resulted in 3D structures with good statistics. $P6_122$ is a high symmetry space group, which allows fast collection of complete, highly redundant data sets. Collected data sets of TcFPPS crystals achieved 17- to 19-fold redundancy, ensuring good data quality. Remarkably, the crystals had a high diffraction limit in spite of a long c-axis of nearly 400 Å in length. An apo TcFPPS structure with a diffraction limit of 1.47 Å was deposited under PDB ID 6R04. The crystal structure of TcFPPS in complex with compound **MCN-1** (chapter **5.5.1**) had even a diffraction limit of 1.28 Å, which is the highest diffraction limit ever obtained for a TcFPPS crystal. Notably, of the approx. 136.000 crystal structures deposited in the PDB there are only seven structures at a resolution of 1.5 Å or better with a least one unit cell axis longer than 390 Å (PDB ID 4UFQ (1.45 Å)[473], 4Y9V (0.90 Å), 3SGZ (1.35 Å)[474], 3PQH (1.30 Å)[475], 3QR7 (0.94 Å)[475], 3GIP (1.50 Å)[476], 1OCY (1.50 Å)[477]) underscoring the superb crystal quality with small reflection spots and data collection setup of the presented experiments. Taken together, this work has identified a novel, reliable, highly reproducible, and well-diffracting crystallization system for TcFPPS that exhibits excellent properties for FBS and therefore paves the way for future studies aiming to identify TcFPPS binders.

## 5.2 Testing of allosteric inhibitors of human FPPS against *T. cruzi* FPPS – A phenylalanine as game changer?

Allosteric inhibitors of a novel scaffold were identified for hFPPS[209a]. More recently, hFPPS was found to be inhibited by FPP, its own product[205]. Product inhibition is governed by FPP binding to the allosteric site, thereby keeping the enzyme in an open and inactive state[205]. This site was also described for FPPS of *P. aeruginosa*[209b] and *P. falciparum*[209c], but has not yet been described for TcFPPS. The only FPPS inhibitors used in the clinic are active site-directed nitrogen-containing bisphosphonates (N-BPs), which exhibit high affinity to bone mineral. Hence, they are ideal to treat bone diseases[213, 244]. Inhibiting FPPS with compounds of a novel scaffold and by a novel mechanism of action has high potential for the treatment of non-bone related diseases[209a]. Here, the allosteric region in TcFPPS is investigated by means of sequence analysis and structural superimposition of various orthologous FPPSs. In an attempt to reposition established inhibitors, eight known allosteric hFPPS inhibitors were tested for their binding affinity to TcFPPS.

### 5.2.1 Results

$TcFPPS_{64-425}$ and $hFPPS_{67-419}$ are homologous proteins that share 34.1% sequence identity and 50.1% sequence similarity as indicated by a global sequence alignment using the Needleman-Wunsch algorithm[478] (Appendix, **Figure 62**). A ClustalX multiple alignment of the amino acid sequence of TcFPPS and 200 homologues of other source organisms with a sequence identity ranging from 35% – 95% was conducted to generate a ConSurf model that illustrates the level of sequence conservation within the enzyme on a scale ranging from high variability (score 1) to high conservation (score 9)[479]. As expected, the generated ConSurf model shows that conservation is very high for residues directly involved in catalysis[162b]. Residues forming the allylic site, which includes the aspartate-rich motifs FARM and SARM, and residues forming the homoallylic site are highly conserved (score 9). However, the residues in the allosteric region are less conserved. The residues forming the pocket in hFPPS[168] differ from the corresponding residues in TcFPPS. Five of the corresponding residues are the same, two are similar and two differ. The polar residue Asn59 and the hydrophobic residue Ile348 of hFPPS are replaced by the aromatic residues Phe50 and the polar residue Thr357 in TcFPPS (**Figure 23**, **Table 17** and Appendix, **Figure 63**).

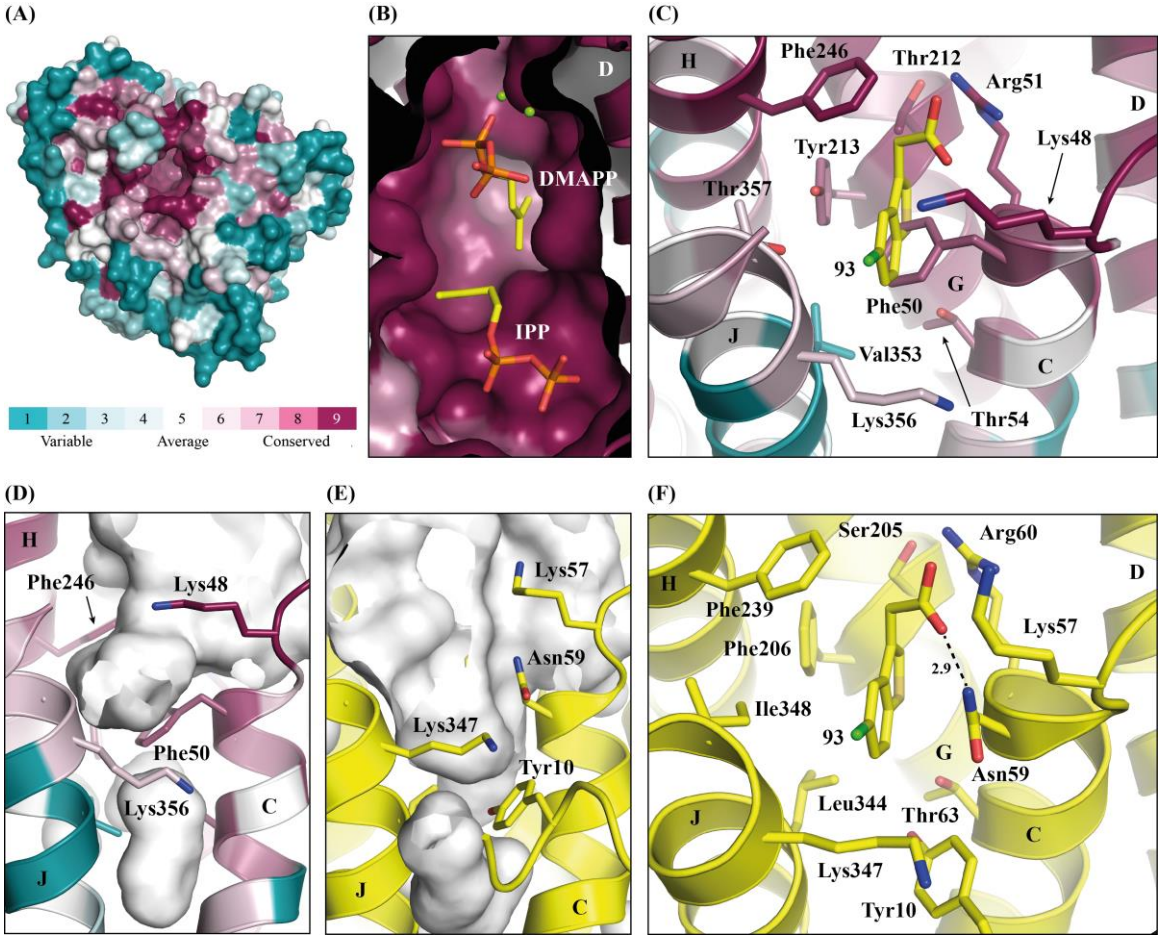



**Figure 23**: ConSurf model of FPPS illustrating sequence variability. **(A)** 3D structure of TcFPPS coloured by sequence conservation. The model was generated using The ConSurf Server[479]. An alignment of the sequence of TcFPPS (PDB ID 6R04, this work) against the sequences of 200 homologues with an identity ranging from 35% to 95% was done using ClustalX[480]. **(B)** Allylic and homoallylic pocket of TcFPPS (PDB ID 6R04, this work). Superimposition with DMAPP, $Mg^{2+}$ (PDB ID 1UBY[206]) and IPP (PDB ID 2F8Z[206]) (backbones not shown). **(C)** Allosteric region of TcFPPS. Superimposition with compound **93** (PDB ID 3N1W[209a], backbone not shown). **(D)** Surface of the pockets and cavities in the allosteric region in TcFPPS (PDB ID 6R04, this work). **(E)** Surface of the pockets and cavities in the allosteric region in hFPPS (PDB ID 3N1W[209a]). **(F)** Allosteric pocket in hFPPS with compound **93** bound. H-bond is indicated with a dashed line. Distance is given in Å (PDB ID 3N1W[209a]).

In open-state hFPPS, the allosteric binding site is a large pocket between helices C, G, H and J that is in close proximity next to the homoallylic site[168]. It was shown to accommodate ligands with up to three aromatic rings that keep the enzyme in the open-state[209a]. The crystal structure of hFPPS in complex with the benzothiophene **93,** which was discovered by a fragment screening campaign using by NMR spectroscopy[205, 209a], shows that residue Asn59 forms a H-bond with the carboxyl function of the inhibitor. In crystal structures of unliganded hFPPS (PDB IDs 2F7M[168], 4XQR, 4XQS and 4XQT), the pocket does already exist and the conformation of Asn59 is nearly the same when compared to structures with an allosteric inhibitor (e.g. PDB ID 3N1W[209a]) or with bound FPP (PDB ID 5JA0[205]). In the apo crystal structure of TcFPPS such a wide pocket does not exist. The space between helices C and J is narrower and the residue Phe50

protrudes perpendicular from helix C into the protein (**Figure 23 (C)**). Hence, Phe50 separates the pocket into two parts, changing the size and properties of the pocket. Comparison of the surface representations of the pockets in TcFPPS and hFPPS show this difference (**Figure 23 (D,E)**). In addition, superimposition of the 3D structures of TcFPPS with hFPPS in complex with compound **93** shows that residue Phe50 of TcFPPS clashes with the hFPPS ligand (**Figure 23 (C)**).

**Table 17**: Comparison of the residues forming the allosteric pocket in TcFPPS and hFPPS.

| Protein | Residues forming the allosteric pocket | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| TcFPPS | - | Lys48 | Phe50 | Arg51 | Thr54 | Tyr213 | Phe246 | Val353 | Lys356 | Thr357 |
| hFPPS | Tyr10 | Lys57 | Asn59 | Arg60 | Thr63 | Phe206 | Phe239 | Leu344 | Lys347 | Ile348 |
| similarity[a] | none | \| | . | \| | \| | : | \| | : | \| | . |

[a] Lines indicate identical residues, colons indicate similar residues, and points indicate mismatch.

Superimposition of the apo structure obtained as part of this work with all 14 published TcFPPS crystal structures[162b, 170a, 211] shows that Phe50 was refined in different conformations, but all conformers have a very similar impact on the pocket (**Figure 24**). This is observed in structures with the natural substrate IPP bound, but also in apo structures or in structures with bisphosphonates bound in the allylic site. Superimposition of the apo structure of TcFPPS with hFPPS in complex with inhibitor **93** shows that the equivalent residue Asn59 is rotated by 112 ° and hence, points towards the protein surface (**Figure 24 (D)**).

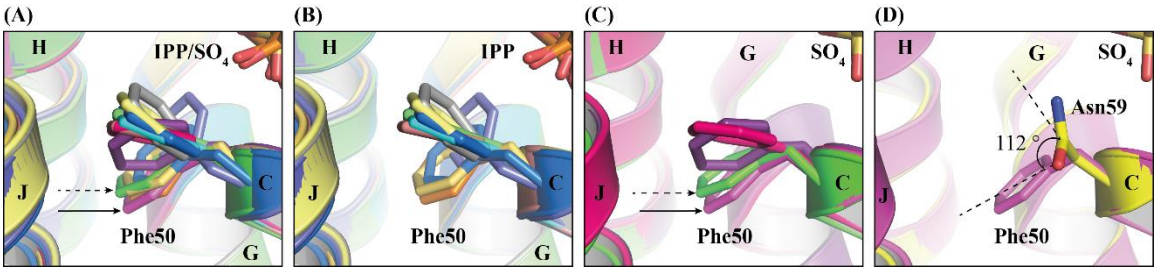



**Figure 24**: TcFPPS crystal structures – focus on residue Phe50. **(A)** Overlay of all deposited X-ray structures (PDB IDs 1YHK (green, position indicated with an arrow with dashed line), 1YHL (cyan), 1YHM (light yellow), 3IBA (nude), 3ICK (dark violet), 3ICM (orange), 3ICN (green), 3ICZ (petrol), 3ID0 (magenta), 4DWB (sand), 4DWG (violet), 4DXJ (grey), 4DZW (blue) and 4E1E (blue)[162b, 170a, 211]) and apo structure generated as part of this work (PDB ID 6R04, pink, position indicated by an arrow). **(B)** TcFPPS with IPP or DMAPP bound only (PDB IDs 1YHL, 1YHM, 3IBA, 3ICK, 3ICM, 3ICN, 3ICZ, 4DWB, 4DXJ, 4DZW and 4E1E[162b, 170a, 211], colours as in **(A)**) **(C)** Apo structures 1YHK[162b] (green, position indicated with an arrow with dashed line) and 6R04 (pink, position indicated with an arrow), as well as 3ID0 (magenta) and 4DWG (violet) (no ligand in the homoallylic binding site) **(D)** Superimposition of apo structure (PDB ID 6R0A, pink) with crystal structure of hFPPS (PDB ID 3N1W, yellow).

Despite the prominent role of Phe50 in TcFPPS as residue blocking the allosteric pocket, this structural variant appears to be an exception. In fact, hFPPS (UniProt ID P14324), TbFPPS

(UniProt ID Q86C09) and a further 190 out of 200 homologues show an asparagine at this position while eight homologues show a deviating residue (Appendix, **Figure 64**). Tyrosine is found in the FPPS of three plant species, glutamine in FPPS of horses and bats and histidine in the FPPS of two monkey species. Thus, TcFPPS is not the only homologue with an aromatic side chain at this position, but the only one with a hydrophobic residue at this position that cannot contribute to H-bonding (**Table 18**).

**Table 18**: FPPSs from organism that show amino acids other than the conserved Asp.

| Position in alignment[a] | Uniprot ID | Organism | Species | Residue |
|---|---|---|---|---|
| 192 | Q8WS26 | *T. cruzi* | *Trypanosoma cruzi* | Phe |
| 105 | A0A140GWW0 | rubber tree | *Hevea brasiliensis* | Tyr |
| 106 | A0A140GWW3 | manioc | *Manihot esculenta* | Tyr |
| 107 | B9S9Y3 | castor oil plant | *Ricinus communis* | Tyr |
| 173 | K9K3N0 | horse | *Equus caballus* | Gln |
| 174 | S7PKH9 | Brandt's bat | *Myotis brandtii* | Gln |
| 175 and 176 | F7GUQ3, B0CM97 | white-tufted-ear marmoset | *Callithrix jacchus* | His |
| 177 | F7FI27 | rhesus macaque | *Macaca mulatta* | His |

[a] Consider **Figure 64** in the Appendix for an excerpt from the alignment of all homologues.

Eight known hFPPS allosteric inhibitors of different size and affinity were selected to test their binding affinity to TcFPPS by protein-observed NMR spectroscopy (**Figure 25**, Appendix, **Table 31**). Among them were the fragments **93**, **94** and **95**, the first allosteric inhibitors, which were discovered using FBS by NMR[209a]. An SBLD campaign resulted in the compounds **118** and **119** that were further optimized to the lead compounds **97** and **98**[209a]. The eighth compound selected, was quinoline **101**, which was discovered by the same team[230].

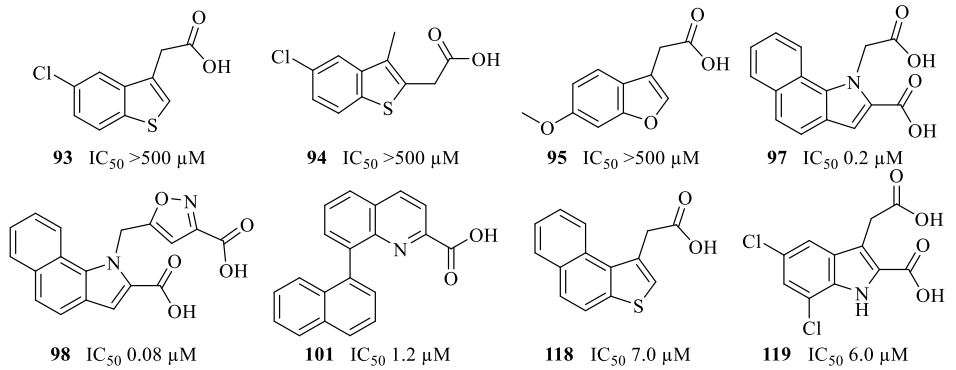



**Figure 25:** Chemical structures of a selection of allosteric inhibitors of hFPPS. $IC_{50}$ values are given, according to Jahnke *et al*.[209a] and Marzinzik *et al*. [230].

Whilst the fragment hits exhibited $IC_{50}$ values >500 µM against hFPPS, the lead compounds exhibited $IC_{50}$ values in the nanomolar range. Crystal structures of hFPPS in complex

with compounds **93**, **94**, **95**, **97** and **101** demonstrate their binding to the allosteric site (PDB IDs 3N1W, 3N1V, 3N3L, 3N6K[209a] and 5DGN[230], respectively).

[$^{13}C^{1}H$]-SOFAST-HMQC experiments were conducted with samples of 1 mM compound and 30 µM $^{13}C^{15}N$-labelled TcFPPS. All compounds showed chemical shift changes of weak to medium strength when compared to the DMSO control containing the equivalent amount of DMSO (**Table 19**, **Figure 26**). Lead **98** showed the strongest chemical shift changes (**Figure 26 (E)**) and fragment **95** showed the weakest chemical shift changes (**Figure 26 (C)**). Information about the binding site of the ligands could not be extracted from the NMR experiments, because the size of the homodimer did not allow any resonance assignments. For binding site determination X-ray crystallography was conducted (**Table 19**).

**Table 19:** Testing allosteric inhibitors of hFPPS against TcFPPS. 2D NMR and soaking experiments.

| | | **2D NMR** | **X-ray crystallography - soaking** | | | |
|---|---|---|---|---|---|---|
| Compound | MW (Da) | Shifting signals Number, strength | Compound (mM) | DMSO (%) | Time | PDB ID |
| **93** | 226.68 | 17, weak to medium | 25 | 9 | overnight | 6R07 |
| **94** | 240.71 | 19, medium | - | - | - | - |
| **95** | 206.20 | few, weak | - | - | - | - |
| **97** | 269.26 | 24, medium to strong | 5 – 50 | 4.5 – 13.5 | 30 min – overnight | - |
| **98** | 336.31 | 30, strong | 10 | 9 | overnight | - |
| **101** | 299.33 | >25, strong | - | - | - | - |
| **118** | 242.30 | 25, medium to strong | 25 | 9 | overnight | - |
| **119** | 288.09 | 14, medium to strong | 25 | 9 | overnight | 6R08 |

Apo TcFPPS crystals were grown using the hanging drop vapour diffusion technique on 24-well VDX plates. Crystallization drops were a mix of 1.2 µL 12.36 mg · $mL^{-1}$ TcFPPS (in 10 mM TRIS, pH 7.4, 25 mM NaCl, 2 mM TCEP · HCl), 0.8 µL reservoir (80 mM MES, pH 6.5, 8.5 mM $ZnSO_4$, 19.42% (v/v) PEG MME 550, 15% (v/v) glycerol) and 0.4 µL TcFPPS micro seeds (in 80 mM NaOAc, pH 5.0, 160 mM $(NH_4)_2SO_4$, 20% (v/v) PEG 400, 20% (v/v) glycerol). The drops were equilibrated against 500 µL reservoir. Soaking was conducted by transferring fresh crystals to a mixture of protein buffer, reservoir solution and seed buffer in a ratio of 3:2:1 (v/v), thus mimicking the mother liquor at the time point of the drop setup. Depending on the concentration of the compound stock solution, crystals were soaked with five of the hFPPS inhibitors at concentrations ranging from 5 mM to 50 mM, which corresponded to 4.5% – 13.5% DMSO. Diffraction data were collected at beamline X10SA of the Swiss Light Source, Villigen, Switzerland. Data sets from crystals soaked with all five compounds were successfully collected,

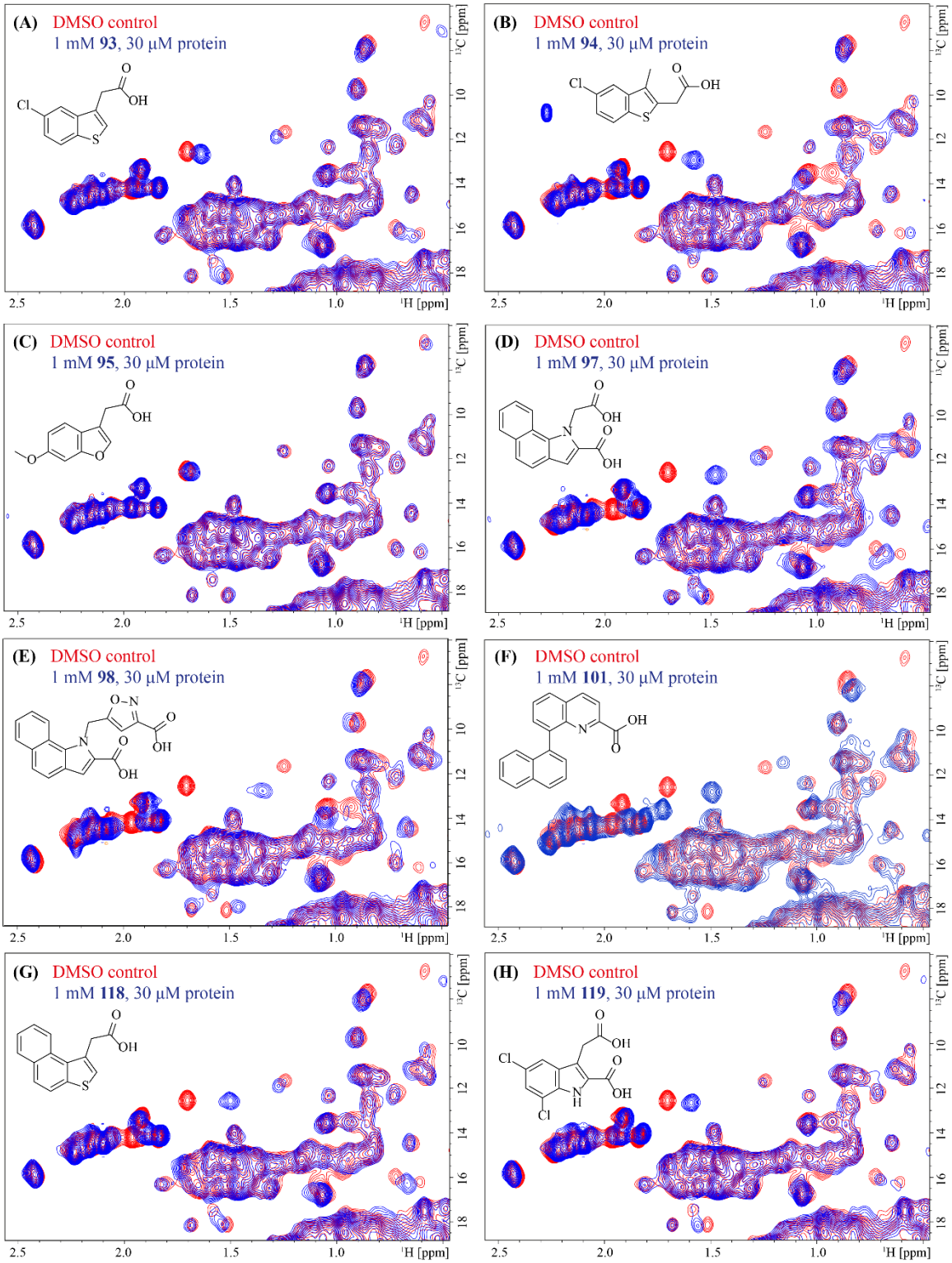



**Figure 26:** [$^{13}C^{1}H$]-SOFAST-HMQC spectra of hFPPS allosteric site binders tested on TcFPPS. **(A)** – **(H)** Compounds **93**, **94**, **95**, **97**, **98**, **101**, **118** and **119**, respectively. Each image shows a cut-out from an overlay of the [$^{13}C^{1}H$]-SOFAST-HMQC spectrum of the DMSO control (red, DMSO concentration equivalent to the DMSO concentration in the sample) and the corresponding sample (blue, 1 mM compound and 30 µM protein in 25 mM BisTris, pH 6.5, 50 mM NaCl, 2 mM TCEP · HCl, 10% $D_2O$, 150 µM DSS) measured at 305 K.

but data processing and refinement revealed that only compounds **93** and **119** had successfully formed a complex with TcFPPS and showed binding to distinct cavities located at the protein surface. Data collection and refinement statistics were comparable to those of the TcFPPS apo structure (PDB ID 6R04) (chapter **5.1**, Appendix, **Table 29**). The diffraction limit for the TcFPPS-**93** complex and the TcFPPS-**119** complex was 1.57 Å and 1.44 Å, respectively. Structural models were deposited in the PDB under PDB IDs 6R07 and 6R08.

Two molecules of fragment **93** (**93-1** and **93-2**) bind per subunit of FPPS in a groove on the protein surface and at the dimer interface (**Figure 27**). The bottom of the cavity is formed by helix G and its sides are formed by helices F and H, and the connecting loop of helices A and B of subunit B. Ligand **93-1** was refined to an occupancy of 0.82 and is well resolved as the unbiased $mF_o - DF_c$ difference electron density map contoured at 3.0 σ indicates (summary of density maps, Appendix, **Figure 65 (A – C)**).

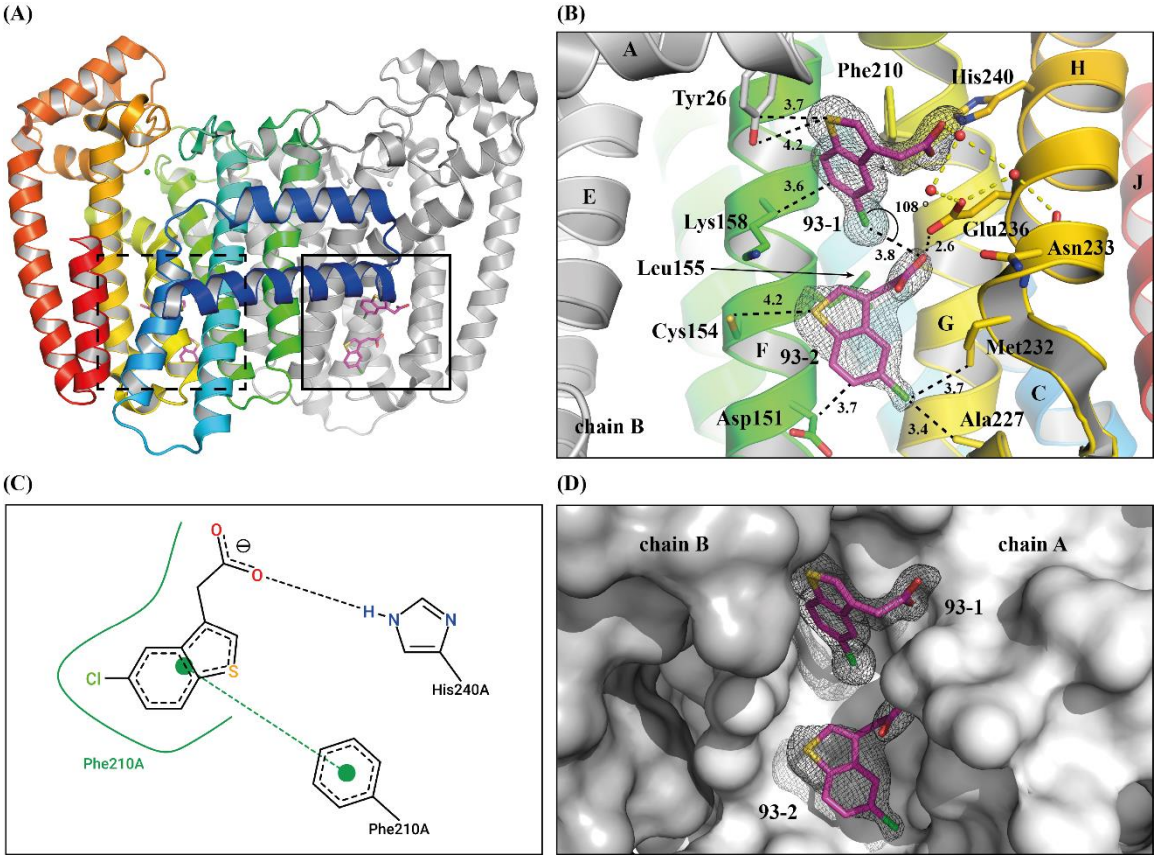



**Figure 27:** Crystal structure of TcFPPS in complex with compound **93** (PDB ID 6R07, this work). **(A)** Homodimer with bound ligands (front view, cartoon representation, subunit A coloured in gradient from blue to red from N-terminus to C-terminus, subunit B coloured in grey. Ligands shown in stick representation. $Zn^{2+}$ ions are shown as green spheres. **(B)** Binding site of ligands **93-1** and **93-2**. The final $2F_o - F_c$ electron density map is contoured at 1.0 σ and represented as liquorice coloured mesh. Waters are shown as red spheres. Interactions are shown as dashed line. Distances are given in Å. **(C)** 2D structure diagram of ligand **93-1** interacting with TcFPPS. Diagram was generated using PoseView[481]. **(D)** Binding site of ligands **93-1** and **93-2**. Subunit A and B shown in surface representation.

The two main interactions of this ligand are π-stacking of the benzothiophene core with Phe210 in helix G and an H-bond of the carboxyl group with His240 in helix H. In addition, the carboxyl group forms a second H-bond to a water molecule that in turn interacts with two additional water molecules that form H-bonds with Glu236 and the carbonyl oxygen of Asn233. Hydrophobic interactions are formed with Lys158. The halogen substituent of ligand **93-1** is interacting in a multipolar interaction with the carboxyl group of ligand **93-1** (bond length 3.8 Å, angle 108 °). The sulphur atom of the benzothiophene core is 4.2 Å away from Tyr26 of subunit B and therefore too far for an H-bond interaction. Ligand **93-2** is a tentative interpretation of the initial difference map. It was refined to an occupancy of 0.59 and is less well defined than ligand **93-1** as seen in the $mF_o - DF_c$ difference electron density map contoured at 3.0 σ (summary of density maps, Appendix, **Figure 65 (A – C)**). It shows hydrophobic interactions with residues Asp151, Leu155, Ala227 and Met232. The respective binding poses and the electron density map for both ligands are depicted in **Figure 27**.

Intermediate **119** binds to a solvent exposed cleft formed by helices H, I and α3. Residues Phe256 and Gln318 form the opposite walls and Phe321 the bottom of the cavity (**Figure 28**).

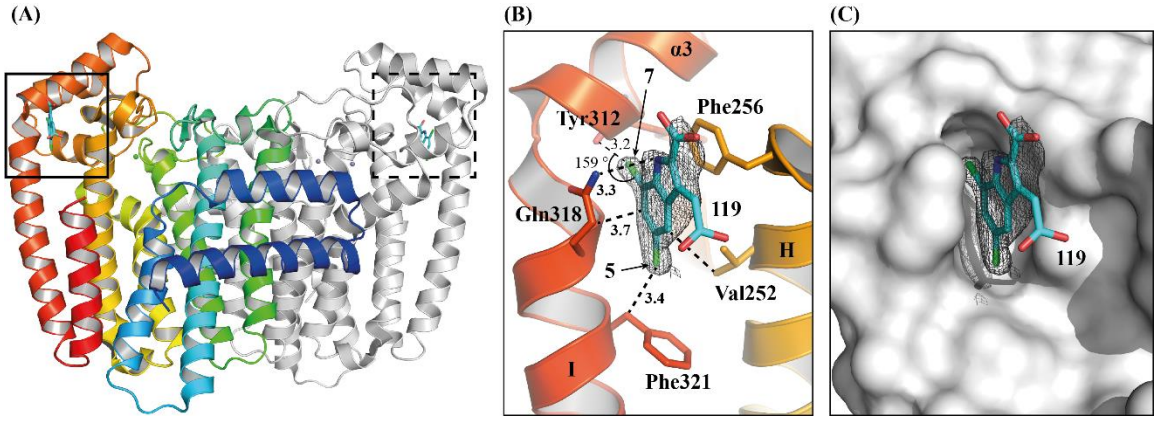



**Figure 28:** Crystal structure of TcFPPS in complex with compound **119** (PDB ID 6R08, this work). **(A)** Homodimer with bound ligands (top view, cartoon representation, subunit A coloured in gradient from blue (N-terminus) to red (C-terminus), subunit B coloured in grey). Ligands shown in stick representation. $Zn^{2+}$ ions are shown as green spheres. **(B)** Binding site of ligand **119**. The final $2F_o - F_c$ electron density map is contoured at 1.0 σ and represented as liquorice coloured mesh. Interactions are shown as dashed line. Distances are given in Å. **(C)** Binding site of ligand **119**. Protein shown in surface representation.

The aromatic core of **119** shows π-stacking with residue Phe256 of helix H and hydrophobic interactions with Val252 and Gln318 of helixes H and I, respectively. Gln318 also forms an H-bond to the indole nitrogen. A halogen bond with a length of 3.2 Å and an angle of 159 ° is formed by the benzothiophene's chlorine substituent in position seven and the carbonyl oxygen of Tyr312. At an occupancy of 0.61, the indole backbone with its chlorine substituents is fully encompassed by the contour at 3.0 σ of the $mF_o - DF_c$ difference electron density map (summary of density maps, Appendix, **Figure 65 (D – F)**). The carboxymethyl group is not defined

in the map. This is likely a consequence of its solvent exposure and a concomitant lack of directional and spatially constraining interactions. The resulting higher flexibility is reflected in higher B-factors further supporting this hypothesis. The binding poses and electron density maps are depicted in **Figure 28**.

The soaking experiments that resulted in the crystal structures of **93** and **119** were conducted as overnight soaks at 25 mM compound and 9% DMSO. Compound **118** was soaked under the same conditions and compound **97** even at a compound concentration of 50 mM. Only ligand **98** was soaked at a lower concentration of 10 mM (**Table 19**). Exact solubility of the compound in the crystallization buffer is not known, but at a protein concentration of 180 µM TcFPPS in the crystallization drop, the nominal compound concentration of 10 mM is estimated to a 56-fold excess.

### 5.2.2 Discussion

Structural comparisons between TcFPPS and hFPPS presented in this work revealed differences among residues in equivalent positions that form the allosteric site. Furthermore, a ConSurf model[479] based on an alignment of 200 FPPS homologues to TcFPPS revealed that the allosteric region is less conserved than the allylic and homoallylic binding site, which is surprising considering that all FPPSs are likely to be product inhibited as shown for hFPPS[209a]. A remarkable finding is that residue Phe50 in TcFPPS is an exception in an otherwise highly conserved position. Asparagine is the most common amino acid in this position and in rare cases histidine, tyrosine and glutamine were observed. The phenyl residue in this position is unique to TcFPPS and is the only residue that cannot contribute to H-bonding. In addition, on first glance it seems to block the allosteric pocket. The only homologues also showing an aromatic side chain at this positon are the FPPS of white-tufted-ear marmoset (UniProt ID F7GUQ3 and B0CM97) and rhesus macaque (UniProt ID F7FI27), which exhibit a histidine in this position and FPPS of rubber tree (UniProt ID A0A140GWW0), manioc (UniProt ID A0A140GWW3) and castor oil plant (UniProt ID B9S9Y3), which show a tyrosine in this position. Crystal structures of these FPPSs are not available, and it can only be speculated that the histidine residue has a similar effect on the pocket landscape as residue Phe50 in TcFPPS. Whether residue Phe50 undergoes conformational changes to give space for the accommodation of binders in this allosteric pocket has yet to be shown. A mutation at this position may affect product inhibition in TcFPPS by tacking impact on the binding at the FPP site. However, future structural and biochemical studies are required to determine these enzyme properties and their putative importance for the function of the protein. In case TcFPPS would be affected by product inhibition, it is likely that an induced-fit mechanism drives conformational rearrangement. In the case of FPP binding, this would also require a widening of the pocket, which is not necessarily important for inhibitors with a different chemical structure or

binding mode. In hFPPS the pocket is wide enough to accommodate FPP, but conformational transformation of residue Tyr10 is required to enable FPP binding[205].

Therefore, it is not surprising that the known allosteric hFPPS inhibitors, **93** and **119**, do not bind in the allosteric site of TcFPPS. Instead two distinct binding cavities were discovered on the protein surface that have not been described in the literature to date. They are remote from any other known binding site of TcFPPS and are not in any way related to the allosteric region. One of these sites is an elongated groove formed at the dimer interface that is large enough to accommodate two copies of fragment **93**. The fragment showed chemical shift changes with weak to medium strength in protein-observed NMR, which is expected for weakly interacting fragments even if they are measured at high concentrations[292, 300]. Compound **119** binds to a small cleft formed by helices H, I and α3 and protrudes partially into the solvent. Only the aromatic moiety forms interactions with the protein and both carboxyl groups of **119** are not involved in any interactions. Thus, the compound is lacking binding affinity generated by H-bonds. In contrast, the crystal structure of hFPPS complex shows that H-bond formation of this carboxyl groups are key interactions of the compound that exhibits an $IC_{50}$ of 6.0 µM against hFPPS[209a]. Whether binding of compounds **93** and **119** in these two cavities on the protein surface have an influence on the activity of TcFPPS has not been investigated in additional experiments but it seems unlikely given their binding position.

It is unclear why the compounds **97**, **98** and **118**, were not visible in the X-ray structure although chemical shift changes in protein-observed NMR had been of similar count and strength. Lead **98** was soaked at a lower concentration (**Table 19**), which still equalled a nominal excess of 56-fold over the protein concentration. Compounds **93** and **119** were already not fully occupied in the structural model, and potentially the protein-fragment interactions of compounds **97**, **98** and **118** are overall too weak, the $k_{off}$ rate is too high or the binding pose is not compatible with the geometry of TcFPPS in crystalline state. Although the current findings did not show conformational changes in the allosteric region to accommodate the aromatic allosteric inhibitors found for hFPPS, it does not necessarily mean that the targeted binding site is not accessible by other molecules. Binding can be conceivable with a well matching ligand or when following a co-crystallization experiments. It cannot be excluded that co-crystallization might have been superior to soaking experiments in the current case, in particular with regard to the enabling of the required conformational changes. This is exemplified in tRNA-guanine transglycosylase (TGT), where only co-crystallization disclosed the ligand-induced conformational changes[482]. However, on the basis of the aforementioned analysis of the allosteric region in TcFPPS and particularly due to the steric hindrance provoked by Phe50, it is dubious whether the known allosteric hFPPS inhibitors could bind to TcFPPS in a similar way and with a high affinity. As structural differences between pathogenic and human proteins can be exploited to engineer inhibitor selectivity[483], this work paves the way for future drug discovery campaigns.

## 5.3 FBS by NMR and hit follow up by X-ray crystallography

Structural superimposition of the allosteric sites of hFPPS and TcFPPS revealed pronounced differences in their architecture and, unsurprisingly, hFPPS inhibitors did not show high affinity binding towards the TcFPPS allosteric site (chapter **5.2**). FBS by NMR is a commonly used method for hit identification that probes all binding sites of a protein in solution[309b]. This method previously resulted in the identification of allosteric hFPPS inhibitors [209a] and was successfully applied also for TcFPPS in this work. Screening of two fragment libraries by ligand-observed NMR and validation by protein-observed NMR revealed 109 novel fragment hits for TcFPPS. 63 of them were subjected to X-ray crystallography, which revealed a novel active site-directed inhibitor of a non-BP scaffold, as well as a binder at the dimer interface. Counter screens against hFPPS were done in collaboration with Lena Muenzker, NIBR, Novartis Pharma AG, Basel, Switzerland. Furthermore, comparisons with TbFPPS, the enzyme investigated by Lena Muenzker, were made.

### 5.3.1 NMR

Ligand-observed NMR T1ρ[360b] and waterLOGSY[353b] experiments were conducted to screen 1336 fragments (in mixtures of 8 compounds) of the Novartis core library against TcFPPS (methods section **4.2.2**). The minimum criterion for hit selection was identification of a positive signal in the waterLOGSY experiments[353b]. A further cut-off criterion was the signal decrease observed in T1ρ[360b] experiments. In total, 81 fragments were identified that showed a signal broadening ranging from >10% to <20% and 216 fragments showed signal broadening >20%. The latter were considered as initial fragment hits corresponding to a hit rate of 19%. In **Figure 29**, waterLOGSY and T1ρ spectra for the fragment hit **CS-18** are depicted as an example. The hit list was reduced for validation in protein-observed NMR by comparing it against a list of fragment hits for TbFPPS, which was screened applying identical conditions. Comparison led to the selection of three sets of fragments: 47 fragment hits formed the intersection of the two hit lists, 46 were selective for TcFPPS and nine fragments were selective for TbFPPS. Rescreening these initial hits as singletons in ligand-observed NMR confirmed 65 fragment hits for TcFPPS. These were further subjected to protein-observed NMR experiments ([$^{13}C^{15}N$]-SOFAST-HMQC) for validation (methods section **4.2.3**).

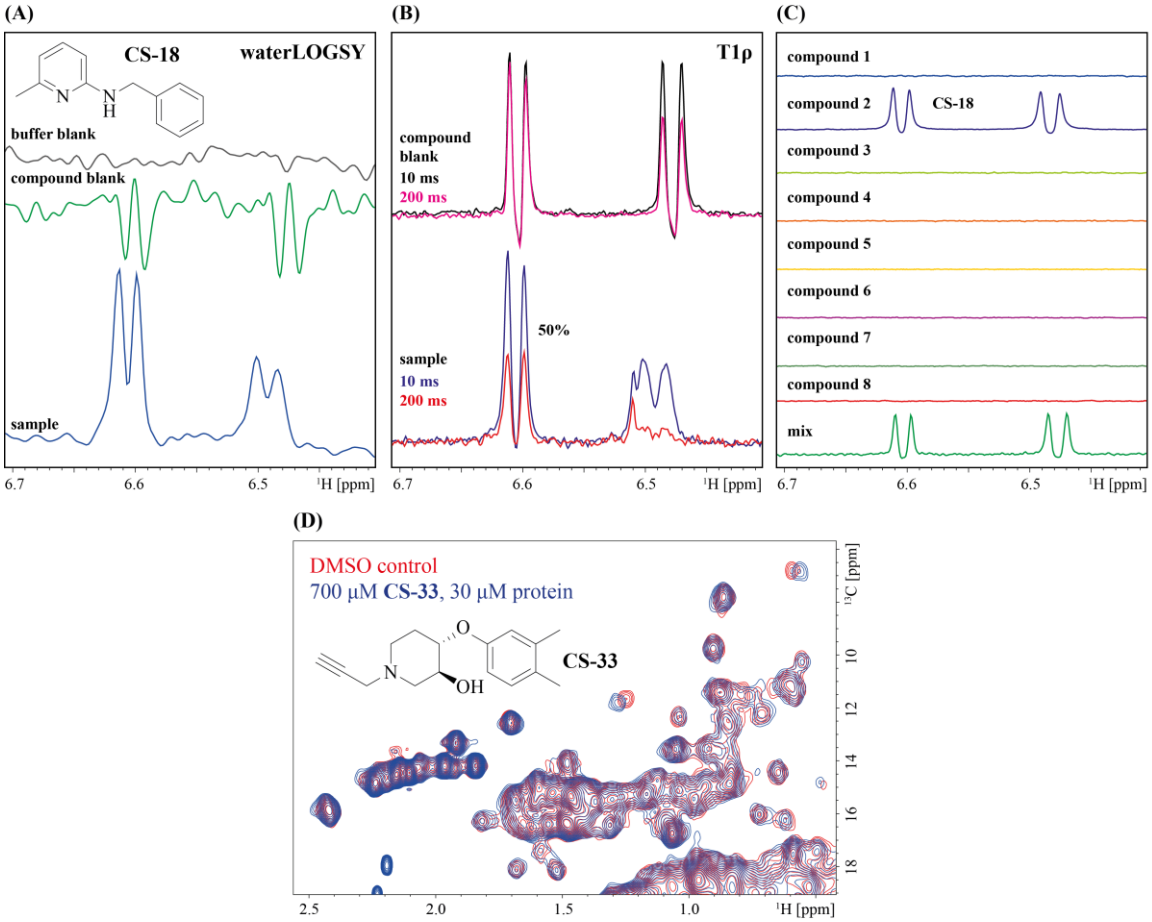



**Figure 29**: NMR experiments that led to hit identification of compound **CS-18** and **CS-33**. **(A)** Positive signal of **CS-18** in waterLOGSY. **(B)** Signal decrease of **CS-18** by 50% in T1ρ. **(C)** Mixture of eight compounds showing two doublets for compound **CS-18**. **(D)** Protein-observed NMR (([$^{13}$C$^{15}$N]-SOFAST-HMQC) of fragment **CS-33**. Here at 700 µM compound concentration. Nevertheless, validation was conducted at 1 mM compound concentration.

The same experiment was performed with TbFPPS. In parallel, counter screening against hFPPS in ligand- and protein-observed NMR was conducted. This led to a validated hit list for both trypanosomal FPPSs and enabled comparison with affinities measured for hFPPS. The hit distribution is visualized in a Venn diagram in **Figure 30 (A)**. Strikingly, TcFPPS has generally more binders (63 fragment hits) than TbFPPS (25 fragment hits) and TbFPPS has few unique hits when compared to TcFPPS. At this early stage of fragment screening, 27 hits are selective for trypanosomal FPPS as they did not show affinity to hFPPS. Notably, all experiments were carried out under identical experimental conditions. Publicly known fragments are listed in **Table 32** and their chemical structures are depicted in **Figure 66** in the Appendix.

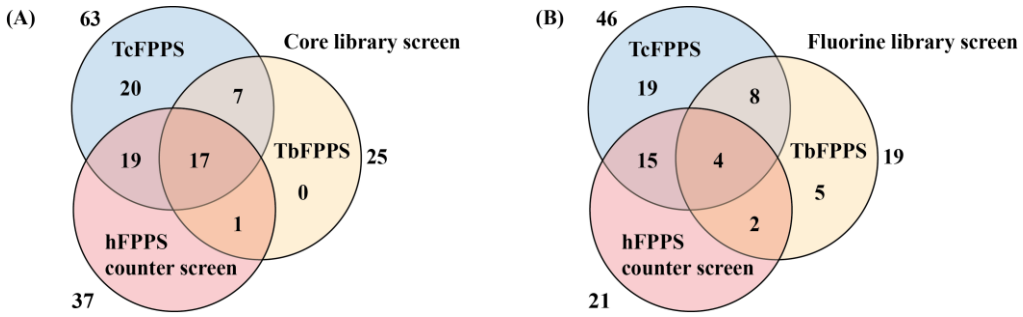



**Figure 30**: Core library screen and fluorine library screen reveal selectivity. **(A)** Core library screen (1336 compounds): Hits identified were validated in ligand-observed NMR after their identification in ligand-observed NMR (hit criteria: effect in T1ρ ≥20% for TcFPPS and hFPPS and >20% for TbFPPS, positive read out in waterLOGSY for all three FPPSs). (**B**) Fluorine library screen (482 compounds): Hits identified were validated in protein-observed NMR after successful testing as singletons on TcFPPS or TbFPPS and counter screen on hFPPS. Hit criteria for signal decrease in $^{19}$F CPMG NMR experiments: ≥40% for TcFPPS, ≥20% for TbFPPS and ≥30% hFPPS.

Screening of the Novartis fluorine library (482 compounds in mixtures) also resulted in the identification of fragment hits for TcFPPS (methods section **4.2.2**). Setting the cut-off criterion for the signal decrease observed in $^{19}$F CPMG NMR experiments to >40% resulted in 52 fragment hits for TcFPPS corresponding to a hit rate of 11%. Repeating the ligand-observed NMR experiments with these 52 hits as singletons and testing in protein-observed NMR experiments ([$^{13}$C$^{15}$N]-SOFAST-HMQC) validated 46 fragment hits for TcFPPS (methods section **4.2.3**). Comparisons with the hits identified for TbFPPS and counter screenings against hFPPS resulted in unique, pairwise and triple binders which are depicted in a Venn diagram in **Figure 30 (B)**). Whilst TcFPPS again shows a higher number of hits than TbFPPS, the number of fragment hits shared by all three FPPSs is smaller when compared to the screen of the core library. Once more, a large number of hits selectively binds to TcFPPS and half of the hits is selective for trypanosomal FPPS. Publicly known fragments are listed in **Table 33** and their chemical structures are depicted in **Figure 67** in the Appendix.

These findings are remarkable, because TcFPPS and TbFPPS are close homologues with approx. 69.0% amino acid sequence identity and 83.2% similarity (**Table 20** and Appendix, **Figure 68**). Alignment to human FPPS reveals an identity to TcFPPS and TbFPPS of 35.13% and 37.13%, respectively (**Table 20** and Appendix, **Figure 69**). Whilst the overall protein architecture is the same for all FPPS enzymes (see chapter **1.4**), one and two insertional loops are found in TcFPPS and TbFPPS, respectively. In TcFPPS, this is an insertion loop of 11 residues, which is located between helices F and G and is formed by residues Lys179-Thr189 with a reverse turn at Pro182[162b, 210]. In TbFPPS, the loops are a 10-residue insertion and an 11-residue insertion formed by residues Ser65-Asp74 and Lys184-Thr194. The latter corresponds to the 11-residue insertion in TcFPPS[204] and is unique to trypanosomal FPPS[162b, 210] (Appendix, **Figure 70**).

**Table 20**: Sequence identity and similarity between TcFPPS, TbFPPS and hFPPS.

| Alignment parameters TcFPPS / TbFPPS[a] | (%) | Enzyme | Identity matrix[b] (%) | | |
|---|---|---|---|---|---|
| | | | TcFPPS | TbFPPS | hFPPS |
| Identity | 69.0 | TcFPPS | 100.00 | 70.36 | 35.28 |
| Similarity | 83.2 | TbFPPS | 70.36 | 100.00 | 37.13 |
| Gaps | 1.9 | hFPPS | 35.28 | 37.13 | 100.00 |

[a] Alignment of TcFPPS and TbFPPS using Emboss Needle, applying the Needleman-Wunsch algorithm[478] (Appendix, **Figure 68**).
[b] Identity matrix of TcFPPS, TbFPPS and hFPPS, which was generated using Clustal (v.12.1) (Appendix, **Figure 69**).

Another notable difference between the three enzymes is that TcFPPS and hFPPS are tight homodimers with a non-observable domain exchange rate in solution, while TbFPPS does exhibit domain exchange in solution (MS studies, Oscar Alba-Hernandez, oral communication). When comparing the chemical structures of the fragment hits, no overrepresentation of certain parent scaffolds or subset of functional groups could be observed (see Appendix, **Figure 66** and **Figure 67**)).

### 5.3.2 Follow up of validated fragment hits

Follow up of the validated fragment hits of the Novartis core library included $K_d$ estimation by NMR spectroscopy (methods section **4.2.4**), as well as soaking and co-crystallisation experiments in order to elucidate their binding position (methods sections **4.3.5** and **4.3.6**). [$^{13}C^{1}H$]-SOFAST-HMQC spectra of a titration series of five of the publicly known compounds that showed strong chemical shift changes in the protein-observed NMR were recorded. Chemical shift changes of selected spectra were plotted and the $K_d$ values were calculated and averaged (**Figure 31**, **Table 21**). $K_d$ values ranged from 61 µM to 1308 µM. Compound **CS-18** showed the best $K_d$ value and was also among the compounds with the largest decrease observed in T1ρ experiments (Appendix, **Table 32**). Based on the estimated $K_d$ values and the heavy atom count (HAC), the ligand efficiencies (LE) were calculated ($LE = RT \cdot \ln(K_d) \cdot HAC^{-1}$)[430a] (**Table 21**). They ranged from 0.23 kcal · mol$^{-1}$ to 0.44 kcal · mol$^{-1}$ with fragments **CS-22** and **CS-18** exhibiting the strongest LE's (0.44 kcal · mol$^{-1}$ and 0.38 kcal · mol$^{-1}$, respectively).

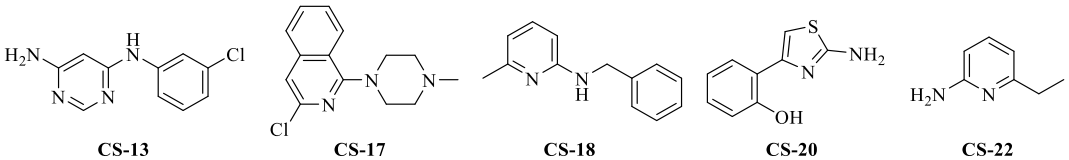



**Figure 31**: Chemical structure of fragment hits listed in **Table 21**.

**Table 21**: $K_d$ estimation by NMR spectroscopy and ligand efficiencies for five selected fragment hits.

| Compound | MW (Da) | Concentration range (µM) | Titration points (count) | Signals considered (count) | Mean $K_d$ value (µM) | HAC (count) | Ligand efficiency (kcal · mol$^{-1}$) |
|---|---|---|---|---|---|---|---|
| **CS-13** | 220.66 | 30 – 4000 | 6 | 5 | 611 | 15 | 0.29 |
| **CS-17** | 261.76 | 51.2 – 5000 | 5 | 2 | 1053 | 18 | 0.23 |
| **CS-18** | 198.27 | 25 – 2000 | 7 | 7 | 61 | 15 | 0.38 |
| **CS-20** | 192.24 | 51.2 – 5000 | 6 | 6 | 1805 | 13 | 0.29 |
| **CS-22** | 122.17 | 51.2 – 5000 | 6 | 4 | 1308 | 9 | 0.44 |

Information about the binding site of the ligands could not be extracted from the protein-observed NMR experiments, because resonance assignment was not available. Therefore, X-ray crystallography was conducted for binding site determination. All 63 validated fragment hits of the core library screen were subjected to co-crystallization experiments (methods section **4.3.6**). Experiments were started with the fragment hits unique for TcFPPS at a compound concentration of 2.6 mM (2.5% DMSO). In crystallization trials with the fragment hits overlapping for TcFPPS, TbFPPS and hFPPS, or overlapping between TcFPPS and hFPPS, a compound concentration of 5.2 mM (5% DMSO) was chosen. Although crystals grew in the presence of most compounds and data sets of crystals co-crystallized with 58 of the compounds exposed could be successfully collected, data processing and visual inspection of the resulting electron density maps did not reveal protein-ligand complexes.

However, soaking experiments were more successful. The set-up of a soaking protocol was previously described in chapter **5.1**. Soaking experiments were started with the compounds listed in **Table 19** and further extended to 40 validated fragment hits. Fragment **CS-18,** termed **JNE** from hereon, formed a protein-ligand complex that was observed in the $mF_o - DF_c$ difference electron density map (summary of density maps, Appendix, **Figure 71 (A-C)**). Data processing and refinement resulted in a 3D structure with good data collection and refinement statistics that were comparable to those of the TcFPPS apo structure (PDB ID 6R04) (chapter **5.1**, Appendix, **Table 29**). The ligand occupancy was refined to 0.86 and a structural model of the TcFPPS-**JNE** complex with a diffraction limit of 1.57 Å was deposited under PDB ID 6R05 (Appendix, **Table 29**). Fragment **JNE** binds at the dimer interface and thus, it forms interactions with subunit A and subunit B of the protein (**Figure 32 (A)**)). As for all previously described TcFPPS crystal structures in this work, the two TcFPPS monomers are related by crystallographic two-fold symmetry. A large dimer interface is formed along the twofold-symmetry axis corresponding to the tightly coupled physiological homodimer[211]. The key interactions formed by the nitrogen atoms of fragment **JNE** are H-bonds with residue Glu183 of subunit B, which is located in the insertion loop. In addition, the ligand shows π-stacking with residue Phe116 and

van der Waals interactions with residues Pro111 and Lys110. These two residues are located in the loop turn D-E of subunit A (**Figure 32 (B)** – **(D)**). Despite the interesting location of the binding site at the dimer interface, the site is otherwise exposed to the solvent and is far from the active centre of TcFPPS. Testing compound **JNE** on hFPPS in protein-observed NMR revealed that **JNE** also interacts with hFPPS, but chemical shift changes were minimal and an estimation of the $K_d$ value was not possible. This is in agreement with structural features of hFPPS, which lacks the insertion loop and therefore has no equivalent binding site.

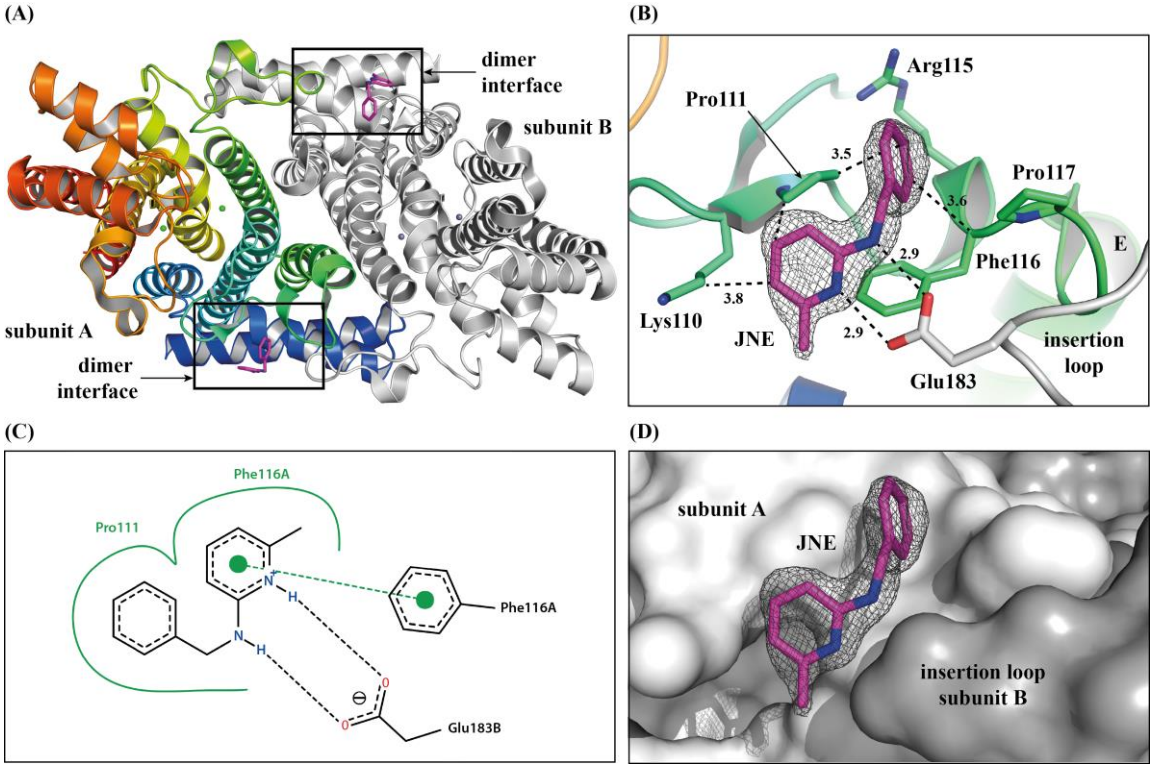



**Figure 32**: Crystal structure of TcFPPS in complex with **JNE** (**CS-18**) (PDB ID 6R05, this work). **(A)** Homodimer with bound ligands (top view, cartoon representation, subunit A coloured in gradient from the N-terminus (blue) to the C-terminus (red), subunit B coloured in grey, ligands shown in stick representation coloured in pink. $Zn^{2+}$ ions are shown as green spheres). **(B)** Binding site of ligand **JNE**. Final $2F_o - F_c$ electron density map is contoured at 1.0 σ and represented as liquorice coloured mesh. Residues forming the binding site are represented as sticks. Interactions are shown as black coloured dashed line. Distances are given in Å. **(C)** 2D structure diagram of ligand **JNE** interacting with TcFPPS. Diagram was generated using PoseView[481]. **(D)** Binding site of ligands **JNE**. Monomer A and B shown in surface representation. Final $2F_o - F_c$ electron density map is contoured at 1.0 σ and represented as liquorice coloured mesh.

Since no further protein-ligand complexes were found by classical data processing and visual inspection of the initial $mF_o - DF_c$ and $2mF_o - DF_c$ density maps, the data sets were additionally analysed by Pan-Dataset Density Analysis (PanDDA). This approach uses statistical methods to identify binding events in a batch of data sets rather than analysing the reflections of single data sets. The developer of PanDDA described protein-ligand complexes as a crystallographic superposition of a ground state (apo form) and a bound-state (any kind of

additional binding event). These two states exist at the same time on a pro rata basis. Thus, the measured electron density is only fully explained by an ensemble model. When running the PanDDA software, the first analysis step is to compute a background electron density estimate as a mean of ground-state measurements based on apo structures and data sets without binding events[393]. In a second step, significant changes from the mean are identified by a weighted subtraction of the background from each electron density map. The resulting partial-difference map is termed event map and corresponds to the bound fraction in the crystal, i.e. a ligand binding event. As PanDDA reveals regions of an individual data set that represent a deviation from the mean, it allows sensitive detection of binding sites[398].

Electron-density background subtraction was conducted using 302 TcFPPS data sets of apo crystals, co-crystals and soaked crystals (data set parameters and quality, Appendix, **Figure 72**). Analysis identified fragment hit **CS-33** as an active site-directed binder that was overlooked by classical data processing and manual inspection of the electron density maps (chemical structure, **Figure 33 (I)**). Whilst partial occupancy and presence of two rotamers of the side chain of residue Tyr94 led to obscured classical electron density maps, PanDDA maps clearly show the binding event of **CS-33**, which is termed **JMN** from hereon. In the crystal structure of apo TcFPPS (PDB ID 6R04), which was elucidated as part of this work, the side chain of Tyr94 is present in two rotamers at nearly equal occupancies (**Figure 33 (A)**), however, **JMN** can only bind when Tyr94 is present as rotamer B.

Conformational changes of Tyr94 have previously been reported upon binding of N-BPs with longer alkyl-chain substituents[170a, 211]. The initial $2mF_o - DF_c$ electron density map shows overlapping density for rotamer A of Tyr94 with the ligand and full coverage of the amino acid and the ligand when contoured at 0.5 σ and 1.0 σ (**Figure 33 (B, C)**). When contouring the initial $2mF_o - DF_c$ electron density map at 1.5 σ the ligand is poorly defined (**Figure 33 (D)**) and also in the $mF_o - DF_c$ difference electron density map contoured at 3.0 σ the ligand is not fully defined (**Figure 33 (G)**). When looking at the PanDDA event map, which represents the bound fraction, and at the map of the ground state, the binding event can be easily identified (**Figure 33 (E, F)**). Nevertheless, the structural model could be refined by splitting Tyr94 in rotamer A and B and accompanying the latter with the ligand to avoid clashes when running a refinement (**Figure 33 (H)**). The ligand occupancy was refined to 0.63 and a structural model of the TcFPPS-**JME** complex with a diffraction limit of 1.56 Å was deposited under PDB ID 6R06. Data collection and refinement statistics are given in **Table 29** in the Appendix.

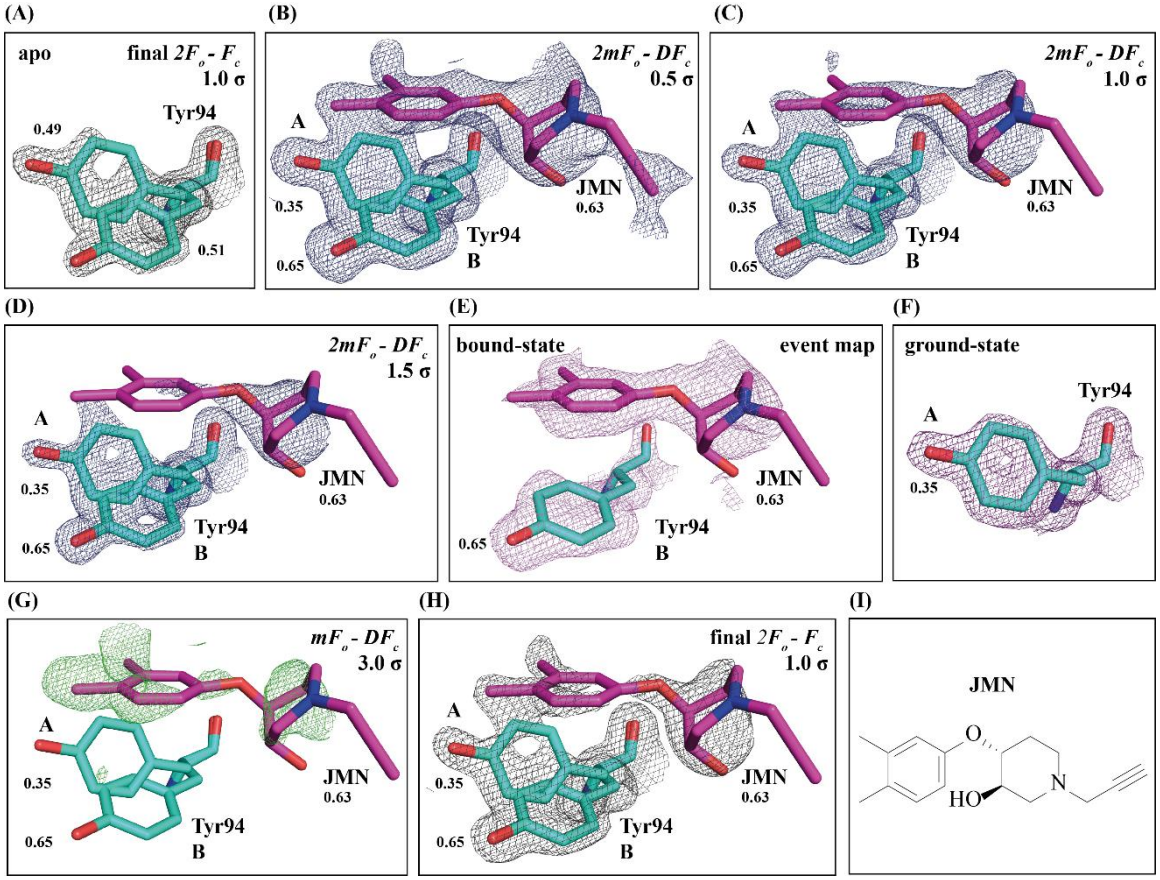



**Figure 33**: PanDDA event maps led to identification of fragment binder **JMN** (**CS-33**) (PDB ID 6R06, this work). **(A)** Crystal structure of apo TcFPPS (PDB ID 6R04). Residue Tyr94 depicted in its final $2F_o - F_c$ electron density map contoured at 1.0 σ. Occupancies are indicated. **(B)** – **(D)** Fragment **JMN** and residue Tyr94 depicted in the inital $2mF_o - DF_c$ electron density map contoured at 0.5 σ, 1.0 σ and 1.5 σ, respectively. Occupancies are indicated. **(E)** Event map contoured at a level equivalent to 2.0 σ reveals that Tyr94 is only found in one conformation in the bound state. **(F)** In the ground state map only rotamer A of Tyr94 is defined. **(G)** Fragment **JMN** and residue Tyr94 depicted in the initial $mF_o - DF_c$ electron density map (green mesh) contoured at 3.0 σ. **(H)** Fragment **JMN** and residue Tyr94 depicted in the final $2F_o - F_c$ electron density map contoured at 0.5 σ, 1.0 σ and 1.5 σ, respectively. Occupancies are indicated. **(I)** Chemical structure of **JMN**.

Ligand **JMN** binds to the allylic binding site of TcFPPS (**Figure 34**), which is the binding site of its cognate substrate DMAPP (see chapter **1.4**). The pocket is located between helix D and F, is terminated by the dimer interface (helix E of subunit B) (**Figure 34 (C, D))** and opens up towards the homoallylic (IPP) binding site (**Figure 34 (E)**). The base of the pocket is formed by the aromatic and polar residues Tyr94, Thr163 and Tyr211. The sides of the pocket are formed by the backbone of Tyr94 and Thr163, residues Leu95, Lys 207 and Thr208 and Ile129 of the opposing dimer mate. The top of the binding site is lined by residues Gln167, Asp98 and Asp250, the latter of which coordinate $Zn^{2+}$ ions (**Figure 34 (B)**). Fragment **JMN** shows π-stacking with Tyr94, H-bonding with Tyr211 and van der Waals interactions with residues Thr163 and Thr208 as well as with Ile129 of the opposing dimer mate (**Figure 34 (B)**). A sulphate ion was modelled in the coordination sphere of one of the $Zn^{2+}$ ions (Zn2), but the angle for H-bonding with the ligand is

not ideal. Superimposition of TcFPPS-**JMN** and TcFPPS in complex with N-BPs (PDB ID 3IBA[211], 4DWG[170a] and 3ICM[211]) shows that **JMN** occupies the same space as the side chains of the N-BPs (**Figure 34 (F)**).

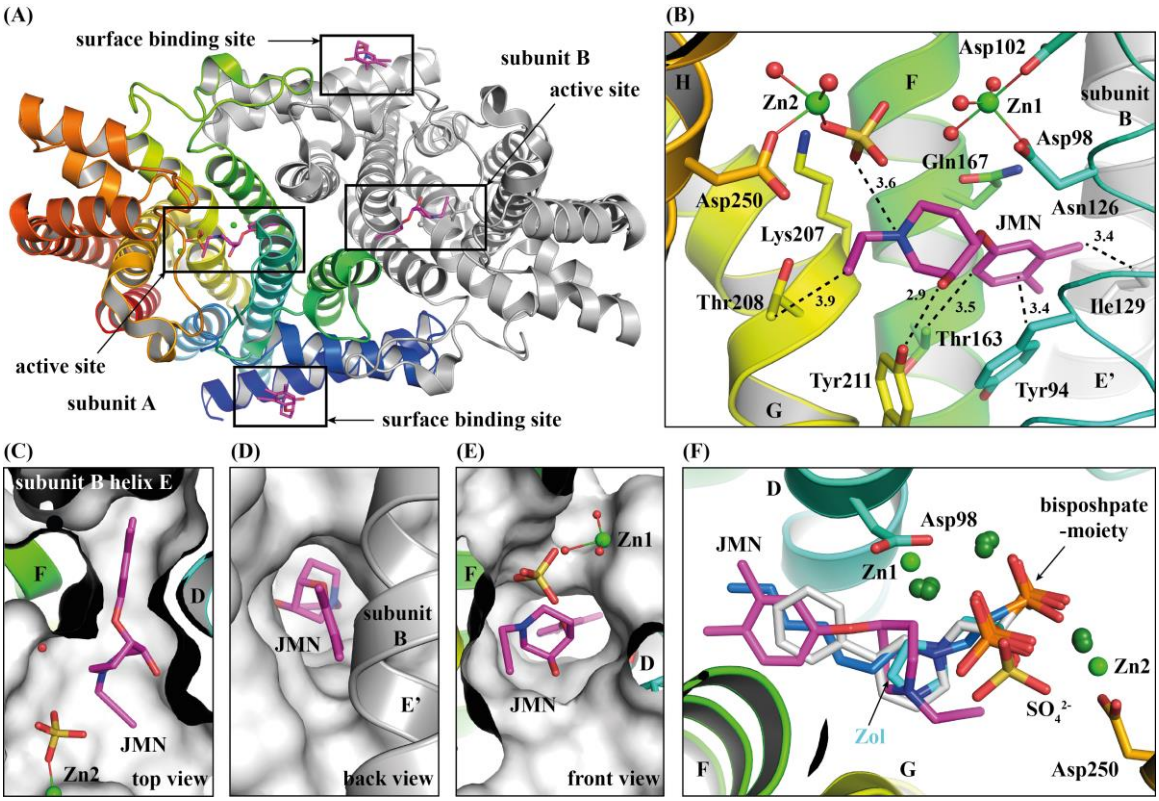



**Figure 34**: Crystal structure of TcFPPS in complex with **JMN** (PDB ID 6R06, this work). **(A)** Homodimer with bound ligands (top view, cartoon representation, subunit A coloured in gradient from the N-terminus (blue) to the C-terminus (red), subunit B coloured in grey. Ligands shown in stick representation coloured in pink. $Zn^{2+}$ ions are shown as green spheres. **(B)** Ligand **JMN** binding in the active site. Residues forming the binding site are represented as sticks. Interactions are shown as black coloured dashed line. Distances are given in Å. **(C)** – **(E) JMN** in its binding pocket **(F)** Superimposition of TcFPPS in complex with **JMN**, ZOL (PDB ID 3IBA[211]) and N-BPs with longer alkyl side chains (PDB IDs 4DWG[170a] and 3ICM[211]).

### 5.3.3 Discussion

Fragment screening by NMR successfully identified TcFPPS binders based on a novel parent scaffold. The results of the NMR screen was different for the two trypanosomal FPPSs, TcFPPS and TbFPPS. Notably, hit rates for TcFPPS were higher than for TbFPPS, however, an underlying structural explanation for this observation is not apparent. Furthermore, the hit lists differed and counter screening indicated TcFPPS selectivity at an early stage. Therefore, binders that were either selective for TcFPPS or TbFPPS were identified. Nevertheless, hits that are not selective at the stage of fragment screening can potentially be optimized into selective binders by

medicinal chemistry efforts, exploiting the differences in local arrangement and composition of residues in both proteins.

Although a large number of compounds was screened, subsequent selection and validation led to a moderate number of TcFPPS fragment hits. Subsequent crystallization experiments resulted only in two crystal structures. The selection criteria employed for the preselection of fragments for testing in protein-observed NMR might have had an impact on this outcome. In this work, comparison with the hit lists of TbFPPS prioritized hits that were either selective for one of the trypanosomal FPPSs or formed their intersection, which is not congruent with the hits that showed the strongest effects in T1ρ experiments. Focusing on these hits might have resulted in a higher hit rate in crystallisation experiments. Furthermore, it cannot be ruled out that experimental soaking conditions were incompatible with at least a subset of compound-protein interactions, e.g. the protonation state of the compounds might have changed due to differences in the pH. However, in the past, NMR, SPR, and/or DSF have proven to be valuable tools for fragment hit selection and prioritization for subsequent X-ray crystallography experiments[378, 384a]. However, biophysical screening methods were shown to not reliably predict the majority of X-ray binders. Vice versa, X-ray crystallography does not necessarily provide 3D structures of proteins in complex with a compound selected by other biophysical methods, although it is a highly sensitive method[315a]. Schiebel *et al*.[384] showed that it is a reliable technique for fragments characterized even by low affinities and advocated X-ray crystallography as primary screening method. This study demonstrates that sophisticated software tools, such as PanDDA, help to identify weakly bound fragment hits that are difficult to identify by manual data analysis. This tools will also help unexperienced crystallographers to reliably identify hits.

Based on its binding mode, fragment **JNE** is unlikely to interfere with enzymatic function of TcFPPS, however, additional experiments to determine this hypothesis are warranted. The crystal structure of TcFPPS in complex with **JMN** gives new impulses for the discovery of novel active site TcFPPS inhibitors, although more new crystal structures would be advantageous to start a structure-based lead discovery campaign.

## 5.4 FBS by X-ray crystallography – The power of PanDDA

X-ray crystallography is a very sensitive method[315a] and was shown to be suitable as hit identification tool in fragment screening[304, 384a, 385]. Therefore, FBS of TcFPPS by X-ray crystallography was used to discover novel ligands and identify potential allosteric ligand-binding sites within the enzyme. Two of such screening campaigns were conducted: The first was run in collaboration with Elliot Nelson, PhD student of Frank von Delft, at the XChem facility, which is located at the Diamond Light Source (DLS), Harwell, UK[400]. The second campaign was run in the HTX lab of Jose Marquez at the EMBL outstation and at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Remote access to this facility was funded under the infrastructure for NMR, EM and X-rays for translational Research (iNEXT) grant agreement number ID 653706 (project number 2847). Both campaigns identified fragment hits in various binding sites spread over the entire protein. The key findings were 10 fragment binders in the active site, 10 fragment binders in the allosteric site and a fragment binding at a novel binding site, which is in close proximity to the active site.

### 5.4.1 Results of the XChem campaign

All experiments, including crystallization trials, soaking experiments and data collection were conducted on site. Therefore, all necessary materials, including the protein formulation, seed solution, buffers and other reagents were shipped to the UK. Crystallization experiments and later large-scale protein-ligand structure determination were assisted by the *XChemExplorer* graphical workflow tool (XCE)[392a]. Crystal plates were set up according to the protocol developed at Novartis laboratories in Basel, Switzerland (methods section **4.4** and chapter **5.1**).

685 apo TcFPPS crystals were individually soaked with 406 compounds from the Diamond-SGC poised library (DSPL) and 279 compounds from the Keymical fragments library (KFL) by EDELRIS. They were soaked for a mean soaking time of 3 h 32 min at 74.5 mM (15% DMSO) and 37.25 mM (15% DMSO), respectively. Whilst the DSPL contains fragments that are poised for straight forward follow up by chemical elaboration[336], the KFL is a 3D-enriched fragment library that contains natural product-like compounds[435]. 666 crystals showed diffraction after the soaking procedure and diffraction data were recorded at beamline I04-1 at the DLS. 573 crystals yielded diffraction data that could be successfully processed (**Table 22**). In addition, eight data sets of apo TcFPPS crystals and 19 data sets of TcFPPS crystals soaked with varying amounts of DMSO were successfully processed, resulting in a total number of 600 data sets suited for analysis (**Table 22**). All crystals belonged to the hexagonal space group $P6_122$, and had unit-cell parameters of approx. a = b = 58 Å, c = 396 Å, and α = β = 90 ° and γ = 120 ° with one TcFPPS monomer per asymmetric unit. The diffraction limit ranged from 1.40 Å to 3.26 Å with an average

diffraction limit of 1.77 Å. 14% of crystals had a diffraction limit ≤1.5 Å and 58% had a diffraction limit ≤1.7 Å. This large number of data sets, their crystallographic homogeneity and good average diffraction limit enabled statistical data analysis by Pan-Dataset Density Analysis (PanDDA)[393] (**Table 22**, Appendix, **Figure 73**). The concept of this data analysis tool was previously described in chapter **5.3.2**.

**Table 22**: Key parameters of the data sets suited for analysis.

| Variable | Value |
|---|---|
| Space group | $P6_122$ (No. 178, International Tables for Crystallography[377b]) |
| Mean unit cell axis a / b / c (Å) | 57.82 (0.10) / 57.82 (0.10) / 395.96 (0.76) |
| Range of diffraction limit (Å) | 1.40 – 3.26 |
| Mean diffraction limit (SD) (Å) | 1.77 (0.25) |
| Mean $R_{free}$ / $R_{work}$ (SD) | 0.261 (0.024) / 0.216 (0.019) |
| Mean ($R_{free}$ - $R_{work}$) (SD) | 0.045 (0.013) |

A ground-state model was built based on 60 data sets that did not indicate any additional binding events. Running PanDDA for all 600 data sets with this ground-state model identified 297 putative binding events that were spread of the whole protein (**Figure 35 (A)**) and distributed over 208 data sets due to multiple binding events in some of the data sets. The events are clustered arbitrarily into 16 non-overlapping sites (S1 – S16) (**Figure 35 (A)** and **(B)**). Based on visual inspection of the events, 85 ligands were manually modelled. Iterative rounds of model building and refinement confirmed 54 ligands that were modelled in 51 events. As 12 structures contain multiply bound copies of the same fragments, the number of ligands modelled corresponds to 35 unique fragment hits that are spread over nine binding sites (**Figure 35 (B)**). Ligand occupancy ranged from 0.53 to 1.00 with an average occupancy of 0.73 (Appendix, **Table 34**), which is comparably high, as PanDDA was shown to identify ligands to a minimal occupancy of 0.26[398]. The diffraction limit of the 35 refined structural models ranged from 1.41 Å to 2.20 Å with a mean diffraction limit of 1.64 Å, hence showing a slightly better mean diffraction limit when compared to all data sets (1.77 Å) (**Figure 35 (C)**). PanDDA identified events in nearly all less resolved structures, none of which led to modellable structures.

The key finding of the fragment screen are the ligands, which are bound to sites that are buried inside the protein. Five fragment hits were identified in the allosteric region of TcFPPS, which was previously described for TcFPPS (chapter **5.2**). Additionally, one ligand in a small and novel cavity was discovered. The event was clustered together with the allosteric site binders. For clarity this binding site was renamed and referred to as binding site SX from hereon. Furthermore, seven ligands with non-BP scaffolds were identified in the active site of TcFPPS. An event on the protein surface was clustered together with the active site binders. For clarity this binding site was

renamed and referred to as binding site SY from here onwards. All remaining sites, S1, S2, S4, S5, S11, S12 and S16, are cavities on the surface and further descriptions will focus on the ligands bound to the active site and to the allosteric region. Interestingly, DSPL fragments account for a 100% of fragments buried by the protein, as well as the majority of all hits (83%). The overall hit rate for the DSPL and the KFL equals 8.8% and 2.5%, respectively.

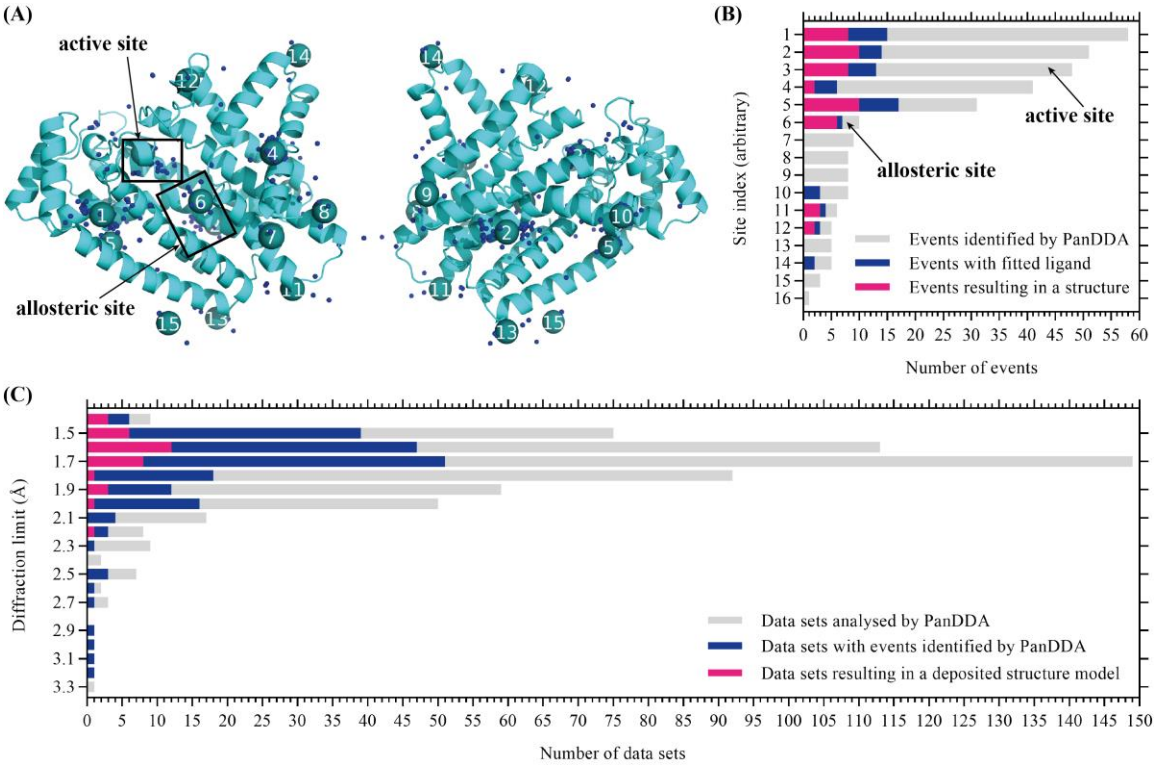



**Figure 35**: Data analysis with PanDDA revealed 35 fragment hits. **(A)** Binding events clustered into 16 non-overlapping binding sites (front view (left), back view (right)). **(B)** Histogram of arbitrary binding sites with events. Superimposition of events: 297 in total, 85 fitted ligands and 51 part of a structural model. **(C)** Histogram of X-ray diffraction limit of soaked TcFPPS crystals. Superimposition of data sets: 600 analysed, 206 with an event, and 35 structural models.

An overview of all 35 crystal structures is depicted in **Figure 36**. The chemical structures of all ligands, their MW, occupancy and binding position in the crystal structure, coverage in the final $2mF_o - DF_c$ electron density maps and residual $mF_o - DF_c$ difference electron density maps and the PanDDA event map that lead to their identification are depicted in **Figure 74** in the Appendix. The bound-state structural models of the TcFPPS-ligand complexes (PDB IDs 5QPD – Z, 5QQ0 – 9, 5QQA – B) and a TcFPPS ground-state model (PDB ID 5QQC) containing all structure features of the unbound models were deposited in the PDB. Data accompanying the PDB deposition, such as log files of the data processing and PanDDA event maps were uploaded to the open archive Zenodo (DOI 10.5281/zenodo.2649077). Data collection and refinement statistics are given in **Table 35** in the Appendix. A summary of density maps for the ligands

discussed in detail in this chapter, including initial $mF_o - DF_c$ difference electron density maps and $2mF_o - DF_c$ electron density maps, are given in **Figure 75** in the Appendix.

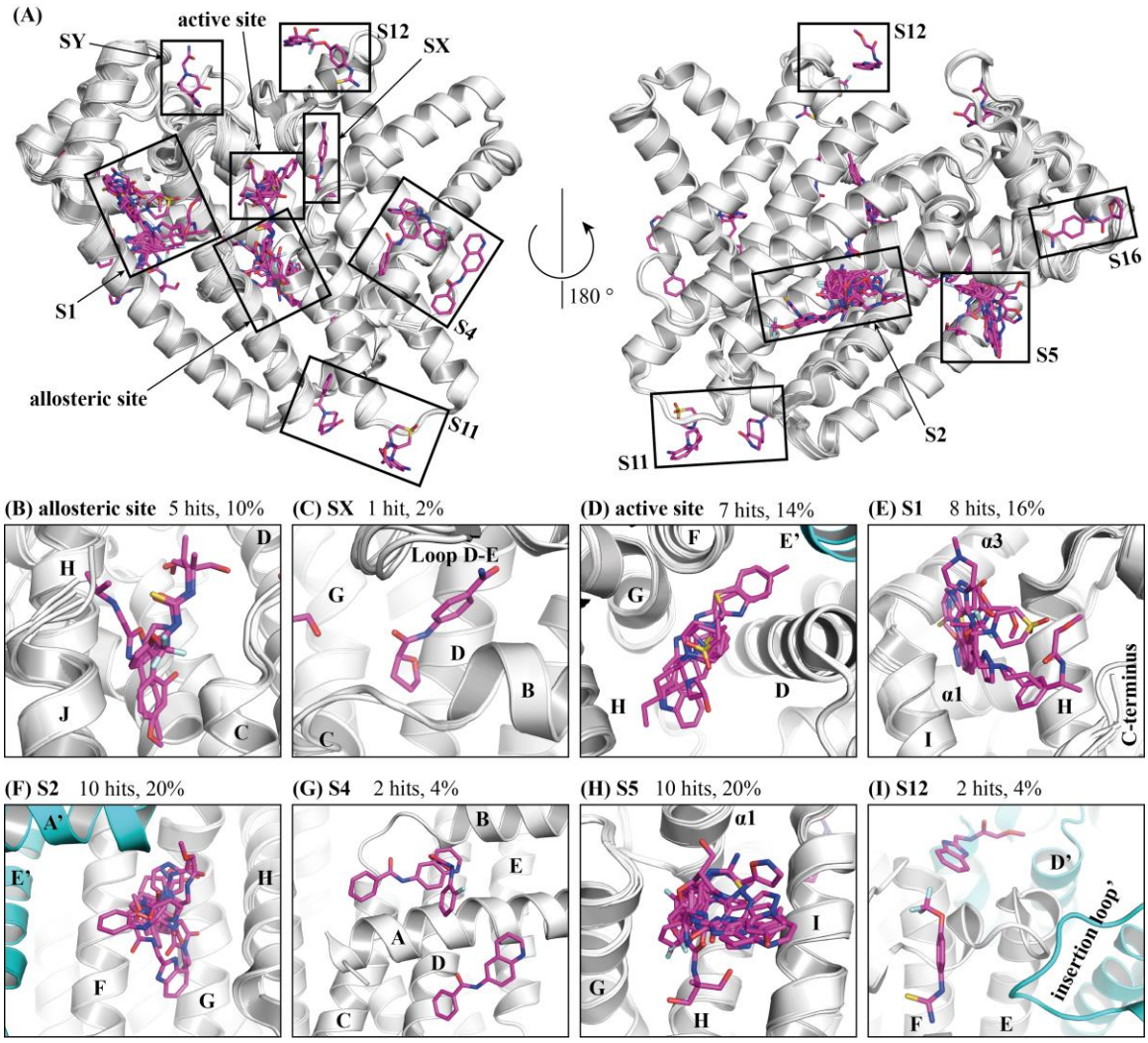
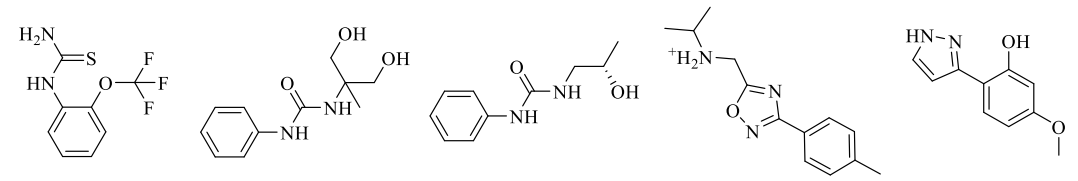



**Figure 36**: Overview of crystal structures obtained by FBS by X-ray crystallography at the XChem facility (PDB IDs 5QPD – Z, 5QQ0 – 9, 5QQA – B, this work). **(A)** Overview of ligands binding across TcFPPS. Front view (left) and back view (right). Binding sites are indicated. **(B)** Fragment cluster at the allosteric site. **(C)** Site SX (renamed, hit was clustered as allosteric site binder). **(D)** Fragment cluster at the active site. **(E) – (I)** Fragment binding sites distal from the active and allosteric site: S1, S2, S4, S5 and S12, respectively. Number of hits and percentage of hits at that site are indicated.

### Allosteric site binders

To identify allosteric binders, the area between helices C, G, H and J and the B-C and H-I loop was examined and the five fragments, **LV4**, **4YV**, **M0J**, **LT7** and **GQM**, were observed (**Figure 37**, Appendix, **Figure 74 (1 – 5)** and summary of density maps **Figure 75 (1 – 5)**).



**LV4**
MW 236.21, occ. 0.67
1-BDC 0.44
5QPP

**AYV**
MW 244.26, occ. 0.72
1-BDC 0.33
5QPZ

**M0J**
MW 194.23, occ. 0.66
1-BDC 0.37
5QPL

**LT7**
MW 232.31, occ. 0.60
1-BDC 0.25
5WPD

**GQM**
MW 190.20, occ. 0.54
1-BDC 0.24
5QPI

**Figure 37**: Chemical structure of the fragment hits identified for the allosteric region. Compounds are derived from the DSPL. MW in Da, ligand occupancy, 1-BDC and PDB ID are indicated.

Three of the fragments, **LV4**, **AYV** and **M0J**, have similar chemical structures and have highly overlapping binding positions. Their key interaction is π-stacking with residue Phe50 (**Figure 38 (A)** – **(D)**) that is found in a conformation that blocks the allosteric pocket by steric hindrance (previously described in chapter **5.2**). The three ligands bind in close proximity to the binding site of IPP, a natural substrate of FPPS. In the absence of IPP, a sulphate ion is bound there, which forms H-bonds with ligands **LV4**, **AYV** and **M0J**.

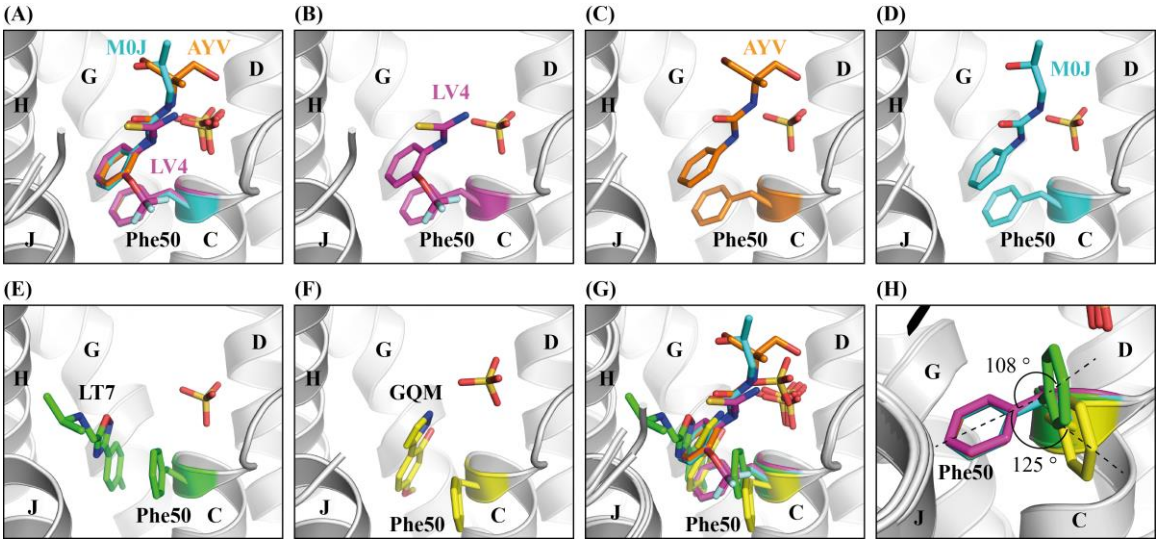



**Figure 38**: Crystal structures of allosteric site binders of TcFPPS. **(A)** Crystal structures of TcFPPS in complex with **LV4**, **4YV** and **M0J** (PDB IDs 5QPP, 5QPZ and 5QPL, this work). **(B)** Ligand **LV4**. **(C)** Ligand **AYV**. **(D)** Ligand **M0J**. **(E)** Ligand **LT7** (PDB ID 5QPD, this work). **(F)** Ligand **GQM** (PDB ID 5QPI, this work). **(G)** Ligands **LV4**, **4YV**, **M0J**, **LT7** and **GQM**. **(H)** Close up of phenyl side-chain of Phe50 in all five crystal structures. In all images the protein backbone is shown in grey coloured cartoon representation. Compounds and residue Phe50 are highlighted in colours.

The two other ligands, **LT7** and **GQM**, caused an induced fit of residue Phe50 that led to the opening of the allosteric pocket and enabled ligand binding (**Figure 38 (E)** and **(F)**). The phenyl moiety of Phe50 is rotated by 108 ° and 125 ° upon binding of **LT7** and **GQM**, respectively (**Figure 38 (H)**). This is the first demonstration that conformational changes of the side chain of

Phe50 can take place and an allosteric site is formed, which corresponds to the allosteric site observed in hFPPS[205, 209a]. Superimposition of all five structures shows the difference between the two types of binding events in the allosteric region (**Figure 38 (G)**). Depicting **LT7** and **GQM** in their binding sites shows that the allosteric pocket formed in TcFPPS is a narrow channel that opens widely towards the catalytic cleft of the protein. The base of the pocket is formed by residues Thr54 and Val353 and the sites are formed by residues Phe50, Arg51, Thr212, Tyr213, Phe246 and Lys356 (**Figure 39 (A)** – **(C)**). The key interactions formed by fragment **LT7** is π-stacking of its phenyl and oxodiazole moiety with residues Phe50 and Tyr213 (**Figure 39 (D)**).

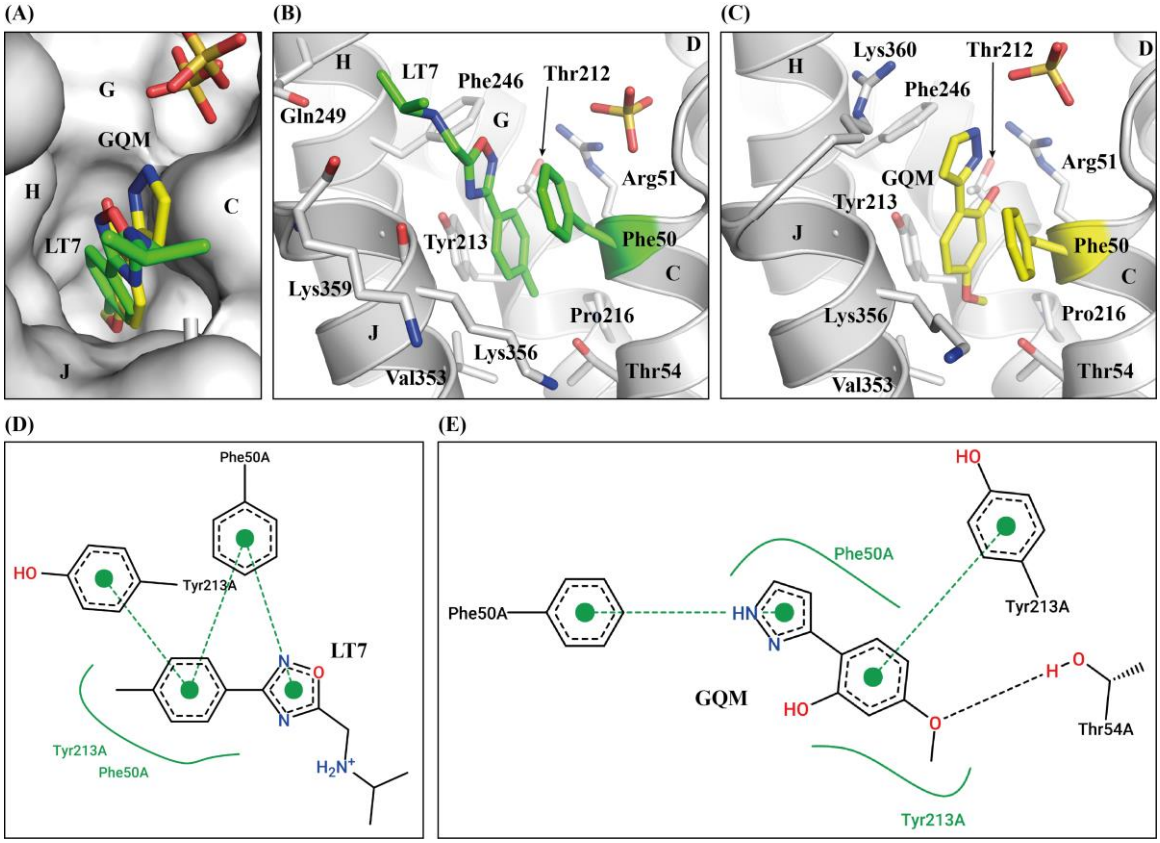



**Figure 39**: Allosteric binding site and allosteric binders **LT7** and **GQM** of TcFPPS. **(A)** Allosteric pocket with **LT7** and **GQM** bound (PDB IDs 5QPD and 5QPI, this work). **(B)** Binding site of **LT7**. **(C)** Binding site of **GQM**. In all images TcFPPS is shown as grey coloured cartoon representation. Compounds and residue Phe50 are highlighted in green and yellow in the comples of TcFPPS-**LT7** and TcFPPS-**GQM**, respectively. **(D)** and **(E)** 2D structure diagram of ligands **LT7** and **GQM**, respectively, interacting with TcFPPS. Diagram was generated using PoseView[481].

Ligand **GQM** forms an H-bond with residue Thr54. In addition, its phenol and pyrazole moiety show π-stacking with residue Phe50 and Tyr213, respectively (**Figure 39 (E)**). **LT7** and **GQM** were refined with occupancies of 0.60 and 0.54, showing that conformational changes and ligand binding did not take place in all protein copies in the crystal. For both ligands, coverage in the final $2F_o - F_c$ map was weaker when compared to the event map and it is also weaker when

compared to the maps of fragments **LV4**, **AYV** and **M0J** (Appendix, **Figure 74**). Whilst ligands **LT7** and **M0J** exclusively bind to the allosteric site, the other three ligands have secondary and/or tertiary binding sites on the protein surface with partly higher occupancies. Fragments **GQM** and **LV4** also bind to site S1 (occ. 0.82 and 0.72, respectively). In addition, fragment **LV4** binds twice to site S2 (occ. 0.54) and **AYV** also binds to site S5 (occ. 0.70) (Appendix, **Figure 74**). Such multiple binding is often seen for fragments[399b] showing that ligands can also bind to cavities on the TcFPPS surface.

Structural comparison of TcFPPS in complex with **LT7** and **GQM** to hFPPS in complex with FPP[205] revealed that the allosteric pocket in TcFPPS is still narrower than in hFPPS (**Figure 40 (A, B)**), where the alkyl chain of FPP protrudes from the pocket towards the protein surface (**Figure 40 (C, D)**). Thus, FPP is only partly embedded by the protein and gets exposed to the solvent. The fragments bound to the allosteric pocket in TcFPPS have exit vectors to the solvent in its open-state (**Figure 40 (E, F)**). However, to accommodate FPP helix J would need to move by approx. 1.3 Å (**Figure 40 (A, B)**). As demonstrated, such widening of the pocket is not necessary for binding of compounds with a different scaffold (**Figure 40 (E – G)**).

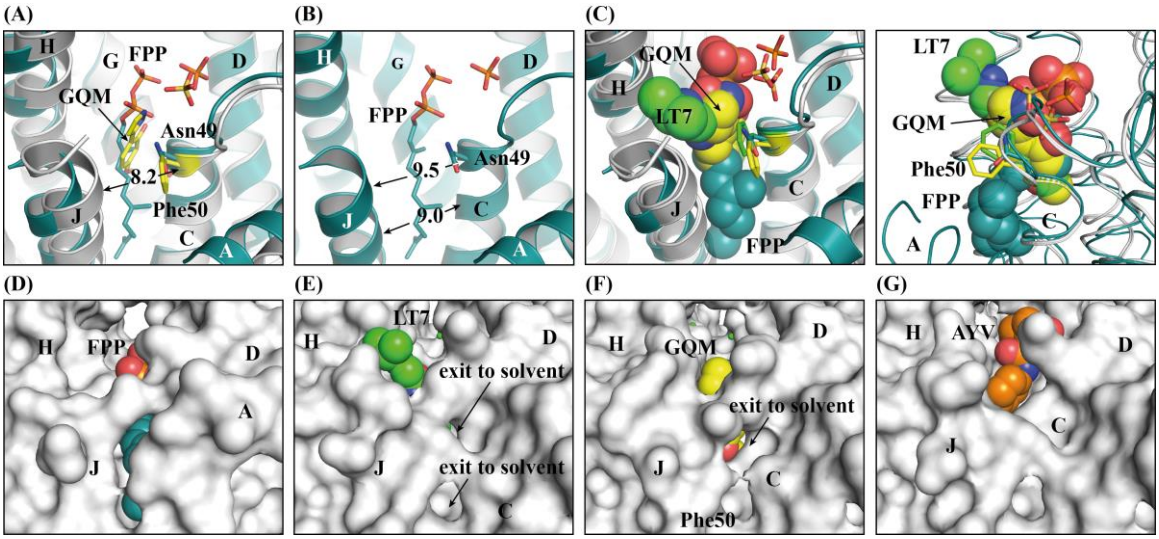



**Figure 40**: Size of allosteric pocket in TcFPPS and comparison with hFPPS. **(A)** Superimposition of TcFPPS in complex with **GQM** (PDB ID 5QPI (this work) grey cartoon, ligand and Phe50 highlighted in yellow) and hFPPS in complex with FPP (PDB ID 50AJ[205], blue-green cartoon) **(B)** Wide allosteric pocket in hFPPS with FPP bound (PDB ID 5JA0[205]) **(C)** Superimposition of TcFPPS in complex with **GQM** (PDB ID 5QPI (this work) yellow) and **LT7** (PDB ID 5QPD (this work) green) and hFPPS in complex with FPP (PDB ID 50AJ[205], cyan). Right site: view rotated by 90 °. **(D)** – **(G)** Surface representation of hFPPS in complex with FPP (cyan) (PDB ID 5JA0[205]) and TcFPPS in complex with **LT7** (PDB ID 5QPD (this work) green), **GQM** (PDB ID 5QPI (this work) yellow) and **AYV** (PDB ID 5QPZ (this work) orange), respectively.

**Novel binding site in close proximity to the IPP binding site**

Fragment **LWA** is the only hit that was identified at a novel site, called SX, that is formed by helix D and B and the two loop turns B-C and D-E (**Figure 41 (A, B, D)**, Appendix, **Figure 74 (6)** and summary of density maps, **Figure 75 (6)**). This binding site is in close proximity to the IPP binding site. The fragment has an exit vector towards the IPP binding site and a small exit vector towards the solvent formed between loop turn D-E and residue Leu45, which resembles the transition point from helix B into the loop turn B-C (**Figure 41 (A)** and **(B)**). Fragment **LWA** forms H-bonds with residues Arg108, Pro111 and Trp113 and van der Waals interactions with Arg107 (**Figure 41 (C)**).

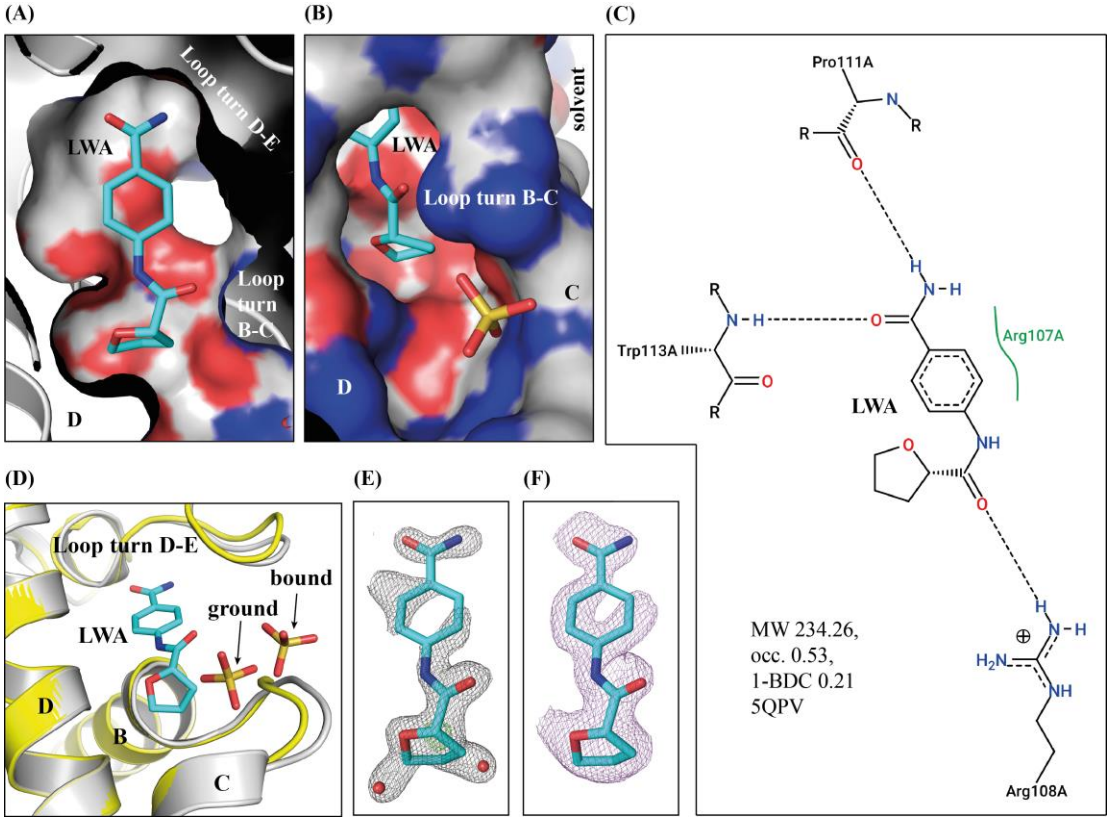


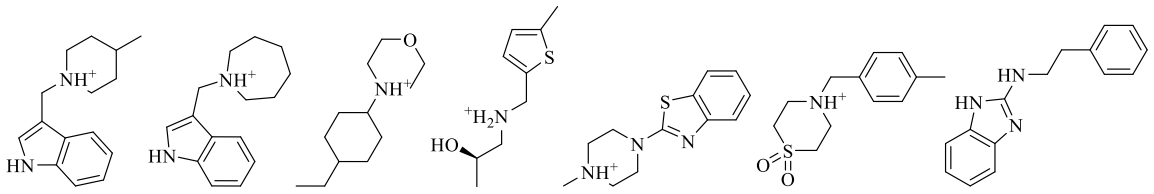

**Figure 41**: Identification of ligand **LWA** with the help of maps generated by PanDDA. **(A)** Crystal structure of FPPS in complex with **LWA**, which is shown in its binding pocket (PDB ID 5QPV (this work)). **(B)** Proximity to IPP binding site. **(C)** 2D structure diagram of ligands **LT7** and **GQM** interacting with TcFPPS. Diagram was generated using PoseView[481]. **(D)** Induced loop shifts upon ligand binding (bound state: yellow cartoon, ground state: grey cartoon). **(E)** Ensemble model: **LWA** and water molecules in the $2F_o - F_c$ electron density map of the ensemble model contoured at 1.0 σ. **(F)** Bound-state model: **LWA** with the background-subtracted PanDDA event map (1-BDC = 0.21) contoured at 0.42 (equals 2.0 σ).

PanDDA analysis was helpful to identify this binding event. The occupancy of **LWA** was refined to 0.53, hence the $2F_o - F_c$ electron density map represents the ground- and bound-state to almost equal shares. In the bound fraction, loop turns D-E and B-C shift, as well as the sulphate

ion located in the IPP binding site. In addition, water ions are present in the ground-state fraction. As a result, the electron density map is difficult to interpret when looking at a classical electron density map of the ensemble model (**Figure 41 (E)**). In contrast, the event map resembles the bound ligand only and allows modelling of the ligand easily (**Figure 41 (F)**), demonstrating the utility of PanDDA.

**Active site binders**

Apart from the discovery of novel ligands for the allosteric binding site and site SX, PanDDA identified 58 putative binding events in the active site. Many of them could be attributed to changes in the binding of metal ions coordinated by Asp98 and Asp250 that are located in the FARM and SARM, respectively. Whilst $Mg^{2+}$ ions had been modelled in this position in the input model and in the ground-state model, well-defined electron density maps of active site-directed binders clearly showed that these positions are occupied by $Zn^{2+}$ ions. In hFPPS the bivalent ions $Mg^{2+}$, $Mn^{2+}$, and $Zn^{2+}$ were shown to have the same octahedral coordination spheres, which are formed by aspartic acids of the FARM and SARM, water ions and N-BPs, such as ALE, PAM and ZOL[168]. Thus, $Zn^{2+}$ ions were modelled at this positions for all 35 deposited structures. Their occupancies were refined and their water spheres restrained if necessary (work of Elliot Nelson). Seven fragments, **LDV**, **AWV**, **LUS**, **MJ4**, **AWM**, **LVV** and **LUY**, were modelled and successfully refined in the active site (**Figure 42** and summary of density maps, Appendix, **Figure 75 (7-13)**).




| **LDV** | **AWV** | **LUS** | **MJ4** | **AWM** | **LVV** | **LUY** |
|---|---|---|---|---|---|---|
| MW 229.35 | MW 229.35 | MW 198.33 | MW 186.29 | MW 234.34 | MW 240.34 | MW 273.31 |
| occ. 0.77 | occ. 0.79 | occ. 0.66 | occ. 0.62 | occ. 0.82 | occ. 0.58 | occ. 0.56 |
| 1-BDC 0.41 | 1-BDC 0.57 | 1-BDC 0.21 | 1-BDC 0.30 | 1-BDC 0.35 | 1-BDC 0.22 | 1-BDC 0.28 |
| 5QPN | 5QPG | 5QPH | 5QQ2 | 5QPF | 5QPT | 5QPK |

**Figure 42**: Chemical structure of the fragment hits identified for the active site. Compounds are derived from the DSPL. MW in Da, ligand occupancy, 1-BDC and PDB ID are indicated.

The binding poses of the seven ligands can be divided into two categories. Ligands **LDV**, **AWV**, **LUS** and **MJ4** protrude from the allylic site (DMAPP binding site) towards the homoallylic binding site (IPP binding site). In their crystal structures the allylic site is not fully accessible due to the conformation of residues Tyr94 and Gln167 (**Figure 43 (A)**). These residues have previously been shown to move, thus forming a channel between helices D and F and allow binding of N-BPs

with longer alkyl chains[170a, 211]. This change is observed here for ligands **AWM**, **LVV** and **LUY** (**Figure 43 (B)** and **(C)**). The previously reported fragment **JMN** which was discovered by FBS by NMR is also binding at this site (chapter **5.3.2**). The difference between the pocket landscapes can be shown by depicting the protein backbone in surface representation (compare **Figure 43 (A)** and **(B)**). The base of the open channel is formed by the aromatic and polar residues Tyr94, Thr163 and Tyr211. Its sites are formed by the backbone of Tyr94 and Thr163, residues Leu95, Lys 207 and Thr208 and Ile129 of the opposing dimer mate. The site at the top of the channel is formed by residue Gln167 and the aforementioned residues Asp98 and Asp250 that coordinate the $Zn^{2+}$ ions (**Figure 43 (C)**).

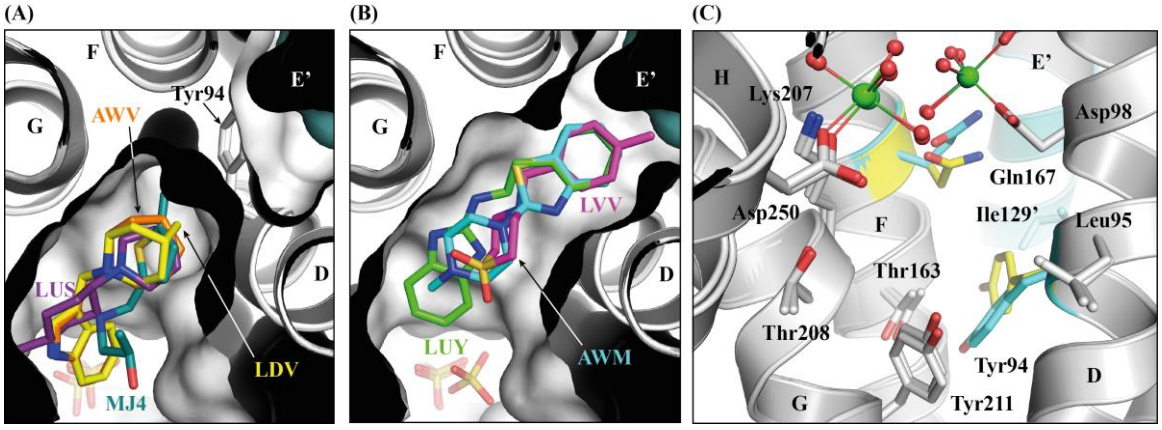



**Figure 43**: Crystal structures of TcFPPS in complex with active site-directed binders. **(A)** Complex with **LDV** (yellow), **AWV** (orange), **LUS** (violet) and **MJ4** (blue-green) (top view) (PDB IDs 5QPN, 5QPG, 5QPH and 5QQ2 (this work)). **(B)** Complex with **AWM** (cyan), **LVV** (magenta) and **LUY** (green) (top view) (PDB IDs 5QPF, 5QPT and 5QPK (this work)). **(C)** Residues forming the active site that accommodates the ligands shown in **(A)** and **(B)** (front view). Backbones of TcFPPS in complex with **LDV** and **AWM** are shown reveal conformational changes of Gln167 and Tyr94 (highlighted in yellow and cyan for binding of **LDV** and **AWM** respectively, ligands not shown).

The benzothiazole moiety of **AWM** and the phenyl moieties of **LVV** and **LUY** show π-stacking with Tyr94. Ligand **AWM** and the fragments **LDV** and **AWV** contribute in H-bonding with a sulphate ion that is part of the coordination sphere of a $Zn^{2+}$ ion (Zn2). The sulphate ion could only be modelled in the presence of these three ligands (**Figure 44 (A)**, **AWV** not shown). The binding position of the sulphate ion is very similar to the position occupied by the phosphate moiety of BPs. In addition, the aromatic moiety of **AWM**, **LVV** and **LUY** occupy the same space then the side chains of the BPs (**Figure 44 (F)** – **(H)**, **LVV** and **LUY** not shown). In addition the indole moiety of **LDV** and **AWV** directly contributes in H-bonding with Asp250 (**Figure 44 (B)**, **AWV** not shown). The methyl substituent of fragment **LVV** in para position contributes in van der Waals interactions with Ile129 of the opposing dimer mate (**Figure 44 (E)**). This interaction has previously been observed for ligands with long alkyl chains (e.g. PDB ID 4EIE[170a]). Fragment

**LUY** can act as hydrogen donor and forms H-bonds with the carbonyl oxygen of Lys207 and the side chains of residues Thr208 and Tyr211 (**Figure 44 (C)**).

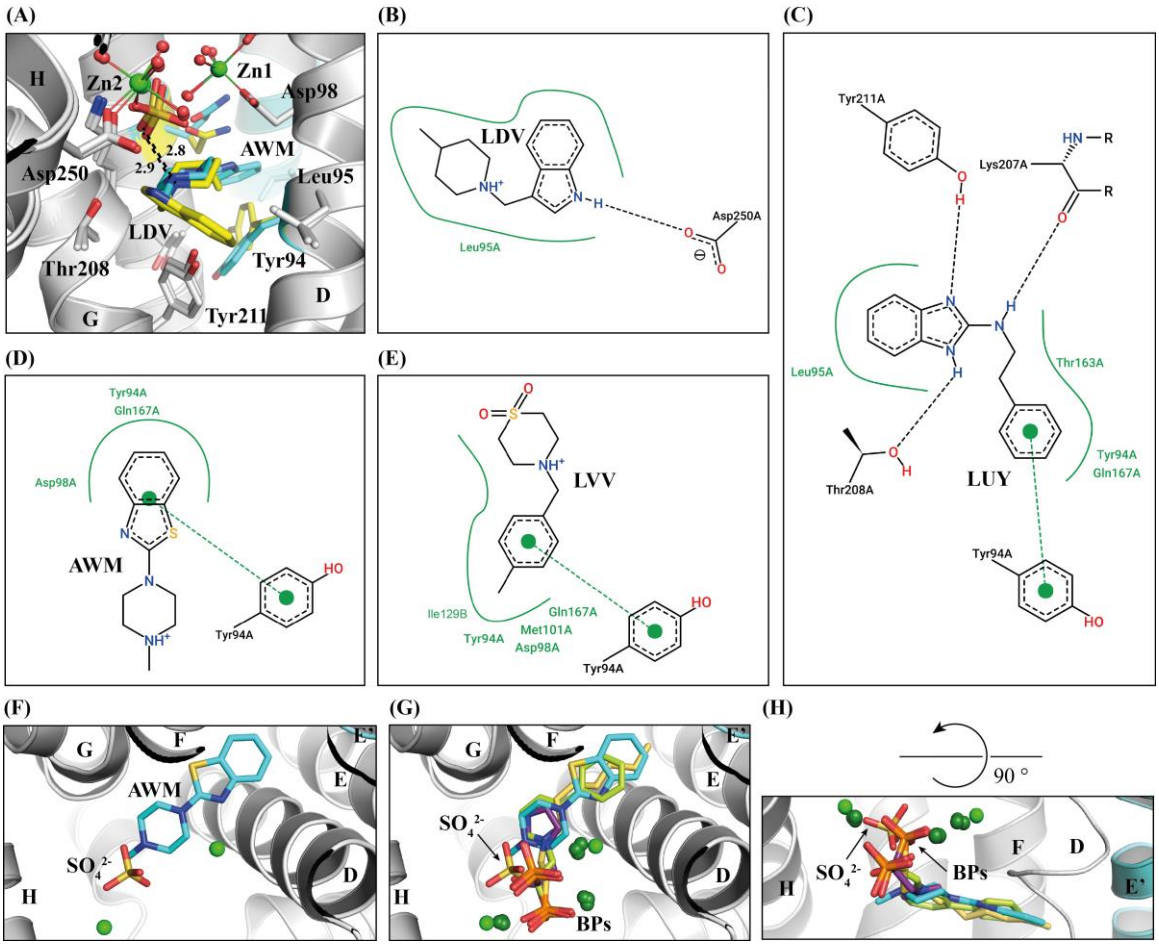



**Figure 44**: Crystal structures of TcFPPS in complex with active site binders. **(A)** Binding site of **AWM** (cyan) and **LDV** (yellow) (front view) (PDB IDs 5QPF and 5QPN (this work)). H-bond with sulphate ion is indicated by dashed lines. Distance is given in Å. **(B) – (E)** 2D structure diagram of ligands **LDV**, **LUY**, **AWM**, and **LVV** respectively, interacting with TcFPPS. Diagrams were generated using PoseView[481] **(F)** Binding site of **AWM** (top view). **(G)** Superimposition of TcFPPS in complex with **AWM** (PDB ID 5QPF (this work)) and TcFPPS in complex with BPs (PDB IDs 3IBA[211] (violet), 3ICM[211] (light green) and 4DWG[170a] (light yellow). **(H)** View of **(G)** rotated by 90 °.

All other ligand binding events are surface binders that are of less interest. Nevertheless, they have been refined and deposited for reasons of completeness and some of them are described in brief here. Site S1 was previously observed as binding site of the hFPPS allosteric inhibitor **119** (PDB ID 6R08, chapter **5.2**). It is a small and surface exposed pocket that is formed by helices H, I and α3. In most cases, it buries the aromatic moiety of the fragments, which show π-stacking with residue Phe256 as key interaction. Sites S2 and S12 are at the dimer interface, and in consequence some of the ligands binding there interact with subunit A and B. Site S2 is an elongated groove formed by helices F, G and H of subunit A and loop turn A-B of subunit B. Two molecules of ligand **LV4** (PDB ID 5QPP) bind to this cavity. This site was previously reported to host the hFPPS

allosteric inhibitor **93**, when bound to TcFPPS (PDB ID 6R07, chapter **5.2**). Ligand **LTX** (PDB ID 5QQA) has a similar binding position as fragment **JNE** (PDB ID 6R05) that was discovered in the previously described FBS by NMR campaign (chapter **5.3.2**). Both ligands form interactions with the loop turn D-E and show H-bonding with residue Gln183, which is located in the insertion loop of subunit B. Thus, the dimer interface is enlarged upon binding of these ligands.

**Follow up of fragment binders with orthogonal biophysical methods**

To test ligand affinity to TcFPPS in an orthogonal biophysical method, protein-observed NMR experiments were conducted with the most interesting fragment hits that could be purchased (**Table 23**). [$^{13}C^{1}H$]-SOFAST-HMQC spectra were recorded from samples containing 30 µM protein and 700 µM compound. Whilst the majority of the compounds showed no or very weak chemical shift changes, spectra of the allosteric site binder **GQM** and the active site binder **LUY** showed medium to strong chemical shift changes (**Table 23**, Appendix, **Figure 76**).

**Table 23**: Results of 2D NMR experiments of fragment hits identified by X-ray crystallography.

| Compound | MW (Da) | Brief description of binding site | Strength of chemical shift changes in 2D NMR |
|---|---|---|---|
| **AYV** | 224.12 | allosteric, stacking with Phe50 | very weak |
| **M0J** | 294.11 | allosteric, stacking with Phe50 | very weak |
| **GQM** | 190.07 | allosteric, open pocket | medium to strong |
| **LWA** | 234.10 | site SX | very weak |
| **LDV** | 228.34 | active, protruding to IPP site | very weak |
| **AWV** | 228.34 | active, protruding to IPP site | no effect |
| **LUS** | 297.18 | active, protruding to IPP site | no effect |
| **AWM** | 233.10 | active, channel to interface | weak |
| **LVV** | 239.10 | active, channel to interface | very weak |
| **LUY** | 237.13 | active, channel to interface | medium to strong |

The active site-directed fragment **LUY** showed the strongest chemical shift (**Figure 45 (A)** and Appendix, **Figure 76 (J)**). [$^{13}C^{1}H$]-SOFAST-HMQC spectra were recorded for a titration series of this compound, ranging from 7.81 µM to 1 mM in order to estimate its binding affinity. Signal shifts were plotted and for two signal series the $K_d$ value was calculated to approx. 54±9 µM and 33±5 µM, respectively (**Figure 45 (B)** – **(D)**).

The compounds listed in **Table 23** were also employed to SPR experiments using a Biacore T200 and a standard buffer system (50 mM Hepes, pH 7.4, 150 mM NaCl, 2 mM TCEP · HCl, 2 mM $MgCl_2$, 0.01% (v/v) Tween). For the SPR experiments, *in vivo* biotinylated avi-tagged TcFPPS had been successfully expressed and purified (chapter **5.1**). Immobilisation on

streptavidin coated chips lead to a stable baseline with a very low baseline shift. Unfortunately, reliable and robust results were not obtained due to super stoichiometric binding. The expected fragment binding response for a 1:1 fragment-protein interaction model was exceeded by several times. In addition, a suitable positive control was lacking, because no compound was available that exhibited high affinity to TcFPPS and an 1:1 interaction model (data not shown).

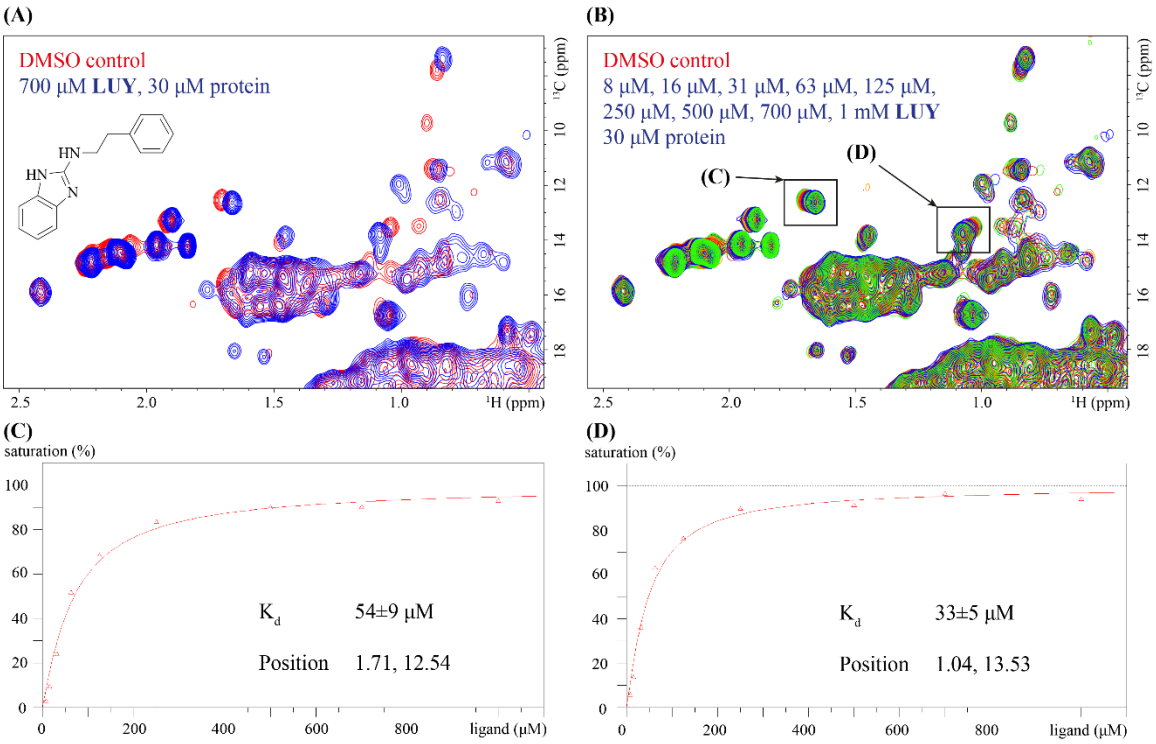



**Figure 45**: Spectra of ligand-observed NMR experiments with fragment hits identified by X-ray crystallography. **(A)** Cut-out from an overlay of the [$^{13}$C$^{1}$H]-SOFAST-HMQC spectrum of the DMSO control (red) and the corresponding sample (blue) at 700 µM **LUY** and 30 µM protein in 25 mM BisTris, pH 6.5, 50 mM NaCl, 2 mM TCEP · HCl, 10% $D_2O$, 150 µM DSS at 31.85 °C. **(B)** Superimposition of spectra recorded for a titration series of compounds. Buffer conditions and cut-out shown as in **(A)**. **(C)** and **(D)** Dose response curves of **LUY** for the signals highlighted in **(B)**.

### 5.4.2 Discussion

The fragment screen by X-ray crystallography at the XChem facility uncovered the first fragments with various scaffolds binding to the allosteric site of TcFPPS. This is remarkable as the pocket appears to be sealed due to steric hindrance by residue Phe50 in all previously published crystal structures of TcFPPS[162b, 170a, 211], as well as in structures elucidated as part of this work so far. Sequence comparison of 200 homologues from various organisms with TcFPPS showed that Phe50 is an exception in a conserved position. In 191 of the homologues an asparagine was found. Among these homologues are hFPPS (UniProt ID P14324) and also FPPSs from other parasitic sources, such as TbFPPS (UniProt ID Q86C09) (chapter **5.2**). Thus it was so far unknown whether

this allosteric pocket, which was first identified in human FPPS[205, 209a], also exists in TcFPPS. Rotation of the phenyl side-chain of Phe50 by 108 ° and 125 ° upon binding of fragments **LT7** and **GQM**, respectively, induces the opening of the pocket. However, the resulting pocket is still narrower than the pocket in hFPPS. It can only be speculated that backbone shifts could take place to allow binding of FPP. Nevertheless, as demonstrated here, this is not necessary for binding of fragments **LT7** and **GQM**. In addition to these two ligands that bind to the allosteric pocket in its open-state, ligands **LV4**, **AYV**, **M0J** bind to the allosteric pocket in its closed-state. All five identified fragments show π-stacking with Phe50 as key interaction with the protein. As this residue resembles a structural difference to the human protein, it can be exploited to design inhibitors with the required specificity[483].

Furthermore, an interesting novel binding site, named SX, was discovered. It is in close proximity to the homoallylic site, which holds also true for the allosteric pocket. Ligand **LWA** shows a high number of H-bonds. However, observed chemical shift changes in [$^{13}C^{1}H$]-SOFAST-HMQC experiments were very weak at a compound concentration of 700 µM. This might be due to the fact that the binding site is more difficult to access in comparison to the others and a series of loop shifts takes place in the event of ligand binding, which also affects the homoallylic binding site. Whether this novel site provides a promising perspective for TcFPPS inhibition has yet to be investigated.

Seven active site-directed fragments of a novel and non-BP scaffold were identified. These fragments can inspire design of active site-directed FPPS inhibitors that can overcome the disadvantages associated with high affinity to bone mineral, which is exhibited by BPs alongside their FPPS binding[226, 238b]. This has limited the usage of BPs in the treatment of CD, but so far nitrogen-containing BPs are the only known FPPS inhibitors used for the treatment of bone diseases[213]. Superimposition of crystal structures of TcFPPS in complex with the fragments **AWM**, **LVV** and **LUY** and TcFPPS in complex with BPs (PDB IDs 3IBA[211], 3ICM[211], 4DWG[170a]) show that their aromatic side-chains share the same binding site. Similar to the natural FPPS substrates, DMAPP and GGPP, the phosphate moieties of the BPs form ionic interactions with $Mg^{2+}$ ions that in turn are coordinated by aspartic acid residues of the conserved regions FARM and SARM. The sulphate ion associated with binding of **AWM**, **LDV** and **AWV** suggest that binding affinity of these active site-directed fragments could be improved by adding a functional group that can coordinate bivalent metal ions. Unfortunately, it is to be expected that the most potent scaffold will be again a bisphosphonate. Previous attempts to remove one phosphate moiety, turning a BP into a mono-phosphate, led to complete loss of potency[225b]. A different and more promising approach to design active-site inhibitors of a novel scaffold is to take the structural information provided by the TcFPPS complexes with fragments **LDV** and **AWV** that protrude from the allylic site. Their indole moieties can directly contribute in H-bonding with Asp250, which is located in the SARM.

Applying FBS by X-ray crystallography on TcFPPS was a success resulting in many more crystal structures than the previously conducted stepwise campaign (chapter **5.3**). This observation is in agreement with reports in the literature[315a, 484]. Screening at high compound concentrations also increases the number of ligands with weak affinity and fragments binding to less specific sites. Crystallization conditions can favour ligands with certain interactions and binding may occur due to extremely high compound concentrations or due to the proteins immobilisation in the crystal lattice[293b]. Screening of the DSPL against TcFPPS was more successful than screening the KFL (hit rates 8.8% and 2.5%, respectively). The latter only accounts for six of the 35 fragment hits. This cannot only be explained by the higher number of data sets analysed of crystals soaked with DSPL fragments (37%). One reason could be that they were soaked at half of the compound concentration than the DSPL fragments. Nevertheless, at an initial concentration of 180 µM TcFPPS in the crystallization drop 37.25 mM equals a 207-fold excess. Therefore, it is more likely that 3D fragments were less suited to bind to TcFPPS, especially to sites buried in the protein.

Analysis with PanDDA accelerated hit identification significantly. The enhanced means of data analysis allowed reliable hit identification of fragments characterised by low affinities and partial occupancy, as well as at binding sites that undergo conformational changes. Good examples are identification of fragments **LWA**, **LT7**, **M0J**, **AWM**, **LVV** and **LUY**. However, the current study also shows that data sets with a diffraction limit ≤2.5 Å are most likely to be identifiable and modellable hits using PanDDA.

The numerous crystal structures of FPPS-fragment complexes, the large diversity of their scaffolds and different binding sites are potential starting point for SBLD, molecular docking and pharmacophore analysis. This includes starting points for allosteric site binders suggesting two different binding modes. Although, whether either starting point is suited to generate a lead series with high-affinity is currently unknown. In addition, starting points for novel and non-BP active site-directed binders are given that might show lower affinities to bone mineral and thus could be suited for the treatment of non-bone related diseases. Some of the ligands have only weak affinities, which was demonstrated by 2D NMR experiments, but the binding affinity has little to say about the suitability of a compound for chemical optimization. Crystal structures of weakly bound fragments or promiscuous binding can provide valuable information about favourable binding poses in each site[399b]. The poised concept of fragments in the DSPL library[336] will enable rapid follow up of these hits. Likewise, a SAR by catalogue study will is possible, as most of the DSPL fragments were purchased from Enamine, who offer analogue libraries for rapid fragment elaboration[485]. The crystal structures provided in this work will pave the way for future drug discovery campaigns for TcFPPS.

### 5.4.3 Results of the HTX lab campaign

Irina Cornaciu and Damien Clavel were in charge of the screening campaign, which encompassed soaking experiments with 531 compounds of the Enamine Golden Library. They were provided with the crystallization protocol and all necessary materials in order to run crystallization and soaking experiments at the HTX lab and data collection at the ESRF. All processes were managed in the Crystallization Information Management System (CRIMS) that was accessible remotely. First on-site crystallization trials resulted in an adjusted protocol for the setup of crystallization drops employing the sitting drop vapour diffusion technique on CrystalDirect™ plates. Rates of 50% – 80% of drops with crystals per plate could only be reproduced when mixing protein formulation, reservoir solution and seed dilution in a ratio of 3:1:2 (v/v)). In the provided protocol a ratio of 3:2:1 (v/v) had been used.

Drops with crystals suitable for soaking were selected by visual inspection. A Mosquito pipetting robot was used to transfer 14.9 mM compound (15% DMSO), and apo TcFPPS crystals were individually soaked for approx. 21 h with 531 compounds. Automated high-throughput crystal harvesting and cryo-cooling was performed with the CrystalDirect™ technology[386b, 387b]. Here, crystals are harvested by laser photolation on the foil the crystals had been growing on. Thus, in some cases several crystals were harvested on a single pin and allowed data collection from single crystals soaked with the same compound. 652 data sets were collected at beamlines ID30A-1[388b, 390b] and ID30B[390a] of the ESRF and 644 of them could be successfully processed. All crystals belonged to the hexagonal space group $P6_122$, and had unit-cell parameters of approx. a = b = 58 Å, c = 397 Å, and $\alpha = \beta = 90$ ° and $\gamma = 120$ ° with one TcFPPS molecule per asymmetric unit. The diffraction limit ranged from 1.41 Å to 3.49 Å with an average diffraction limit of 2.15 Å. 2% had a diffraction limit ≤1.5 Å and 13% had a diffraction limit ≤1.7 Å. This large number of data sets, their crystallographic homogeneity and good average diffraction limit enabled statistical data analysis with PanDDA[393, 398] (**Table 24**, Appendix, **Figure 77**).

PanDDA identified 309 putative binding sites that got clustered arbitrarily into 21 non-overlapping sites (**Figure 47 (B)**). Based on manual inspection of all events in Coot, 54 ligands were modelled into the bound-state models of 45 data sets and their ensemble models were generated using *pandda.export*. Nevertheless, model building and refinement work focused on the most interesting eight fragment hits **HTX-1** to **HTX-8**. **HTX-1** to **HTX-5** bind in the active site and **HTX-6** to **HTX-8** in the allosteric site (**Figure 47 (A, C, D)**). Their chemical structures are depicted in **Figure 46**. The diffraction limit of the eight structural models ranged from 1.71 Å to 2.10 Å with a mean diffraction limit of 1.82 Å. Hence, they show a slightly better mean diffraction limit than compared to all data sets (2.15 Å) (**Table 36**). All eight ligands were well resolved as the PanDDA event map contoured at a contour-level equivalent to 2.0 σ indicates (summary of density maps, Appendix, **Figure 78 (1 – 8)**). When looking at the density maps of the

single data sets only, ligands **HTX-1** and **HTX-8** were also well resolved in the unbiased $mF_o - DF_c$ difference electron density map contoured at 3.0 σ and the $2mF_o - DF_c$ electron density map of the unfitted model contoured at 1.0 σ (summary of density maps, Appendix, **Figure 78 (1, 8)**). The ligands occupancies were refined to 0.82 and 0.58, respectively, and structural models were deposited under PDB IDs 6SI5 and 6SHV (Appendix, **Table 37**).

**Table 24**: Key parameters of the data sets suited for analysis that were obtained from the HTX campaign.

| Variable | Value |
|---|---|
| Space group | $P6_122$ (No. 178, International Tables for Crystallography[377b]) |
| Mean unit cell axis a / b / c (SD) (Å) | 57.98 (0.21) / 57.94 (0.21) / 396.77 (0.88) |
| Diffraction limit (Å) | 1.41 – 3.49 |
| Mean diffraction limit (SD) (Å) | 2.15 (0.43) |
| Mean $R_{free}$ / $R_{work}$ (SD) | 0.260 (0.030) / 0.191 (0.015) |
| Mean ($R_{free}$ - $R_{work}$) (SD) | 0.143 (0.041) |

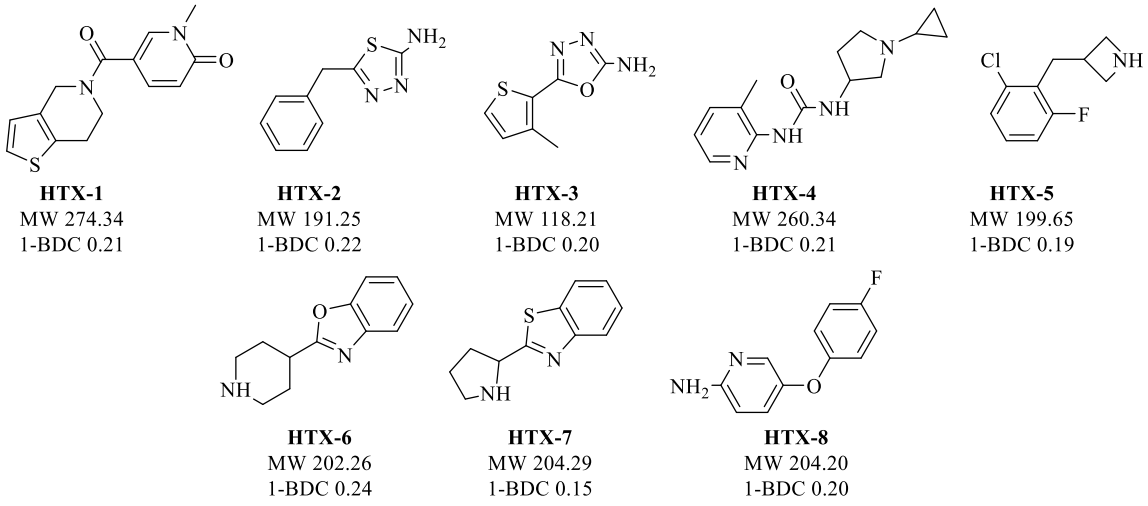



**Figure 46**: Chemical structure of fragment hits identified for the allosteric and active site of TcFPPS. Compounds are derived from the Enamine Golden Library. MW in Da and 1-BDC are indicated.

The allosteric binders **HTX-1** to **HTX-5** show π-stacking with residue Phe50 as key interaction, but have two different binding modes. Fragments **HTX-1**, **HTX-2** and **HTX-3** bind to the allosteric pocket in its open-state (**Figure 47 (E – H)**). To accommodate these ligands, the allosteric pocket opens by rotation of the phenyl side chain of residue Phe50 by approx. 120 ° (**Figure 47 (L)**). This type of conformation change and binding pose of ligands was previously observed for fragments **LT7** and **GQM** (**Figure 38 (E, F, H)**) and was described in more detail in chapter **5.4.1**. Fragments **HTX-4** and **HTX-5** bind to the pocket in its closed-state (**Figure 47 (I – K)**). This binding pose was previously described for fragments **LV4**, **AYV** and

**M0J** (**Figure 38 (A – D)**). In addition, the chemical structure of **HTX-4** is similar to fragments **AYV** and **M0J** (**Figure 46** and **Figure 37**).

The active site binders **HTX-6** to **HTX-8** show the same binding pose as earlier described for fragments **AWM**, **LVV** and **LUY** (chapter **5.4.1**, **Figure 43 (B)**). Their key interaction is a π-stacking with residue Tyr94. Compounds **HTX-6**, **HTX-7** show very similar chemical structures and are chemical analogues of fragment **AWM** (**Figure 46** and **Figure 42**).

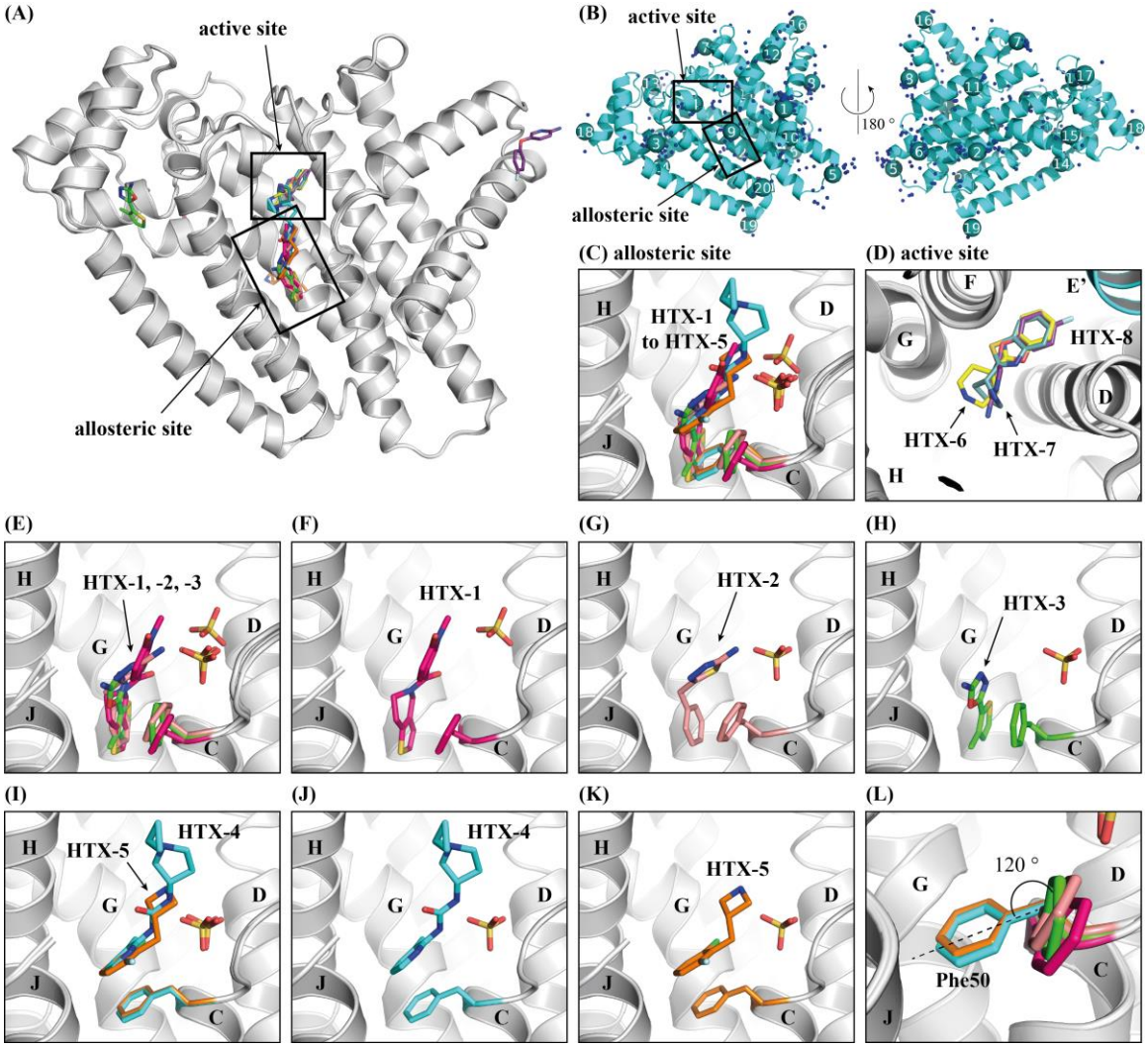



**Figure 47**: Overview of crystal structures obtained by FBS by X-ray crystallography at the HTX lab. **(A)** Overview of ligands binding across TcFPPS. Front view. Binding sites are indicated. **(B)** Data analysis with PanDDA: Binding events clustered into 21 non-overlapping binding sites (front view (left). Back view (right)). **(C)** Fragments **HTX-1** to **HTX-5** binding in the allosteric site (**HTX-1**: PDB ID 6SI5 (this work)). Residue Phe50 is shown in stick representation. **(D)** Fragments **HTX-6** to **HTX-8** binding to the active site (**HTX-8**: PDB ID 6SHV (this work)). **(E)** – **(H)** Allosteric binders **HTX-1** to **HTX-5**: Superimposition and single views **(I) – (K)** Allosteric binders **HTX-4** and **HTX-5**: Superimposition and single views. **(H)** Rotamers found for residue Phe50 in all structures with allosteric binders.

### 5.4.4 Discussion

The fragment screen at the HTX lab identified additional allosteric site binders of TcFPPS in the open- and closed-state of the pocket, hence, adding potential starting points for medicinal chemistry campaigns to develop a TcFPPS inhibitor with a novel mode of action. The same holds true for the active site. Interestingly, some of the fragments from the HTX lab campaign have very similar scaffolds to the fragment hits identified in the XChem screen. Thus, they provide a positive control, which is only possible in crystallization experiments when chemical analogous compounds are tested.

## 5.5 SBLD by virtual screening in ANCHOR.QUERY

Here the virtual screening tool ANCHOR.QUERY was used to discover novel active site-directed binders of a different and non-BP scaffold for TcFPPS. Such compounds could exhibit lower affinities to bone mineral and might be able to overcome the inappropriate PK properties of BPs[244]. Starting point were fragment binders, which were recently discovered within the scope of this work in an FBS by X-ray crystallography at the XChem facility, Harwell, UK (chapter **5.4**). When using the ANCHOR.QUERY approach, the proposed compounds can be synthesised by multi component reactions (MCR)[486]. Eleven scaffolds were generated using the fragment hit **LUY** as a template, but unfortunately they could not be validated experimentally. The herein described work was conducted in collaboration with Markella Konstantinidou, who is a PhD student on the AEGIS project in the group of Prof. Alexander Dömling, during a one month visit at the University of Groningen, the Netherlands.

### 5.5.1 Results of virtual screening and synthesis by MCR

ANCHOR.QUERY is an interactive web-based pharmacophore search technology, which virtually screens a library of more than 31 million compounds and approx. 2 billion preformed conformers[462, 487]. These compounds can be easily synthesised from cheap and commercially available starting material by one-step, one-pot MCRs[462]. The application and use of ANCHOR.QUERRY were described elsewhere[432, 462]. In brief: ANCHOR.QUERY was developed to target protein-protein interactions. Therefore, all compounds in the library contain an anchor motif that is bioisosteric to an amino acid. This anchor biases interactions with the protein and allows rapid pharmacophore searches. A 3D structure of a protein-ligand complex is required to generate a query. The software identifies an anchor motif within the ligand and a pharmacophore query is created by adding additional features to the anchor, such as charged groups, hydrogen donors/acceptors or hydrophobic rings. Then, the library is screened for matching compounds, which are spatially aligned by a root mean square fit and energy-minimized. Additional filters, such as selecting certain reaction types and threshold limitations of the MW can be applied[462].

Screening was focused on ligand **LUY**, which is an active site-directed binder with a non-BP scaffold that showed convincing binding signals in protein-observed NMR spectroscopy and its $K_d$ was estimated to be 40 µM (chapter **5.4**). All ANCHOR.QUERY-generated pharmacophore models based on ligand **LUY** included the phenyl anchor and various additional motifs including a nitrogen group as donor, aromatic rings with positive charge and hydrophobic rings. In all queries, three scaffolds occurred more frequently: *N*-bridgehead hetero-bicyclic compounds, such as **AQ-1** to **AQ-3**, tetrazoles, such as **AQ-4** to **AQ-6**, and β-lactams, such as **AQ-7** to **AQ-9** (**Figure 48**).

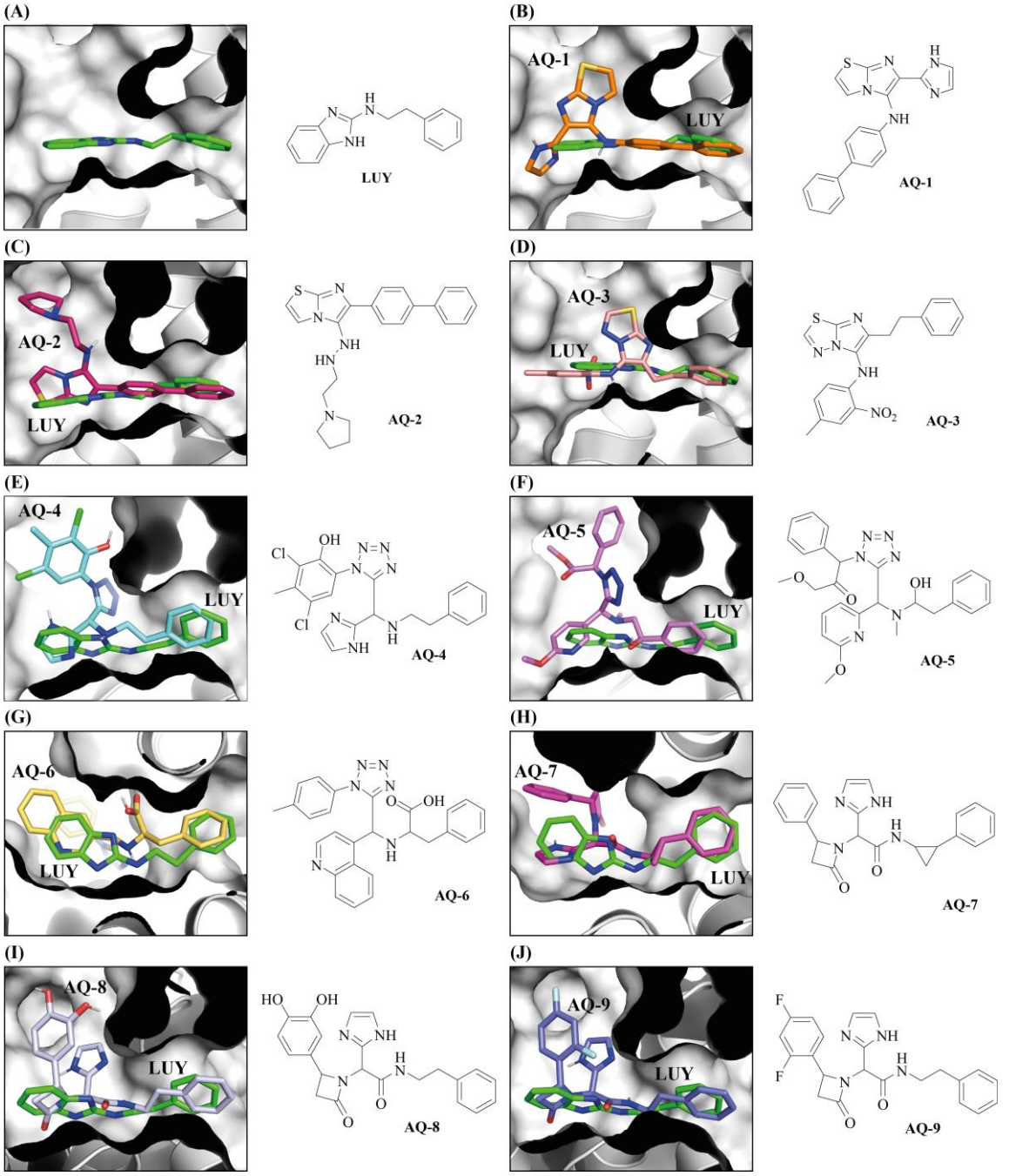



**Figure 48**: Compounds suggested by ANCHOR.QUERY. **(A)** TcFPPS in complex with **LUY**. **(B)** – (**J)** Docked binding poses of compounds are shown in superimposition with TcFPPS in complex with **LUY**: **(B)** – **(D)** *N*-bridgehead thiazoles **(E)** – **(G)** tetrazoles **(H) – (J)** β-lactams.

A small library of eleven compounds was synthesised including all three scaffolds. Among them are the *N*-bridgehead thiazole **MCR-1** and imidazole **MCR-2**, seven tetrazoles **MCR 3** to **MCR-9** and the β-lactams **MCR-10** and **MCR-11**. Synthesised scaffolds deviated from the scaffolds suggested by ANCHOR.QUERY, because starting materials that were already available in the laboratory were used. Compounds **MCR-1** and **MCR-2** are based on **AQ-1** to **AQ-3**. They

were synthesised by the one-pot reaction of aldehyde, amidine and isocyanide using scandium trifluoromethanesulfonate as a catalyst and under microwave radiation. Purification by flash column chromatography (FCC) yielded the products **MCR-1** and **MCR-2** with yields of 55% and 79%, respectively. The corresponding MCR is called Groebke-Blackburn-Bianaymé reaction (GBBR)[488] (**Figure 49**).

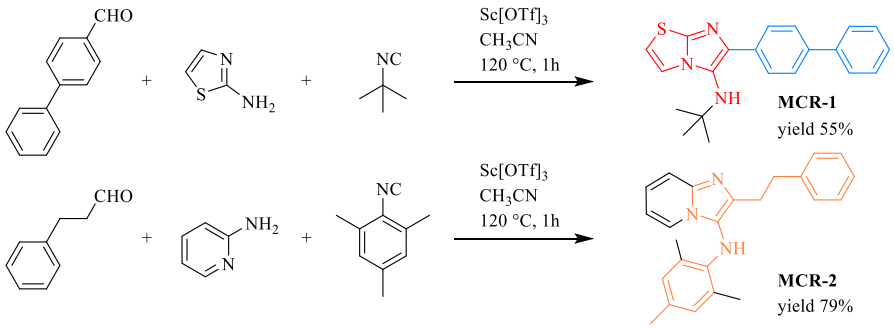



**Figure 49**: GBBR to give compounds **MCR-1** and **MCR-2**. Moieties that are in common with **AQ-1** and **AQ-2**, **AQ-2** and **AQ-3** are highlighted in red, blue and orange, respectively.

A series of tetrazoles was synthesised by a four-compound condensation, the Ugi-tetrazole reaction[489]. It is a one-pot reaction of aldehyde, amine, isocyanide and TMS azide, which is stirred at room temperature, overnight. Purification by FCC yielded the products **MCR-3** to **MCR-9** with yields ranging from 24% to 89% (**Figure 50**).

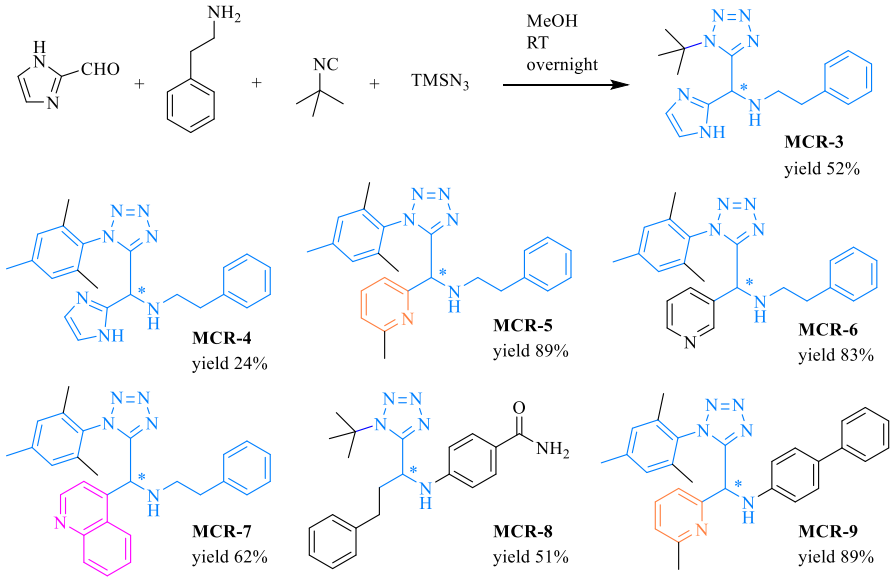



**Figure 50**: Ugi-4CR to give compound **MCR-3** to **MCR-9**. Reaction scheme shown for **MCR-3**. **MCR-4** to **MCR-9** were also synthesised by an Ugi-4CR. Yields are indicated. Moieties that are in common with **AQ-4**, **AQ-5** and **AQ-6** are highlighted in blue, orange and pink, respectively.

Two β-lactams were synthesised in a one-pot reaction of β-amino acid, aldehyde and isocyanide under microwave radiation. Purification by FCC yielded the product **MCR-10** with a high yield of 93% and **MCR-11** as a racemate with a yield of 47% (**Figure 51**).

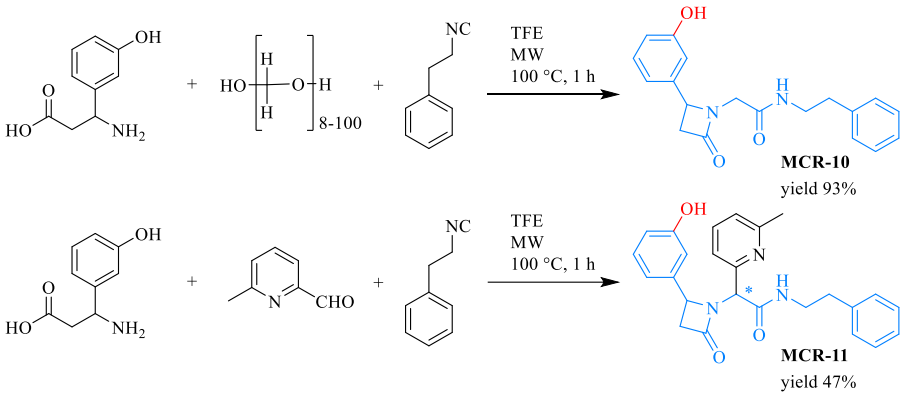



**Figure 51**: β-lactams **MCR-10** and **MCR-11** synthesised by an one-pot MCR. Yields are indicated. Moieties that are in common with compounds **AQ-8** and **AQ-9** or only **AQ-8** are highlighted in blue and red, respectively.

The synthesised compounds were employed for affinity testing and structure determination experiments. First, solubility tests in SPR buffer were conducted by 1D NMR spectroscopy using DSS as an internal standard. Whilst some of the compounds were insoluble others showed a solubility ranging from 40 µM to 1 mM. The compounds with good to high solubility were tested at a single concentration in SPR. Unfortunately, no binding could be detected or the compounds were even misbehaving and showed a negative signal on the reference channel. Then, 2D NMR experiments were conducted with all compounds measuring [$^{13}C^{1}H$]-SOFAST-HMQC spectra of samples that contained 700 µM compound and 30 µM $^{13}C^{15}N$-labelled TcFPPS. In consequence, most compounds were measured at saturation. However, spectra showed only a few (three to seven) and weak chemical shift changes for all compounds. Thus, the readout was weaker when compared to compounds that were previously measured in protein-observed NMR (chapter **5.2**).

Regardless of these unsatisfactory results, the compounds were employed to soaking experiments. TcFPPS apo crystals were grown on 96-well SwissCi/MRC plates employing the sitting-drop vapour diffusion technique. Crystallization drops were mixed of 300 nL 12.36 mg · mL$^{-1}$ TcFPPS (in 10 mM TRIS, pH 7.4, 25 mM NaCl, 2 mM TCEP · HCl), 100 nL reservoir (80 mM MES, pH 6.5, 4 mM $ZnSO_4$, 12.36% (v/v) PEG MME 550, 11.57% (v/v) glycerol) and 200 nL seed dilution (in 80 mM NaOAc, pH 5.0, 160 mM $(NH_4)_2SO_4$, 20% (v/v) PEG 400, 20% (v/v) glycerol). Apo crystals were soaked at a nominal compound concentration of 75 mM or 37.5 mM, which both corresponded to 15% DMSO. As all 11 compounds showed precipitation in the crystallization drop, the actual concentration in solution was lower. Crystals were fished after 4 h and 21 h of soaking. Data sets could be collected of all 11 crystals soaked for 21 h. They showed good diffraction limits of approx. 1.7 Å. Nevertheless, data processing and

visual inspection of the electron density maps showed that none of the compounds had formed a protein-ligand complex.

Furthermore, fragment **LWA**, which binds to the novel binding site SY, was used as a starting point for ANCHOR.QUERY. The software could not identify an anchor scaffold and hence, the benzamide moiety was run with a tyrosine anchor or phenylalanine anchor instead. ANCHOR.QUERRY proposed compound **AQ-10**, which makes optimal use of the pocket. To synthesise the compound with an amide substituent (**MCR-12**) in accordance to ligand **LWA**. It can be synthesised running a Groebke-Blackburn-Bienaymé reaction (**Figure 52**). Unfortunately, synthesis of the required isocyanide failed (communication with Markella Konstantinidou). Hence the synthesis of the compound was not pursued further.

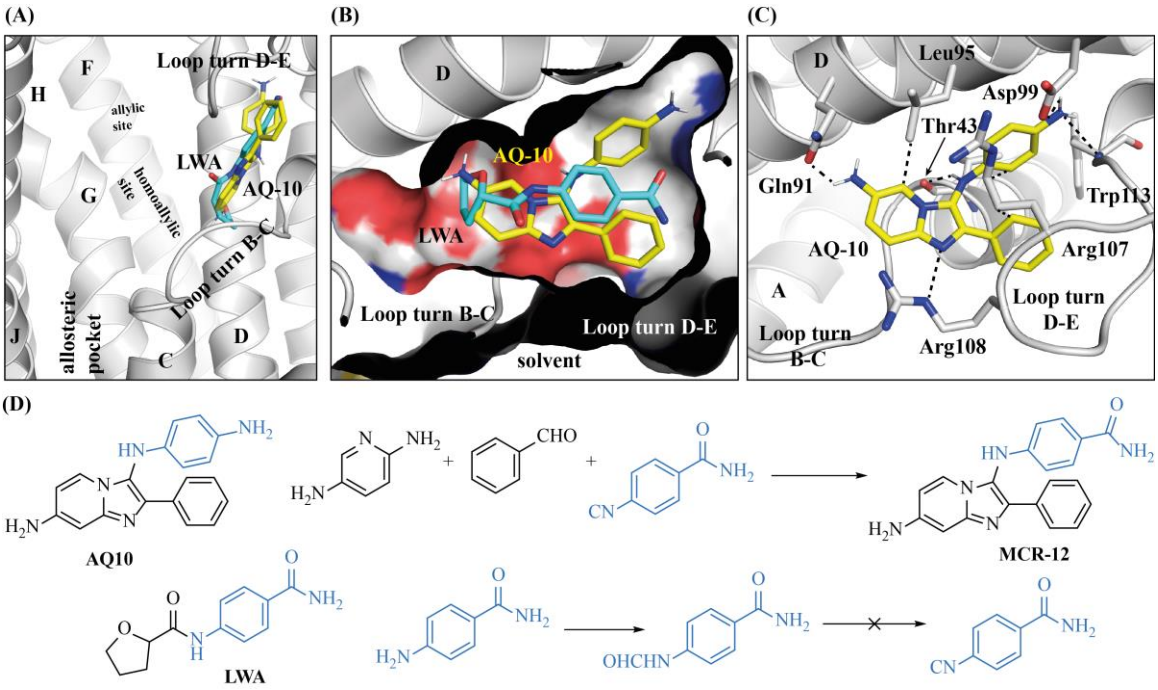



**Figure 52**: Virtual screening proposed ligand binding site SX. **(A)** Relative position of binding site SX to the three known binding sites in TcFPPS. Superimposition of ligand **LWA** (cyan) and compound **AQ-10** (yellow). Binding pose of **AQ-10** originated from ANCHOR.QUERRY. **(B)** Ligands shown in its binding pocket SX. **(C)** Binding pose and possible interactions formed by **AQ-10**. **(D)** Chemical structure of **LWA** and reaction scheme for a compound similar to **AQ-10**.

We also tested the 3D structures of TcFPPS in complex with ligands **LT7** and **GQM** as starting point for ANCHOR.QUERY. Running a meaningful virtual screen was difficult here, because the allosteric pocket opens up widely towards the homoallylic site and is close to the protein surface. All pharmacophore searches identified compounds that showed poor binding poses. Most of them showed a small number of interactions with the protein and protruded wide into the solvent space. Therefore, these queries were not considered any further.

### 5.5.2 Discussion

Unfortunately, none of the eleven compounds synthesised using fragment hit **LUY** as a template for ANCHOR.QUERRY showed increased affinity to TcFPPS. Soaking experiments at high compound concentrations, which benefit weakly bound fragments did not lead to a 3D structure of a protein-ligand complex. Possible reasons for the failure of the presented attempts could be that the synthesised compounds were increased in size too ambiguously. While the starting compound **LUY** had an MW of 237.3 Da the average MW of **MCR-1** to **MCR-11** was 386.8 Da. Poor solubility of some of the compounds was clearly an issue and thus, solubility should be a selection criterion for future compound synthesis. In case, channel size in the preformed apo crystals was a limitation for the large compounds to be soaked, co-crystallization experiments could have circumvent this issue. With a MW of 295.34, compound **LXM** (PDB ID 5QQ7) was the largest compound soaked into a preformed TcFPPS crystal (chapter **5.4**).

In addition, the synthesised compounds deviated from the compounds proposed from ANCHOR.QUERY, because starting materials that were available in the working group were used. This might also have an impact on the results. Nevertheless, ANCHOR.QUERY was previously demonstrated to be a useful tool to morph weakly bound fragments into potent tool compounds[432] and it was already successfully applied to query fragments identified by X-ray crystallography[487]. Thus, further investment and synthesis of additional compounds might be worthwhile for this project as well.

## 5.6 Fragment-to-lead optimization using fragment merging

As described in chapter **5.4**, active site-directed binders were identified in an FBS campaign by X-ray crystallography at the XChem facility, Harwell, UK. These binders have a novel, non-BP scaffold and therefore, compounds based on them likely exhibit lower affinities to bone mineral, hence, overcoming the inappropriate PK properties of BPs[244]. Fragment merging was employed as an optimization strategy for these fragments and twelve compounds were synthesised. Unfortunately, they did not show increased potency and X-ray crystallography revealed that they were binding to a previously discovered cavity on the protein surface.

### 5.6.1 Results of the fragment-to-lead optimization

The crystal structures of all seven active site-directed fragment hits obtained in the FBS campaign were examined manually in Coot. The binding modes of ligands **LDV**, **AWV**, **AWM**, **LUY** and **LVV** suggested that fragment merging is the best strategy for chemical optimization of the fragment hits. Superimposition of the crystal structures of TcFPPS in complex with **LDV** and **AWV** showed that these two ligands with very similar chemical structures show a perfect overlap (**Figure 53 (A)** and **(D)**). Superimposing these two with the crystal structure of TcFPPS in complex with **AWM** demonstrates that the mehtylpiperidine moiety of **LDV** and the azacycloheptan moiety of **AWV** overlap with the methylpiperazine moiety of **AWM** (**Figure 53 (B)**). Furthermore, superimposition of the TcFPPS-**AWM** complex with crystal structures of the TcFPPS-**LUY** complex and the TcFPPS-**LVV** complex confirmed the binding position of an aromatic moiety, either benzothiazole (**AWM**) or phenyl rings (**LUY**, **LVV**) (**Figure 53 (C)**). Based on compounds **LDV** and **AWM,** either a piperidine or a piperazine would be possible to link the benzothiazole and the indole moiety (**Figure 53 (D)**). The piperazine was chosen, as it makes the desired compounds easily accessible by amination reactions. Moreover, the piperazine scaffold in ligand **AWM** was already masked by a methyl group and thus, a different carbon substituent in this positon should be easily tolerated without major changes of its physicochemical properties.

In addition, fragment **LUY** suggested that aliphatic substituents can be accommodated in the pocket and might promote additional interactions. Therefore, the commercially available benzothiazoles with chloro- and trifluoro-methyl substituents might be good starting points. The crystal structures did not provide any concrete hints for substituents at the indole moiety, but adding polar groups or aliphatic substituents in order to form additional interactions, e.g. with residues Gln91 or Leu95 could be a promising approach to improve the interactions. Commercially available are three indole carbaldehydes with the following substituents: 7-methoxy, 7-chloro and 5-hydroxy.

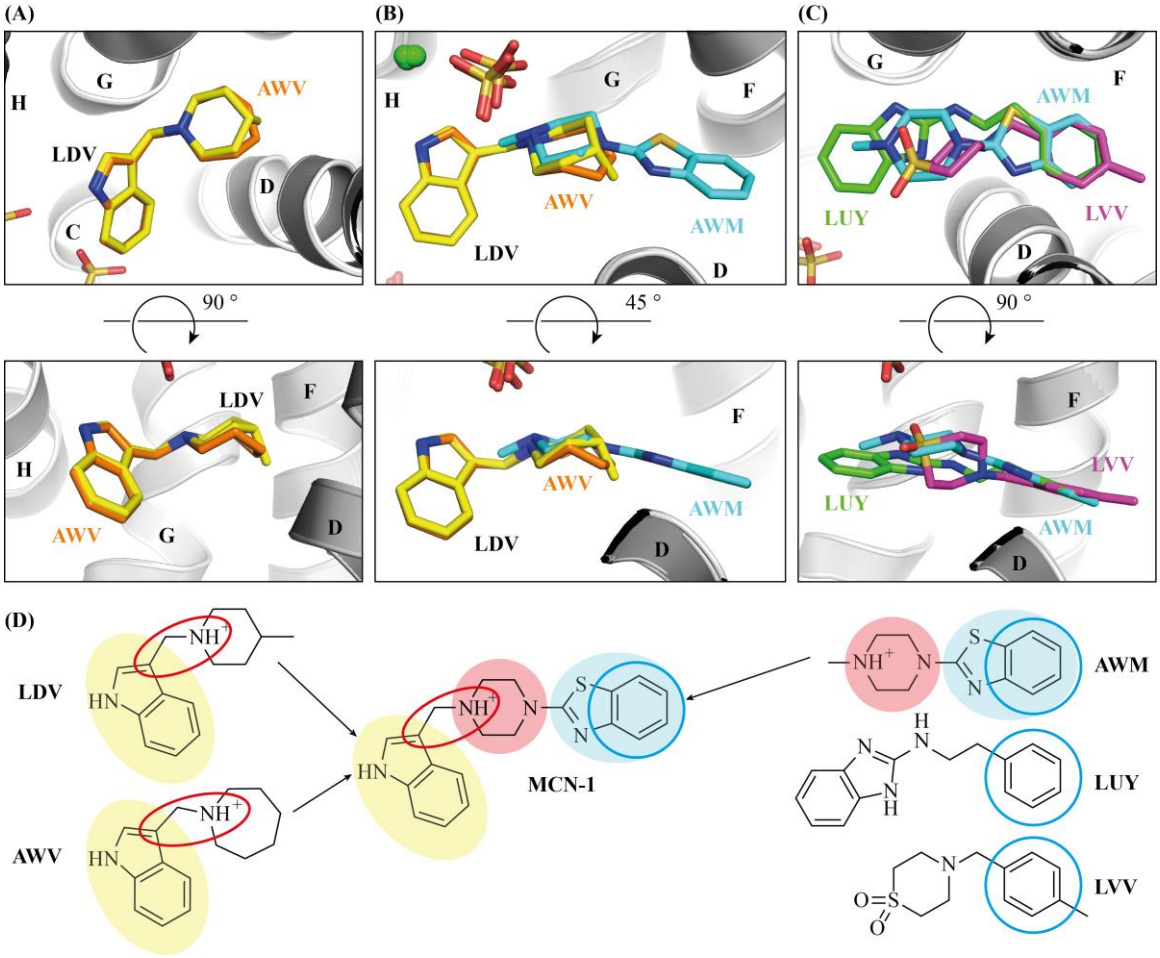



**Figure 53**: Binders of TcFPPS that were starting points for fragment merging. **(A)** Superimposition of **LDV** (yellow) and **AWV** (orange) (PDB IDs 5QPG and 5QPN (this work)). **(B)** Superimposition of **LDV**, **AWV** and **AWM** (cyan) (PDB IDs 5QPF, 5QPN and 5QPG (this work)). **(C)** Superimposition of **AWM**, **LVV** (violet) and **LUY** (green) (PDB IDs 5QPF, 5QPT and 5QPK (this work)). **(D)** Chemical structures of the fragments **LDV**, **AWV**, **AWM**, **LVV** and **LUY** that led to merger **MCN-1**. Certain moieties are highlighted with an ellipsoid: indole in light yellow, piperazine in light red, benzothiazole in light blue, phenyl in cyan, spacer in red.

Considering all possible combinations of building blocks with and without substituents, a fragment follow-up library of 12 compounds, **MCN-1** to **MCN-12**, was chosen for synthesis (**Figure 54**). All 12 compounds were synthesised by reductive amination in a one-pot reaction conducted in two steps. In this reaction, first the 2-(piperazin-1-yl)benzo[*d*]thiazole formed an imine with the indole-3-carbaldehyde and was protonated to an iminium cation under acidic conditions. For the second reaction step, sodium triacetoxyborohydride was added, forming the desired amine **MCN-1** under reduction of the iminium cation (**Figure 54**).

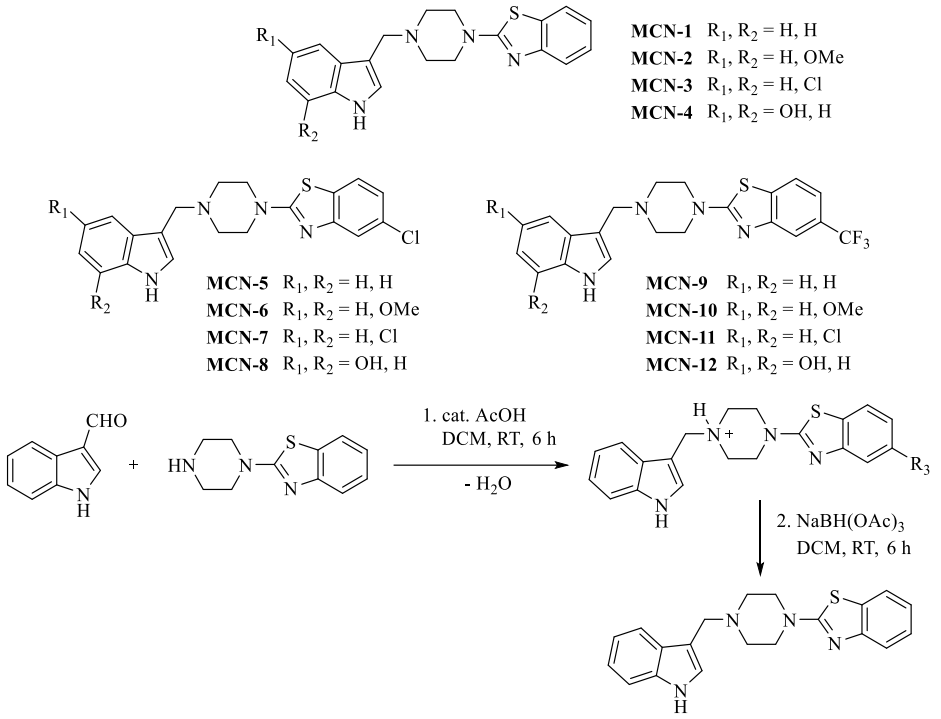



**Figure 54**: Chemical structures of compounds **MCN-1** to **MCN-12**, which were chosen for synthesis and reaction scheme of reductive amination.

Two out of three desired 2-piperazinebenzothiazoles had to be synthesised, as they were not commercially available (**Figure 55**).

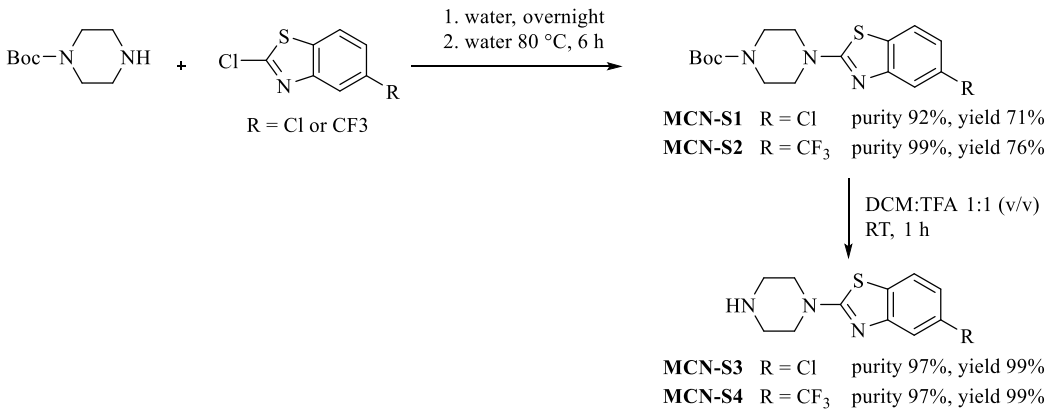
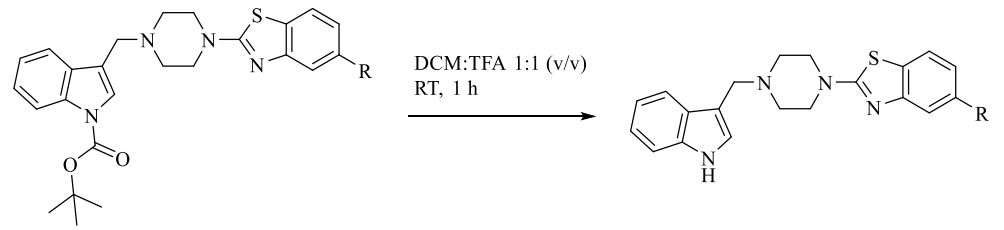



**Figure 55**: Synthesis of **MCN-S3** and **MCN-S4**. C-N bond formation and deprotection of **MCN-S1** and **MCN-S2** yielded the 2-pierazinebenzothiazoles **MCN-S3** and **MCN-S4** respectively. Purities and yields are given for each compound.

A sustainable chemistry approach was chosen for C-N bond formation, which was published by Kumar *et al*.[467] The reaction was performed with slight adaptations (methods section **4.8.3**). In brief, no product formation was observed after overnight stirring and thus, the reaction mixture was stirred for an additional 6 h at 80 °C in a pressure tube. This led to complete

product formation. Purification by flash column chromatography yielded the products in high purity and good yield. The free amines were obtained with high purity and in high yields after cleavage of the Boc protecting group **Figure 55**).

After all starting materials were available, all twelve products were synthesised by reductive amination (**Figure 54**). For this purpose, the piperazine derivatives were mixed with the aldehydes and dissolved in DCM. Catalytic amounts of acetic acid were added and the reaction mixture stirred for 6 h at RT in a pressure tube under argon atmosphere. Then, the reaction mixture was cooled to 0 °C, the reducing agent was added, and the mixture was stirred for another 6 h at RT. Impurities were extracted with water and the organic phases were combined and concentrated *in vacuo* for final purification by preparative LC-MS. The unsubstituted indole-3-carbaledhyde was used with Boc protecting group in the reductive amination, resulting in intermediates **MCN-S5**, **MCN-S6** and **MSN-S7**, which were obtained with high purities and medium to high yields **(Figure 56)**. The final products **MCN-1**, **MCN-5**, **MCN-9** were obtained after cleavage of the Boc protecting group in DCM:TFA 1:1 (v/v) and subsequent purification by preparative LC. All three products were obtained at high purities and good yields. (**Figure 56**).

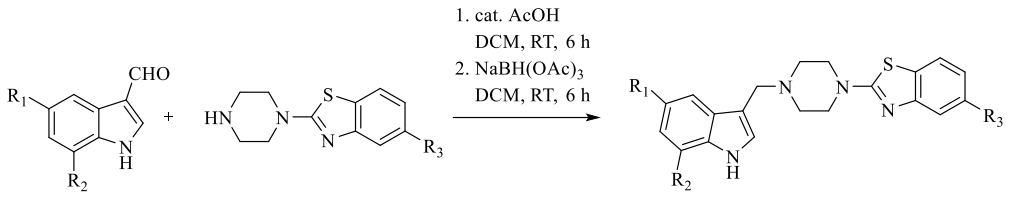



**MCN-S5** R = H purity 92%, yield 93%
**MCN-S6** R = Cl purity 92%, yield 56%
**MCN-S7** R = $CF_3$ purity 92%, yield 45%

**MCN-1** R = H purity 89%, yield 63%
**MCN-5** R = Cl purity 99%, yield 61%
**MCN-9** R = $CF_3$ purity 97%, yield 64%

**Figure 56**: Synthesised compound **MCN-1**, **MCN-5** and **MCN-9** Deprotection of the intermediates **MCN-S5**, **MCN-S6** and **MCN-S7** yielded the final products **MCN-1**, **MCN-5** and **MCN-9**, respectively. Purities and yields are given for each compound.

The three remaining indole-3-carbaldehydes could not be purchased with BOC-protection group at the indole amine. Nevertheless, the desired products formed and the introduction of a protective group was not necessary. Hence, products **MSN-2** to **MSN-4**, **MSN-6** to **MSN-8** and **MSN-10** to **MSN-12** could be obtained in a one-step synthesis. While purities were good to excellent, the yields were modest with exception of **MCN-8** (**Figure 57**).




| | | |
|---|---|---|
| **MCN-2** | $R_1$, $R_2$ = H, OMe $R_3$ = H | purity 99% yield 30% |
| **MCN-3** | $R_1$, $R_2$ = H, Cl $R_3$ = H | purity 96%, yield 71% |
| **MCN-4** | $R_1$, $R_2$ = OH, H $R_3$ = H | purity 92%, yield 22% |
| **MCN-6** | $R_1$, $R_2$ = H, OMe $R_3$ = Cl | purity 85%, yield 48% |
| **MCN-7** | $R_1$, $R_2$ = H, Cl $R_3$ = Cl | purity 98%, yield 44% |
| **MCN-8** | $R_1$, $R_2$ = OH, H $R_3$ = Cl | purity 88%, yield 2% |
| **MCN-10** | $R_1$, $R_2$ = H, OMe $R_3$ = $CF_3$ | purity 94%, yield 42% |
| **MCN-11** | $R_1$, $R_2$ = H, Cl $R_3$ = $CF_3$ | purity 94%, yield 42% |
| **MCN-12** | $R_1$, $R_2$ = OH, H $R_3$ = $CF_3$ | purity 88%, yield 19% |

**Figure 57**: Synthesised compounds **MCN-2** – **MCN-4**, **MCN-6** – **MCN-8** and **MCN-10** – **MCN-12**. Purities and yields are given for each compound.

The synthesised compounds were utilised in several experiments to investigate their binding affinities (**Table 25**).

**Table 25**: Results of 2D NMR experiments of the compound series **MCN-1** to **MCN-12**.

| Compound | Chemical shift changes in 2D NMR<br>count | Strength of chemical shift changes in 2D NMR<br>count | Solubility in SPR buffer<br>(µM) |
|---|---|---|---|
| **MCN-1** | 30 | strong | 60 |
| **MCN-2** | 30 | mostly weak | <10 |
| **MCN-3** | 10 | very weak | 15 |
| **MCN-4** | 30 | medium to strong | 180 |
| **MCN-5** | 2 | weak | <10 |
| **MCN-6** | 5 | weak | 65 |
| **MCN-7** | 2 | weak | <10 |
| **MCN-8** | 30 | medium to strong | 20 |
| **MCN-9** | 10 | weak | <10 |
| **MCN-10** | none | - | 13 |
| **MCN-11** | 2 | weak | <10 |
| **MCN-12** | 30 | medium | 50 |

First, solubility tests in SPR buffer were conducted by 1D NMR spectroscopy using DSS as an internal standard. Whilst five compounds showed poor solubility in the SPR buffer, six showed solubility only in double digit µM range. Compound **MCN-4** showed the best solubility at 180 µM. To determine binding affinities by SPR, it is desired to measure dilution series of compounds up to 5-fold or 10-fold higher than the expected $K_d$ in order to reach a plateau/saturation. Due to their low solubility in SPR buffer, it was unlikely to collect spectra that

would result in sufficient dose-response curves unless the $K_d$ values would be in the single digit µM range. Thus, 2D NMR experiments were conducted to test affinity to TcFPPS and to assess the strength of the chemical shift changes. [$^{13}C^{1}H$]-SOFAST-HMQC were conducted in aqueous buffer with samples that contained 700 µM compound and 30 µM $^{13}C^{15}N$-labelled TcFPPS. Therefore, all compounds were measured at saturation. Whilst seven compounds showed chemical shift changes for a number of signals, the remaining five showed weak chemical shift changes for a small number of signals or for none at all (**Table 25**).

A series of samples at different concentrations of compound **MCN-4** were collected, ranging from 31.25 µM to 1.00 mM. 1D spectra showed that its solubility in the aqueous NMR buffer is ≥500 µM, but ≤1 mM (**Figure 58**).

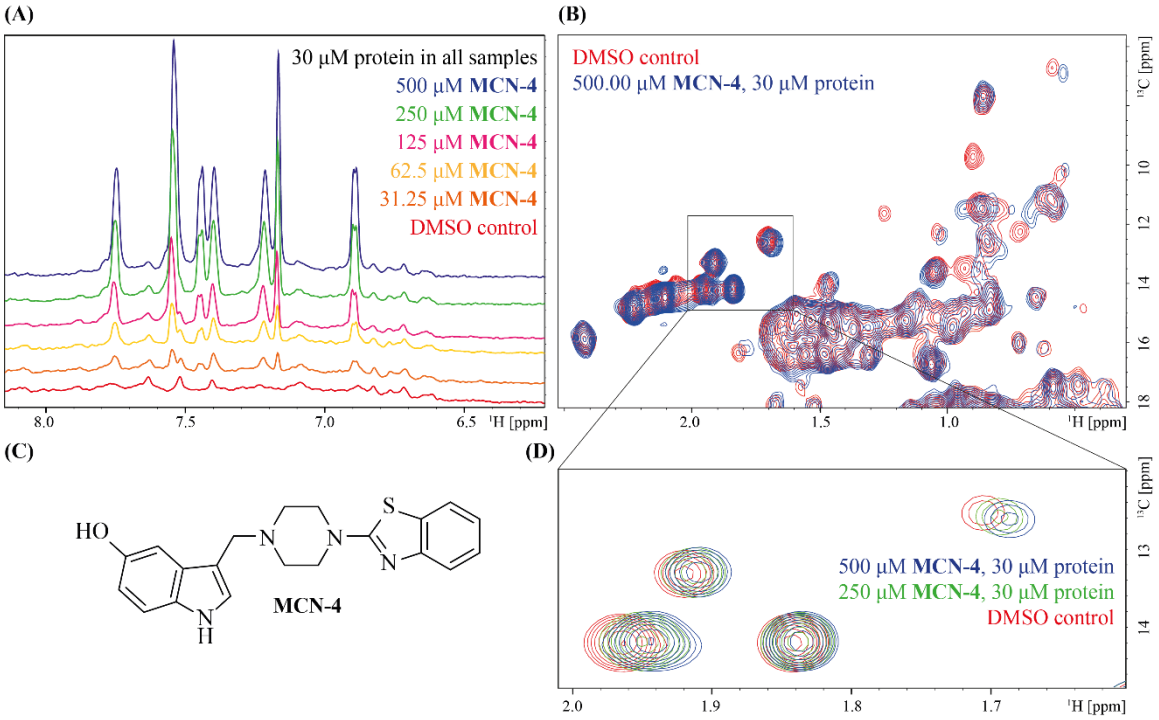



**Figure 58**: **MCN-4** is a binder of TcFPPS that shows a clear dose response in 2D-NMR. **(A)** 1D spectra that demonstrate compound solubility up to 500 µM. **(B)** Cut-out from an overlay of the [$^{13}C^{1}H$]-SOFAST-HMQC spectrum of the DMSO control (red) and the corresponding sample (blue) at 500 µM compound and 30 µM protein in 25 mM BisTris, pH 6.5, 50 mM NaCl, 2 mM TCEP · HCl, 10% $D_2O$, 150 µM DSS at 305 K. **(C)** Chemical structure of **MCN-4**. **(D)** Chemical shift changes demonstrating dose response for clarity three spectra are shown.

Superimposition of the NMR spectra of samples with different compound concentration showed a clear dose-response proving that compound **MCN-4** binds to TcFPPS. Nevertheless, saturation was not reached and the binding affinity could not be determined. Therefore, the $K_d$ was determined to have a lower limit of >500 µM rendering SPR experiments unfeasible (**Figure 58**). Utilizing all 12 compounds, soaking experiments were performed with apo TcFPPS crystals that

were grown on 96-well plates (methods section **4.3.5**). High-resolution diffraction data were collected after two rounds of crystallization trials for 10 compounds (**Table 26**).

**Table 26**: Soaking experiments with **MCN-1** to **MCN-12**.

| Compound | TFA salt yes / no | Data collection soaking trial 1[a] | Data collection soaking trial 2[b] | PDB ID |
|---|---|---|---|---|
| **MCN-1** | yes | crystal showed no diffraction | collected (4 h and 7 h soak) | 6R09 |
| **MCN-2** | yes | crystal showed no diffraction | collected (4 h and 7 h soak) | - |
| **MCN-3** | yes | crystal showed no diffraction | crystal showed no diffraction | - |
| **MCN-4** | no | collected (17 h soak) | - | 6R0A |
| **MCN-5** | yes | crystal showed no diffraction | collected (4 h and 7 h soak) | - |
| **MCN-6** | yes | crystal showed no diffraction | crystal showed no diffraction | - |
| **MCN-7** | yes | crystal showed no diffraction | collected (4 h and 7 h soak) | - |
| **MCN-8** | no | collected (17 h soak) | - | 6R0B |
| **MCN-9** | yes | crystal showed no diffraction | collected (4 h and 7 h soak) | - |
| **MCN-10** | yes | collected (4 h soak) | - | - |
| **MCN-11** | yes | crystal showed no diffraction | collected (4 h and 7 h soak) | - |
| **MCN-12** | no | collected (17 h soak) | - | - |

[a] Soaking trial 1 was performed at a nominal compound concentration of 75 mM (15% DMSO).
[b] Soaking trial 2 was performed after neutralizing the TFA salts by equimolar amounts of base at a compound concentration of 37.5 mM (15% DMSO).

Data processing and refinement lead to crystal structures of TcFPPS in complex with three of the synthesised compounds, **MCN-1**, **MCN-4** and **MCN-8**. Strikingly, these compounds were not harboured in the enzyme's active site. This result was unexpected, because the fragment merging approach was based on an almost perfect overlap suggesting that the binding site of the merged compound should be retained. In addition, docking using the software tool Glide[464] showed that compound **MCN-1** could bind to the desired binding site without steric clashes. The binding poses differed slightly when docking the molecule with a protonated piperazine moiety or when docked in protonated state, but in both cases the binding poses were in good agreement with the binding poses of the fragment hits **LDV** and **AWM**. Nevertheless, all three compounds bind in the region corresponding to the binding site S1 (**Figure 59 (A)**), which is a small cavity on the protein surface formed by helices H, I and α3. The site was previously described in this work as binding site for lead compound **119** (PDB ID 6R08, chapter **5.2**). In addition, several fragments identified in the FBS campaign were binding to this site. This included compounds **AWM**, **LVV** and **AWV**, which show site S1 as secondary binding site.

Ligands **MCN-1**, **MCN-4** and **MCN-8** were refined to an occupancy of 0.77, 0.85 and 0.73, respectively. Whilst ligand **MCN-4** is excellently resolved as the $mF_o - DF_c$ difference electron density map contoured at 3.0 σ indicates (summary of density maps, Appendix, **Figure 79 (D – F)**), ligands **MCN-1** and **MCN-8** were not entirely resolved (summary of density

maps, Appendix, **Figure 79 (A – C)** and **(J – L)**). The diffraction limits of the complexes are 1.28 Å, 1.32 Å and 1.61 Å, and the coordinates have been deposited under PDB IDs 6R09, 6R0A, and 6R0B, respectively. Notably, the crystals structure of TcFPPS in complex with **MCN-1** has the best diffraction limit ever obtained for a TcFPPS crystal. For collection and refinement statistics see **Table 29** in the Appendix.

Ligands **MCN-1**, **MCN-4** and **MCN-8** mainly show unspecific hydrophobic interactions and π-stacking with residue Phe256. The angles and distances to residue Gln318 are not ideal for H-bonding (**Figure 59 (C)** – **(E)**).

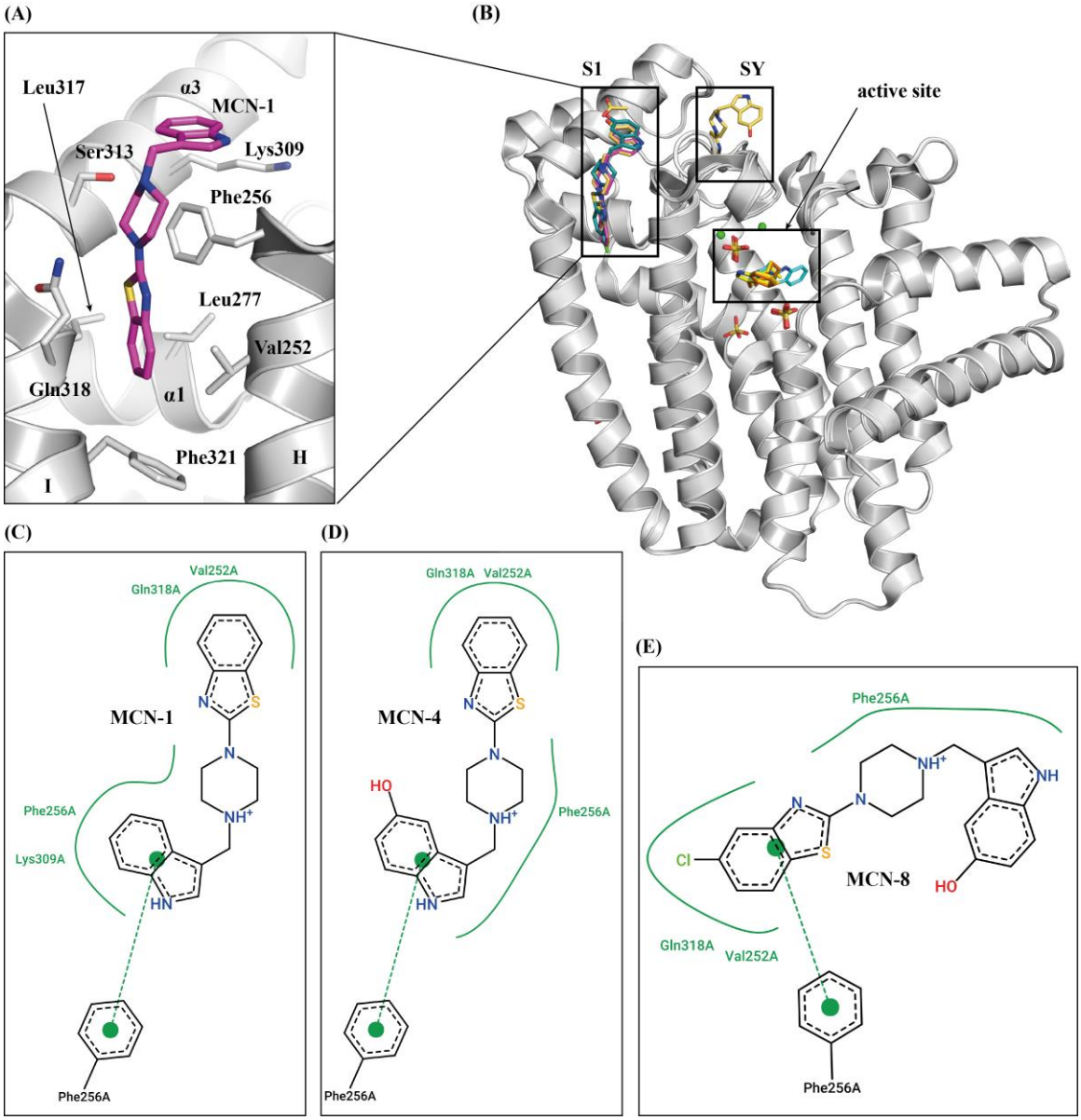



**Figure 59**: Merged compounds bind to the surface-directed site S1 of TcFPPS. Figure is continued on the next page. For legend also see next page.

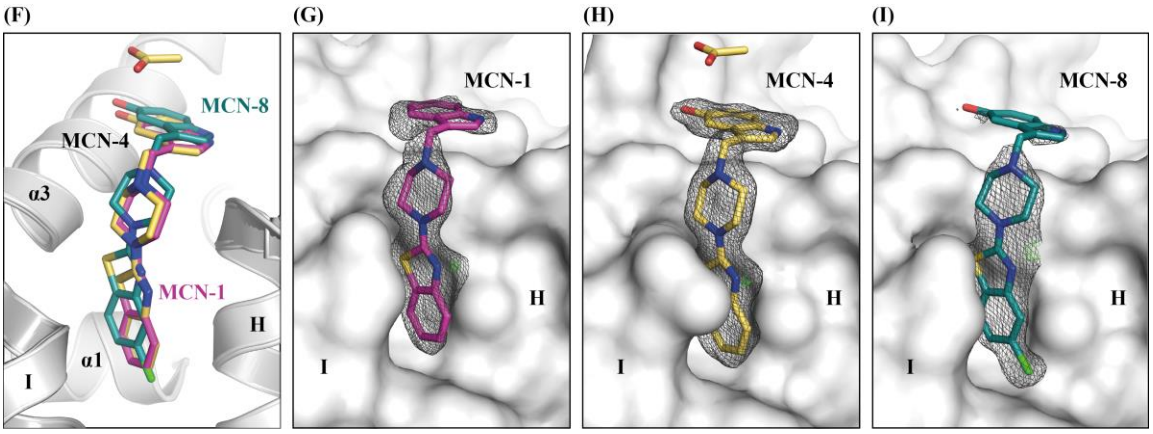



**Figure 59:** See also previous page. **(A)** Crystal structure of TcFPPS in complex with **MCN-1** (PDB ID 6R09 (this work) residues interacting with the ligand are also depicted in stick representation). **(B)** Superimposition of the crystal structures of TcFPPS in complex with the compounds **MCN-1**, **MCN-4**, **MCN-8**, **LDV**, **AWV** and **AWM (PDB IDs** PDB IDs 6R09, 6R0A, 6R0B, 5QPN, 5QPG and 5QPF, respectively (this work) ligands and $SO_4^{2+}$ ions are shown in stick representation, the protein backbone is shown in cartoon representation and $Zn^{2+}$ ions are shown as green coloured spheres. **(C) – (E)** 2D structure diagram of ligands **MCN-1**, **MCN-4**, **MCN-8**, respectively, interacting with TcFPPS. Diagram was generated using PoseView[481] **(F)** Superimposition of the crystal structures of **MCN-1**, **MCN-4** and **MCN-8**. **(G)** – **(I)** Binding site of ligands **MCN-1**, **MCN-4** and **MCN-8**, respectively. Protein shown in surface representation. Refined $2F_o - F_c$ electron density map is contoured and represented as liquorice coloured mesh. Positive and negative $F_o - F_c$ electron density map is contoured at 3.5 σ and represented as green and red coloured mesh, respectively.

Superimposition shows a perfect overlap of ligands **MCN-1** and **MCN-4**. Ligand **MCN-8** has a similar binding mode to compounds **MCN-1** and **MCN-4**, but due to its chlorine substituent it is shifted upwards (**Figure 59 (D)** – **(G)**). For the derivative **MCN-4** a secondary binding site at a crystal contact was observed corresponding to the SY site (summary of density maps, Appendix, **Figure 79 (G – I)**) that was previously observed for compound **M0D**, which was also identified in the FBS campaign by X-ray crystallography.

### 5.6.2 Literature review revealed promising compounds with similar scaffolds

An in-depth literature review revealed a compound with nanomolar activity for *T. cruzi* in the ChEMBL database. It is *N*-((1*H*-imidazol-2-yl)methyl)-*N*-(4-(benzo[*d*]thiazol-2-yl)phenyl) methanesulfon-amide (**DND*i*-1**, CHEMBL2448735), which had been tested *in vitro* against TcFPPS and exhibited an $IC_{50}$ of 0.1 nM (Assay ID CHEMBL2448754). It was developed and tested by Keenan *et al.*[490] from the Drugs for Neglected Disease *initiative* (DND*i*) and belongs to a series of compounds named CM74. A selection of compounds of this series is depicted in **Figure 60**. Whilst **DND*i*-1** has the benzothiazole scaffold in common with the herein synthesised compound series **MCN-1** to **MCN-12**, a phenyl moiety forms the central moiety instead of a piperazine. In case these compounds bind to TcFPPS and are active site-directed, its sulphonamide moiety, might interact with the FARM and SARM in the way bisphosphonates interact with these

conserved regions. The hypothesis that the scaffolds might bind to the active site of FPPS is underscored by the fact that sulphates were previously observed to bind to this position. Additionally, crystal structures of TcFPPS described in this work accommodate a sulphate ion in close proximity to the active site-directed binders **AWM**, **AWV** and **LDV** (chapter **5.4**).

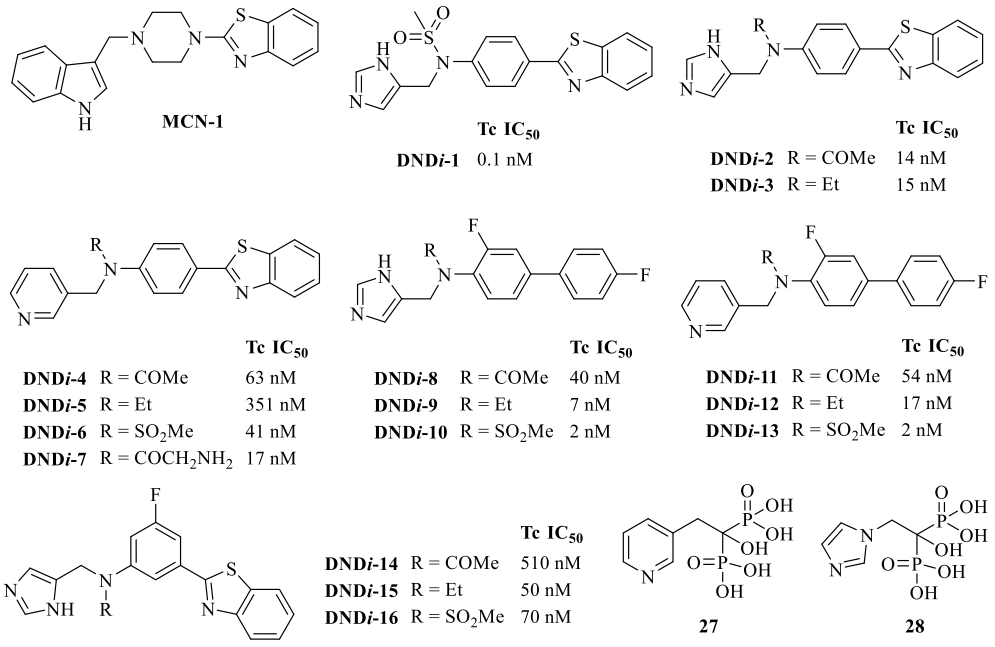



**Figure 60**: Chemical structure of **MCN-1**, compounds from the CM74 series[491] and RIS (**27**) and ZOL (**28**). Smiles codes and numbering of the compounds from the CM74 series in the Keenan publication[490] are given in **Table 38** in the Appendix.

Indeed, molecular docking of compound **DND*i*-1** into the closed-state TcFPPS (PDB ID 1YHL[162b]) using the software tool Amber10:EHT[466] suggested its binding to the active site. The docking pose shows that the sulphonamide moiety interacts with the $Mg^{2+}$ ions that are coordinated by Asp98 and Asp102 of FARM and Asp250 of the SARM. Whilst the benzothiazole moiety of fragment **AWM** is accommodated by a channel formed by helices D and F, the imidazole moiety of **DND*i*-1** is protruding into this channel. Overlays with the crystal structures of TcFPPS in complex with the N-BPs **RIS** and **ZOL** show that their aromatic side chains occupy the same space. The benzothiazole moiety of **DND*i*-1** is suggested to protrude into the IPP binding site where the nitrogen can form H-bonds with Lys48 and Gln91 (**Figure 61**). Structural comparison of TcFPPS in complex with N-BPs, which are strong and rapid active site-directed inhibitors, shows that the aromatic moieties of the N-BPs risedronate (**27**, RIS, Actonel®, Merck)[223] and zoledronate (**28**, ZOL, Zometa®, Novartis)[220] occupy the same space as the imidazole moiety of **DND*i*-1**. In addition, the benzothiophene moiety of compound **DND*i*-1** protrudes from the DMAPP binding site into the IPP binding site (**Figure 61**). Superimposition of the ligands reveals that the compound

would be able to occupy the DMAPP and IPP binding site. Whether these compounds bind to TcFPPS awaits testing.

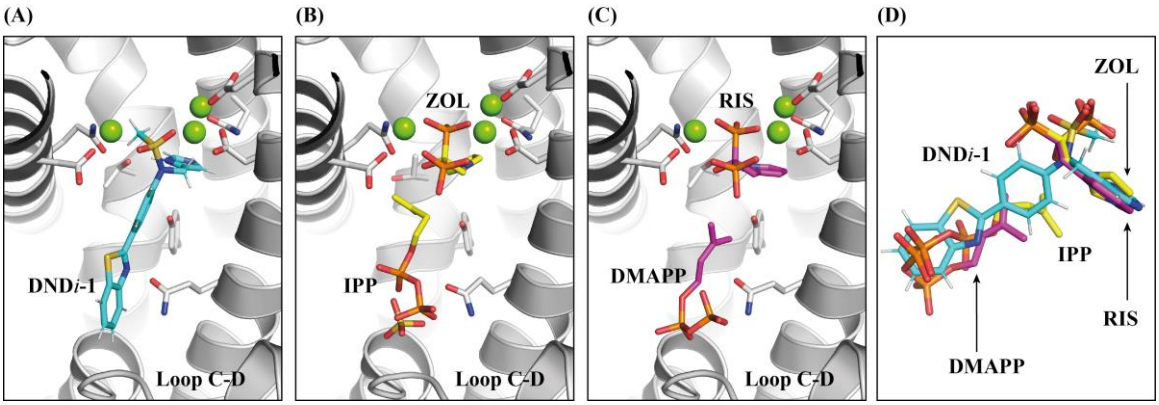



**Figure 61**: Docking model of **DNDi-1** binding to TcFPPS and comparison to crystal structures of TcFPPS in complex with the N-BPs ZOL and RIS and its natural substrate IPP and DMAPP (bound in the IPP site). **(A)** Docking pose of **DNDi-1** binding to TcFPPS (PDB ID 1YHL[162b]). **(B)** Crystal structure of TcFPPS in complex with ZOL and IPP bound (PDB ID 3IBA[211]). **(C)** Crystal structure of TcFPPS in complex with RIS and DMAPP (PDB ID 1YHL[162b]). **(D)** Superimposition with crystal structure of TcFPPS in complex with ZOL and IPP (PDB ID 3IBA[211]) and RIS and DMAPP (PDB ID 1YHL[162b]) (backbones not shown).

### 5.6.3 Discussion

The binding modes of compounds **MCN-1**, **MCN-4** and **MCN-8** were unexpected. Although the compounds remained flat and unbranched scaffolds, it is tempting to speculate that the compounds did not bind to the active site because their increased size after merging might provoke steric clashes with the protein channel directing the compound to the active site. Co-crystallization experiments could have circumvent this issue. Nevertheless, in such experiments, steric crowding might be further exacerbated by the lack of electrostatic charges and the resulting poor compound solubility. Competition with other ligands occupying the binding site is unlikely, because the applied crystallization conditions were very similar to the ones which led to the identification of the fragment hits. Another reason why the compounds do not target the active site could be that the interactions formed by the merged fragments are not specific enough. Drwal *et al*.[492] reported that the binding mode between a fragment and a related drug-like ligands is conserved. Polar interactions are better conserved. When looking at the fragments **LDV** and **AWV** that have been uses as starting points, there is only one energetically favoured H-bond formed by the indole moiety of fragments to Asp250. A second H-bond is formed by the piperidine, azepane and piperazine moiety of **LDV**, **AWV** and **AWM**, respectively. Nevertheless, this bond is not formed directly with the protein, but via a sulphate and $Zn^{2+}$ ion. As the piperazine moiety is now part of the linker, its pKa values should differ from the ones found for the fragments. Even if docking experiments of **MCN-1** in the protonated and un-protonated forms suggested very similar

binding poses, the un-protonated state would lead to the loss of the H-bond. The third important interaction is the π-stacking observed for the benzothiazole moiety of ligand **AWM**. However, this interaction is not specific. At binding site S1, π-stacking of the ligands with residue Phe256 is the key interaction.

Nevertheless, these findings are also vital for the design of novel compounds in a second iteration. One starting point is to vary the pKa of the linking moiety. Determining the correct pKa value is difficult, because protonation and pKa values change in protein-ligand binding[493]. An increased pKa value leads to higher basicity and therefore will assure protonation and the ability to contribute as a charge-assisted H-bond. In addition, higher basicity will increase solubility in aqueous buffers. Other linking moieties that have higher pKa values are pyrrolidine or a piperidine. Another option could be an open-chain spacer, such as *N*-methylethane-1,2-diamine, which has a tertiary amine which is protonated in neutral aqueous buffer systems. Another option is to add nitrogen groups to the molecule to enable additional H-bond formation. Instead of an indole moiety, a *1H*-pyrrolo[3,2-c]pyridine could contribute in a second H-bond. The findings related to compounds of the CM74 series suggest that a substituent that could directly interacting via $Zn^{2+}$ ions and the aspartate-rich motifs FARM and SARM should be considered. These options require iterative exploration to ultimately lead to a high-affinity binder. In addition, compounds with higher solubility in aqueous buffer systems should be prioritized to enable testing with biophysical methods.

# 6. Concluding remarks and outlook

As demonstrated in this work, unlabelled, $^{13}C^{15}N$-labelled and biotinylated avi-tagged *T. cruzi* farnesyl pyrophosphate synthase (TcFPPS) can be obtained in sufficient amounts and purity for fragment screening campaigns, structural experiments, and biophysical characterisation. Furthermore, the novel, reliable, highly reproducible, and well-diffracting crystallization system that was established for TcFPPS exhibits excellent properties for fragment-based screening (FBS). This crystallization system had significant impact on this work but will also pave the way for future studies aiming to identify TcFPPS binders and contribute to structure-based lead design of TcFPPS inhibitors.

The FBS by NMR campaigns identified 109 validated fragment hits. Several of them were further exploited by X-ray crystallography and revealed a first active site binder of a novel, non-bisphosphonate (non-BP) scaffold. Its identification underscored the power of Pan-Dataset Density Analysis (PanDDA) when dealing with partially bound fragments that require conformational changes of amino acid side chains. In addition, PanDDA accelerated analysis of the large batches of diffraction data sets obtained throughout this work. FBS by X-ray crystallography revealed several binders of a novel scaffold in the active site and also in additional binding sites in TcFPPS, which are spread over the entire protein. Thus, applying FBS by X-ray crystallography on TcFPPS was superior to previously conducted stepwise screening by NMR spectroscopy and follow-up in crystallisation experiments. The fragments identified by FBS by X-ray crystallography could provide opportunities to develop novel inhibitors for TcFPPS and will give new ideas for the drug discovery for Chagas disease. This applies in particular to the binders identified in the allosteric site of TcFPPS. All ligands that have been identified in this region show π-stacking with the phenyl side chain of residue Phe50 as key interaction with the protein but show two different binding modes. As this residue resembles a structural difference between the pathogenic FPPS and the human FPPS, it can be exploited to engineer inhibitor specificity. In addition, a novel mode of action and different physicochemical properties of inhibitors such as lower affinities to bone mineral might help to overcome the limitations related to the BP scaffold.

Even though a potent lead compound was not discovered in the first cycle of fragment-to-lead optimization employing fragment merging and by virtual design, the fragment hits and the 50 crystal structures of TcFPPS-fragment complexes provided in this work will pave the way for future lead discovery campaigns. The large diversity of scaffolds and the accommodation in different binding sites are potential starting point for SBLD, molecular docking and pharmacophore analysis. Hence, they may result in a tool compound that could prove the concept of allosteric inhibition of TcFPPS.

# Appendix

**Table 27**: Latest review articles on current efforts in drug discovery against CD.

| First author | Year | Topics covered | Citation |
|---|---|---|---|
| Paucar | 2016 | Overview of collaborative alliances; overview of clinical trials; target product profile (TPP) | [7c] |
| Ferreira | 2016 | Overview of collaborative partnerships; target-based approaches: focusing on cruzain and CYP51; phenotypic-based approaches | [98] |
| Moraes | 2016 | Methodology: HTS, HCS, Target-based screening | [67] |
| Salomão | 2016 | Detailed CD portrait, HCS, proteomics, drug repositioning, target enzymes: CYP51, cruzain, trypanothione reductase, flashlight on nitro compounds, clinical trials | [68d] |
| Bermudez | 2016 | Treatment: BNZ, NFX; targets: nitroreductase type I, ergosterol synthesis, toposisomerase inhibitors, cruzain, trans-sialidase, New compounds: repositioning | [37] |
| Scarim | 2018 | Phenotypic-based and target-based screening: nitroreductase, cruzain, SQS, FPPS | [99a] |
| Alonso-Padilla | 2014 | Short review on HTS | [75] |
| Zingales | 2014 | Stain diversity | [30] |
| Keenan | 2015 | CYP51, clinical trials, sterol biosynthesis | [491] |
| Duschak | 2016 | Extremely detailed review on targets and patented drugs for CD in the last 15 years | [156b] |
| Sanchez-Sanchez | 2016 | Targets: triosephosphate isomerase, glyceraldehyde 3-phosphate dehydrogenase, trypanothione reductase, cruzain, squalene synthase, FPPS and CYP51 | [494] |
| Rodriguez | 2016 | Patent review | [104] |
| Gilbert | 2013 | Target-based and phenotypic-based | [136c] |
| Clayton | 2010 | List of clinical trials and target-based approaches | [111b] |

**Table 28**: Crystal structures of TcFPPS and TcFPPS complexes published by 2019.

| No | PDB ID | citation | Diffraction limit (Å) | Ligands |
|---|---|---|---|---|
| 1 | IYHK | [162b] | 2.10 | apo protein, $SO_4^{2-}$ |
| 2 | IYHL | [162b] | 1.95 | risedronate, DMAPP, $SO_4^{2-}$, $Mg^{2+}$, |
| 3 | IYHM | [162b] | 2.50 | alendronate, IPP, $Mg^{2+}$, $SO_4^{2-}$ |
| 4 | 3IBA | [211] | 2.40 | zoledronate, IPP, $Mg^{2+}$, $SO_4^{2-}$ |
| 5 | 3ICK | [211] | 2.40 | minodronate, IPP, $Mg^{2+}$, $SO_4^{2-}$ |
| 6 | 3ICM | [211] | 2.20 | 1-(2-hydroxy-2,2-bis-phosphono-ethyl)-3-penyl-pyridinium, IPP, $Mg^{2+}$, $SO_4^{2-}$ |
| 7 | 3ICN | [211] | 2.40 | 3-fluoro-1-(2-hydroxy-2,2-bis-phosphonoethyl)pyridinium, IPP, $Mg^{2+}$, $SO_4^{2-}$ |
| 8 | 3ICZ | [211] | 2.15 | 3-[(1E)-but-1-en-1-yl]-1-(2,2-diphosphonoethyl)pyridinium, IPP, $Mg^{2+}$ |
| 9 | 3ID0 | [211] | 2.81 | 3-fluoro-1-(2-hydroxy-2,2-diphosphonoethyl)pyridinium, $Mg^{2+}$, $SO_4^{2-}$ |
| 10 | 4DWB | [170a] | 2.10 | [2-(pentylamino)ethane-1,1-diyl]bis(phosphonicacid), IPP, $Mg^{2+}$, $Na^{+}$, $SO_4^{2-}$, acetate ion |
| 11 | 4DWG | [170a] | 2.01 | [2-(heptylamino)ethane-1,1-diyl]bis(phosphonicacid), $Mg^{2+}$, $Na^{+}$, $SO_4^{2-}$, acetate ion, di(hydroxyethyl)ether |
| 12 | 4DXJ | [170a] | 2.35 | [2-(propylamino)ethane-1,1-diyl]bis(phosphonicacid), IPP, $Mg^{2+}$, $Na^{+}$, $SO_4^{2-}$, acetate ion, triethylene glycol, di(hidroxyethyl)ether |
| 13 | 4DZW | [170a] | 3.05 | [2-(cyclohexylamino)ethane-1,1-diyl]bis(phosphonicacid), IPP, $Mg^{2+}$, $SO_4^{2-}$ |
| 14 | 4E1E | [170a] | 2.65 | [2-(hexylamino)ethane-1,1-diyl]bis(phosphonicacid), IPP, $Mg^{2+}$, $Na^{+}$ |

**Table 29:** Data collection and refinement statistics of TcFPPS crystal structures.

| **PDB ID** | **6R04** | **6R05** | **6R06** | **6R07** |
|---|---|---|---|---|
| **compound** | | | | |
| PDB identifier | apo | **JNE** | **JMN** | **3N2** |
| naming in this thesis | apo | **CS-18** | **CS-33** | **93** |
| | | | | |
| **Data collection** | | | | |
| X-ray source | X10SA, SLS | X10SA, SLS | X10SA, SLS | X10SA, SLS |
| Wavelength (Å) | 0.99991 | 0.99981 | 0.99999 | 0.99984 |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | |
| *a* = *b*, *c* (Å) | 57.65, 397.59 | 58.11, 396.69 | 58.065, 397.51 | 58.07, 397.51 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 66.26 – 1.47 | 66.12 – 1.57 | 50.29 – 1.56 | 65.75 – 1.57 |
| | (1.50 – 1.47)[a] | (1.60 – 1.57)[a] | (1.59 – 1.56)[a] | (1.60 – 1.57)[a] |
| $R_{merge}$ | 0.060 (2.813)[a] | 0.099 (3.102)[a] | 0.097 (3.043)[a] | 0.082 (4.446)[a] |
| Unique reflections | 68742 (3327)[a] | 57627 (2762)[a] | 58053 (2851)[a] | 56620 (2739)[a] |
| $I$ / σ$I$ | 20.8 (0.9)[a] | 15.1 (0.8)[a] | 15.5 (0.8)[a] | 19.4 (0.7)[a] |
| Completeness (%) | 100 (100)[a] | 100 (100)[a] | 99 (100)[a] | 99.7 (100)[a] |
| Redundancy | 18.1 (17.6)[a] | 18.0 (18.3)[a] | 17.8 (18.1)[a] | 18.8 (18.8)[a] |
| $CC_{1/2}$ | 1.000 (0.380 )[a] | 0.998 (0.417)[a] | 1.000 (0.342)[a] | 1.000 (0.342)[a] |
| | | | | |
| **Refinement** | | | | |
| Resolution (Å) | 49.930 – 1.469 | 22.62 – 1.57 | 50.286 – 1.559 | 50.09 – 1.57 |
| No. reflections | 68734 | 57590 | 58053 | 56619 |
| $R_{work}$ / $R_{free}$ | 0.1868 / 0.2123 | 0.1832 / 0.2085 | 0.1954 / 0.2236 | 0.1885 / 0.2139 |
| No. atoms | | | | |
| Protein | 2863 | 2843 | 2781 | 2843 |
| Ligand/ion | 16 | 34 | 54 | 39 |
| Water | 287 | 262 | 249 | 268 |
| B-factors Protein | | | | |
| Protein (Å$^2$) | 35.85 | 34.20 | 36.35 | 35.62 |
| Ligand/ion (Å$^2$) | 44.85 | 37.27 | 38.32 | 36.56 |
| Water (Å$^2$) | 45.46 | 45.39 | 43.41 | 47.25 |
| R.m.s. deviations | | | | |
| Bond length (Å) | 0.01 | 0.01 | 0.01 | 0.01 |
| Bond angles (°) | 0.91 | 0.89 | 0.88 | 0.88 |
| Molprobity statistics | | | | |
| Ramachandran | | | | |
| Favoured (%) | 98.04 | 98.31 | 98.60 | 99.15 |
| Outliers (%) | 0.00 | 0.00 | 0.00 | 0.00 |
| Allowed (%) | 1.96 | 1.69 | 1.4 | 0.85 |
| All-atom clash score | 1.04 | 1.05 | 1.36 | 1.04 |
| Solvent content (%) | 47.54 | 47.54 | 47.40 | 45.59 |

[a] Values in parentheses are for the highest resolution shell.

Table is continued on the next page.

**Table 29** continued.

| **PDB ID** | **6R08** | **6R09** | **6R0A** | **6R0B** |
|---|---|---|---|---|
| **compound** | | | | |
| PDB identifier | **GO1** | **JMK** | **JMT** | **JMW** |
| naming in this thesis | **119** | **MCN-1** | **MCN-4** | **MCN-8** |
| | | | | |
| **Data collection** | | | | |
| X-ray source | X10SA, SLS | X10SA, SLS | X10SA, SLS | X10SA, SLS |
| Wavelength (Å) | 0.99985 | 1.00000 | 1.00003 | 1.00003 |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | |
| *a* = *b*, *c* (Å) | 57.90, 398.23 | 58.11, 397,07 | 58.19, 395,89 | 58.09, 396.95 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 49.75 – 1.44 | 66.19 – 1.28 | 65.98 – 1.32 | 50.31 – 1.61 |
| | (1.46- 1-44)[a] | (1.30 – 1.28)[a] | (1.34 – 1.32)[a] | (1.64 – 1.61)[a] |
| $R_{merge}$ | 0.053 (4.737)[a] | 0.103 (4.077)[a] | 0.074 (4.580)[a] | 0.092 (3.595)[a] |
| Unique reflections | 74408 (3629)[a] | 96670 (5120)[a] | 95142 (4654)[a] | 52572 (2540)[a] |
| $I$ / σ$I$ | 23.1 (0.6)[a] | 11.7 (0.7)[a] | 16.7 (0.6)[a] | 16.9 (0.7)[a] |
| Completeness (%) | 100.0 (100.0)[a] | 92.3 (100.0)[a] | 99.3 (99.6)[a] | 98.9 (100.0)[a] |
| Redundancy | 18.7 (19.5)[a] | 18.6 (18.6)[a] | 18.7 (19.0)[a] | 18.0 (17.5)[a] |
| $CC_{1/2}$ | 1.000 (0.362)[a] | 0.998 (0.462)[a] | 1.000 (0.337)[a] | 1.000 (0.318)[a] |
| | | | | |
| **Refinement** | | | | |
| Resolution (Å) | 48.62 – 1.44 | 66.179 – 1.28 | 65.98 – 1.32 | 50.310 – 1.612 |
| No. reflections | 74407 | 96670 | 95139 | 52297 |
| $R_{work}$ / $R_{free}$ | 0.1906 / 0.2130 | 0.2090 / 0.2319 | 0.1998 / 0.2151 | 0.1906 / 0.2295 |
| No. atoms | | | | |
| Protein | 2868 | 2863 | 2852 | 2852 |
| Ligand/ion | 25 | 37 | 63 | 38 |
| Water | 335 | 264 | 363 | 278 |
| B-factors Protein | | | | |
| Protein (Å$^2$) | 34.49 | 27.91 | 26.86 | 35.29 |
| Ligand/ion (Å$^2$) | 49.06 | 39.66 | 36.58 | 55.20 |
| Water (Å$^2$) | 47.65 | 37.75 | 40.82 | 45.03 |
| R.m.s. deviations | | | | |
| Bond length (Å) | 0.01 | 0.01 | 0.01 | 0.01 |
| Bond angles (°) | 0.90 | 0.93 | 0.94 | 0.90 |
| Molprobity statistics | | | | |
| Ramachandran | | | | |
| Favoured (%) | 98.60 | 98.32 | 98.88 | 98.88 |
| Outliers (%) | 0.00 | 0.00 | 0.00 | 0.00 |
| Allowed (%) | 1.40 | 1.68 | 1.12 | 1.12 |
| All-atom clash score | 1.02 | 1.56 | 1.20 | 0.69 |
| Solvent content (%) | 47.19 | 47.42 | 47.42 | 47.38 |

[a] Values in parentheses are for the highest resolution shell.

**Table 30**: Overview of crystallization experiments with TcFPPS.

| Experiment | Protein | Plate | Drop (µL) | Volumes prot, res, seed (µL) | Ratio prot:res:seed (v/v) | Protein in crystallization drop (mg · mL$^{-1}$) |
|---|---|---|---|---|---|---|
| Condition screen I | Formulation I[a] | 2-drop 96-well SwissCi/MRC | 0.5 | 0.3, 0.2, - | 3:2 | 4.09 |
| Optimization I (Round 1) | Formulation I[a] | 24-well VDX 18 mm | 2.0 | 1.2, 0.8, - | 3:2 | 4.09 |
| Optimization I (Round 2, Round 3) and seed crystals | Formulation I[a] | 24-well VDX 18 mm | 1.5 | 1.0, 0.5, - | 2:1 | 4.54 |
| Condition screen II | Formulation II[b] | 2-drop 96-well SwissCi/MRC | 0.5 | 0.3, 0.2, - | 3:2 | 7.57 |
| Condition screen II with MMS | Formulation II[b] | 2-drop 96-well SwissCi/MRC | 0.6 | 0.3.0.2, 0.1 | 3:2:1 | 6.31 |
| Optimization II (Round 1) and apo crystals | Formulation II[b] | 24-well VDX 18 mm | 2.4 | 1.2, 0.8, 0.4 | 3:2:1 | 6.31 |
| Optimization II (Round 2) and apo crystals | Formulation II[b] | 2-drop 96-well SwissCi/MRC | 0.6 | 0.3, 0.2, 0.1 | 3:2:1 | 6.31 |
| Optimization II (Round 2) and apo crystals | Formulation II[b] | 3-drop 96-well SwissCi/MRC | 0.6 | 0.3, 0.2, 0.1 | 3:2:1 | 6.31 |
| Optimization II, (Round 3) and apo crystals | Formulation II[b] | 2-drop 96-well SwissCi/MRC | 0.6 | 0.3, 0.1, 0.2 | 3:1:2 | 6.31 |
| Optimization II, (Round 3) and apo crystals | Formulation II[b] | CrystalDirect™ plates | 0.3 | 0.15, 0.05, 0.1 | 3:1:2 | 6.31 |

[a] Protein formulation I is 6.81 mg · mL$^{-1}$ TcFPPS in high salt buffer (50 mM TRIS, pH 8.0, 200 mM NaCl, 2 mM TCEP · HCl)

[b] Protein formulation II is 12.20 mg · mL$^{-1}$ – 12.70 mg · mL$^{-1}$ in low salt buffer (10 mM TRIS, pH 7.4, 25 mM NaCl, 2 mM TCEP · HCl). For comparison, all experiments with Formulation II were calculated with at 12.62 mg · mL$^{-1}$.

**Table 31**: Allosteric inhibitors of hFPPS that were selected for binding test to TcFPPS.

| No. | Smiles string | MW | CAS | Reaxys ID | Ref. |
|---|---|---|---|---|---|
| **93** | O=C(CC1=CSC2=CC=C(C=C12)Cl)O | 226.68 | 17266-30-7 | 12244795 | [209a] |
| **94** | CC(C(C=C1Cl)=C(C=C1)S2)=C2CC(O)=O | 240.71 | 51527-19-6 | 8057736 | [209a] |
| **95** | COC1=CC2=C(C=C1)C(CC(O)=O)=CO2 | 206.2 | 69716-05-8 | 384045 | [209a] |
| **97** | OC(C1=CC2=CC=C3C=CC=CC3=C2N1CC(O)=O)=O | 269.26 | | 24135224 | [209a] |
| **98** | OC(C1=CC2=CC=C3C=CC=CC3=C2N1CC4=CC(C(O)=O)=NO4)=O | 336.31 | | 24889801 | [209a] |
| **101** | OC(C1=NC(C(C2=CC=CC3=CC=CC=C23)=CC=C4)=C4C=C1)=O | 299.33 | 1185407-78-6 | 19750428 | [230] |
| **118** | OC(CC1=CSC2=CC=C3C(C=CC=C3)=C21)=O | 242.3 | 108900-25-0 | 14690 | [209a] |
| **119** | ClC1=CC(Cl)=CC2=C1NC(C(O)=O)=C2CC(O)=O | 288.09 | | 24889800 | [209a] |

```
|Q8WS26|TcFPPS     1 ----------MASMERFLSVYDEVQAFLLDQLQSKYEIDPNRARYLRIMM     40
                         ....:.|:..:.::...|.:......||....|| |:.::
|P14324|hFPPS      1 MNGDQNSDVYAQEKQDFVQHFSQIVRVLTEDEMGHPEIGDAIAR-LKEVL     49

|Q8WS26|TcFPPS    41 DTTCLGGKYFRGMTVVNVAEGFLAVTQHDEATKERILHDACVGGWMIEFL     90
                      :...:||||.||:|||......:...:.|..:.:|    |...||.:|.|
|P14324|hFPPS     50 EYNAIGGKYNRGLTVVVAFRELVEPRKQDADSLQR----AWTVGWCVELL     95

|Q8WS26|TcFPPS    91 QAHYLVEDDIMDGSVMRRGKPCWYRFPGVTTQCAINDGIILKSWTQIMAW    140
                      ||.:||.|||||.|:.|||:.|||:.|||... ||||..:|::....:..
|P14324|hFPPS     96 QAFFLVADDIMDSSLTRRGQICWYQKPGVGLD-AINDANLLEACIYRLLK    144

|Q8WS26|TcFPPS   141 HYFADRPFLKDLLCLFQKVDYATAVGQMYDVTSMCDSNKLDPEVAQPMTT    190
                      .|..::|:..:|:.||.:..|.|.:||..|:.:....|           .
|P14324|hFPPS    145 LYCREQPYYLNLIELFLQSSYQTEIGQTLDLLTAPQGN-----------V    183

|Q8WS26|TcFPPS   191 DFAEFTPAIYKRIVKYKTTFYTYLLPLVMGLLVSEAAASVEMNLVERVAH    240
                      |...||...||.||||||.||::.||:...:.::.....|....:::..
|P14324|hFPPS    184 DLVRFTEKRYKSIVKYKTAFYSFYLPIAAAMYMAGIDGEKEHANAKKILL    233

|Q8WS26|TcFPPS   241 LIGEYFQVQDDVMDCFTPPEQLGKVGTDIEDAKCSWLAVTFLGKANAAQV    290
                      .:||:||:|||.:|.|..|...||:||||:|.|||||.|..|.:|...|.
|P14324|hFPPS    234 EMGEFFQIQDDYLDLFGDPSVTGKIGTDIQDNKCSWLVVQCLQRATPEQY    283

|Q8WS26|TcFPPS   291 AEFKANYGEKDPAKVAVVKRLYSKANLQADFAAYEAEVVREVESLIEQLK    340
                      ...|.|||:|:..|||.||.||.:.:|.|.|..||.:....:.:||||..
|P14324|hFPPS    284 QILKENYGQKEAEKVARVKALYEELDLPAVFLQYEEDSYSHIMALIEQYA    333

|Q8WS26|TcFPPS   341 VKSPTFAESVAV---VWEKTHKRKK    362
                      ...|.     ||   :..|.:||:|
|P14324|hFPPS    334 APLPP-----AVFLGLARKIYKRRK    353

Length: 375

Identity:     128/375 (34.1%)
Similarity:   188/375 (50.1%)
Gaps:          35/375 ( 9.3%)

Score: 537.5
```

**Figure 62**: Pairwise sequence alignment of TcFPPS and hFPPS. Lines indicate identical residues, colons indicate similar residues, and points indicate mismatch. Sequence alignment was made using Emboss Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/), which uses the Needleman-Wunsch algorithm[478].

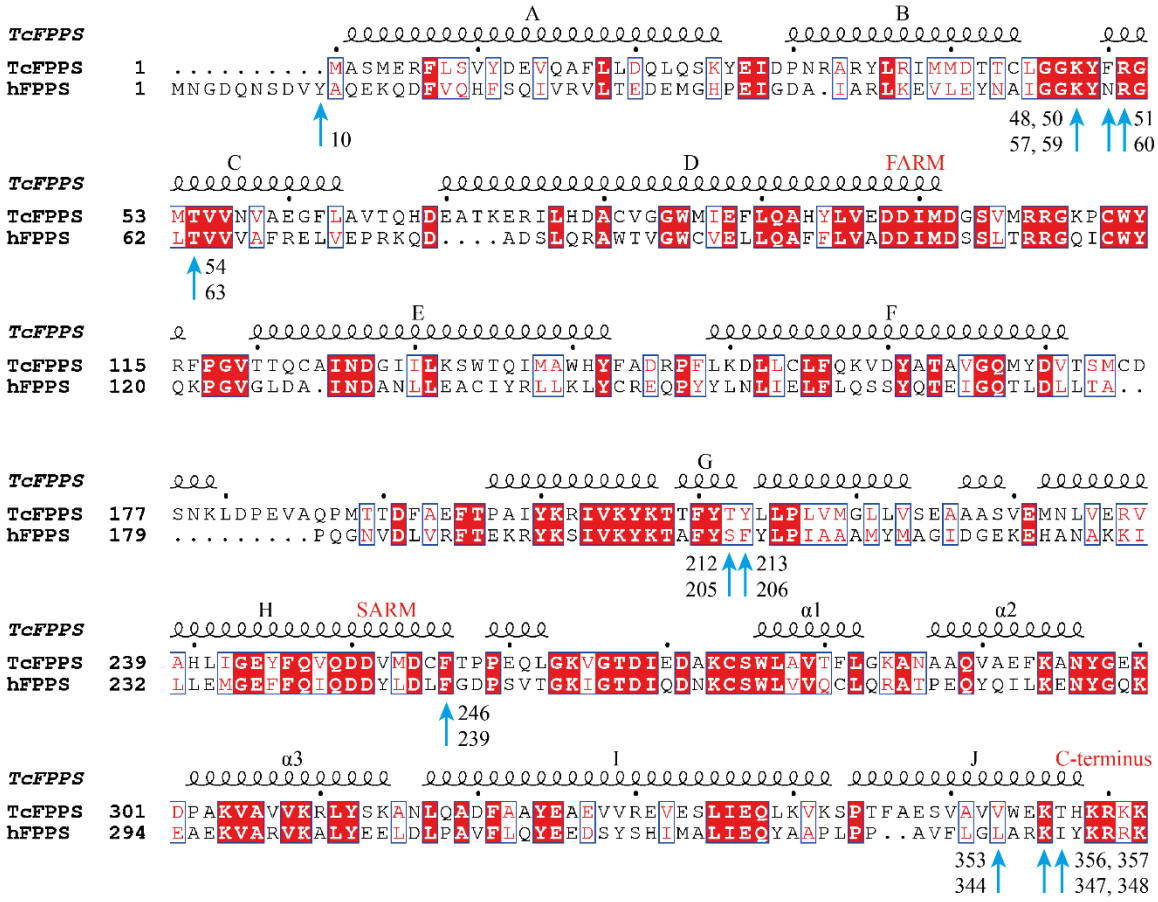



**Figure 63**: Sequence alignment of TcFPPS and hFPPS. The secondary structure is shown for TcFPPS (PDB ID 1YHK [162b]). Spirals indicate α-helices. Red background, red letters and blue boxes indicate identical residues, similar residues and conserved positions, respectively. Blue arrows highlight residues that form the allosteric pocket in hFPPS and numbers indicate the corresponding residues in TcFPPS and hFPPS. Alignment was made using Clustal Omega[495] and EsPrit (v.3.0)[496].


**Figure 64**: Alignment and ConSurf model reveal that Phe50 in TcFPPS is an exception. FPPS homologues included in the alignment against TcFPPS$_{64\text{-}425}$ and conservation score for the residues in position 49 to 60. Position of Phe50 is marked with a yellow box. Residues in this position that are an exception are marked with a yellow circle. Figure is continued on the next page. Model for level of sequence variability was generated on the website http://consurf.tau.ac.il/2016/. Running Parameter: PDB file 6R04, chain identifier A, Alignment: Multiple sequence alignment was built using CLUSTALW, the homologues were collected from CLEAN_UNIPROT, homologue search algorithm: HMMER, HMMER E value: 0.0001, number of HMMER iterations: 1, 200 sequences that sample the list of homologues to the query were selected by the user, maximal %ID between sequences: 95, minimal %ID for homologues: 35, Conservation Scores: Method of calculation: Bayesian, model of substitution for proteins: best fit.


**Figure 64** continued.

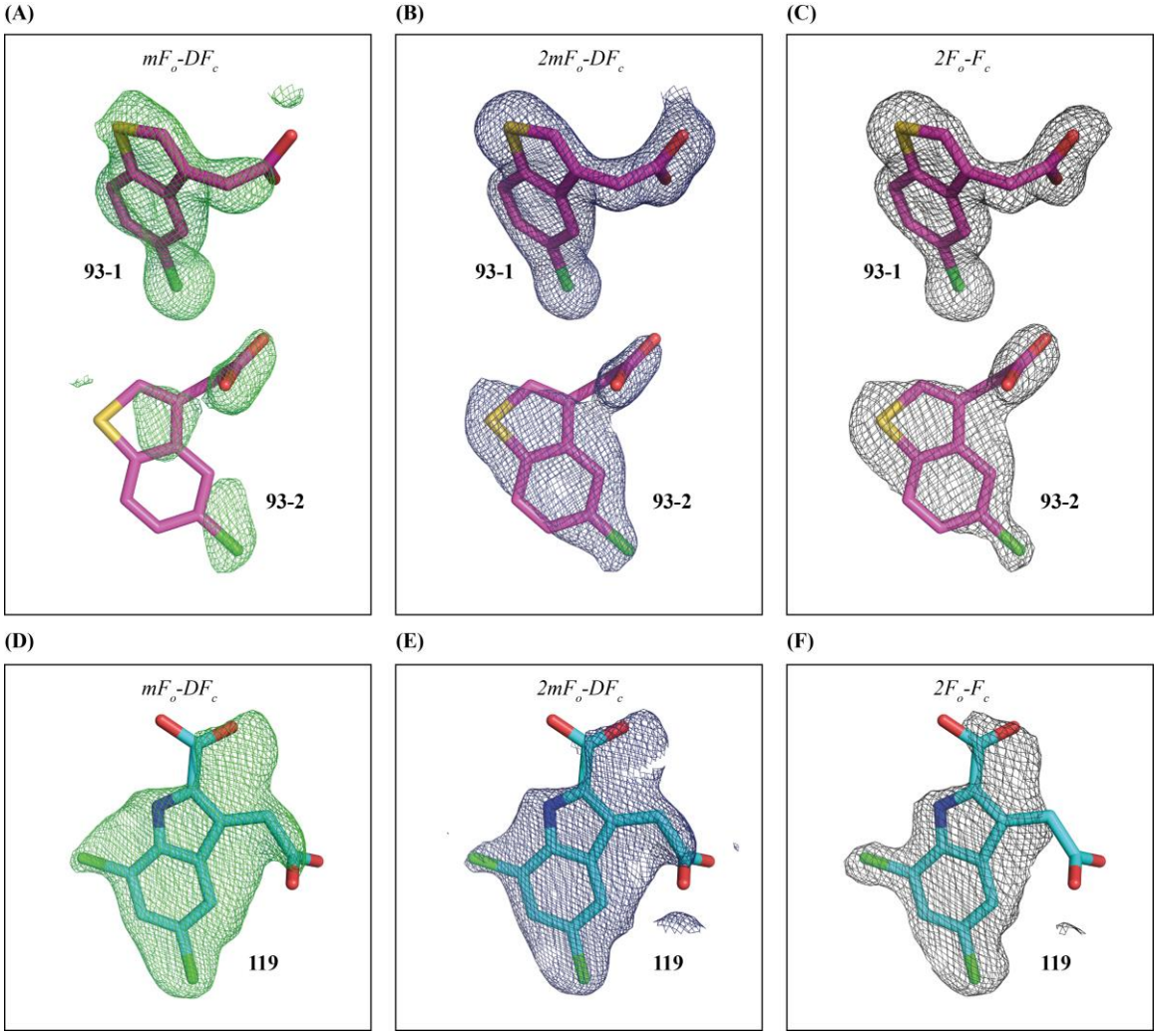



**Figure 65**: Summary of density maps of the ligands **93-1**, **93-2** and **119**. **(A) – (C)** Ligand **93-1** and **93-2** (PDB ID 6R07): $mF_o - DF_c$ difference electron density map contoured at 3.00 σ, $2mF_o - DF_c$ electron density map contoured at 1.0 σ and $2F_o - F_c$ electron density map contoured at 1.00 σ, respectively. **(D)** – **(F)** Ligand **119** (PDB ID 6R08): $mF_o - DF_c$ difference electron desity map contoured at 3.0 σ, $2mF_o - DF_c$ electron density map contoured at 1.0 σ and $2F_o - F_c$ electron density map contoured at 1.0 σ, respectively.

**Table 32**: Hits derived from the Novartis core library screen against TcFPPS.

| No[a] | Smiles string | MW | CAS<br>*PubChem ID*<br>**Reaxys ID** | Hit[b]<br>Tc | Hit[b]<br>Tb | Hit[b]<br>h | TcFPPS: Effect in T1ρ[c] (%) | TbFPPS: Effect in T1ρ[c] (%) | hFPPS: Effect in T1ρ[c] (%) |
|---|---|---|---|---|---|---|---|---|---|
| **CS-1** | C(C1C=C(C(NC(C)=O)=CC=1)Cl)(C2CC2)C(OC)=O | 281.72 | 63061-43-8 | 1 | 0 | 0 | 25, 23, 20, 20 | 6, 6 | <5 |
| **CS-2** | C2(C1C=C(C(OC)=CC=1)OC)=C(N(C)N=C2C)N | 247.30 | **8986549** | 1 | 0 | 0 | 40, 35, 34 | <5 | 14, 13, 12 |
| **CS-3** | C2(C(C1=CC=NC=C1)=O)N(N=CC=2)C | 187.20 | *63960949* | 1 | 0 | 0 | 26, 18, 14 | <5 | <5 |
| **CS-4** | C2(C(N1CCOCC1)=O)=C(C=C(N)C=C2)Cl | 240.70 | 926201-77-6 | 1 | 0 | 0 | 26, 24, 16 | 7, 6 | <5 |
| **CS-5** | C2(N1CCN(C)CC1)=NC(=C(C(OCC)=O)C=N2)N | 265.31 | 682789-28-2 | 1 | 0 | 0 | 23, 22, 14 | 10, 9, 7 | 7 |
| **CS-6** | N2=NC(C1=CC=C(C(=O)N)C=C1)=CS2 | 205.22 | **28274622** | 1 | 0 | 0 | 25, 18 | 7, 5 | 10, 8, 6 |
| **CS-7** | C3(N1CCN(C)CC1)=CC2C(=CC=CC=2)N=C3 | 227.31 | 78641-22-2 | 1 | 0 | 0 | 40, 32, 27 | 8, 8, 7 | 12, 9, 9 |
| **CS-8** | C2(NC(NC1CCCCC1)=O)=CC(=NC(=N2)C)C | 248.32 | *4412025* | 1 | 0 | 0 | 45, 40 | 6, 0 | 7, 7 |
| **CS-9** | C23N(CC1C=CC(Cl)=CC=1)C=NC=2N=CN=C3NC | 273.72 | *110171704* | 1 | 1 | 0 | 22, 21, 18, 8 | 30, 15, 13, 11 | <5 |
| **CS-10** | C1(C3C(N(C)C2C(N=1)=CC=CC=2)=CC=CC=3)N4CCNCC4 | 292.40 | *69041412* | 1 | 1 | 0 | 42,38 | 27, 24, 32, 22 | 5 |
| **CS-11** | C1(CS(C)(=O)=O)C(=CC=CC=1Cl)Cl | 239.11 | 20018-02-4 | 0 | 1 | 1 | | 32, 14, 11 | 37, 36, 23 |
| **CS-12** | C12(CC3CC(C1)CC(C2)C3)NCCN | 194.31 | 37818-93-2 | 1 | 1 | 1 | 35, 31, 19 | 26, 21, 16 | 30, 19, 18 |
| **CS-13** | N(C1=CC(=CC=C1)Cl)C2C=C(N=CN=2)N | 220.66 | 872511-13-2 | 1 | 1 | 1 | 30, 30 | 31, 23, 22, 14 | 57, 43, 33 |
| **CS-14** | C12C(=C(C=C(C=1)OC)N)N=CC=C2 | 174.20 | 90-52-8 | 1 | 1 | 1 | 35, 30, 28 | 26 ,21, 15 | 50, 33, 30, 28 |
| **CS-15** | C1(=CC=C(Cl)C=C1)OCCCN2C=CN=C2 | 236.70 | *3599333* | 1 | 1 | 1 | 46, 45, 22 | 33, 26, 21, 9, 9 | 32, 27, 26, 26 |
| **CS-16** | C23C(CN1C=CN=C1N2)=CC(OC)=C(C=3OC)OC | 261.29 | **13345518** | 1 | 1 | 1 | 50, 46 | 35, 28, 26, 20 | 22, 22 |
| **CS-17** | C23C(N1CCN(C)CC1)=NC(Cl)=CC=2C=CC=C3 | 261.78 | **11441642** | 1 | 1 | 1 | 27, 20, 14 | 29, 28, 25,13,13 | 29, 26, 14 |
| **CS-18** | C1(N=C(C=CC=1)C)NCC2C=CC=CC=2 | 198.27 | 70644-47-2 | 1 | 1 | 1 | 62, 60, 50 | 29, 27, 14 | 39, 38, 33 |
| **CS-19** | C23C1=C(CCCC1)SC=2N=CN=C3N(C)C | 233.32 | 871807-58-8 | 1 | 1 | 1 | 60, 32, 30 | 23, 16 | 59, 23 |
| **CS-20** | C2(C1N=C(N)SC=1)=C(C=CC=C2)O | 192.22 | 60135-72-0 | 1 | 1 | 1 | 51, 50, 40 | 22, 22, 17, 8 | 65 |
| **CS-21** | C1=CC=C2C(=C1)CC(N2C(=O)OCC3=CC=CC=C3)CO | 283.32 | 135829-04-8 | 1 | 1 | 1 | 60, 27, 22 | 28, 21 | 29, 21 |
| **CS-22** | C1(C=CC=C(N=1)N)CC | 122.17 | 21717-29-3 | 1 | 1 | 1 | 37, 36, 34, 21 | 20, 18, 17, 14 | 26, 23, 21, 13, 8 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **CS-23** | C2(OC1C=CC(O)=CC=1)C(=CC(Cl)=CN=2)Cl | 256.10 | 60075-03-8 | 1 | 1 | 1 | 45, 40, 38 | 48, 37, 33 | 45, 32 |
| **CS-24** | C23(N(C(C1=CC=CC=C1C2)=O)C)CCCCC3 | 229.31 | 21868-94-0 | 1 | 1 | 1 | 38, 23, 22, 21, 18 | 21, 20, 19, 15, 12 | 38, 21, 20, 7 |
| **CS-25** | C2(C1C=CN=CC=1)N=C(N=CC=2)NC | 186.21 | 66522-26-7 | 1 | 0 | 1 | 33, 21, 19 | <5 | 36, 25, 21, 21 |
| **CS-26** | C1(C(=CC(CCNC(C)=O)=CC=1)OC)OCC2=CC=CC=C2 | 299.37 | 39731-97-0 | 1 | 0 | 1 | 35, 28, 26, 25 | 10, 7, 6 | 24, 20, 19, 18 |
| **CS-27** | C2(N1C(CCC1)=O)C(=CC=C(C=2)Cl)C | 209.69 | *82077906* | 1 | 0 | 1 | 21, 18 | 10, 8 | 35, 27, 21 |
| **CS-28** | C12=C(SC=C1C(O)=O)C=CC=C2 | 178.21 | 5381-25-9 | 1 | 0 | 1 | 39, 34, 28 | 9 | 44, 40, 26 |
| **CS-29** | C2(NC(N1CCOCC1)=O)C=C(C(C)=CC=2)C | 234.30 | *4464839* | 1 | 0 | 1 | 27, 24, 20, 17 | 5, 4 | 39, 28, 20, 19 |
| **CS-30** | C12=C(C=C(N)C=C1)OCC2 | 135.19 | 57786-34-2 | 1 | 0 | 1 | 35, 14, 7, 6 | <5 | 55, 55, 46 |
| **CS-31** | C12=C(C=CC=C1OCC(NC)=O)C=CC(=N2)C | 230.27 | 3135-42-0 | 1 | 0 | 1 | 36, 29, 28, 25 | 6, 3 | 42, 36, 35, 31 |
| **CS-32** | C1(C(=CC=CC=1Cl)Cl)OCC(O)=O | 221.02 | 575-90-6 | 1 | 0 | 1 | 26, 24 | 6 | 50 |
| **CS-33** | [C@@H]2(OC1C=C(C(C)=CC=1)C)[C@H](CN(CC#C)CC2)O | 259.33 | **13315827** | 1 | 0 | 1 | 27, 22, 20 | 10, 9, 7 | 39, 35, 33 |
| **CS-34** | C1(C(=CC(Cl)=CC=1)N)CO | 157.60 | 37585-16-3 | 1 | 0 | 1 | 21, 20, 19 | <5 | 30, 22, 20 |
| **CS-35** | C2(C1C(=CC=CC=1)Cl)=C(N=CO2)C(OC)=O | 237.62 | 89204-91-1 | 1 | 0 | 1 | 38, 35, 15 | 5 | 44, 12, |
| **CS-36** | N2=C(C=C(OC1=CC=C(C=C1)C)N=C2)Cl | 220.66 | 124040-99-9 | 1 | 0 | 1 | 24, 23, 18, 4 | 5, 5 | 32, 19, 8 |
| **CS-37** | C12=C(C=CC(=C1)S(NC(NC)=O)(=O)=O)C=CC=C2 | 264.30 | *23548392* | 1 | 0 | 1 | 48, 47, 2 x 29 | 6 | 2 x 46, 24, 22 |
| **CS-38** | C12C(NC(=C1)C(OC)=O)=CC=CC=2 | 175.20 | 1202-04-6 | 1 | 0 | 1 | 45, 34, 29 | 5, 5, 0 | 39, 33, 31, 30 |
| **CS-39** | C2(C1=CC=CC=C1Br)NN=NN=2 | 225.05 | 73096-42-1 | 1 | 0 | 1 | 33, 32, 19 | <5 | 42, 40, 39 |
| **CS-40** | C1(=NC(=CS1)C(O)=O)C2C=CC(C)=CC=2 | 219.28 | 17228-99-8 | 1 | 0 | 1 | 46, 42, 38, 14, | neg, 13, 7 | 54, 21, 17 |

[a] Only publically known compounds are listed.
[b] Fragment hits that are ranked with a one for the corresponding protein showed a positive effect in waterLOGSY experiments.
[c] Signal decrease for several signals is given ranked by its strength.

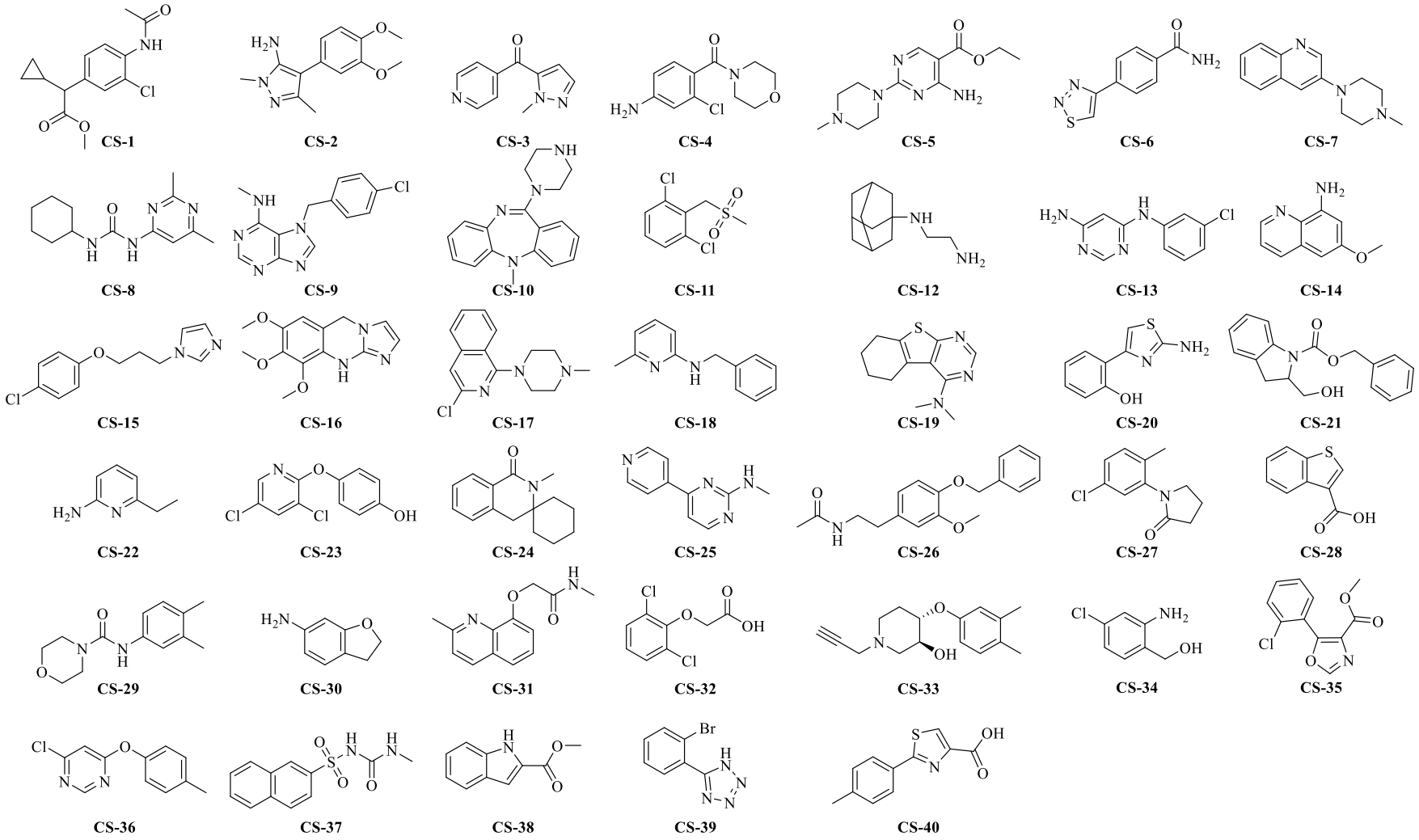



**Figure 66**: Chemical structures of fragment hits from the Novartis core library. Only publically known hits are shown. See **Table 32**.

**Table 33:** Hits derived from the Novartis fluorine library screen against TcFPPS.

| No[a] | Smiles string | MW | CAS<br>*PubChem ID*<br>**Reaxys ID** | Hit TcFPPS | Hit TbFPPS | Hit hFPPS | TcFPPS: Effect in $^{19}$F CPMG[b]<br>mix (%), single (%) |
|---|---|---|---|---|---|---|---|
| **FS-1** | FC(F)(F)C1=CC=C(NC(OC)=O)C=C1 | 219.16 | 23794-77-6 | 1 | 0 | 0 | 46, 47 |
| **FS-2** | FC(F)(F)C1=NN=C(NC2=CN=CC=C2)S1 | 246.21 | *2725838* | 1 | 0 | 0 | 44, 51 |
| **FS-3** | O=C(C1=CN=CC=C1)NC2=NC=C(C(F)(F)F)C=C2 | 267.21 | *84350204* | 1 | 0 | 0 | 48, 50 |
| **FS-4** | O=C(NC)C1=CC(C(C(F)(F)F)=N2)=C(S1)N2C | 263.24 | *44769570* | 1 | 0 | 0 | 47, 51 |
| **FS-5** | O=C(NC1=CC=C(C=C1)C(F)(F)F)C2=C(C)ON=C2C | 284.24 | **13679250** | 1 | 0 | 0 | 60, 42 |
| **FS-6** | FC(C1=CC=C(NC(C2=CN=CC=C2)=O)C=C1)(F)F | 266.22 | 25617-45-2 | 1 | 0 | 0 | 53, 52 |
| **FS-7** | NC1=CC(C(F)(F)F)=CC=C1N2CCOCC2 | 246.23 | 784-57-6 | 1 | 0 | 0 | 41, 54 |
| **FS-8** | N#CC(C(C(F)(F)F)=C1)=CC=C1NC(C)=O | 228.19 | 97760-99-1 | 1 | 0 | 0 | 52, 58 |
| **FS-9** | OC1=C(C(F)(F)F)C=C([N+]([O-])=O)C=C1C(F)(F)F | 275.10 | *13784430* | 1 | 1 | 0 | 27, 31 |
| **FS-10** | FC(F)(F)C1=CC(/C=N/NC(N)=N)=CC(C(F)(F)F)=C1 | 298.20 | 23557-66-6 | 1 | 1 | 0 | 45, 47 |
| **FS-11** | FC(F)(F)C(C=C1)=CC(N)=C1[N+]([O-])=O | 206.13 | 402-14-2 | 1 | 1 | 1 | 66, 50 |
| **FS-12** | FC(F)(F)C1=NC(NC2=CC=C(Cl)C=C2)=NC(N)=N1 | 289.63 | 53387-70-5 | 1 | 1 | 1 | 44, 42 |
| **FS-13** | FC(F)(F)C1=NN2C(C(C3=CC=CS3)=C1)=NN=C2 | 270.24 | *760142* | 1 | 1 | 1 | 47, 50 |
| **FS-14** | NC1=CC(C(F)(F)F)=CC=C1C(OC2)=NC2(C)C | 258.22 | 1361005-81-3 | 1 | 1 | 1 | 60, 66 |
| **FS-15** | NC1=C(C(NC2=CC(C(F)(F)F)=CC=C2)=O)C=CC=C1 | 280.25 | 20878-52-8 | 1 | 1 | 1 | 51, 45 |
| **FS-16** | FC(F)(F)/C(C1=CSC=C1)=N/O | 195.18 | 138395-47-8 | 1 | 1 | 1 | 23, 32 |
| **FS-17** | ClC1=C(C(F)(F)F)C=C([N+]([O-])=O)C(N)=C1 | 240.59 | 35375-74-7 | 1 | 1 | 1 | 34, 36 |
| **FS-18** | CC1=CC([N+]([O-])=O)=C(C(F)(F)F)C=C1N | 220.18 | *129319121* | 1 | 1 | 1 | 37, 34 |
| **FS-19** | NC1=NC2=CC=C(C(F)(F)F)C=C2S1 | 218.20 | 777-12-8 | 1 | 1 | 1 | 27, 44 |
| **FS-20** | FC(F)(F)C1=CC(Cl)=C(C2=CN=C(N=C2)N)N=C1 | 274.63 | *1483168* | 1 | 0 | 1 | 60, 64 |
| **FS-21** | OC(CO1)=C(C1=O)C2=CC=CC(C(F)(F)F)=C2 | 244.19 | **28370057** | 1 | 0 | 1 | 55, 70 |
| **FS-22** | CC1=NN(C=C2C3=CC(C(F)(F)F)=CC=C3)C(N=C2)=N1 | 278.24 | **13322897** | 1 | 0 | 1 | 41, 64 |
| **FS-23** | FC(F)(F)C(C=C1Br)=CC(N)=C1N | 255.02 | 113170-72-2 | 1 | 0 | 1 | 41, 23 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **FS-24** | NC1=CC(N2C=NC(C)=C2)=CC(C(F)(F)F)=C1 | 241.21 | 641571-11-1 | 1 | 0 | 1 | 13, 41 |
| **FS-25** | FC(F)(F)C1=C(C=CC=C1)C2=NC(C3=CC=NC=C3)=NO2 | 291.23 | **1486742** | 1 | 0 | 1 | 40, 42 |
| **FS-26** | FC(F)(F)C1=CC(Cl)=C(N=C1)C2=CNN=C2 | 247.61 | *1473368* | 1 | 0 | 1 | 47, 43 |
| **FS-27** | FC(F)(F)C1=CC=CC(NC2=NC=NN2)=C1 | 228.19 | *2766475* | 1 | 0 | 1 | 60, 60 |
| **FS-28** | NC1=CC(C2=CC=CC=N2)=CC(C(F)(F)F)=C1 | 238.21 | **11405061** | 1 | 0 | 1 | 18, 26 |
| **FS-29** | O=C1N(CC2=CC(C(F)(F)F)=CC=C2)C(CSC1)=O | 289.28 | *1478957* | 1 | 0 | 1 | 50, 58 |
| **FS-30** | FC(F)(F)C1=CC=C(C2=NC(C)=C(CO)S2)C=C1 | 273.29 | 317318-96-0 | 1 | 0 | 1 | 50, 62 |
| **FS-31** | O=C1N2C(N(CC=C)CC2)=NC3=CC=C(C(F)(F)F)C=C13 | 295.29 | 85964-93-8 | 1 | 0 | 1 | 56, 65 |
| **FS-32** | N#CC1=CC(C2=CC=CC(C(F)(F)F)=C2)=CNC1=O | 264.21 | 76053-36-6 | 1 | 0 | 1 | 50, 47 |
| **FS-33** | FC(F)(F)C1=CC=C(NC2=NC(C)=CC(C)=N2)C=C1 | 267.28 | *4644876* | 1 | 0 | 1 | 50, 50 |
| **FS-34** | FC(F)(F)C1=CC(N)=CC=C1N2C=CN=C2 | 227.20 | 351324-53-3 | 1 | 0 | 1 | 50, 70 |
| **FS-35** | N#CC1=NNC2=CC=C(C(F)(F)F)C=C21 | 211.13 | *72218411* | 1 | 0 | 1 | 32, 21 |

[a] Only publically known compounds are listed.
[b] Hit criteria for effect in $^{19}F$ CPMG ≥20%

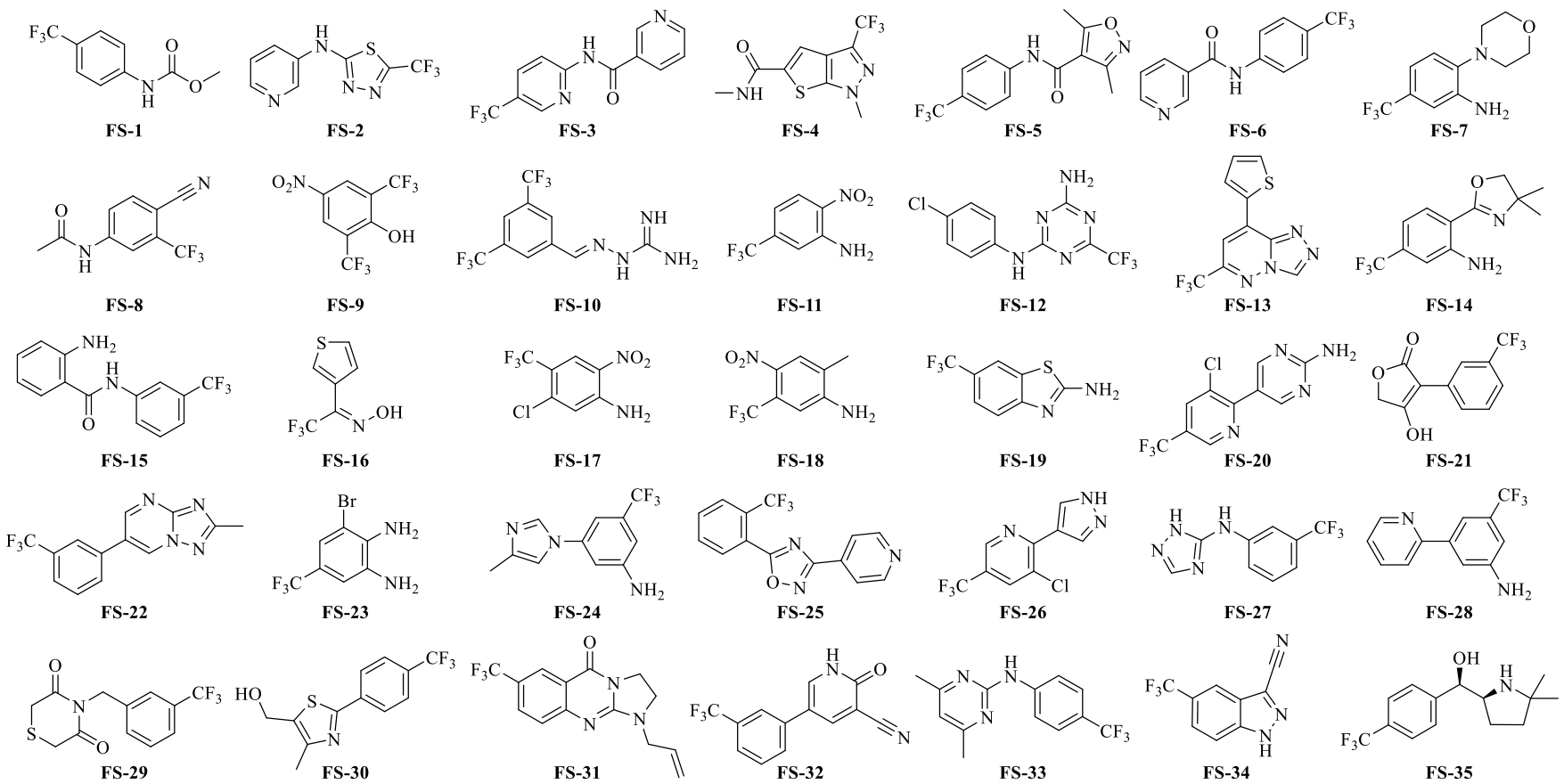



**Figure 67:** Chemical structures of fragment hits from the Novartis fluorine library. Only publically known hits are shown. See **Table 33**.

```
|Q8WS26|TcFPPS      1 MASMERFLSVYDEVQAFLLDQLQSKYEIDPNRARYLRIMMDTTCLGGKYF     50
                       ..|:.|:.||||:|.|||::|:.|:::|||||.||||.||||||||||.
|Q86C09|TbFPPS      1 -MPMQMFMQVYDEIQMFLLEELELKFDMDPNRVRYLRKMMDTTCLGGKYN     49

|Q8WS26|TcFPPS     51 RGMTVVNVAEGFLAVT------QHDEATKERILHDACVGGWMIEFLQAHY     94
                       ||:||::|||..|:::      :.|.|.::|:||||||.||||||||||
|Q86C09|TbFPPS     50 RGLTVIDVAESLLSLSPNNNGEEDDGARRKRVLHDACVCGWMIEFLQAHY     99

|Q8WS26|TcFPPS     95 LVEDDIMDGSVMRRGKPCWYRFPGVTTQCAINDGIILKSWTQIMAWHYFA    144
                       ||||||||.||.|||||||||.|.||.|||||||::||||||.:||.|:||
|Q86C09|TbFPPS    100 LVEDDIMDNSVTRRGKPCWYRHPDVTVQCAINDGLLLKSWTHMMAMHFFA    149

|Q8WS26|TcFPPS    145 DRPFLKDLLCLFQKVDYATAVGQMYDVTSMCDSNKLDPEVAQPMTTDFAE    194
                       |||||:||||.|.:|||.|||||:||||||.|||||||:|:||.||||||
|Q86C09|TbFPPS    150 DRPFLQDLLCRFNRVDYTTAVGQLYDVTSMFDSNKLDPDVSQPTTTDFAE    199

|Q8WS26|TcFPPS    195 FTPAIYKRIVKYKTTFYTYLLPLVMGLLVSEAAASVEMNLVERVAHLIGE    244
                       ||.:.|||||||||.:|||||||||||:||||..:|:|.:.|.:|.|:||
|Q86C09|TbFPPS    200 FTLSNYKRIVKYKTAYYTYLLPLVMGLIVSEALPTVDMGVTEELAMLMGE    249

|Q8WS26|TcFPPS    245 YFQVQDDVMDCFTPPEQLGKVGTDIEDAKCSWLAVTFLGKANAAQVAEFK    294
                       ||||||||||||||:||||||||:||||||||||||.||::|||||||
|Q86C09|TbFPPS    250 YFQVQDDVMDCFTPPERLGKVGTDIQDAKCSWLAVTFLAKASSAQVAEFK    299

|Q8WS26|TcFPPS    295 ANYGEKDPAKVAVVKRLYSKANLQADFAAYEAEVVREVESLIEQLKVKSP    344
                       ||||..|..|||.|:|||.:|:||.|:.||||.|..:|:.|||:|::.||
|Q86C09|TbFPPS    300 ANYGSGDSEKVATVRRLYEEADLQGDYVAYEAAVAEQVKELIEKLRLCSP    349

|Q8WS26|TcFPPS    345 TFAESVAVVWEKTHKRKK    362
                       .||.||..:|.||:||:|
|Q86C09|TbFPPS    350 GFAASVETLWGKTYKRQK    367

Length: 368       Identity:     254/368 (69.0%)
                  Similarity:   306/368 (83.2%)
                  Gaps:           7/368 ( 1.9%)
Score: 1374.5
```

**Figure 68**: Pairwise sequence alignment of TcFPPS and TbFPPS. Lines indicate identical residues, colons indicate similar residues, and points indicate mismatch. Sequence alignment was made using Emboss Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/), which applies the Needleman-Wunsch algorithm[478].

```
|Q8WS26|TcFPPS ----------MASMERFLSVYDEVQAFLLDQLQSKYEIDPNRARYLRIMMDTTCLGGKYF 50
|Q86C09|TbFPPS -----------MPMQMFMQVYDEIQMFLLEELELKFDMDPNRVRYLRKMMDTTCLGGKYN 49
|P14324|hFPPS  MNGDQNSDVYAQEKQDFVQHFSQIVRVLTEDEMGHPEI-GDAIARLKEVLEYNAIGGKYN 59
                          : *:. :.::  .* ::   : ::  :    *: ::: ..:****

|Q8WS26|TcFPPS RGMTVVNVAEGFLAVTQH------DEATKERILHDACVGGWMIEFLQAHYLVEDDIMDGS 104
|Q86C09|TbFPPS RGLTVIDVAESLLSLSPNNNGEEDDGARRKRVLHDACVCGWMIEFLQAHYLVEDDIMDNS 109
|P14324|hFPPS  RGLTVVVAFRELVEPRK----------QDADSLQRAWTVGWCVELLQAFFLVADDIMDSS 109
               **:**: . . ::                  *: * . ** :*:***.:** *****.*

|Q8WS26|TcFPPS VMRRGKPCWYRFPGVTTQCAINDGIILKSWTQIMAWHYFADRPFLKDLLCLFQKVDYATA 164
|Q86C09|TbFPPS VTRRGKPCWYRHPDVTVQCAINDGLLLKSWTHMMAMHFFADRPFLQDLLCRFNRVDYTTA 169
|P14324|hFPPS  LTRRGQICWYQKPGVGLDA-INDANLLEACIYRLLKLYCREQPYYLNLIELFLQSSYQTE 168
               : ***: ***: *.*  :. ***. :*::    :   :  ::*:  :*:  * : .* *

|Q8WS26|TcFPPS VGQMYDVTSMCDSNKLDPEVAQPMTTDFAEFTPAIYKRIVKYKTTFYTYLLPLVMGLLVS 224
|Q86C09|TbFPPS VGQLYDVTSMFDSNKLDPDVSQPTTTDFAEFTLSNYKRIVKYKTAYYTYLLPLVMGLIVS 229
|P14324|hFPPS  IGQTLDLLTA-----------PQGNVDLVRFTEKRYKSIVKYKTAFYSFYLPIAAAMYMA 217
               :**  *: :              ..*:..**   ** ******::*:: **:. .: ::

|Q8WS26|TcFPPS EAAASVEMNLVERVAHLIGEYFQVQDDVMDCFTPPEQLGKVGTDIEDAKCSWLAVTFLGK 284
|Q86C09|TbFPPS EALPTVDMGVTEELAMLMGEYFQVQDDVMDCFTPPERLGKVGTDIQDAKCSWLAVTFLAK 289
|P14324|hFPPS  GIDGEKEHANAKKILLEMGEFFQIQDDYLDLFGDPSVTGKIGTDIQDNKCSWLVVQCLQR 277
                     :   .:.:   :**:**:*** :* *  *.  **:****:* *****.*  * :

|Q8WS26|TcFPPS ANAAQVAEFKANYGEKDPAKVAVVKRLYSKANLQADFAAYEAEVVREVESLIEQLKVKSP 344
|Q86C09|TbFPPS ASSAQVAEFKANYGSGDSEKVATVRRLYEEADLQGDYVAYEAAVAEQVKELIEKLRLCSP 349
|P14324|hFPPS  ATPEQYQILKENYGQKEAEKVARVKALYEELDLPAVFLQYEEDSYSHIMALIEQYAAPLP 337
               *.  *   :* ***. :  *** *: **.: :* . :  **     .:  ***:     *

|Q8WS26|TcFPPS TFAESVAVVWEKTHKRKK 362
|Q86C09|TbFPPS GFAASVETLWGKTYKRQK 367
|P14324|hFPPS  P--AVFLGLARKIYKRRK 353
                    .  :  * :**:*

Percent Identity  Matrix

|Q8WS26|TcFPPS   35.28  100.00   70.36
|Q86C09|TbFPPS   37.13   70.36  100.00
|P14324|hFPPS   100.00   35.28   37.13
```

**Figure 69**: Sequence alignment and identity matrix of hFPPS, TcFPPS and TbFPPS. Asterisks indicate fully conserved residues, colons indicate conserved substitutions of residues (strongly similar properties), and points indicate semi-conserved substitutions of residues (weakly similar properties). Sequence alignment was made using Clustal Omega (v.1.2.4.) (https://www.ebi.ac.uk/Tools/msa/clustalo/). Identity matrix was generated using Clustal (v.12.1).

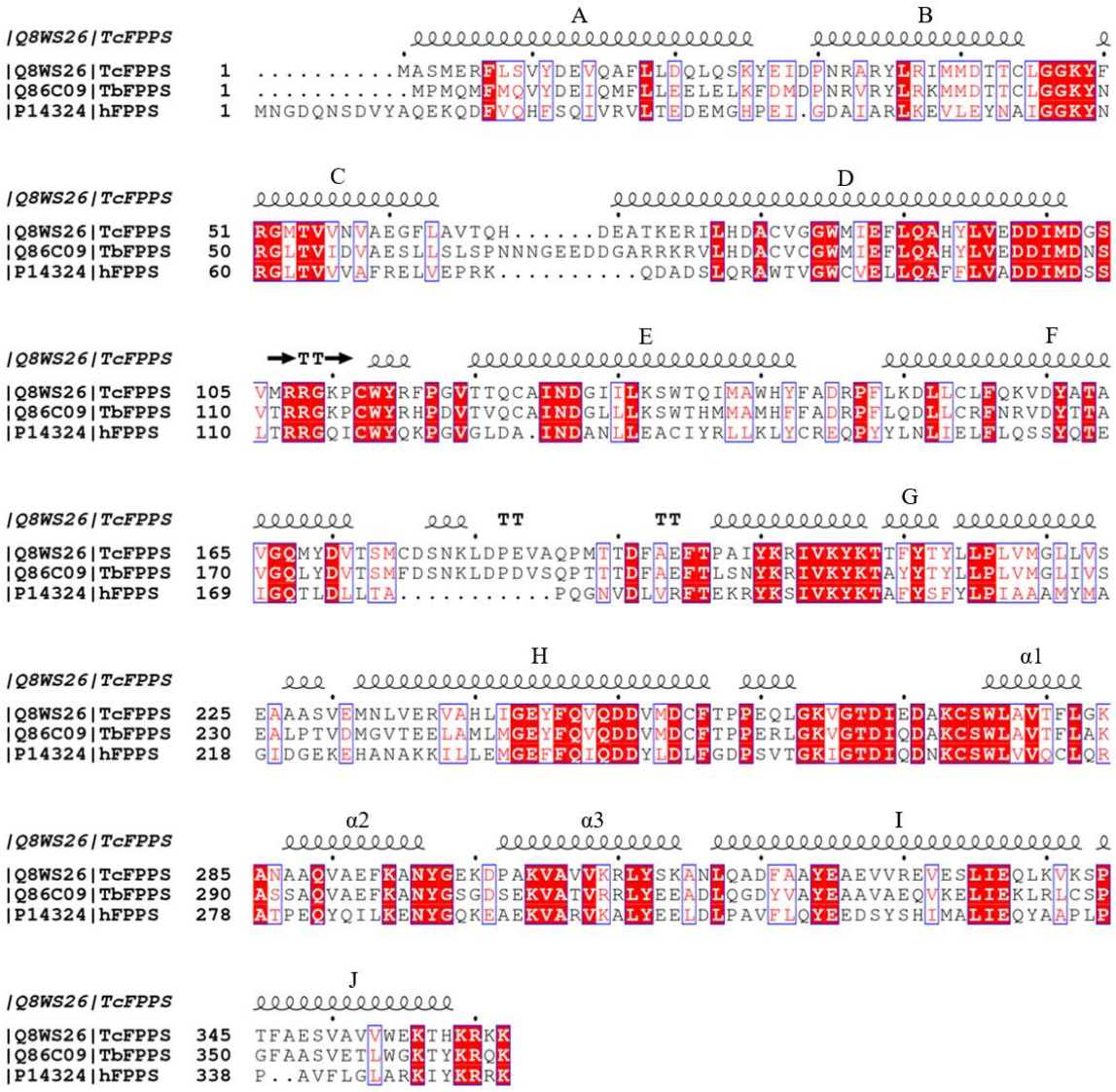



**Figure 70**: Sequence alignment of TcFPPS with the sequence of TbFPPS and hFPPS. In the top the secondary structure of TcFPPS (PDB ID 1YHK) is shown. Spirals indicate α-helices, TT strict β-turns and arrows indicate β-strands. Red background indicates identical residues, red letters indicate similar residues, and blue boxes indicate conserved positions. Sequence alignment was made using EsPrit (v.3.0)[496].

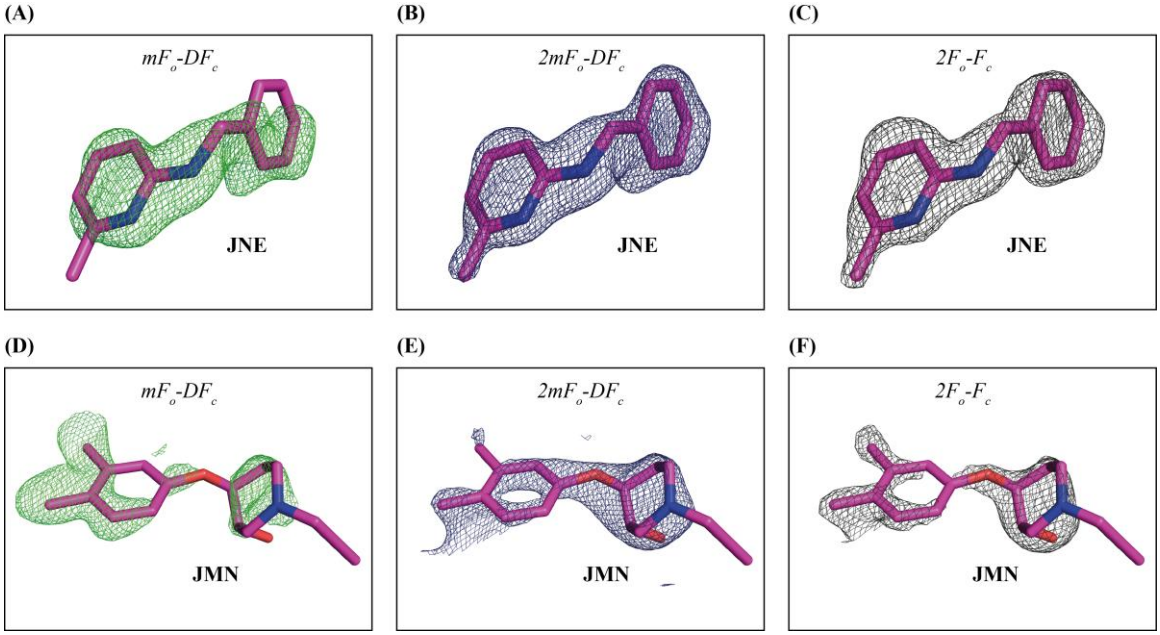



**Figure 71**: Summary of density maps of the ligands **JNE** and **JMN**. **(A) – (C)** Ligand **JNE** (PDB ID 6R05): $mF_o - DF_c$ difference electron desity map contoured at 3.0 σ, $2mF_o - DF_c$ electron density map contoured at 1.0 σ and $2F_o - F_c$ electron density map contoured at 1.0 σ, respectively. **(D)** – **(F)** Ligand **JMN** (PDB ID 6R06): $mF_o - DF_c$ difference electron desity map contoured at 3.0 σ, $2mF_o - DF_c$ electron density map contoured at 1.0 σ and $2F_o - F_c$ electron density map contoured at 1.0 σ, respectively.


**Figure 72**: 302 in-house datasets analysed with PanDDA. **(A)** Resolution limit. **(B)** R-free and R-work. **(C)** RMSD to reference structure. **(D)** Unit cell volume variation. **(E)** Cell axis variation. **(F)** Cell angle variation. On the y-axis the count is plotted.

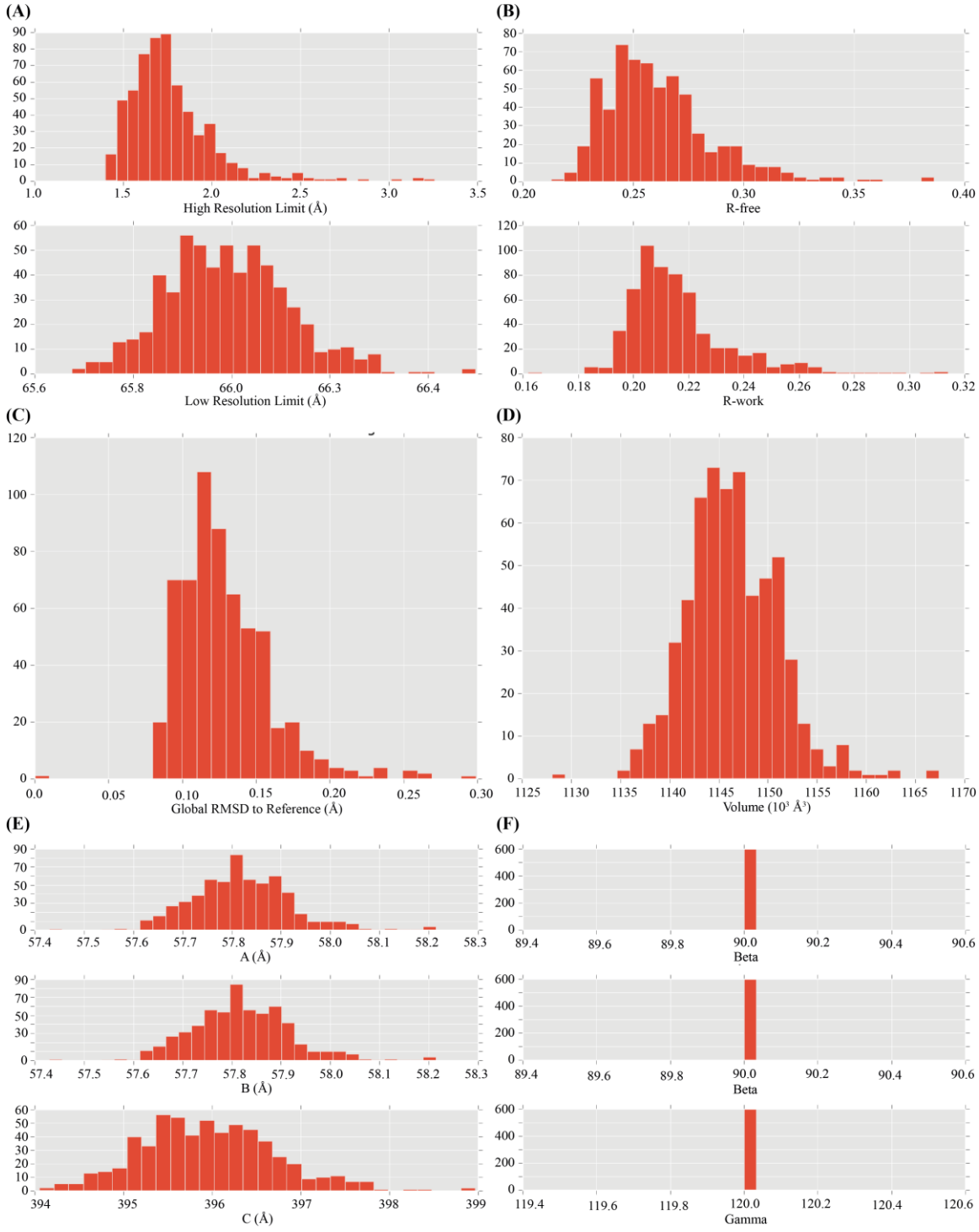



**Figure 73**: Datasets analysed of the XChem campaign with PanDDA. **(A)** Resolution limit. **(B)** R-free and R-work. **(C)** RMSD to reference structure. **(D)** Unit cell volume variation. **(E)** Cell axis variation. **(F)** Cell angle variation. On the y-axis the count is plotted.

**Table 34:** Overview of 35 structural models that resulted from the XChem campaign.

| No | X-tal | PDB ID | PDB ID ligand[a] | XChem identifier | Smiles string | MW (Da) | Site[b] | Occupancy | Soaking time (hh:mm) | Diffraction limit (Å) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | x0051 | 5QPD | LT7 | FMOPL000293a | Cc1ccc(cc1)c2nc(on2)C[NH2+]C(C)C | 232.31 | 6 | 0.60 | 03:22 | 1.93 |
| 2 | x0058 | 5QPE | AWG | FMOPL000295a | c1ccc(c(c1)N)Nc2[nH]c3ccccc3n2 | 224.27 | 5 | 0.82 | 03:27 | 1.77 |
| 3 | x0064 | 5QPF | AWM | FMOPL000478a | C[NH+]1CCN(CC1)c2nc3ccccc3s2 | 234.34 | 1 / 3 | 0.082 / 0.82 | 03:31 | 1.50 |
| 4 | x0076 | 5QPG | AWV | FMOPL000291a | c1ccc2c(c1)c(c[nH]2)C[NH+]3CCCCCC3 | 229.35 | 1 / 3 | 0.95 / 0.79 | 03:38 | 1.58 |
| 5 | x0086 | 5QPH | LUS | FMOPL000315a | CCC1CCC(CC1)[NH+]2CCOCC2 | 198.33 | 3 | 0.66 | 03:46 | 1.86 |
| 6 | x0106 | 5QPI | GQM | FMOPL000554a | COc1ccc(c(c1)O)c2cc[nH]n2 | 190.20 | 6 | 0.54 | 03:58 | 1.67 |
| 7 | x0129 | 5QPJ | JGJ | FMOPL000465a | C[C@H](c1ccccc1)NC(=O)COC | 193.25 | 1 / 6 | 0.82 / 0.66 | 03:21 | 1.41 |
| 8 | x0163 | 5QPK | LUY | FMOPL000586a | c1ccc(cc1)CCNc2[nH]c3ccccc3n2 | 237.31 | 3 | 0.56 | 03:39 | 1.50 |
| 9 | x0165 | 5QPL | M0J | FMOPL000464a | C[C@@H](CNC(=O)Nc1ccccc1)O | 194.23 | 6 | 0.66 | 03:40 | 1.41 |
| 10 | x0168 | 5QPM | LV1 | FMOPL000500a | c1cc(cc(c1)O)NC(=O)Nc2ccc(cc2)F | 246.24 | 5 | 0.72 | 03:24 | 1.68 |
| 11 | x0196 | 5QPN | LDV | FMOPL000576a | CC1CC[NH+](CC1)Cc2c[nH]c3c2cccc3 | 229.35 | 3 | 0.77 | 03:43 | 1.45 |
| 12 | x0197 | 5QPO | GQP | FMOPL000574a | c1ccc2c(c1)ncn2Cc3ccc(cc3)F | 226.25 | 2 | 0.94 | 03:43 | 1.60 |
| 13 | x0231 | 5QPP | LV4 | FMOPL000512a | c1ccc(c(c1)NC(=S)N)OC(F)(F)F | 236.21 | 1/2/2/6 | 0.72/0.54/0.54/0.67 | 03:52 | 1.48 |
| 14 | x0232 | 5QPQ | LV7 | FMOPL000631a | CC(=O)Nc1ccccc1C(=O)NN | 193.21 | 1 | 0.94 | 03:52 | 1.49 |
| 15 | x0246 | 5QPR | LVD | XST00001145b | c1ccc(cc1)CONC(=O)N | 166.18 | 2 | 0.86 | 04:00 | 1.67 |
| 16 | x0273 | 5QPS | LVP | FMOPL000644a | c1ccc(c(c1)NNC(=O)c2ccno2)F | 221.19 | 4 / 5 | 0.68 / 0.82 | 02:56 | 1.61 |
| 17 | x0284 | 5QPT | LVV | FMOPL000642a | Cc1ccc(cc1)C[NH+]2CCS(=O)(=O)CC2 | 240.34 | 1 / 3 / 11 | 0.68/0.58/0.74 | 03:04 | 1.46 |
| 18 | x0286 | 5QPU | JHS | FMOPL000733a | CC(=O)NCC1(CCOCC1)c2ccccc2 | 233.31 | 2 | 0.88 | 03:04 | 1.44 |
| 19 | x0304 | 5QPV | LWA | FMOPL000416a | c1cc(ccc1C(=O)N)NC(=O)[C@@H]2CCCO2 | 234.26 | SX (6) / 16 | 0.53 / 1.00 | 03:21 | 1.60 |
| 20 | x0310 | 5QPW | JH7 | FMOPL000632a | Cn1c(cc(=O)[nH]1)Nc2ccccc2 | 189.22 | 1 / 5 | 0.88 / 0.87 | 03:24 | 1.72 |
| 21 | x0316 | 5QPX | LWD | FMOPL000534a | c1ccc(cc1)C(=O)Nc2ccc3c(c2)cccn3 | 248.29 | 4 / 5 | 0.70 / 0.70 | 03:01 | 1.67 |
| 22 | x0321 | 5QPY | JH1 | FMOPL000449a | CCn1cc(cn1)C(=O)NCc2ccc(cc2)F | 247.27 | 5 | 0.70 | 03:36 | 1.67 |
| 23 | x0325 | 5QPZ | AYV | FMOPL000524a | CC(CO)(CO)NC(=O)Nc1ccccc1 | 244.26 | 5 / 6 | 0.70 / 0.72 | 03:38 | 1.62 |
| 24 | x0336 | 5QQ0 | LWV | XST00000046b | c1ccc(c(c1)N)N2CCOCC2 | 178.24 | 11 | 0.78 | 03:46 | 1.60 |
| 25 | x0355 | 5QQ1 | LX4 | FMOPL000699a | C1CCC(C1)C(=O)N2CCNC(=O)C2 | 196.25 | 2 | 0.70 | 03:58 | 1.97 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 26 | x0359 | 5QQ2 | MJ4 | FMOPL000693a | Cc1ccc(s1)C[NH2+]C[C@@H](C)O | 186.29 | 3 | 0.62 | 04:06 | 1.73 |
| 27 | x0360 | 5QQ3 | LXA | FMOPL000672a | COCC(=O)NCc1[nH]c2ccccc2n1 | 219.24 | 2 / 2 / 12 | 0.84 / 0.84 / 0.64 | 04:06 | 1.60 |
| 28 | x0406 | 5QQ4 | LX7 | FMOPL000276a | c1cc(ccc1NC(=S)N)OC(F)(F)F | 236.21 | 2 / 5 / 12 | 0.78 / 0.66 / 0.74 | 01:56 | 1.58 |
| 29 | x0411 | 5QQ5 | JJM | PKTTA024495b | Cc1cccc(c1)Nc2c3cnn(c3ncn2)C | 239.28 | 5 | 0.78 | 02:00 | 1.68 |
| 30 | x0483 | 5QQ6 | LXJ | FMOOA000530a | O=C(N1CC[C@@H]2CNC(=O)[C@H]2CC1)c3ccccc3 | 258.32 | 11 | 0.66 | 03:01 | 1.94 |
| 31 | x0515 | 5QQ7 | LXM | FMOOA000562a | OC[C@@H]1CN([C@@H]2[C@H]1Oc3ccccc23)C(=O)c4ccccc4 | 295.34 | 5 | 0.64 | 03:20 | 1.61 |
| 32 | x0516 | 5QQ8 | LXS | FMOOA000563a | CCN1C(=O)CN([C@H]2[C@H](O)[C@@H]3O[C@H]2c4ccccc34)C1=O | 288.30 | 2 | 0.78 | 03:21 | 1.62 |
| 33 | x0520 | 5QQ9 | M0D | FMOOA000567a | NC(=O)CN1C[C@H](O)[C@@H]2C[C@@H]1Cc3c2[nH]c4ccccc34 | 285.35 | SY (3) | 0.76 | 03:24 | 1.61 |
| 34 | x0596 | 5QQA | LZV | FMOOA000648a | CCOC(=O)C1CN2N(CCC2=O)C13CCCC3 | 252.31 | 2 | 0.70 | 04:07 | 2.20 |
| 35 | x0623 | 5QQB | LZY | FMOOA000676a | O=C1CCNc2ccccc2CN1CC3CC3 | 230.31 | 2 | 0.66 | 03:45 | 1.58 |

[a] Fragment binder of crystal structures 1 – 29 and 30 – 35 derived from the Diamond-SGC poised library (DSPL) library and Keymical fragments library (EDELRIS).
[b] Binding sites 3 and 6 correspond to the active and allosteric binding site, respectively

**Table 35**: XChem campaign: Data collection and refinement statistics of the TcFPPS structural models.

| **PDB ID** | **5QPD** | **5QPE** | **5QPF** | **5QPG** | **5QPH** |
|---|---|---|---|---|---|
| **compound**[a] | | | | | |
| PDB identifier | **LT7** | **AWG** | **AWM** | **AWV** | **LUS** |
| XChem identifier | FMOPL000293a | FMOPL000295a | FMOPL000478a | FMOPL000291a | FMOPL000315a |
| **Data collection** | | | | | |
| X-ray source | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | | |
| *a* = *b*, *c* (Å) | 57.83, 395.30 | 57.86, 395.38 | 57.90, 395.16 | 58.03, 396.05 | 57.85, 396.78 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 197.67 – 1.93 | 197.66 – 1.77 | 98.82 – 1.50 | 198.04 – 1.58 | 132.29 – 1.86 |
| | (1.98 – 1.93)[b] | (1.82 – 1.77)[b] | (1.54 – 1.50)[b] | (1.62 – 1.58)[b] | (1.91 – 186)[b] |
| $R_{merge}$ | 0.188 (2.834)[b] | 0.201 (2.981)[b] | 0.122 (2.130)[b] | 0.095 (1.845)[b] | 0.18 (3.638)[b] |
| Unique reflections | 31230 | 40215 | 65050 | 54689 | 34907 |
| $I$ / σ$I$ | 12.2 (2.49)[b] | 11.5 (1.86)[b] | 12.9 (1.64)[b] | 16.6 (2.72)[b] | 11.5 (2.40)[b] |
| Completeness (%) | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 98.2 (95.6)[b] | 100 (100)[b] |
| Redundancy | 18.2 (19.3)[b] | 18.3 (17.5)[b] | 16.6 (11.8)[b] | 17.8 (14.7)[b] | 18.3 (19.4)[b] |
| **Refinement** | | | | | |
| Resolution (Å) | 65.88 – 1.93 | 65.90 – 1.77 | 49.74 – 1.50 | 66.01 – 1.58 | 66.13 – 1.86 |
| No. reflections | 29365 | 37756 | 64258 | 51049 | 32708 |
| $R_{work}$ / $R_{free}$ | 0.201, 0.259 | 0.198, 0.247 | 0.213, 0.248 | 0.198, 0.237 | 0.202, 0.256 |
| No. atoms | | | | | |
| Protein | 2860 | 2871 | 2880 | 2871 | 2871 |
| Ligand/ion | 33 | 37 | 63 | 60 | 30 |
| Water | 301 | 317 | 330 | 336 | 304 |
| B-factors overall (Å$^2$) | 33.27 | 24.58 | 22.18 | 24.51 | 31.71 |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.019 | 0.021 | | 0.011 | 0.020 |
| Bond angles (°) | 1.646 | 1.854 | | 1.711 | 1.833 |
| Molprobity statistics | | | | | |
| Ramachandran | | | | | |
| Favoured (%) | 98 | 98 | 98 | 98 | 98 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 |
| Allowed (%) | 2 | 2 | 2 | 2 | 2 |
| Solvent content (%) | 46.68 | 46.75 | 46.79 | 47.15 | 46.91 |

[a] Fragment binders were previously listed in **Table 34**.
[b] Values in parentheses are for the highest resolution shell.

Table is continued on the next pages.

**Table 35** continued.

| **PDB ID** | **5QPI** | **5QPJ** | **5QPK** | **5QPL** | **5QPM** |
|---|---|---|---|---|---|
| **compound**[a] | | | | | |
| PDB identifier | **GQM** | **JGJ** | **LUY** | **M0J** | **LV1** |
| XChem identifier | FMOPL000554a | FMOPL000465a | FMOPL000586a | FMOPL000464a | FMOPL000500a |
| **Data collection** | | | | | |
| X-ray source | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | | |
| *a* = *b*, *c* (Å) | 57.82, 396.38 | 57.96, 396.75 | 58.13, 397.40 | 58.02, 395.55 | 57.98, 395.53 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 198.22 – 1.67 | 98.97 – 1.41 | 198.72 – 1.50 | 197.85 – 1.41 | 131.84 – 1.68 |
| | (1.71 – 1.67)[b] | (1.45 – 1.41)[b] | (1.54 – 1.50)[b] | (1.45 – 1.41) | (1.72 – 1.68)[b] |
| $R_{merge}$ | 0.122 (2.504)[b] | 0.098 (1.884)[b] | 0.075 (2.110)[b] | 0.095 (1.992) | 0.206 (3.175)[b] |
| Unique reflections | 47494 | 78126 | 65934 | 78228 | 46981 |
| $I$ / σ$I$ | 11.5 (1.83)[b] | 13.7 (1.84)[b] | 16.8 (2.16) | 13.1 (1.55)[b] | 11.0 (1.66)[b] |
| Completeness (%) | 100 (100)[b] | 100 (100)[b] | 100 (100) | 100 (100)[b] | 100 (100)[b] |
| Redundancy | 18.0 (15.6)[b] | 15.7 (9.7)[b] | 16.7 (11.9) | 15.5 (9.6)[b] | 18.1 (15.7)[b] |
| **Refinement** | | | | | |
| Resolution (Å) | 66.06 – 1.67 | 65.96 – 1.41 | 66.23 – 1.50 | 65.93 – 1.41 | 65.92 – 1.68 |
| No. reflections | 43318 | 73169 | 61504 | 73370 | 44134 |
| $R_{work}$ / $R_{free}$ | 0.203, 0.254 | 0.197, 0.226 | 0.203, 0.240 | 0.197, 0.228 | 0.209, 0.254 |
| No. atoms | | | | | |
| Protein | 2871 | 2889 | 2871 | 2871 | 2871 |
| Ligand/ion | 44 | 44 | 33 | 30 | 38 |
| Water | 304 | 331 | 305 | 331 | 321 |
| B-factors overall (Å$^2$) | 32.16 | 22.44 | 28.57 | 22.55 | 24.62 |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.030 | 0.031 | 0.028 | 0.026 | 0.029 |
| Bond angles (°) | 1.884 | 2.515 | 2.241 | 2.440 | 2.033 |
| Molprobity statistics | | | | | |
| Ramachandran | | | | | |
| Favoured (%) | 99 | 99 | 99 | 99 | 98 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 |
| Allowed (%) | 1 | 1 | 1 | 1 | 2 |
| Solvent content (%) | 46.80 | 46.97 | 47.51 | 47.06 | 46.96 |

[a] Fragment binders were previously listed in **Table 34**.
[b] Values in parentheses are for the highest resolution shell.

**Table 35** continued.

| **PDB ID** | **5QPN** | **5QPO** | **5QPP** | **5QPQ** | **5QPR** |
|---|---|---|---|---|---|
| **compound**[a] | | | | | |
| PDB identifier | **LDV** | **GQP** | **LV4** | **LV7** | **LVD** |
| XChem identifier | FMOPL000576a | FMOPL000574a | FMOPL000512a | FMOPL000613a | XST00001145b |
| **Data collection** | | | | | |
| X-ray source | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | | |
| *a* = *b*, *c* (Å) | 58.13, 395.33 | 57.89, 395.79 | 57.99, 395.77 | 57.91, 396.13 | 58.20, 397.61 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 197.65 – 1.45 (1.49 – 1.45)[b] | 198.04 – 1.60 (1.64 – 1.60)[b] | 197.91 – 1.48 (1.52 – 1.48)[b] | 198.00 – 1.19 (1.53 – 1.49)[b] | 29.37 – 1.67 (1.71 – 167)[b] |
| $R_{merge}$ | 0.143 (2.190)[b] | 0.212 (2.642)[b] | 0.222 (2.035)[b] | 0.082 (2.045)[b] | 0.090 (0.829)[b] |
| Unique reflections | 72358 | 54030 | 67934 | 66315 | 48552 |
| $I$ / σ$I$ | 12.9 (2.26)[b] | 7.6 (1.75)[b] | 7.3 (1.49)[b] | 16.2 (1.99)[b] | 18.8 (1.83)[b] |
| Completeness (%) | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] |
| Redundancy | 16.2 (10.9)[b] | 17.4 (14.5)[b] | 16.4 (11.5)[b] | 16.7 (12.0)[b] | 17.8 (15.2)[b] |
| **Refinement** | | | | | |
| Resolution (Å) | 50.35 – 1.45 | 65.96 – 1.60 | 65.96 – 1.48 | 66.02 – 1.49 | 66.36 – 1.67 |
| No. reflections | 71542 | 49443 | 63377 | 61689 | 45886 |
| $R_{work}$ / $R_{free}$ | 0.209, 0.229 | 0.227, 0.269 | 0.205, 0.240 | 0.200, 0.235 | 0.176, 0.216 |
| No. atoms | | | | | |
| Protein | 2871 | 2871 | 2880 | 2871 | 2871 |
| Ligand/ion | 38 | 31 | 76 | 30 | 28 |
| Water | 316 | 327 | 317 | 333 | 333 |
| B-factors overall (Å$^2$) | 19.02 | 18.81 | 22.50 | 26.65 | 30.23 |
| R.m.s. deviations | | | | | |
| Bond length (Å) | | 0.022 | 0.022 | 0.027 | 0.023 |
| Bond angles (°) | | 2.193 | 2.193 | 2.411 | 1.946 |
| Molprobity statistics | | | | | |
| Ramachandran | | | | | |
| Favoured (%) | 99 | 98 | 98 | 99 | 98 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 |
| Allowed (%) | 1 | 2 | 2 | 1 | 2 |
| Solvent content (%) | 47.24 | 46.86 | 47.04 | 46.94 | 47.67 |

[a] Fragment binders were previously listed in **Table 34**.
[b] Values in parentheses are for the highest resolution shell.

**Table 35** continued.

| **PDB ID** | **5QPS** | **5QPT** | **5QPU** | **5QPV** | **5QPW** |
|---|---|---|---|---|---|
| **compound**[a] | | | | | |
| PDB identifier | **LVP** | **LVV** | **JHS** | **LWA** | **JH7** |
| XChem identifier | FMOPL000644a | FMOPL000642a | FMOPL000733a | FMOPL000416a | FMOPL000632a |
| **Data collection** | | | | | |
| X-ray source | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | | |
| *a* = *b*, *c* (Å) | 57.83, 395.38 | 57.88, 395,53 | 57.98, 395.14 | 57.83, 397.47 | 57.82, 395.41 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 131.77 – 1.61 (1.65 – 1.61)[b] | 197.79 – 1.46 (1.50 – 1.46)[b] | 197.64 – 1.44 (1.48 – 1.44)[b] | 132.49 – 1.60 (1.64 – 1.60)[b] | 131.81 – 1.72 (1.76 – 1.72)[b] |
| $R_{merge}$ | 0.181 (2.804)[b] | 0.092 (1.860)[b] | 0.100 (1.545)[b] | 0.144 (2.381)[b] | 0.220 (3.863)[b] |
| Unique reflections | 52840 | 70386 | 73456 | 54109 | 43644 |
| $I$ / σ$I$ | 11.1 (1.88)[b] | 14.5 (1.83)[b] | 13.5 (2.00)[b] | 14.1 (1.89)[b] | 9.8 (1.62)[b] |
| Completeness (%) | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] |
| Redundancy | 17.7 (17.8)[b] | 16.3 (11.1)[b] | 16.1 (10.7)[b] | 17.6 (14.6)[b] | 18.2 (16.4)[b] |
| **Refinement** | | | | | |
| Resolution (Å) | 65.90 – 1.61 | 65.92 – 1.46 | 65.86 – 1.44 | 66.24 – 1.60 | 65.90 – 1.72 |
| No. reflections | 49652 | 65872 | 68819 | 50836 | 41036 |
| $R_{work}$ / $R_{free}$ | 0.198, 0.233 | 0.192, 0.229 | 0.196, 0.225 | 0.196, 0.237 | 0.187, 0.231 |
| No. atoms | | | | | |
| Protein | 2871 | 2871 | 2871 | 2871 | 2871 |
| Ligand/ion | 53 | 68 | 38 | 45 | 44 |
| Water | 307 | 328 | 332 | 312 | 325 |
| B-factors overall (Å$^2$) | 21.52 | 23.84 | 22.64 | 28.93 | 25.66 |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.023 | 0.029 | 0.027 | 0.020 | 0.042 |
| Bond angles (°) | 2.044 | 2.229 | 2.422 | 2.000 | 1.994 |
| Molprobity statistics | | | | | |
| Ramachandran | | | | | |
| Favoured (%) | 99 | 99 | 98 | 97 | 99 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 |
| Allowed (%) | 1 | 1 | 2 | 3 | 1 |
| Solvent content (%) | 46.68 | 46.80 | 46.94 | 46.97 | 46.68 |

[a] Fragment binders were previously listed in **Table 34**.
[b] Values in parentheses are for the highest resolution shell.

**Table 35** continued.

| **PDB ID** | **5QPX** | **5QPY** | **5QPZ** | **5QQ0** | **5QQ1** |
|---|---|---|---|---|---|
| **compound**[a] | | | | | |
| PDB identifier | **LWD** | **JH1** | **AYV** | **LWV** | **LX4** |
| XChem identifier | FMOPL000523a | FMOPL000449a | FMOPL000524a | XST0000046b | FMOPL000699a |
| **Data collection** | | | | | |
| X-ray source | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | | |
| *a* = *b*, *c* (Å) | 57.92, 395.05 | 57.82, 395.68 | 57.85, 394.63 | 57.78, 395.05 | 57.78, 396.81 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 131.68 – 1.67 (1.71 – 1.67)[b] | 197.84 – 1.67 (1.71 – 1.67)[b] | 131.53 – 1.62 (1.66 – 1.62)[b] | 131.6 – 1.60 (1.64 – 1.60)[b] | 198.39 – 1.97 (2.02 – 1.97)[b] |
| $R_{merge}$ | 0.139 (3.259)[b] | 0.180 (2.838)[b] | 0.117 (2.724)[b] | 0.229 (5.786)[b] | 0.239 (3.101)[b] |
| Unique reflections | 47624 | 47553 | 51803 | 53687 | 29547 |
| $I$ / σ$I$ | 12.2 (1.79)[b] | 10.5 (1.89)[b] | 14.0 (1.67)[b] | 10.3 (1.67)[b] | 9.0 (1.71)[b] |
| Completeness (%) | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] |
| Redundancy | 17.8 (15.4)[b] | 17.9 (15.5)[b] | 17.7 (14.9)[b] | 17.6 (14.7)[b] | 18.2 (18.9)[b] |
| **Refinement** | | | | | |
| Resolution (Å) | 65.84 – 1.67 | 65.95 – 1.67 | 65.77 – 1.62 | 65.84 – 1.60 | 66.14 – 1.97 |
| No. reflections | 44598 | 44692 | 48675 | 50494 | 27717 |
| $R_{work}$ / $R_{free}$ | 0.204, 0.255 | 0.194, 0.238 | 0.200, 0.244 | 0.189, 0.220 | 0.199, 0.254 |
| No. atoms | | | | | |
| Protein | 2871 | 2871 | 2871 | 2871 | 2871 |
| Ligand/ion | 73 | 34 | 53 | 42 | 30 |
| Water | 311 | 325 | 316 | 322 | 312 |
| B-factors overall (Å$^2$) | 29.69 | 22.12 | 27.77 | 25.48 | 33.86 |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.021 | 0.020 | 0.021 | 0.025 | 0.018 |
| Bond angles (°) | 1.899 | 1.967 | 2.024 | 2.219 | 1.696 |
| Molprobity statistics | | | | | |
| Ramachandran | | | | | |
| Favoured (%) | 98 | 98 | 99 | 99 | 98 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 |
| Allowed (%) | 2 | 2 | 1 | 1 | 2 |
| Solvent content (%) | 46.81 | 46.72 | 46.62 | 46.54 | 46.78 |

[a] Fragment binders were previously listed in **Table 34**.
[b] Values in parentheses are for the highest resolution shell.

**Table 35** continued.

| **PDB ID** | **5QQ2** | **5QQ3** | **5QQ4** | **5QQ5** | **5QQ6** |
|---|---|---|---|---|---|
| **compound**[a] | | | | | |
| PDB identifier | **MJ4** | **LXA** | **LX7** | **JJM** | **LXJ** |
| XChem identifier | FMOPL000693a | FMOPL000672a | FMOPL000276a | PKTTA024495b | FMOOA000530a |
| **Data collection** | | | | | |
| X-ray source | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | | |
| *a* = *b*, *c* (Å) | 57.83, 395.66 | 57.72, 3.96.82 | 57.82, 396.08 | 57.86, 395.19 | 57.82, 396.68 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 131.90 – 1.73 (1.77 – 1.73)[b] | 198.40 – 1.60 (1.64 – 1.60)[b] | 396.07 – 1.58 (1.62 – 1.58)[b] | 197.60 – 1.68 (1.72 – 1.68)[b] | 198.30 – 1.94 (1.99 – 1.94)[b] |
| $R_{merge}$ | 0.192 (3.355)[b] | 0.149 (3.642)[b] | 0.145 (2.439)[b] | 0.243 (3.272)[b] | 0.332 (4.033)[b] |
| Unique reflections | 42945 | 53792 | 55931 | 46752 | 30877 |
| $I$ / σ$I$ | 11.4 (1.81)[b] | 10.6 (1.61)[b] | 13.5 (2.24)[b] | 9.6 (1.77)[b] | 8.0 (1.70)[b] |
| Completeness (%) | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] |
| Redundancy | 18.2 (16.5)[b] | 17.6 (14.7)[b] | 17.5 (14.1)[b] | 18.0 (15.7)[b] | 18.3 (19.3)[b] |
| **Refinement** | | | | | |
| Resolution (Å) | 65.94 – 1.73 | 66.14 – 1.60 | 66.01 – 1.58 | 65.86 – 1.68 | 66.11 – 1.94 |
| No. reflections | 40240 | 50142 | 52019 | 43993 | 28988 |
| $R_{work}$ / $R_{free}$ | 0.198, 0.238 | 0.196, 0.242 | 0.205, 0.245 | 0.193, 0.237 | 0.201, 0.249 |
| No. atoms | | | | | |
| Protein | 2871 | 2871 | 2871 | 2871 | 2871 |
| Ligand/ion | 33 | 65 | 60 | 39 | 35 |
| Water | 308 | 319 | 320 | 329 | 302 |
| B-factors overall (Å$^2$) | 26.05 | 29.3 | 22.91 | 20.43 | 26.76 |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.023 | 0.038 | 0.046 | 0.023 | 0.020 |
| Bond angles (°) | 1.879 | 2.009 | 2.606 | 1.971 | 1.727 |
| Molprobity statistics | | | | | |
| Ramachandran | | | | | |
| Favoured (%) | 98 | 98 | 99 | 98 | 98 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 |
| Allowed (%) | 2 | 2 | 1 | 2 | 2 |
| Solvent content (%) | 46.72 | 46.68 | 46.77 | 46.71 | 46.84 |

[a] Fragment binders were previously listed in **Table 34**.
[b] Values in parentheses are for the highest resolution shell.

**Table 35** continued.

| **PDB ID** | **5QQ7** | **5QQ8** | **5QQ9** | **5QQA** | **5QQB** |
|---|---|---|---|---|---|
| **compound**[a] | | | | | |
| PDB identifier | **LXM** | **LXS** | **M0D** | **LZV** | **LZY** |
| XChem identifier | FMOPL000562a | FMOPL000563a | FMOOA000567a | FMOOA000648a | FMOOA000676a |
| **Data collection** | | | | | |
| X-ray source | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å | 0.91587 Å |
| Space group | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ | $P6_122$ |
| Cell dimensions | | | | | |
| *a* = *b*, *c* (Å) | 57.76, 396.04 | 57.78, 396.94 | 57.68, 395.94 | 57.81, 396.97 | 57.76, 394.81 |
| α, β, γ (°) | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 | 90, 90, 120 |
| Resolution (Å) | 198.02 – 1.61 (1.65 – 1.61)[b] | 132.37 – 1.62 (1.66 – 1.62)[b] | 131.99 – 1.61 (1.65 – 1.61)[b] | 198.52 – 2.20 (2.26 – 2.20)[b] | 78.96 – 1.58 (1.62 – 1.58)[b] |
| $R_{merge}$ | 0.144 (3.170)[b] | 0.136 (2.343)[b] | 0.097 (2.772)[b] | 0.537 (4.882)[b] | 0.124 (2.533)[b] |
| Unique reflections | 52786 | 52032 | 52625 | 21537 | 55626 |
| $I$ / σ$I$ | 10.4 (1.71)[b] | 10.8 (1.78)[b] | 14.8 (1.81)[b] | 7.9 (2.13)[b] | 11.6 (2.03)[b] |
| Completeness (%) | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] | 100 (100)[b] |
| Redundancy | 17.7 (14.8)[b] | 17.6 (14.9)[b] | 17.7 (14.9)[b] | 18.2 (18.4)[b] | 17.3 (14.00)[b] |
| **Refinement** | | | | | |
| Resolution (Å) | 66.01 – 1.61 | 66.16 – 1.62 | 65.99 – 1.61 | 66.16 – 2.20 | 65.8 – 1.58 |
| No. reflections | 48936 | 18732 | 49393 | 20151 | 51823 |
| $R_{work}$ / $R_{free}$ | 0.192, 0.233 | 0.198, 0.235 | 0.191, 0.238 | 0.204 – 0.276 | 0.195 – 0.230 |
| No. atoms | | | | | |
| Protein | 2871 | 2871 | 2871 | 2871 | 2871 |
| Ligand/ion | 38 | 37 | 37 | 34 | 31 |
| Water | 308 | 307 | 314 | 306 | 310 |
| B-factors overall (Å$^2$) | 28.21 | 28.26 | 30.61 | 37.5 | 27.35 |
| R.m.s. deviations | | | | | |
| Bond length (Å) | 0.023 | 0.024 | 0.024 | 0.016 | 0.363 |
| Bond angles (°) | 2.062 | 2.095 | 2.191 | 1.538 | 2.146 |
| Molprobity statistics | | | | | |
| Ramachandran | | | | | |
| Favoured (%) | 99 | 98 | 99 | 98 | 98 |
| Outliers (%) | 0 | 0 | 0 | 0 | 0 |
| Allowed (%) | 1 | 2 | 1 | 2 | 2 |
| Solvent content (%) | 46.65 | 46.81 | 46.48 | 46.87 | 46.49 |

[a] Fragment binders were previously listed in **Table 34**.
[b] Values in parentheses are for the highest resolution shell.

**Table 35** continued.

| **PDB ID** | **5QQC** |
|---|---|
| **compound** | |
| compound | apo (ground-state) |
| **Data collection** | |
| X-ray source | I04-1, DLS |
| Wavelength (Å) | 0.91587 Å |
| Space group | $P6_122$ |
| Cell dimensions | |
| *a* = *b*, *c* (Å) | 57.68, 396.51 |
| α, β, γ (°) | 90, 90, 120 |
| Resolution (Å) | 132.19 – 1.62 |
| | (1.66 – 1.62)[a] |
| $R_{merge}$ | 0.100 (2.734)[a] |
| Unique reflections | 51799 |
| $I / \sigma I$ | 15.0 (1.81)[a] |
| Completeness (%) | 100 (100)[a] |
| Redundancy | 17.7 (14.9)[a] |
| **Refinement** | |
| Resolution (Å) | 66.09 – 1.62 |
| No. reflections | 48589 |
| $R_{work}$ / $R_{free}$ | 0.192 – 0.227 |
| No. atoms | |
| Protein | 2871 |
| Ligand/ion | 16 |
| Water | 291 |
| B-factors overall (Å$^2$) | 30.73 |
| R.m.s. deviations | |
| Bond length (Å) | 0.022 |
| Bond angles (°) | 2.015 |
| Molprobity statistics | |
| Ramachandran | |
| Favoured (%) | 99 |
| Outliers (%) | 0 |
| Allowed (%) | 1 |
| Solvent content (%) | 46.57 |

[a] Values in parentheses are for the highest resolution shell.

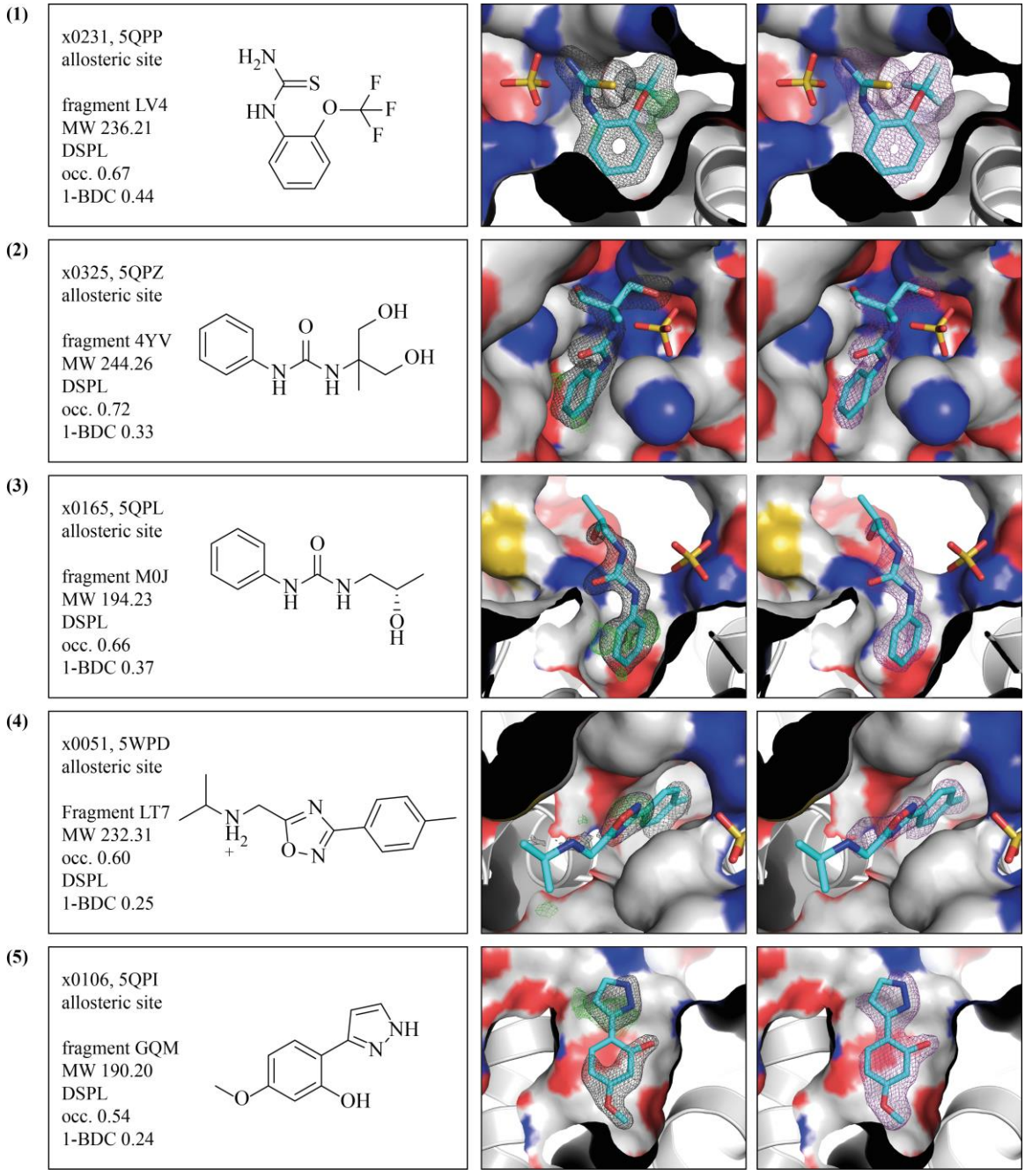



**Figure 74**: Overview of all ligands modelled in 35 crystal structures. Fragment hits are spread over several different sites in TcFPPS and multiple binding events lead to a total of 51 events with ligands. Fragment binders were previously listed in **Table 34**. **(1) – (5)** Allosteric site binders. (**6)** Novel binding site SX. **(7) – (13)** Active site binders. **(14) – (21)** Binding site S1. **(22) – (32)** Binding site S2. **(33)** and **(34)** Binding site S4. **(35)** – **(44)** Binding site S5. **(45)** Binding site S3. **(46)** – **(48)** Binding sites S11. **(49)** and **(50)** Binding site S12. **(51)** Binding site S16. All images follow the same scheme: On the left the PDB ID is given for the crystal and the PDB ID, MW, origin (DSPL or EDELRIS library), occupancy in the crystal and chemical structure are given for the ligand. In the middle the ligand is depicted in stick representation with its final $2F_o - F_c$ electron density map shown as liquorice coloured mesh contoured at 1.0 σ and its $F_o - F_c$ positive and negative difference electron density map shown as green and red coloured mesh, respectively, contoured at 3.0 σ. On the right the event map that lead to ligand identification in PanDDA inspect is shown as violet coloured mesh at twice the 1-BDC value which corresponds to a level of 2.0 σ TcFPPS is shown in cartoon and surface representation and is coloured by element: C, O, N and S in grey, red, blue and yellow respectively. Figure is continued on the next pages.

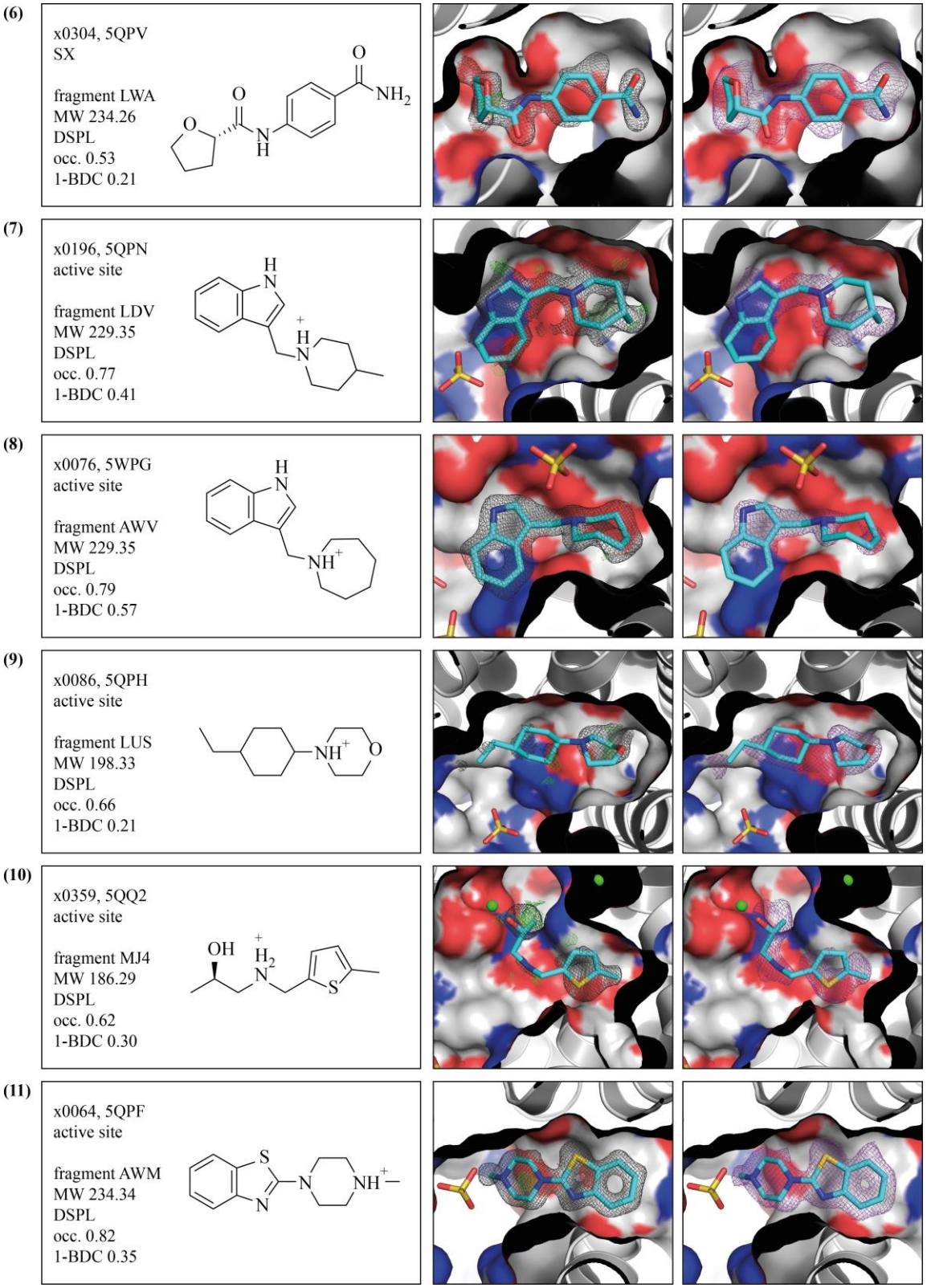



**Figure 74** continued.

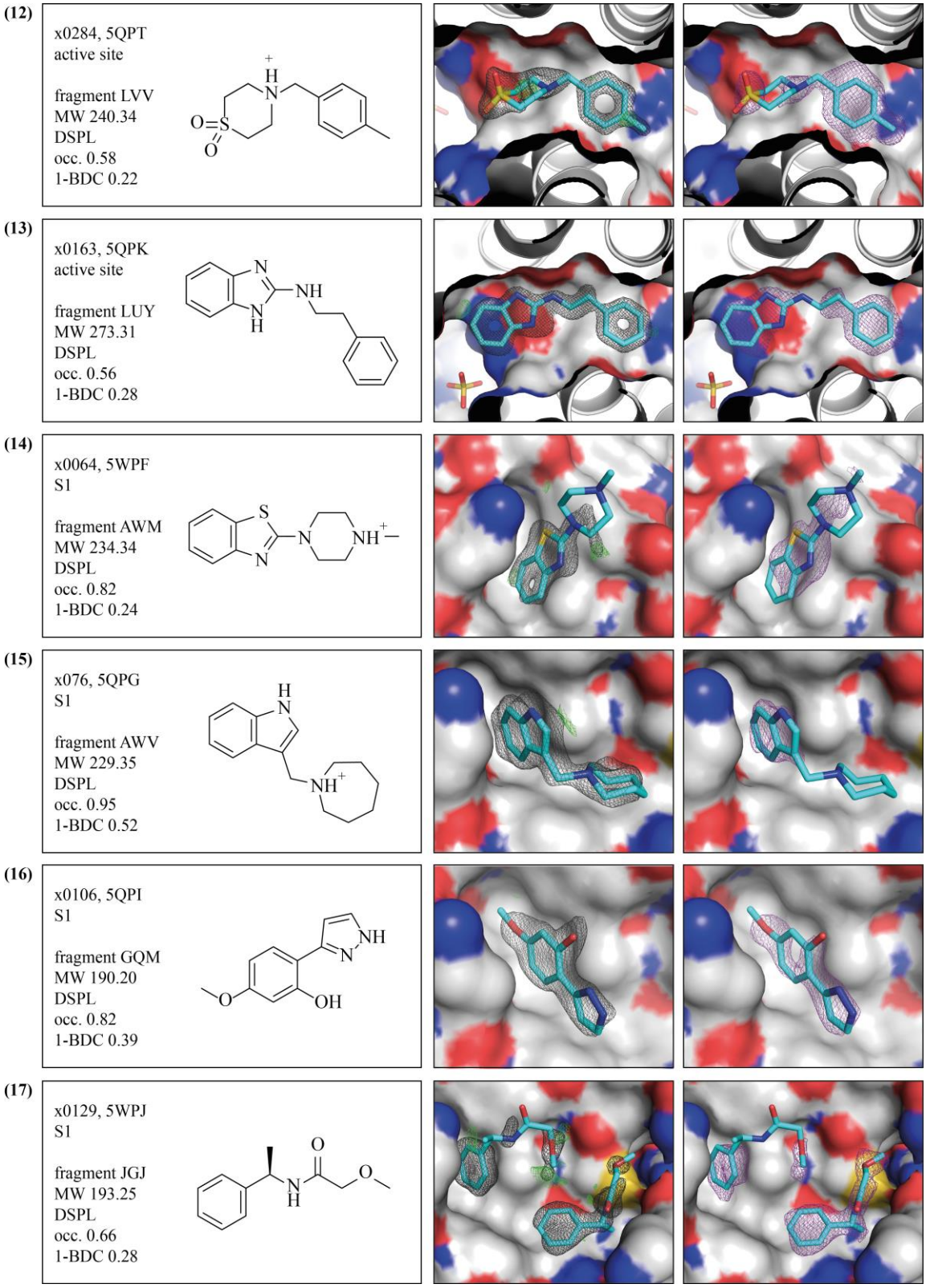



**Figure 74** continued.

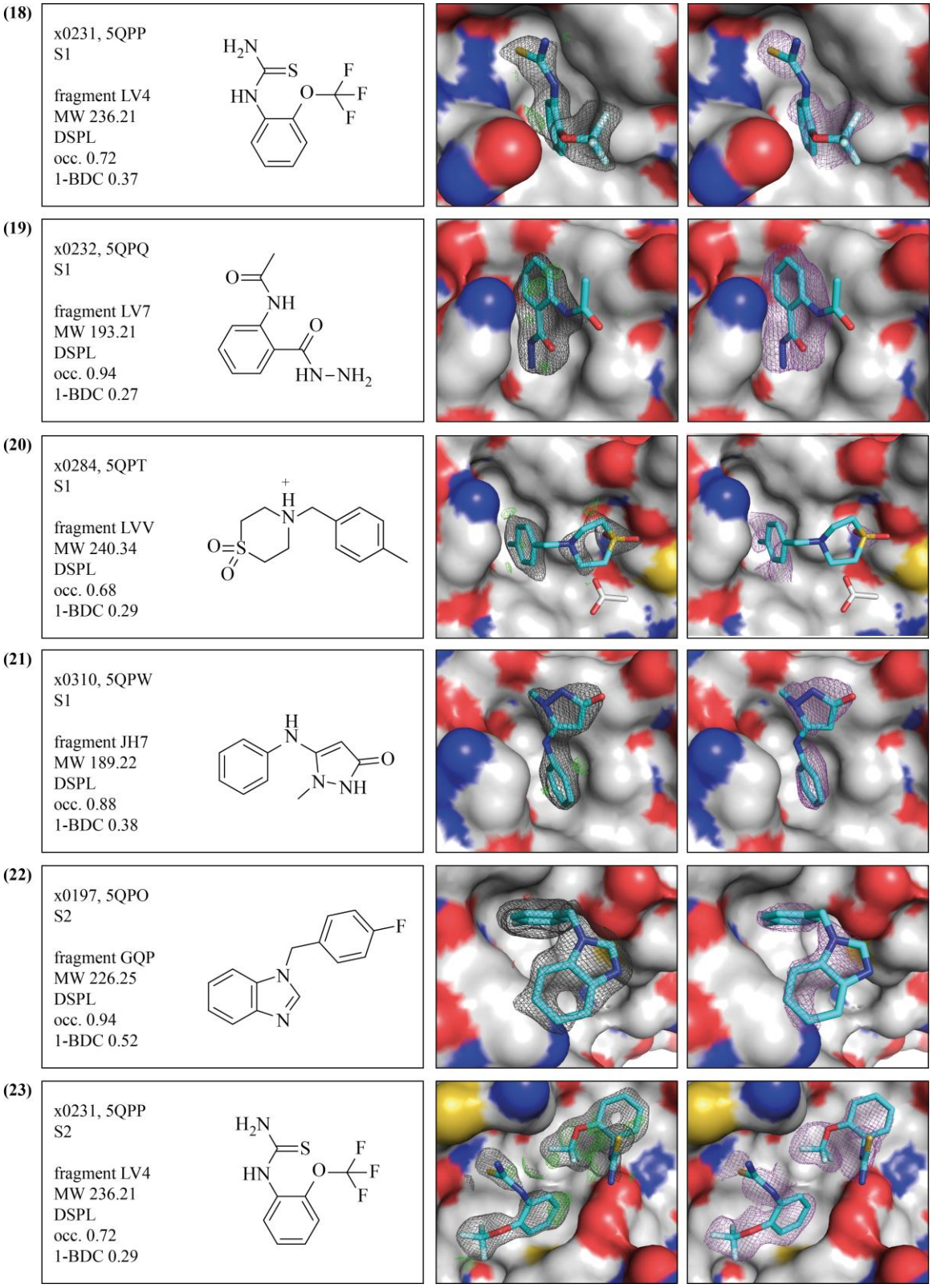



**Figure 74** continued.

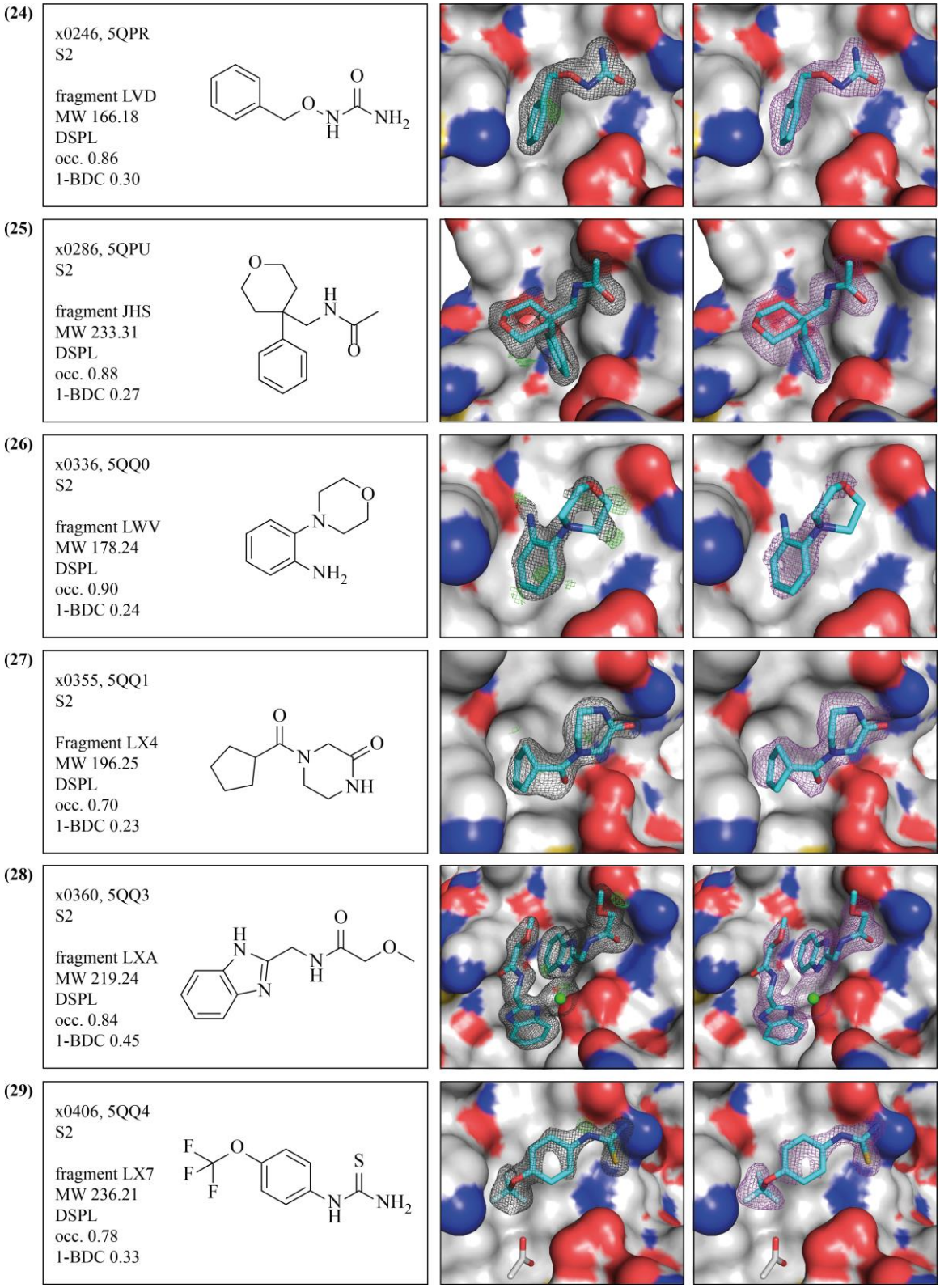



**Figure 74** continued.

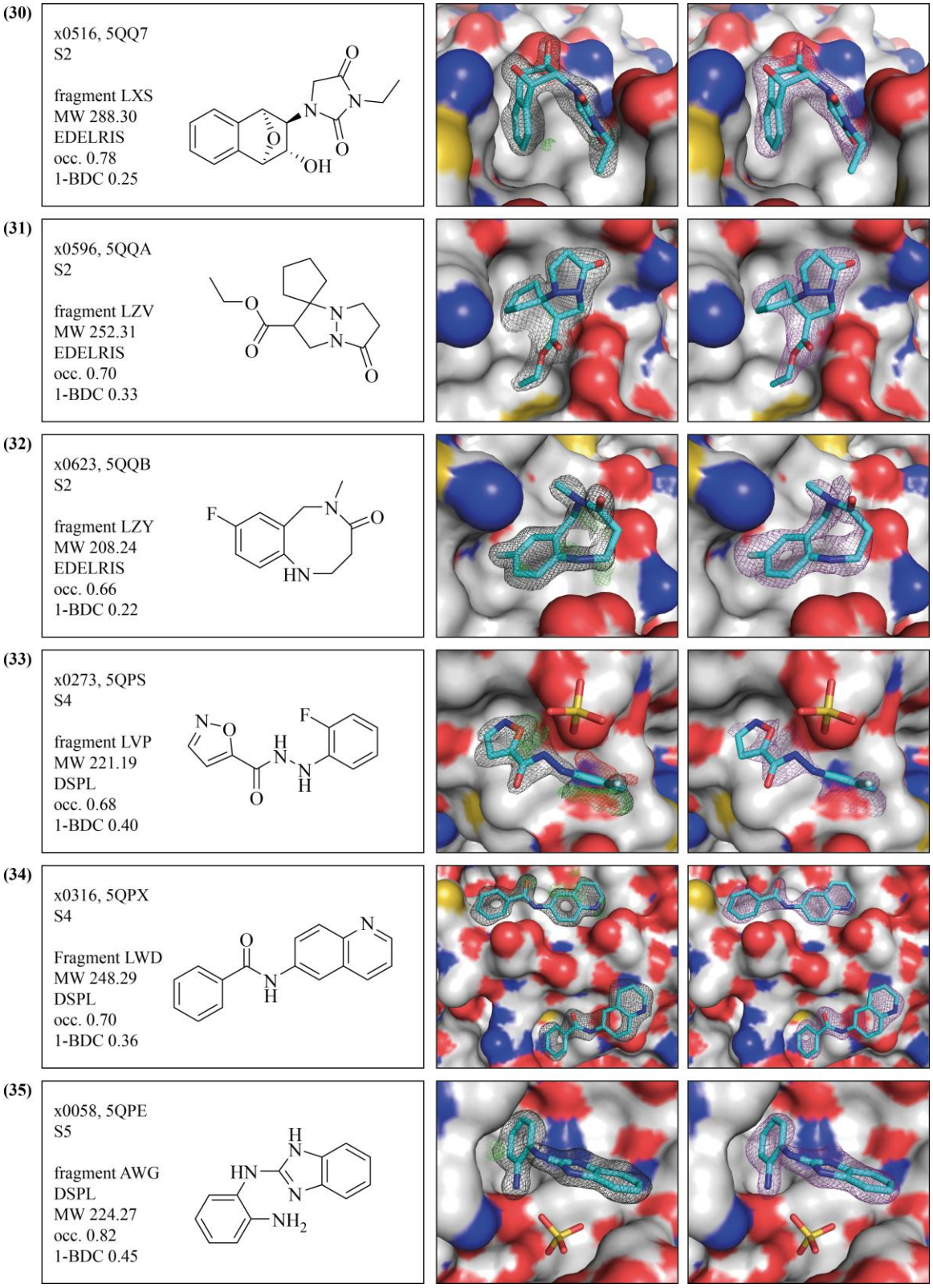



**Figure 74** continued.

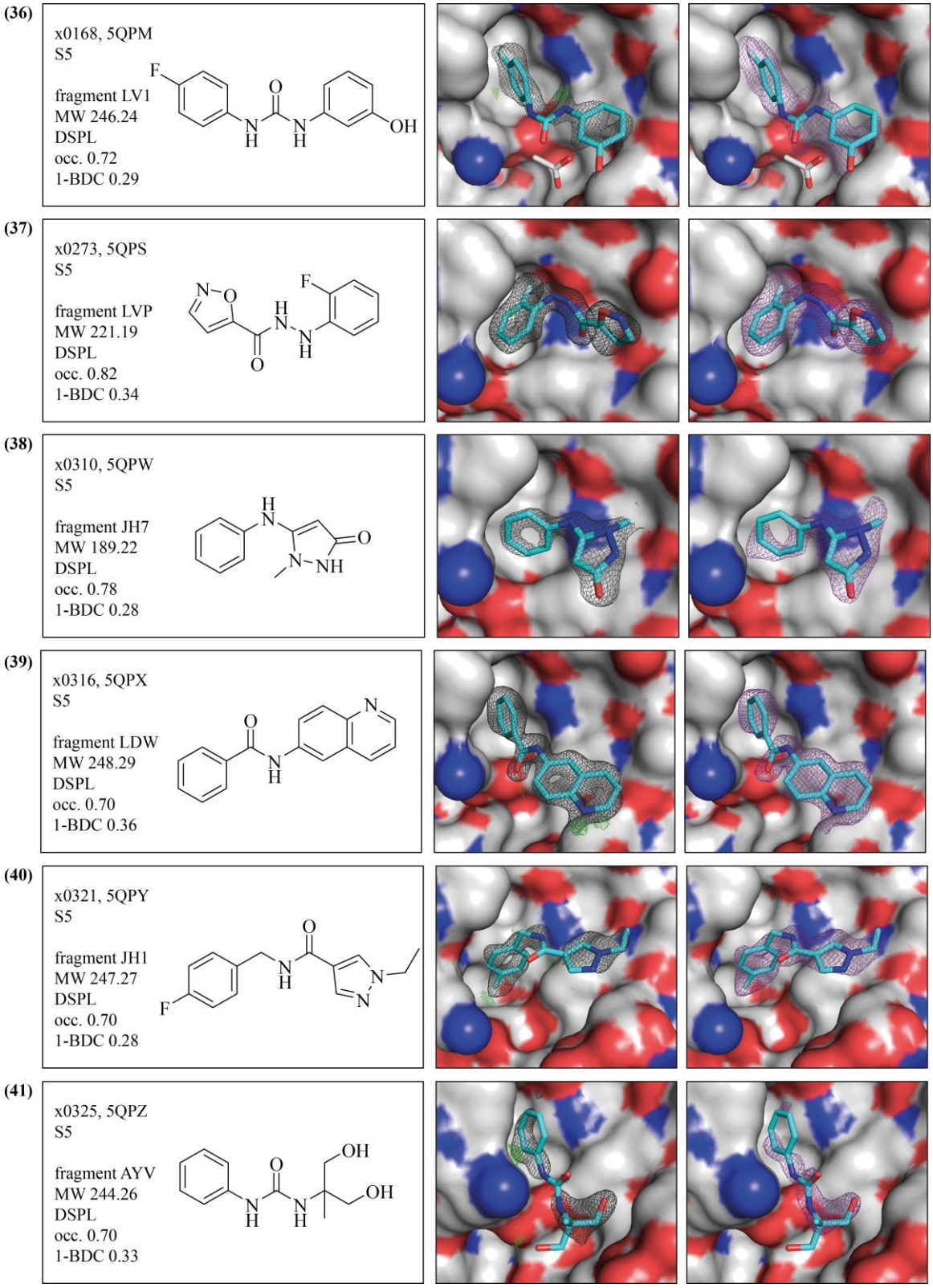



**Figure 74** continued.

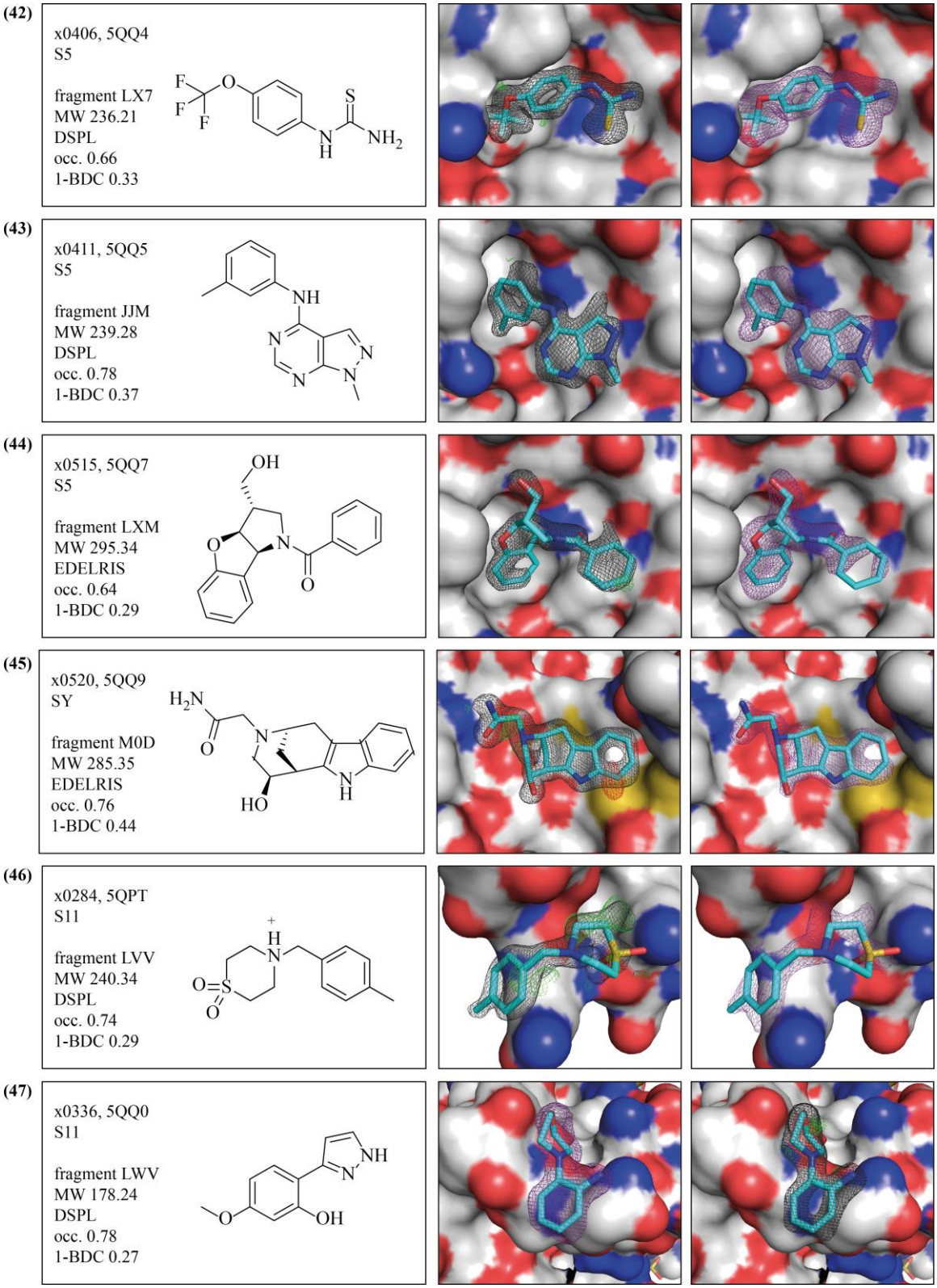



**Figure 74** continued.

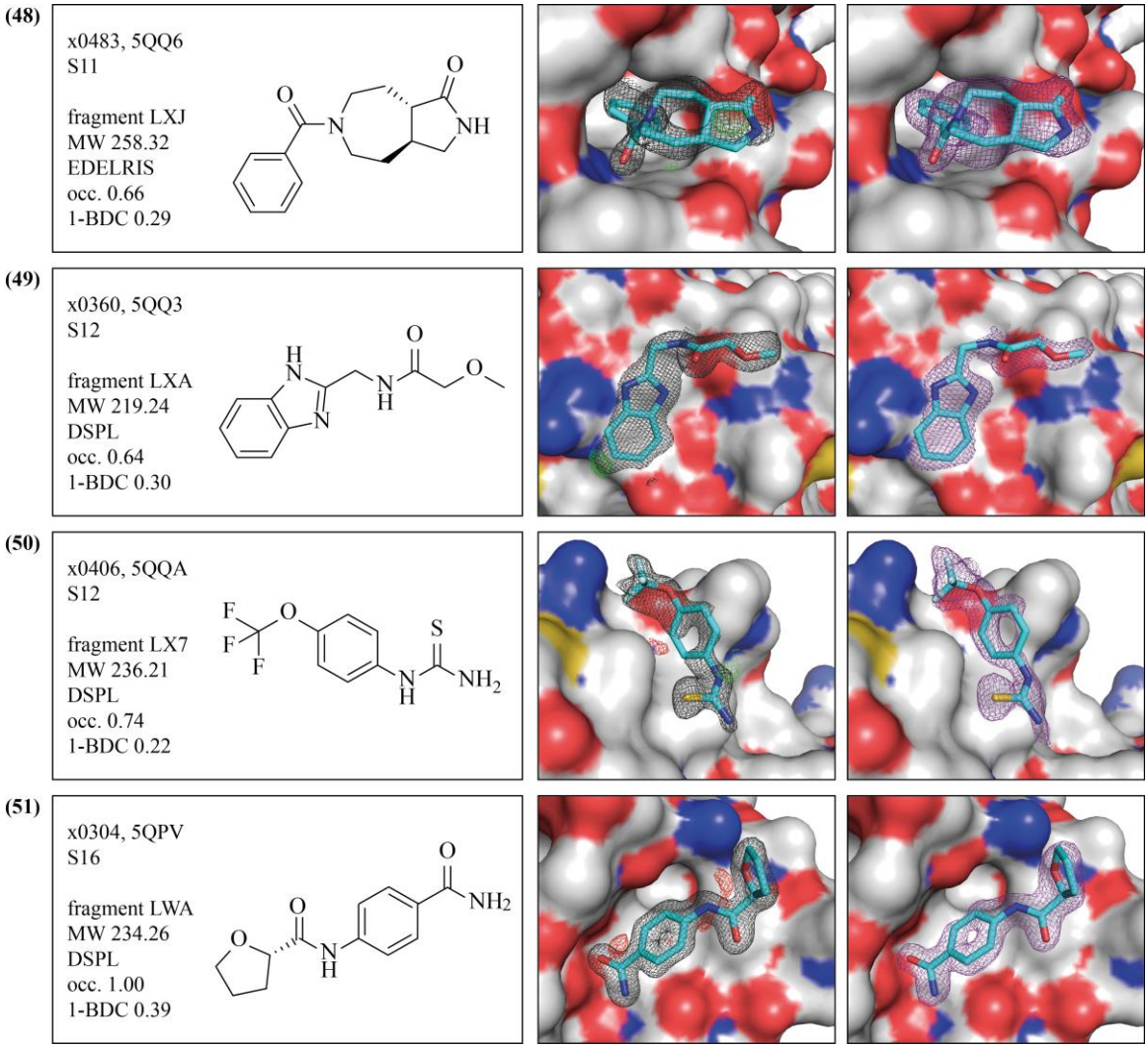



**Figure 74** continued.

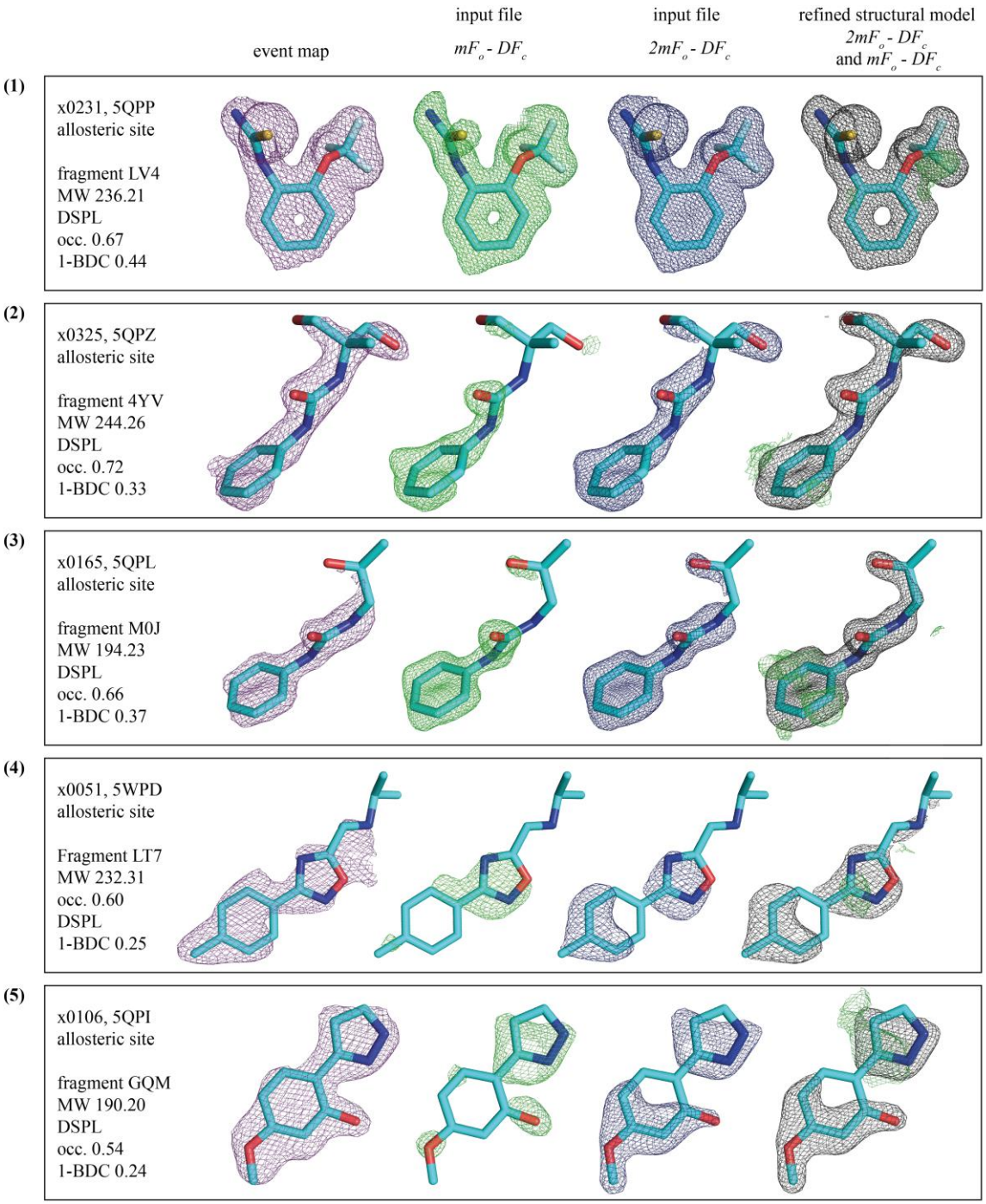



**Figure 75**: Summary of density maps of ligands binding in the allosteric, SX and active site of TcFPPS. **(1) – (5)** Maps of allosteric site binders **LV4**, **AYV**, **M0J**, **LT7** and **GQM**, respectively. **(6)** Maps of ligand **LWA**. **(7)** – **(13)** Maps of active site binders **LDV**, **AWV**, **LUS**, **MJ4**, **AWM**, **LVV** and **LUY**, respectively. On the left, the event maps are shown contoured at twice the 1-BDC value, which corresponds to a level of 2.0 σ (violet mesh). In the middle the maps of the input files are shown: the $mF_o – DF_c$ positive difference electron density maps contoured at 3.0 σ (green mesh) and the $2mF_o – DF_c$ electron density maps contoured at 1.0 σ (blue mesh). On the right, the refined maps of the final model are shown: the $2mF_o – DF_c$ electron density maps contoured at 1.0 σ (liquorice mesh) in superimposition with the $mF_o – DF_c$ positive and negative difference electron density maps (green and red coloured mesh, respectively) contoured at 3.0 σ, respectively. Figure is continued on the next pages.

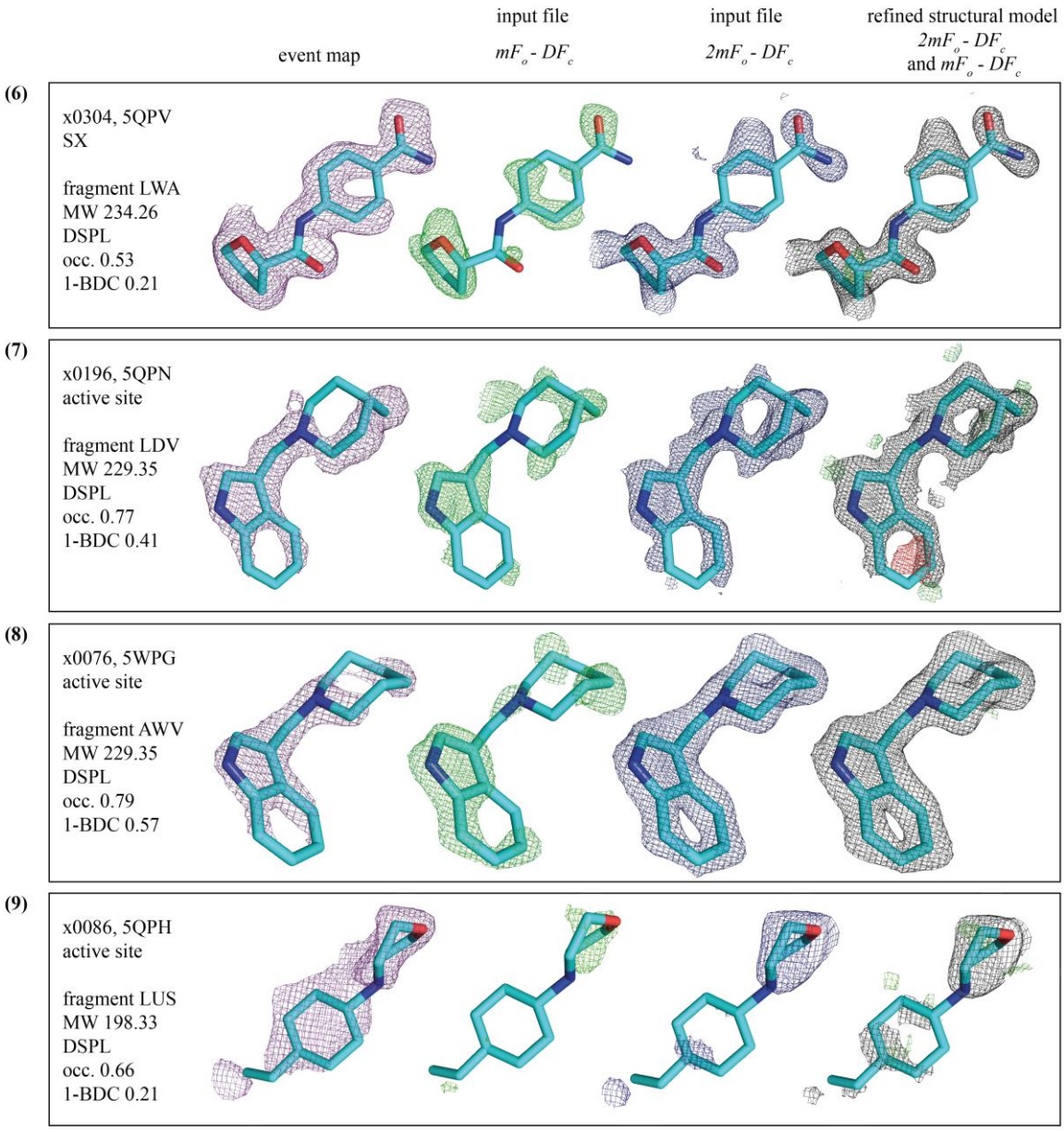



**Figure 75** continued.

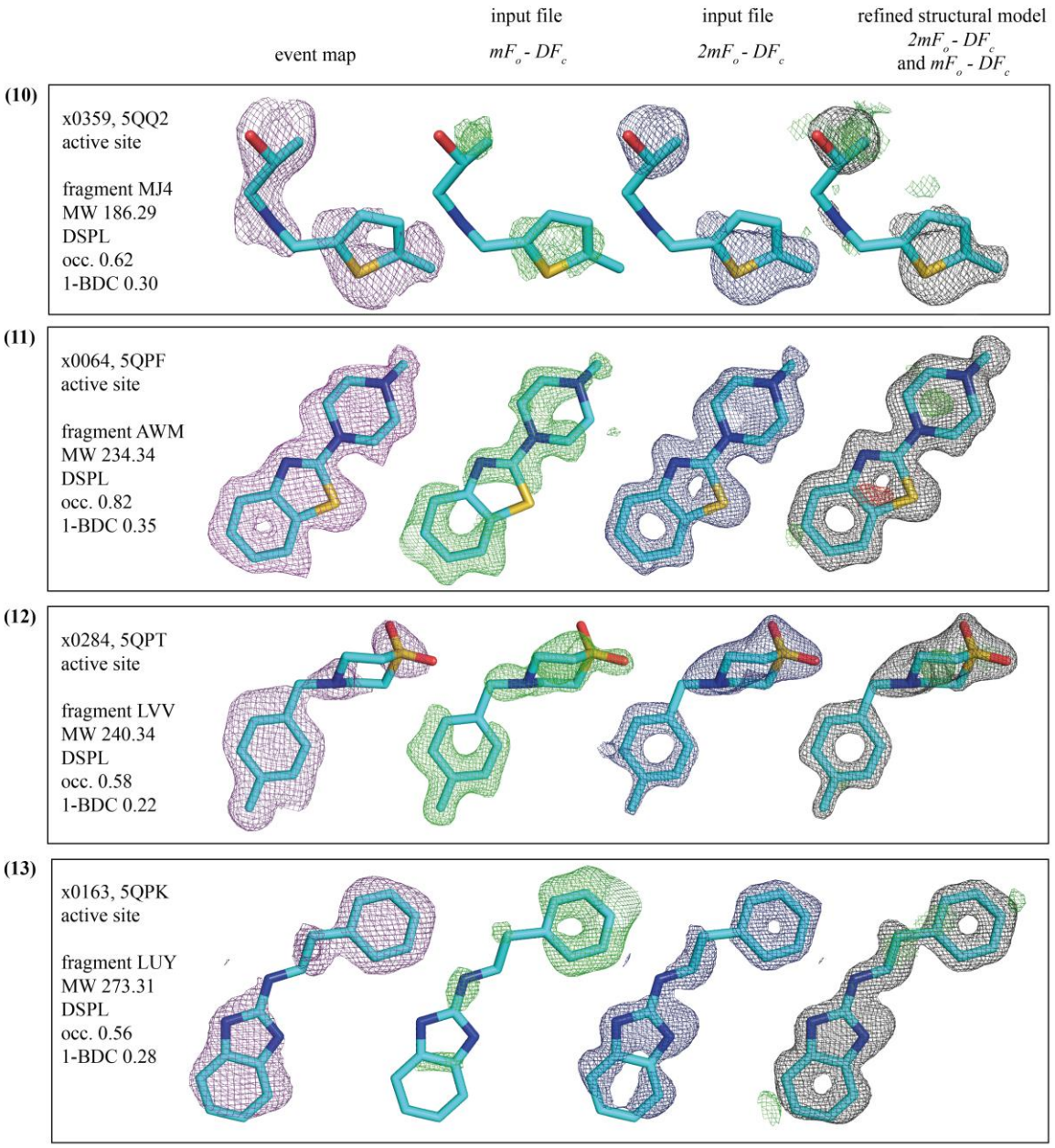



**Figure 75** continued.

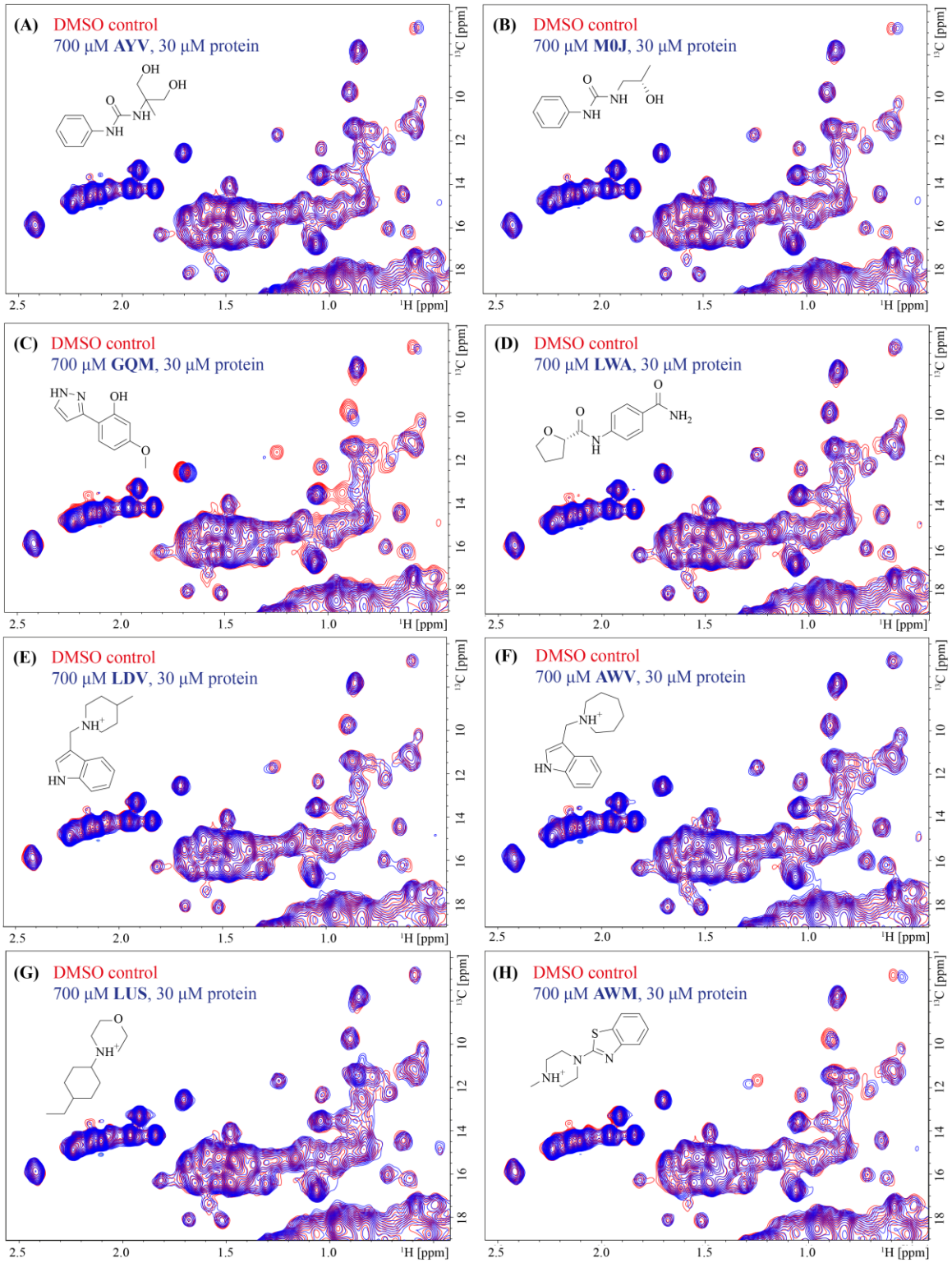



**Figure 76**: Spectra of ligand-observed NMR experiments with fragment hits identified by X-ray crystallography. **(A)** – **(J)** Compounds **AYY**, **M0J**, **GQM**, **LWA**, **LDV**, **AWV**, **LUS**, **AWM**, **LVV** and **LUY** respectively. Each image shows a cut-out from an overlay of the [$^{13}$C$^{1}$H]-SOFAST-HMQC spectrum of the DMSO control (red) and the corresponding sample (blue) at 700 µM compound and 30 µM protein in 25 mM BisTris, pH 6.5, 50 mM NaCl, 2 mM TCEP · HCl, 10% $D_2O$, 150 µM DSS at 31.85 °C. Image is continued on the next page.

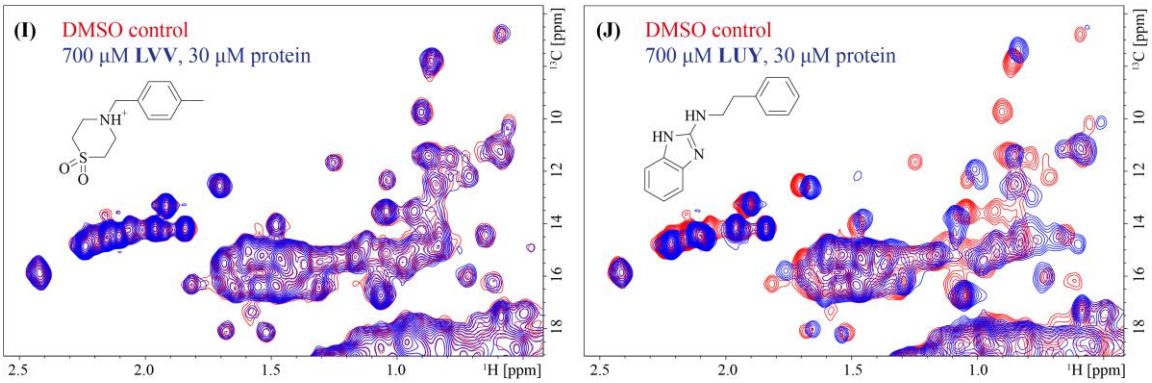



**Figure 76** continued.

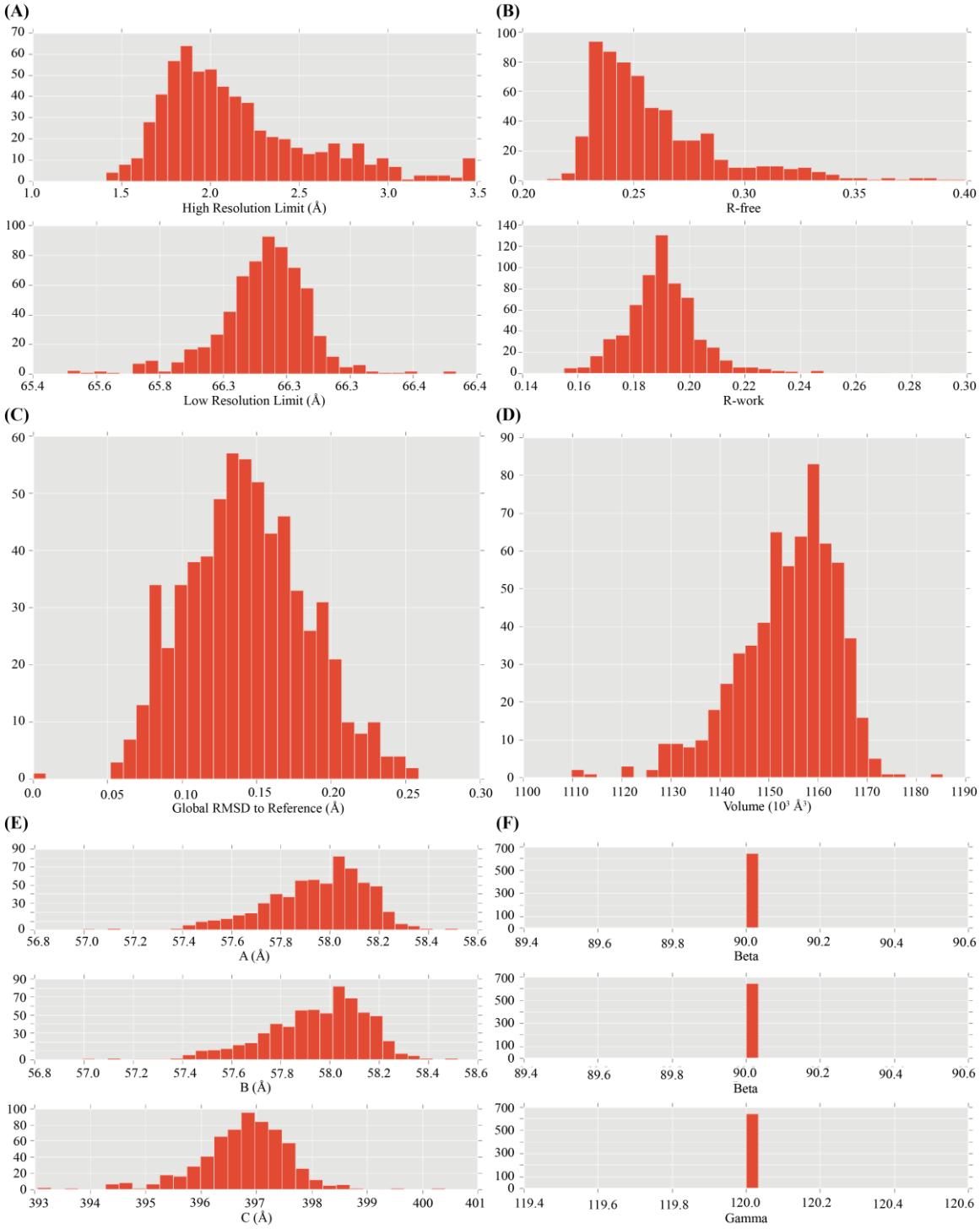



**Figure 77**: Datasets of the HTX campaign analysed with PanDDA. **(A)** Resolution limit. **(B)** R-free and R-work. **(C)** RMSD to reference structure. **(D)** Unit cell volume variation. **(E)** Cell axis variation. **(F)** Cell angle variation. On the count is plotted on the y-axis.

**Table 36**: Ligands identified in the FBS by X-ray crystallography at the HTX lab.

| Compound[a] | Smiles string | MW (Da) | Binding site | Crystal ID | Soak (hh:mm) | Diffraction limit (Å) | 1-BDC |
|---|---|---|---|---|---|---|---|
| **HTX-1** | Cn1cc(C(=O)N2CCc3sccc3C2)ccc1=O | 274.34 | allosteric, open | CRU-CD022463_F08-2_x1 | 23:32 | 2.10 | 0.21 |
| **HTX-2** | Nc1nnc(Cc2ccccc2)s1 | 191.25 | allosteric, open | CRU-CD022464_F10-3 | 26:11 | 1.76 | 0.22 |
| **HTX-3** | Cc1ccsc1-c1nnc(N)o1 | 118.21 | allosteric, open | CRU-CD022834_H11-2 | 21:44 | 1.86 | 0.20 |
| **HTX-4** | Cc1cccnc1NC(=O)NC1CCN(C2CC2)C1 | 260.34 | allosteric, close | CRU-CD022480_D03-2 | 22:20 | 1.71 | 0.21 |
| **HTX-5** | Cl.Fc1cccc(Cl)c1CC1CNC1 | 199.65 | allosteric, close | CRU-CD022779_C02-2 | 24:34 | 1.71 | 0.19 |
| **HTX-6** | c1ccc2oc(C3CCNCC3)nc2c1 | 202.26 | active | CRU-CD022464_B12-2 | 25:09 | 1.76 | 0.24 |
| **HTX-7** | c1ccc2sc(C3CCCN3)nc2c1 | 204.29 | active | CRU-CD022464_H12-2 | 26:11 | 1.81 | 0.15 |
| **HTX-8** | Nc1ccc(Oc2ccc(F)cc2)cn1 | 204.20 | active | CRU-CD022833_D02-2 | 29:15 | 1.81 | 0.20 |

[a] All listed compounds are part of the Enamine Golden Library.

**Table 37**: Data collection and refinement statistics of TcFPPS crystal structures from the HTX campaign.

| **PDB ID** | **6SI5** | **6SHV** |
|---|---|---|
| **compound**[a] | | |
| PDB identifier | **LEQ** | **LDW** |
| naming in this thesis | **HTX-1** | **HTX-8** |
| **Data collection** | | |
| X-ray source | id30b, ESRF | id30b, ESRF |
| Wavelength [Å] | 0.97625 Å | 0.97625 Å |
| Space group | $P6_122$ | $P6_122$ |
| Cell dimensions | | |
| *a* = *b*, *c* [Å] | 58.06, 396.79 | 58.21, 398.31 |
| α, β, γ [°] | 90, 90, 120 | 90, 90, 120 |
| Resolution [Å] | 29.31 – 2.10 | 44.98 – 1.81 |
| | (2.13– 2.10)[b] | (1.84 – 1.81)[b] |
| $R_{merge}$ | 0.172 (2.109)[b] | 0.099 (1.223)[b] |
| Unique reflections | 24665 (1185) | 38338 (1850) |
| $I / \sigma I$ | 16.6 (2.3)[b] | 15.1 (2.1)[b] |
| Completeness [%] | 99.7 (99.2)[b] | 100 (100)[b] |
| Redundancy | 35.9 (37.5)[b] | 12.3 (12.2)[b] |
| **Refinement** | | |
| Resolution [Å] | 66.13 – 2.10 | 44.98 – 1.80 |
| No. reflections | 24665 | 38338 |
| $R_{work}$ / $R_{free}$ | 0.200, 0.260 | 0.196, 0.230 |
| No. atoms | | |
| Protein | 2788 | 2801 |
| Ligand/ion | 29 | 36 |
| Water | 159 | 206 |
| B-factors overall [Å$^2$] | | 34.05 |
| R.m.s. deviations | | |
| Bond length [Å] | 0.01 | 0.01 |
| Bond angles [°] | 0.92 | 0.87 |
| Molprobity statistics | | |
| Ramachandran | | |
| Favoured [%] | 98.86 | 98.85 |
| Outliers [%] | 0.00 | 0.00 |
| Allowed [%] | 1.14 | 1.15 |
| Solvent content [%] | | 47.8 |

[a] Fragment binders were previously listed in **Table 36**.
[b] Values in parentheses are for the highest resolution shell.

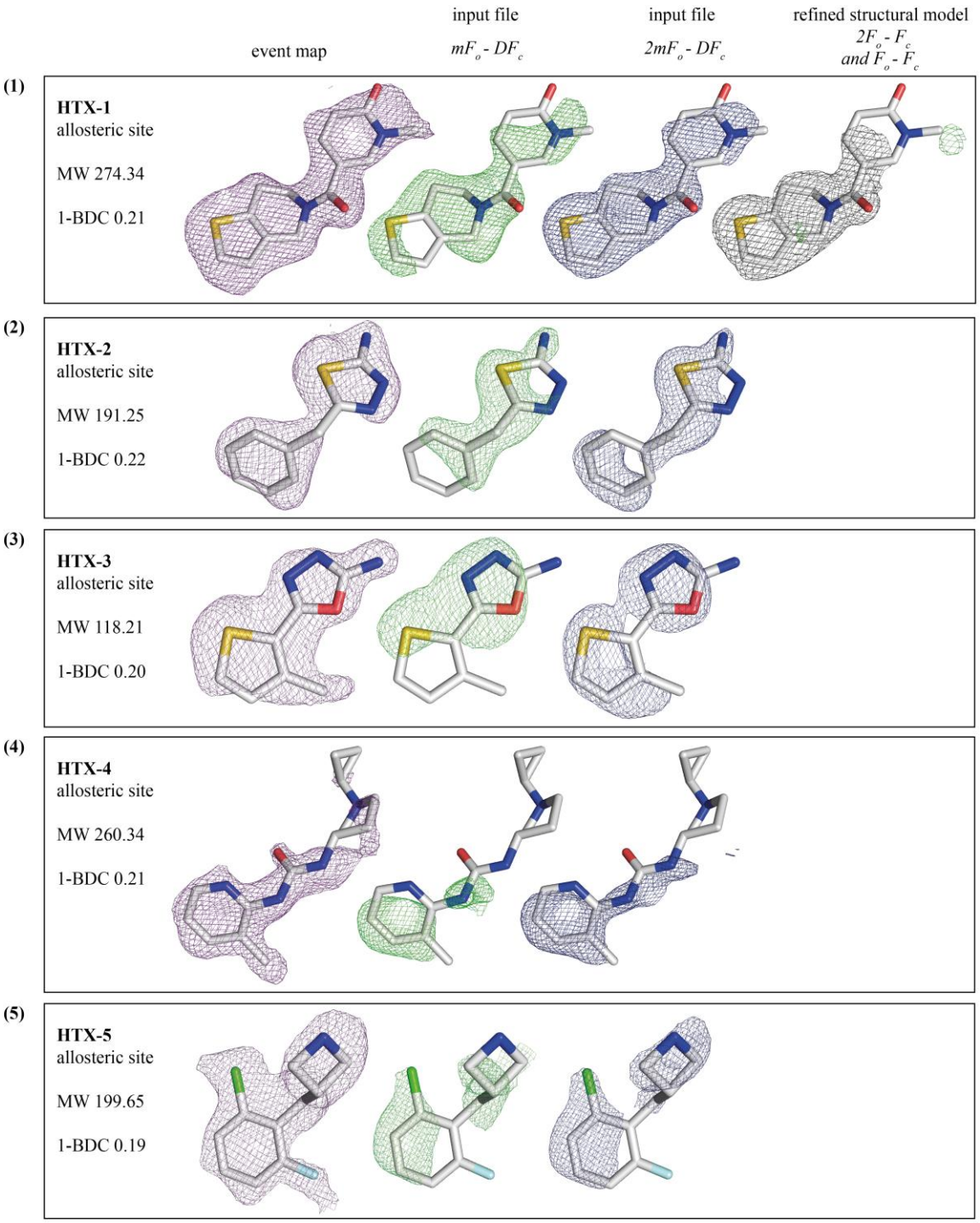



**Figure 78**: Summary of density maps of ligands identified in the HTX campaign: allosteric and active site binders of TcFPPS. **(1) – (5)** Maps of allosteric site binders **HTX-1** to **HTX-5**. **(6)** – **(8)** Maps of active site binders **HTX-6** to **HTX-8**, respectively. On the right site, the event maps are shown contoured at twice the 1-BDC value, which corresponds to a level of 2.0 σ. In the middle the $mF_o - DF_c$ difference electron density maps contoured at 3.0 σ and the $2mF_o - DF_c$ electron density maps contoured at 1.0 σ of the input files are shown. On the right site, the $2F_o - F_c$ electron density maps contoured at 1.0 σ of the final structural model of **HTX-1** and **HTX-8** is shown in superimposition with the $F_o - F_c$ difference electron density maps contoured at 3.0 σ, respectively. Figure is continued on the next pages.

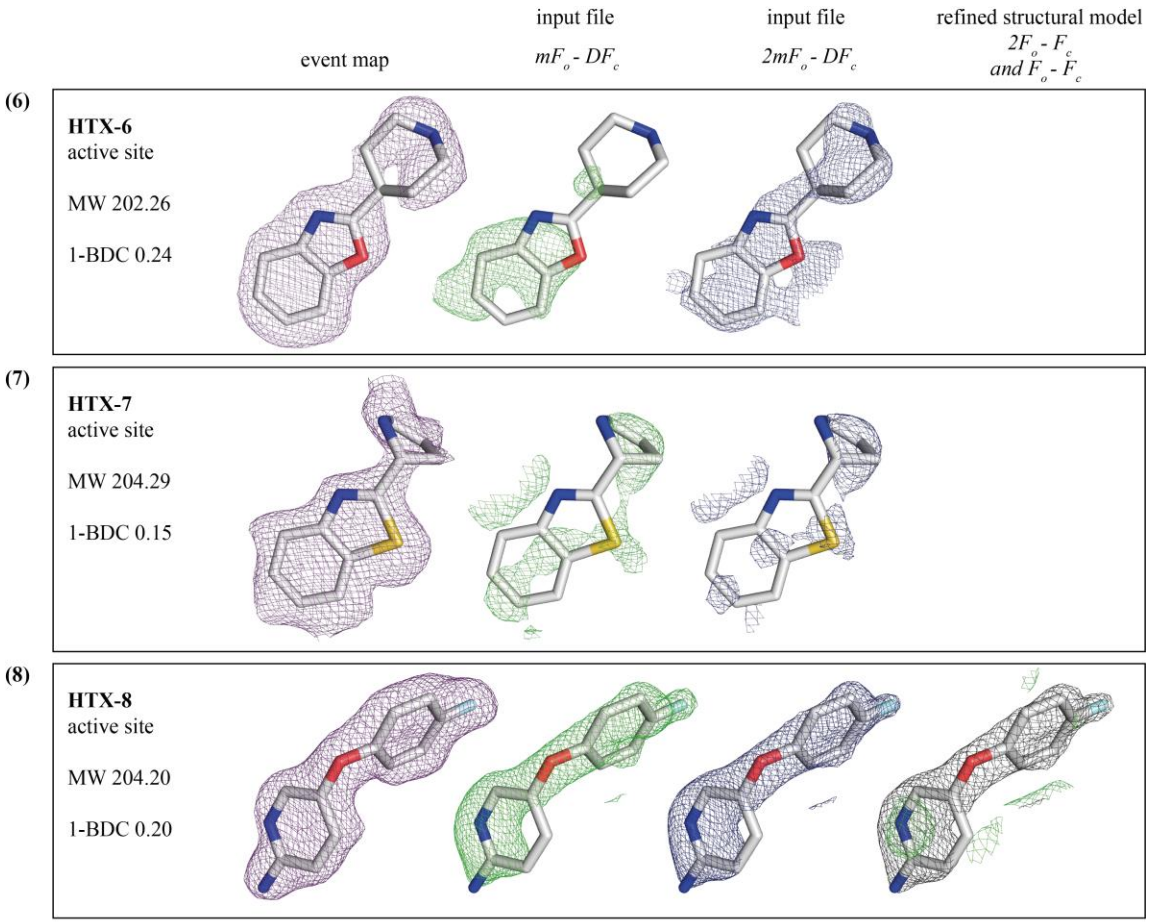



**Figure 78** continued.

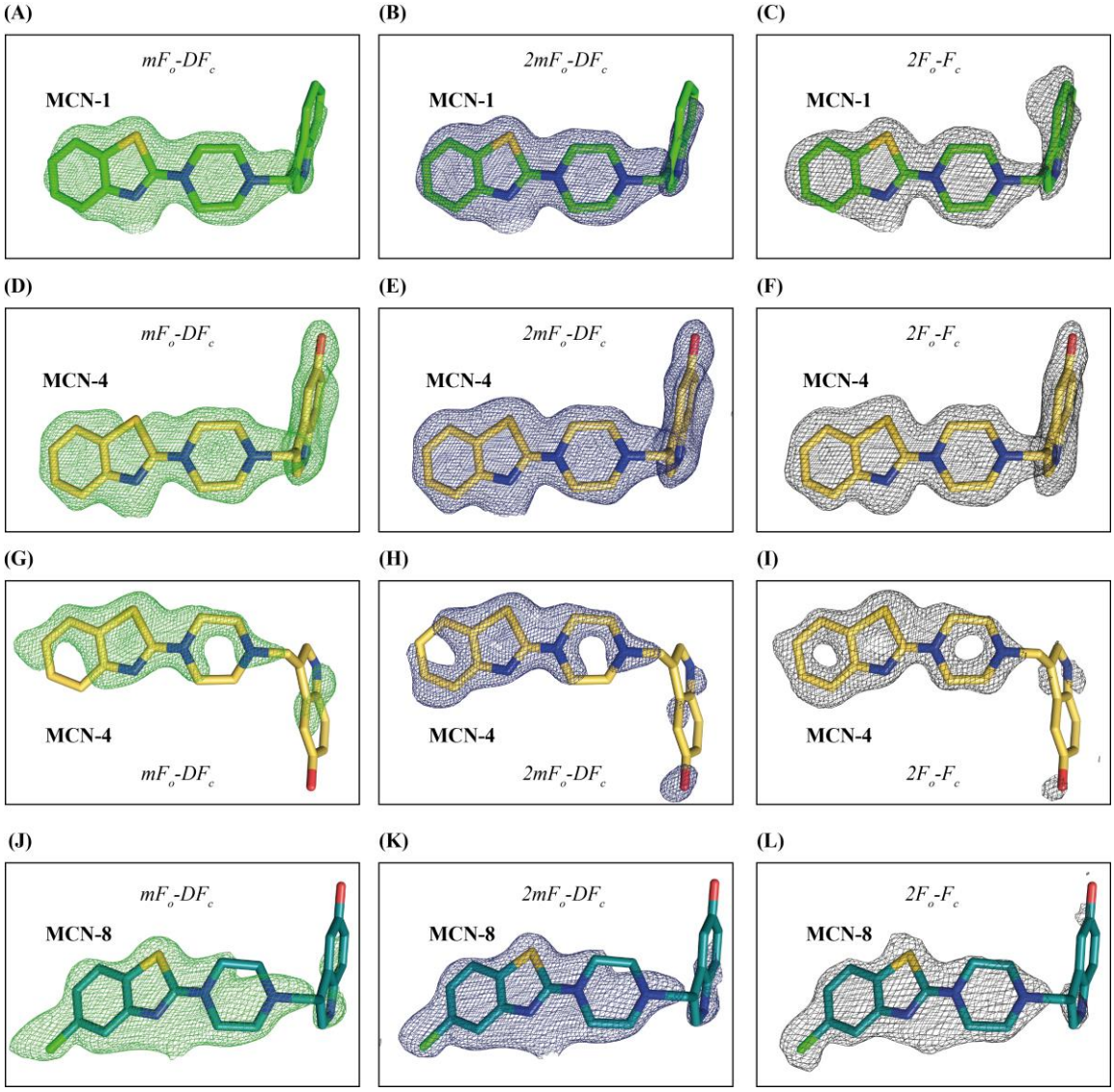



**Figure 79**: Summary of density maps of the ligands **MCN-1**, **MCN-4** and **MCN-8**. **(A) – (C)** Ligand **MCN-1** (**JMK**) in site S1 (PDB ID 6R09): $mF_o - DF_c$ difference electron desity map contoured at 3.0 σ, $2mF_o - DF_c$ electron density map contoured at 1.0 σ and $2F_o - F_c$ electron density map contoured at 1.0 σ, respectively. **(D)** – **(F)** Ligand **MCN-4** (**JMT**) in site S1 (PDB ID 6R0A): $mF_o - DF_c$ difference electron desity map contoured at 3.0 σ, $2mF_o - DF_c$ electron density map contoured at 1.00 σ and $2F_o - F_c$ electron density map contoured at 1.0 σ, respectively. **(G) – (H)** Ligand **MCN-4** (**JMT**) in site SY (PDB ID 6R0A): $mF_o - DF_c$ difference electron desity map contoured at 3.0 σ, $2mF_o - DF_c$ electron density map contoured at 1.0 σ and $2F_o - F_c$ electron density map contoured at 1.0 σ, respectively. **(J)** – **(L)** Ligand **MCN-8** (**JMW**) in site S1 (PDB ID 6R0B): $mF_o - DF_c$ difference electron desity map contoured at 3.0 σ, $2mF_o - DF_c$ electron density map contoured at 1.00 σ and $2F_o - F_c$ electron density map contoured at 1.0 σ, respectively.

**Table 38**: List of compounds of the CM74 series.

| No | Smiles string | MW | Number in Keenan *et al.*[490] |
|---|---|---|---|
| **DND*i*-1** | O=S(N(CC1=CN=CN1)C2=CC=C(C=C2)C3=NC(C=CC=C4)=C4S3)(C)=O | 384.47 | 26 |
| **DND*i*-2** | CC(N(CC1=CN=CN1)C2=CC=C(C=C2)C3=NC(C=CC=C4)=C4S3)=O | 348.42 | 24 |
| **DND*i*-3** | CCN(CC1=CN=CN1)C2=CC=C(C=C2)C3=NC(C=CC=C4)=C4S3 | 334.44 | 25 |
| **DND*i*-4** | O=C(N(CC1=CC=CN=C1)C2=CC=C(C=C2)C3=NC(C=CC=C4)=C4S3)C | 359.45 | 14 |
| **DND*i*-5** | CCN(CC1=CC=CN=C1)C2=CC=C(C=C2)C3=NC(C=CC=C4)=C4S3 | 345.46 | 15 |
| **DND*i*-6** | O=S(N(CC1=CC=CN=C1)C2=CC=C(C=C2)C3=NC(C=CC=C4)=C4S3)(C)=O | 395.50 | 16 |
| **DND*i*-7** | O=C(CN)N(CC1=CC=CN=C1)C2=CC=C(C=C2)C3=NC(C=CC=C4)=C4S3 | 374.46 | 17 |
| **DND*i*-8** | FC1=CC(C2=CC=C(F)C=C2)=CC=C1N(CC3=CN=CN3)C(C)=O | 327.33 | 27 |
| **DND*i*-9** | FC1=CC(C2=CC=C(F)C=C2)=CC=C1N(CC3=CN=CN3)CC | 313.35 | 26 |
| **DND*i*-10** | FC1=CC(C2=CC=C(F)C=C2)=CC=C1N(CC3=CN=CN3)S(C)(=O)=O | 363.38 | 29 |
| **DND*i*-11** | FC1=CC(C2=CC=C(F)C=C2)=CC=C1N(CC3=CC=CN=C3)C(C)=O | 338.36 | 18 |
| **DND*i*-12** | FC1=CC(C2=CC=C(F)C=C2)=CC=C1N(CC3=CC=CN=C3)CC | 324.37 | 19 |
| **DND*i*-13** | FC1=CC(C2=CC=C(F)C=C2)=CC=C1N(CC3=CC=CN=C3)S(C)(=O)=O | 374.41 | 20 |
| **DND*i*-14** | FC1=CC(C2=NC3=C(S2)C=CC=C3)=CC(N(CC4=CN=CN4)C(C)=O)=C1 | 366.41 | 21 |
| **DND*i*-15** | FC1=CC(C2=NC3=C(S2)C=CC=C3)=CC(N(CC4=CN=CN4)CC)=C1 | 352.43 | 22 |
| **DND*i*-16** | FC1=CC(C2=NC3=C(S2)C=CC=C3)=CC(N(CC4=CN=CN4)S(=O)(C)=O)=C1 | 402.46 | 23 |

# Literature


[1] a A. Rassi, Jr., A. Rassi, J. A. Marin-Neto, Chagas disease, *Lancet* **2010**, *375*, 1388-1402; b C. Chagas, Nova entidade morbida do homem: rezumo geral de estudos etiolojicos e clinicos, *Memórias do Instituto Oswaldo Cruz* **1911**, *3*, 219-275; c C. Chagas, Nova tripanozomiaze humana: estudos sobre a morfolojia e o ciclo evolutivo do Schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem, *Memórias do Instituto Oswaldo Cruz* **1909**, *1*, 159-218.

[2] World Health Organisation, *Chagas disease (American trypanosomiasis) Fact sheet*, http://www.who.int/mediacentre/factsheets/fs340/en/, accesses date 11. January 2019

[3] World Health Organisation, *Chagas disease (American trypanosomiasis) - Epidemiology*, https://www.who.int/chagas/epidemiology/en/, accesses date 14. January 2019

[4] K. M. Bonney, Chagas disease in the 21st century: a public health success or an emerging threat?, *Parasite* **2014**, *21*, 11.

[5] a C. J. Murray, T. Vos, R. Lozano, et al., Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010, *Lancet* **2012**, *380*, 2197-2223; b B. Y. Lee, K. M. Bacon, M. E. Bottazzi, et al., Global economic burden of Chagas disease: a computational simulation model, *Lancet Infect Dis* **2013**, *13*, 342-348.

[6] L. J. Robertson, B. Devleesschauwer, B. Alarcon de Noya, et al., Trypanosoma cruzi: Time for International Recognition as a Foodborne Parasite, *PLoS Negl Trop Dis* **2016**, *10*, e0004656.

[7] a J. D. Stanaway, G. Roth, The burden of Chagas disease: estimates and challenges, *Glob Heart* **2015**, *10*, 139-144; b C. Bern, Chagas' Disease, *N Engl J Med* **2015**, *373*, 1882; c R. Paucar, E. Moreno-Viguri, S. Perez-Silanes, Challenges in Chagas Disease Drug Discovery: A Review, *Curr Med Chem* **2016**, *23*, 3154-3170.

[8] A. Araujo, A. M. Jansen, K. Reinhard, et al., Paleoparasitology of Chagas disease--a review, *Mem Inst Oswaldo Cruz* **2009**, *104 Suppl 1*, 9-16.

[9] a A. C. Aufderheide, W. Salo, M. Madden, et al., A 9,000-year record of Chagas' disease, *Proc Natl Acad Sci U S A* **2004**, *101*, 2034-2039; b J. R. Coura, P. A. Vinas, Chagas disease: a new worldwide challenge, *Nature* **2010**, *465*, 6-7.

[10] I. Maudlin, M. C. Eisler, S. C. Welburn, Neglected and endemic zoonoses, *Philos Trans R Soc Lond B Biol Sci* **2009**, *364*, 2777-2787.

[11] C. Chagas, Aspecto clínico geral da nova entidade mórbida produzida pelo Schizotrypanum cruzi, *Brazil-Médico* **1910**, *24*, 263-265.

[12] E. Brumpt, Le Trypanosoma cruzi évolue chez Conorhinus megistus, Cimex lectularis, Cimex boueti et Ornithodorus moubata. Cycle évolutif de ce parasite. , *Bull Soc Pathol Exot* **1912**, *5*, 360–367.

[13] a M. A. Rajao, C. Furtado, C. L. Alves, et al., Unveiling benznidazole's mechanism of action through overexpression of DNA repair proteins in Trypanosoma cruzi, *Environ Mol Mutagen* **2014**, *55*, 309-321; b C. Bern, Antitrypanosomal therapy for chronic Chagas' disease, *N Engl J Med* **2011**, *364*, 2527-2534; c J. A. Urbina, R. Docampo, Specific chemotherapy of Chagas disease: controversies and advances, *Trends Parasitol* **2003**, *19*, 495-501.

[14] L. E. Villanueva-Lizama, J. V. Cruz-Chan, A. D. C. Aguilar-Cetina, et al., Trypanosoma cruzi vaccine candidate antigens Tc24 and TSA-1 recall memory immune response associated with HLA-A and -B supertypes in Chagasic chronic patients from Mexico, *PLoS Negl Trop Dis* **2018**, *12*, e0006240.

[15] a J. Alonso-Padilla, I. Cotillo, J. L. Presa, et al., Automated high-content assay for compounds selectively toxic to Trypanosoma cruzi in a myoblastic cell line, *PLoS Negl Trop Dis* **2015**, *9*, e0003493; b C. L. Bellera, M. L. Sbaraglini, D. E. Balcazar, et al., High-throughput drug repositioning for the discovery of new treatments for Chagas disease, *Mini Rev Med Chem* **2015**, *15*, 182-193.

[16] L. Stevens, P. L. Dorn, J. O. Schmidt, et al., Kissing bugs. The vectors of Chagas, *Adv Parasitol* **2011**, *75*, 169-192.

[17] S. Antinori, L. Galimberti, R. Bianco, et al., Chagas disease in Europe: A review for the internist in the globalized world, *Eur J Intern Med* **2017**, *43*, 6-15.

[18] Z. Brener, Biology of Trypanosoma cruzi, *Annu Rev Microbiol* **1973**, *27*, 347-382.

[19] a B. Zingales, S. G. Andrade, M. R. Briones, et al., A new consensus for Trypanosoma cruzi intraspecific nomenclature: second revision meeting recommends TcI to TcVI, *Mem Inst Oswaldo Cruz* **2009**, *104*, 1051-1054; b B. Zingales, M. A. Miles, D. A. Campbell, et al., The revised Trypanosoma cruzi subspecific nomenclature: rationale, epidemiological relevance and research applications, *Infect Genet Evol* **2012**, *12*, 240-253; c J. C. Ramirez, C. Torres, M. L. A. Curto, et al., New insights into Trypanosoma cruzi evolution, genotyping and molecular diagnostics from satellite DNA sequence analysis, *PLoS Negl Trop Dis* **2017**, *11*, e0006139; d A. Marcili, L. Lima, M. Cavazzana, et al., A new genotype of Trypanosoma cruzi associated

with bats evidenced by phylogenetic analyses using SSU rDNA, cytochrome b and Histone H2B genes and genotyping based on ITS1 rDNA, *Parasitology* **2009**, *136*, 641-655.

[20] a L. S. Filardi, Z. Brener, Susceptibility and natural resistance of Trypanosoma cruzi strains to drugs used clinically in Chagas disease, *Trans R Soc Trop Med Hyg* **1987**, *81*, 755-759; b R. A. Neal, J. van Bueren, Comparative studies of drug susceptibility of five strains of Trypanosoma cruzi in vivo and in vitro, *Trans R Soc Trop Med Hyg* **1988**, *82*, 709-714; c M. D. Lewis, M. S. Llewellyn, M. Yeo, et al., Recent, independent and anthropogenic origins of Trypanosoma cruzi hybrids, *PLoS Negl Trop Dis* **2011**, *5*, e1363.

[21] a P. M. Henrique, T. Marques, M. V. da Silva, et al., Correlation between the virulence of T. cruzi strains, complement regulatory protein expression levels, and the ability to elicit lytic antibody production, *Exp Parasitol* **2016**, *170*, 66-72; b A. M. Macedo, C. R. Machado, R. P. Oliveira, et al., Trypanosoma cruzi: genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease, *Mem Inst Oswaldo Cruz* **2004**, *99*, 1-12.

[22] K. L. Caradonna, B. A. Burleigh, Mechanisms of host cell invasion by Trypanosoma cruzi, *Adv Parasitol* **2011**, *76*, 33-61.

[23] R. O. Cosentino, F. Aguero, A simple strain typing assay for Trypanosoma cruzi: discrimination of major evolutionary lineages from a single amplification product, *PLoS Neglected Trop. Dis.* **2012**, *6*, e1777.

[24] F. S. Buckner, Experimental chemotherapy and approaches to drug discovery for Trypanosoma cruzi infection, *Adv Parasitol* **2011**, *75*, 89-119.

[25] a R. W. Costa, J. F. da Silveira, D. Bahia, Interactions between Trypanosoma cruzi Secreted Proteins and Host Cell Signaling Pathways, *Frontiers in Microbiology* **2016**, *7*, 388; b K. M. Tyler, D. M. Engman, The life cycle of Trypanosoma cruzi revisited, *Int J Parasitol* **2001**, *31*, 472-481.

[26] A. Rodriguez, E. Samoff, M. G. Rioult, et al., Host cell invasion by trypanosomes requires lysosomes and microtubule/kinesin-mediated transport, *J Cell Biol* **1996**, *134*, 349-362.

[27] J. C. Engel, P. S. Doyle, J. A. Dvorak, Trypanosoma cruzi: biological characterization of clones derived from chronic chagasic patients. II. Quantitative analysis of the intracellular cycle, *J Protozool* **1985**, *32*, 80-83.

[28] a B. A. Burleigh, A. M. Woolsey, Cell signalling and Trypanosoma cruzi invasion, *Cell Microbiol* **2002**, *4*, 701-711; b N. Gil-Jaramillo, I. M. D. Bastos, J. M. Santana, et al., Dendritic Cells: A Double-Edged Sword in Immune Responses during Chagas Disease, *Front Microbiol* **2016**, *7*, 1076; c A. F. Nardy, C. G. Freire-de-Lima, A. Morrot, Immune Evasion Strategies of Trypanosoma cruzi, *J Immunol Res* **2015**, *2015*, 178947.

[29] A. M. Canavaci, J. M. Bustamante, A. M. Padilla, et al., In vitro and in vivo high-throughput assays for the testing of anti-Trypanosoma cruzi compounds, *PLoS Negl Trop Dis* **2010**, *4*, e740.

[30] B. Zingales, M. A. Miles, C. B. Moraes, et al., Drug discovery for Chagas disease should consider Trypanosoma cruzi strain diversity, *Memorias Do Instituto Oswaldo Cruz* **2014**, *109*, 828-833.

[31] J. A. Perez-Molina, I. Molina, Chagas disease, *Lancet* **2018**, *391*, 82-94.

[32] O. K. Giddings, C. S. Eickhoff, T. J. Smith, et al., Anatomical route of invasion and protective mucosal immunity in Trypanosoma cruzi conjunctival infection, *Infect Immun* **2006**, *74*, 5549-5560.

[33] A. M. Cevallos, R. Hernandez, Chagas' disease: pregnancy and congenital transmission, *Biomed Res Int* **2014**, *2014*, 401864.

[34] a V. Nussenzweig, R. Sonntag, A. Biancalana, et al., Effect of triphenylmethane dyes on Trypanosoma cruzi in vitro; use of gentian violet in prevention of transmission of Chagas disease by blood transfusion, *Hospital (Rio J)* **1953**, *44*, 731-744; b G. A. Schmunis, Trypanosoma cruzi, the etiologic agent of Chagas' disease: status in the blood supply in endemic and nonendemic countries, *Transfusion* **1991**, *31*, 547-557; c R. J. Benjamin, S. L. Stramer, D. A. Leiby, et al., Trypanosoma cruzi infection in North America and Spain: evidence in support of transfusion transmission, *Transfusion* **2012**, *52*, 1913-1921; quiz 1912.

[35] B. L. Herwaldt, Laboratory-acquired parasitic infections from accidental exposures, *Clin Microbiol Rev* **2001**, *14*, 659-688, table of contents.

[36] D. V. Andrade, K. J. Gollob, W. O. Dutra, Acute chagas disease: new global challenges for an old neglected disease, *PLoS Negl Trop Dis* **2014**, *8*, e3010.

[37] J. Bermudez, C. Davies, A. Simonazzi, et al., Current drug therapy and pharmaceutical challenges for Chagas disease, *Acta Trop.* **2016**, *156*, 1-16.

[38] C. J. Bastos, R. Aras, G. Mota, et al., Clinical outcomes of thirteen patients with acute chagas disease acquired through oral transmission from two urban outbreaks in northeastern Brazil, *PLoS Negl Trop Dis* **2010**, *4*, e711.

[39] M. L. Calvo-Mendez, B. Nogueda-Torres, R. Alejandre-Aguilar, et al., [Experimental Trypanosoma cruzi infection via contaminated water and food], *Rev Latinoam Microbiol* **1994**, *36*, 67-69.

[40] C. Santos Ferreira, V. Amato Neto, E. Gakiya, et al., Microwave treatment of human milk to prevent transmission of Chagas disease, *Rev Inst Med Trop Sao Paulo* **2003**, *45*, 41-42.

[41] R. L. Barbosa, V. L. Dias, K. S. Pereira, et al., Survival in vitro and virulence of Trypanosoma cruzi in acai pulp in experimental acute Chagas disease, *J Food Prot* **2012**, *75*, 601-606.

[42] E. J. Howard, X. Xiong, Y. Carlier, et al., Frequency of the congenital transmission of Trypanosoma cruzi: a systematic review and meta-analysis, *BJOG* **2014**, *121*, 22-33.

[43] F. F. Norman, R. Lopez-Velez, Mother-to-child transmission of Trypanosoma cruzi infection (Chagas disease): a neglected problem, *Trans R Soc Trop Med Hyg* **2014**, *108*, 388-390.

[44] Y. Carlier, S. Sosa-Estani, A. O. Luquetti, et al., Congenital Chagas disease: an update, *Mem Inst Oswaldo Cruz* **2015**, *110*, 363-368.
[45] J. H. Diaz, Recognizing and reducing the risks of Chagas disease (American trypanosomiasis) in travelers, *J Travel Med* **2008**, *15*, 184-195.
[46] L. Zhang, R. L. Tarleton, Parasite persistence correlates with disease severity and localization in chronic Chagas' disease, *J Infect Dis* **1999**, *180*, 480-486.
[47] a K. M. Bonney, D. M. Engman, Autoimmune pathogenesis of Chagas heart disease: looking back, looking ahead, *Am J Pathol* **2015**, *185*, 1537-1547; b F. R. Gutierrez, P. M. Guedes, R. T. Gazzinelli, et al., The role of parasite persistence in pathogenesis of Chagas heart disease, *Parasite Immunol* **2009**, *31*, 673-685.
[48] A. R. Teixeira, N. Nitz, M. C. Guimaro, et al., Chagas disease, *Postgrad Med J* **2006**, *82*, 788-798.
[49] a X. Deng, E. C. Sabino, E. Cunha-Neto, et al., Genome wide association study (GWAS) of Chagas cardiomyopathy in Trypanosoma cruzi seropositive subjects, *PLoS One* **2013**, *8*, e79629; b F. del Puerto, J. E. Nishizawa, M. Kikuchi, et al., Protective human leucocyte antigen haplotype, HLA-DRB1*01-B*14, against chronic Chagas disease in Bolivia, *PLoS Negl Trop Dis* **2012**, *6*, e1587.
[50] M. J. Pinazo, G. Espinosa, C. Cortes-Lletget, et al., Immunosuppression and Chagas disease: a management challenge, *PLoS Negl Trop Dis* **2013**, *7*, e1965.
[51] World Health Organisation, *Chagas disease (American trypanosomiasis) - More on Chagas disease diagnosis*, https://www.who.int/chagas/disease/home_diagnosis_more/en/, accesses date 11. Januar 2019
[52] Y. Jackson, C. Myers, A. Diana, et al., Congenital transmission of Chagas disease in Latin American immigrants in Switzerland, *Emerg Infect Dis* **2009**, *15*, 601-603.
[53] M. Flores-Chavez, B. Fernandez, S. Puente, et al., Transfusional chagas disease: parasitological and serological monitoring of an infected recipient and blood donor, *Clin Infect Dis* **2008**, *46*, e44-47.
[54] a M. Piron, R. Fisa, N. Casamitjana, et al., Development of a real-time PCR assay for Trypanosoma cruzi detection in blood samples, *Acta Trop* **2007**, *103*, 195-200; b A. G. Schijman, M. Bisio, L. Orellana, et al., International study to evaluate PCR methods for detection of Trypanosoma cruzi DNA in blood samples from Chagas disease patients, *PLoS Negl Trop Dis* **2011**, *5*, e931.
[55] J. C. Ramirez, C. I. Cura, O. da Cruz Moreira, et al., Analytical Validation of Quantitative Real-Time PCR Methods for Quantification of Trypanosoma cruzi DNA in Blood Samples from Chagas Disease Patients, *J Mol Diagn* **2015**, *17*, 605-615.
[56] Z. Moure, A. Angheben, I. Molina, et al., Serodiscordance in chronic Chagas disease diagnosis: a real problem in non-endemic countries, *Clin Microbiol Infect* **2016**, *22*, 788-792.
[57] H. B. Tanowitz, P. E. Scherer, M. M. Mota, et al., Adipose Tissue: A Safe Haven for Parasites?, *Trends Parasitol* **2016**.
[58] Y. Sguassero, C. B. Cuesta, K. N. Roberts, et al., Course of Chronic Trypanosoma cruzi Infection after Treatment Based on Parasitological and Serological Tests: A Systematic Review of Follow-Up Studies, *PLoS One* **2015**, *10*, e0139363.
[59] Y. Sguassero, K. N. Roberts, G. B. Harvey, et al., Course of serological tests in treated subjects with chronic Trypanosoma cruzi infection: A systematic review and meta-analysis of individual participant data, *Int J Infect Dis* **2018**, *73*, 93-101.
[60] E. Granjon, M. L. Dichtel-Danjoy, E. Saba, et al., Development of a Novel Multiplex Immunoassay Multi-cruzi for the Serological Confirmation of Chagas Disease, *PLoS Negl Trop Dis* **2016**, *10*, e0004596.
[61] M. Zrein, E. Granjon, L. Gueyffier, et al., A novel antibody surrogate biomarker to monitor parasite persistence in Trypanosoma cruzi-infected patients, *PLoS Negl Trop Dis* **2018**, *12*, e0006226.
[62] C. Santamaria, E. Chatelain, Y. Jackson, et al., Serum biomarkers predictive of cure in Chagas disease patients after nifurtimox treatment, *BMC Infect Dis* **2014**, *14*, 302.
[63] M. E. Pereira, L. M. Santos, M. S. Araujo, et al., Recrudescence induced by cyclophosphamide of chronic Trypanosoma cruzi infection in mice is influenced by the parasite strain, *Mem Inst Oswaldo Cruz* **1996**, *91*, 71-74.
[64] F. M. Dos Santos, S. Caldas, S. B. de Assis Cau, et al., Trypanosoma cruzi: Induction of benznidazole resistance in vivo and its modulation by in vitro culturing and mice infection, *Exp Parasitol* **2008**, *120*, 385-390.
[65] J. M. Bustamante, L. M. Bixby, R. L. Tarleton, Drug-induced cure drives conversion to a stable and protective CD8+ T central memory response in chronic Chagas disease, *Nat Med* **2008**, *14*, 542-550.
[66] M. J. Pinazo, M. C. Thomas, J. Bustamante, et al., Biomarkers of therapeutic responses in chronic Chagas disease: state of the art and future perspectives, *Mem Inst Oswaldo Cruz* **2015**, *110*, 422-432.
[67] C. B. Moraes, C. H. Franco, Novel drug discovery for Chagas disease, *Expert Opin Drug Discov* **2016**, *11*, 447-455.
[68] a J. C. Pinto Dias, J. R. Rodrigues Coura, M. A. Shikanai Yasuda, The present situation, challenges, and perspectives regarding the production and utilization of effective drugs against human Chagas disease, *Rev Soc Bras Med Trop* **2014**, *47*, 123-125; b J. A. Urbina, Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches, *Acta Trop* **2010**, *115*, 55-68; cWorld Health Organisation, *Chagas disease (American trypanosomiasis) - Treatment of Chagas disease*, https://www.who.int/chagas/disease/home_treatment/en/, accesses date 11. January 2019; d K. Salomao, R.

F. Menna-Barreto, S. L. de Castro, Stairway to Heaven or Hell? Perspectives and Limitations of Chagas Disease Chemotherapy, *Curr Top Med Chem* **2016**, *16*, 2266-2289.

[69] a A. Trochine, D. J. Creek, P. Faral-Tello, et al., Benznidazole biotransformation and multiple targets in Trypanosoma cruzi revealed by metabolomics, *PLoS Negl Trop Dis* **2014**, *8*, e2844; b S. R. Wilkinson, J. M. Kelly, Trypanocidal drugs: mechanisms, resistance and new targets, *Expert Rev Mol Med* **2009**, *11*, e31; c M. Bock, A. Haberkorn, H. Herlinger, et al., The structure-activity relationship of 4-(5'-nitrofurfurylidene-amino)-tetrahydro-4H-1,4-thiazine-1,1-dioxides active against Trypanosoma cruzi, *Arzneimittelforschung* **1972**, *22*, 1564-1569; d J. D. Maya, B. K. Cassels, P. Iturriaga-Vasquez, et al., Mode of action of natural and synthetic drugs against Trypanosoma cruzi and their interaction with the mammalian host, *Comp Biochem Physiol A Mol Integr Physiol* **2007**, *146*, 601-620.

[70] C. Bern, S. P. Montgomery, B. L. Herwaldt, et al., Evaluation and treatment of chagas disease in the United States: a systematic review, *JAMA* **2007**, *298*, 2171-2181.

[71] M. I. Walton, P. Workman, Nitroimidazole bioreductive metabolism. Quantitation and characterisation of mouse tissue benznidazole nitroreductases in vivo and in vitro, *Biochem Pharmacol* **1987**, *36*, 887-896.

[72] U. A. Boelsterli, H. K. Ho, S. Zhou, et al., Bioactivation and hepatotoxicity of nitroaromatic drugs, *Curr Drug Metab* **2006**, *7*, 715-727.

[73] a C. A. Morillo, J. A. Marin-Neto, A. Avezum, et al., Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy, *N Engl J Med* **2015**, *373*, 1295-1306; b J. A. Castro, M. M. de Mecca, L. C. Bartel, Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis), *Hum Exp Toxicol* **2006**, *25*, 471-479.

[74] M. Castillo-Riquelme, Chagas disease in non-endemic countries, *Lancet Glob Health* **2017**, *5*, e379-e380.

[75] J. Alonso-Padilla, A. Rodriguez, High throughput screening for anti-Trypanosoma cruzi drug discovery, *PLoS Negl Trop Dis* **2014**, *8*, e3259.

[76] B. Pecoul, C. Batista, E. Stobbaerts, et al., The BENEFIT Trial: Where Do We Go from Here?, *PLoS Negl Trop Dis* **2016**, *10*, e0004343.

[77] R. Viotti, B. Alarcon de Noya, T. Araujo-Jorge, et al., Towards a paradigm shift in the treatment of chronic Chagas disease, *Antimicrob Agents Chemother* **2014**, *58*, 635-639.

[78] E. L. Camandaroba, E. A. Reis, M. S. Goncalves, et al., Trypanosoma cruzi: susceptibility to chemotherapy with benznidazole of clones isolated from the highly resistant Colombian strain, *Rev Soc Bras Med Trop* **2003**, *36*, 201-209.

[79] a B. Zingales, R. G. Araujo, M. Moreno, et al., A novel ABCG-like transporter of Trypanosoma cruzi is involved in natural resistance to benznidazole, *Mem Inst Oswaldo Cruz* **2015**, *110*, 433-444; b J. Franco, R. C. Ferreira, S. Ienne, et al., ABCG-like transporter of Trypanosoma cruzi involved in benznidazole resistance: gene polymorphisms disclose inter-strain intragenic recombination in hybrid isolates, *Infect Genet Evol* **2015**, *31*, 198-208.

[80] M. C. O. Campos, L. L. Leon, M. C. Taylor, et al., "Benznidazole-resistance in Trypanosoma cruzi: Evidence that distinct mechanisms can act in concert" [Mol. Biochem. Parasit. (2014) 193, 17-19], *Mol Biochem Parasitol* **2015**, *201*, 83.

[81] I. Quijano-Hernandez, E. Dumonteil, Advances and challenges towards a vaccine against Chagas disease, *Hum Vaccin* **2011**, *7*, 1184-1191.

[82] R. L. Tarleton, R. E. Gurtler, J. A. Urbina, et al., Chagas disease and the london declaration on neglected tropical diseases, *PLoS Negl Trop Dis* **2014**, *8*, e3219.

[83] G. Mougabure-Cueto, M. I. Picollo, Insecticide resistance in vector Chagas disease: evolution, mechanisms and management, *Acta Trop* **2015**, *149*, 70-85.

[84] M. D. Germano, P. Santo-Orihuela, G. Roca-Acevedo, et al., Scientific evidence of three different insecticide-resistant profiles in Triatoma infestans (Hemiptera: Reduviidae) populations from Argentina and Bolivia, *J Med Entomol* **2012**, *49*, 1355-1360.

[85] P. L. Santo-Orihuela, C. V. Vassena, E. N. Zerba, et al., Relative contribution of monooxygenase and esterase to pyrethroid resistance in Triatoma infestans (Hemiptera: Reduviidae) from Argentina and Bolivia, *J Med Entomol* **2008**, *45*, 298-306.

[86] P. L. Santo-Orihuela, G. Carvajal, M. I. Picollo, et al., Toxicological and biochemical analysis of the susceptibility of sylvatic Triatoma infestans from the Andean Valley of Bolivia to organophosphate insecticide, *Mem Inst Oswaldo Cruz* **2013**, *108*, 790-795.

[87] a J. A. Perez-Molina, F. Norman, R. Lopez-Velez, Chagas disease in non-endemic countries: epidemiology, clinical presentation and treatment, *Curr Infect Dis Rep* **2012**, *14*, 263-274; b Y. Jackson, A. Pinto, S. Pett, Chagas disease in Australia and New Zealand: risks and needs for public health interventions, *Trop Med Int Health* **2014**, *19*, 212-218.

[88] a J. Gascon, C. Bern, M. J. Pinazo, Chagas disease in Spain, the United States and other non-endemic countries, *Acta Trop* **2010**, *115*, 22-27; b E. M. Shelly, R. Acuna-Soto, K. C. Ernst, et al., A Critical Assessment of Officially Reported Chagas Disease Surveillance Data in Mexico, *Public Health Rep* **2016**, *131*, 59-66.

[89] L. Basile, J. M. Jansa, Y. Carlier, et al., Chagas disease in European countries: the challenge of a surveillance system, *Euro Surveill* **2011**, *16*.

[90] A. Soriano-Arandes, N. Serre-Delcor, B. Trevino-Maruri, et al., Control and management of congenital Chagas disease in Europe and other non-endemic countries: current policies and practices, *Trop Med Int Health* **2016**, *21*, 590-596.
[91] A. Requena-Mendez, S. Bussion, E. Aldasoro, et al., Cost-effectiveness of Chagas disease screening in Latin American migrants at primary health-care centres in Europe: a Markov model analysis, *Lancet Glob Health* **2017**, *5*, e439-e447.
[92] J. Manne-Goehler, C. A. Umeh, S. P. Montgomery, et al., Estimating the Burden of Chagas Disease in the United States, *PLoS Negl Trop Dis* **2016**, *10*, e0005033.
[93] P. J. Hotez, The rise of neglected tropical diseases in the "new Texas", *PLoS Negl Trop Dis* **2018**, *12*, e0005581.
[94] a M. N. Garcia, L. Woc-Colburn, S. N. Rossmann, et al., Trypanosoma cruzi screening in Texas blood donors, 2008-2012, *Epidemiol Infect* **2016**, *144*, 1010-1013; b M. N. Garcia, P. J. Hotez, K. O. Murray, Potential novel risk factors for autochthonous and sylvatic transmission of human Chagas disease in the United States, *Parasit Vectors* **2014**, *7*, 311; c S. Meymandi, S. Hernandez, S. Park, et al., Treatment of Chagas Disease in the United States, *Curr Treat Options Infect Dis* **2018**, *10*, 373-388.
[95] E. Chatelain, J. R. Ioset, Drug discovery and development for neglected diseases: the DNDi model, *Drug Des Devel Ther* **2011**, *5*, 175-181.
[96] a P. Trouiller, E. Torreele, P. Olliaro, et al., Drugs for neglected diseases: a failure of the market and a public health failure?, *Trop Med Int Health* **2001**, *6*, 945-951; b P. Chirac, E. Torreele, Global framework on essential health R&D, *Lancet* **2006**, *367*, 1560-1561.
[97] M. Moran, J. Guzman, A. L. Ropars, et al., Neglected disease research and development: how much are we really spending?, *PLoS Med* **2009**, *6*, e30.
[98] L. G. Ferreira, M. T. de Oliveira, A. D. Andricopulo, Advances and Progress in Chagas Disease Drug Discovery, *Curr. Top. Med. Chem. (Sharjah, United Arab Emirates)* **2016**, *16*, 2290-2302.
[99] a C. B. Scarim, D. H. Jornada, R. C. Chelucci, et al., Current advances in drug discovery for Chagas disease, *Eur J Med Chem* **2018**, *155*, 824-838; b S. Nwaka, A. Hudson, Innovative lead discovery strategies for tropical diseases, *Nat Rev Drug Discov* **2006**, *5*, 941-955.
[100] V. J. Haupt, J. E. Aguilar Uvalle, S. Salentin, et al., Computational Drug Repositioning by Target Hopping: A Use Case in Chagas Disease, *Curr Pharm Des* **2016**, *22*, 3124-3134.
[101] a M. C. Field, D. Horn, A. H. Fairlamb, et al., Anti-trypanosomatid drug discovery: an ongoing challenge and a continuing need, *Nat Rev Microbiol* **2017**, *15*, 447; b M. Kaiser, L. Maes, L. P. Tadoori, et al., Repurposing of the Open Access Malaria Box for Kinetoplastid Diseases Identifies Novel Active Scaffolds against Trypanosomatids, *J Biomol Screen* **2015**, *20*, 634-645; c S. N. Deftereos, C. Andronis, E. J. Friedla, et al., Drug repurposing and adverse event prediction using high-throughput literature analysis, *Wiley Interdiscip Rev Syst Biol Med* **2011**, *3*, 323-334; d M. H. Gelb, W. C. Van Voorhis, F. S. Buckner, et al., Protein farnesyl and N-myristoyl transferases: piggy-back medicinal chemistry targets for the development of antitrypanosomatid and antimalarial therapeutics, *Mol Biochem Parasitol* **2003**, *126*, 155-163.
[102] M. L. Sykes, V. M. Avery, Approaches to protozoan drug discovery: phenotypic screening, *Journal of medicinal chemistry* **2013**, *56*, 7727-7740.
[103] I. H. Gilbert, D. Leroy, J. A. Frearson, Finding new hits in neglected disease projects: target or phenotypic based screening?, *Curr Top Med Chem* **2011**, *11*, 1284-1291.
[104] J. B. Rodriguez, B. N. Falcone, S. H. Szajnman, Detection and treatment of Trypanosoma cruzi: a patent review (2011-2015), *Expert Opin Ther Pat* **2016**, *26*, 993-1015.
[105] Drugs for Neglected Disease initiative DNDi, *Chagas disease target product profile*, https://www.dndi.org/diseases-projects/chagas/chagas-target-product-profile/, accesses date 16. January 2019
[106] a I. C. Ribeiro, E., Research and developement landscape in Chagas disease. , *9th Congress on Tropical Medicine and International Health* **2015**; b K. Katsuno, J. N. Burrows, K. Duncan, et al., Hit and lead criteria in drug discovery for infectious diseases of the developing world, *Nat Rev Drug Discov* **2015**, *14*, 751-758.
[107] C. A. Lipinski, F. Lombardo, B. W. Dominy, et al., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Advanced Drug Delivery Reviews* **1997**, *23*, 3-25.
[108] D. F. Veber, S. R. Johnson, H. Y. Cheng, et al., Molecular properties that influence the oral bioavailability of drug candidates, *Journal of medicinal chemistry* **2002**, *45*, 2615-2623.
[109] D. Law, C. B. Moore, D. W. Denning, Activity of SCH 56592 compared with those of fluconazole and itraconazole against Candida spp, *Antimicrob Agents Chemother* **1997**, *41*, 2310-2311.
[110] Y. Ueda, J. D. Matiskella, J. Golik, et al., Phosphonooxymethyl prodrugs of the broad spectrum antifungal azole, ravuconazole: synthesis and biological properties, *Bioorg Med Chem Lett* **2003**, *13*, 3669-3672.
[111] a F. S. Buckner, N. Navabi, Advances in Chagas disease drug development: 2009-2010, *Curr Opin Infect Dis* **2010**, *23*, 609-616; b J. Clayton, Chagas disease: pushing through the pipeline, *Nature* **2010**, *465*, S12; c J. A. Urbina, Ergosterol biosynthesis and drug development for Chagas disease, *Mem Inst Oswaldo Cruz* **2009**, *104 Suppl 1*, 311-318; d C. W. Roberts, R. McLeod, D. W. Rice, et al., Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa, *Mol Biochem Parasitol* **2003**, *126*, 129-142.

[112] M. J. Pinazo, G. Espinosa, M. Gallego, et al., Successful treatment with posaconazole of a patient with chronic Chagas disease and systemic lupus erythematosus, *Am J Trop Med Hyg* **2010**, *82*, 583-587.
[113] I. Molina, J. Gomez i Prat, F. Salvador, et al., Randomized trial of posaconazole and benznidazole for chronic Chagas' disease, *N Engl J Med* **2014**, *370*, 1899-1908.
[114] C. A. Morillo, H. Waskin, S. Sosa-Estani, et al., Benznidazole and Posaconazole in Eliminating Parasites in Asymptomatic T. Cruzi Carriers: The STOP-CHAGAS Trial, *J Am Coll Cardiol* **2017**, *69*, 939-947.
[115] A. F. Francisco, M. D. Lewis, S. Jayawardhana, et al., Limited Ability of Posaconazole To Cure both Acute and Chronic Trypanosoma cruzi Infections Revealed by Highly Sensitive In Vivo Imaging, *Antimicrob Agents Chemother* **2015**, *59*, 4653-4661.
[116] F. Torrico, J. Gascon, L. Ortiz, et al., Treatment of adult chronic indeterminate Chagas disease with benznidazole and three E1224 dosing regimens: a proof-of-concept, randomised, placebo-controlled trial, *Lancet Infect Dis* **2018**, *18*, 419-430.
[117] L. F. Diniz, A. L. Mazzeti, I. S. Caldas, et al., Outcome of E1224-Benznidazole Combination Treatment for Infection with a Multidrug-Resistant Trypanosoma cruzi Strain in Mice, *Antimicrob Agents Chemother* **2018**, *62*, e00401-00418.
[118] ClinicalTrials.gov U.S. National Library of Medicine, *BENDITA BEnznidazole New Doses Improved Treatment and Associations (BENDITA)*, https://clinicaltrials.gov/ct2/show/NCT03378661?term=chagas+dndi, accesses date 15. January 2019
[119] a J. R. Coura, J. Borges-Pereira, Chagas disease: 100 years after its discovery. A systemic review, *Acta Trop* **2010**, *115*, 5-13; b I. Ribeiro, A. M. Sevcsik, F. Alves, et al., New, improved treatments for Chagas disease: from the R&D pipeline to the patients, *PLoS Negl Trop Dis* **2009**, *3*, e484.
[120] J. M. Bustamante, J. M. Craft, B. D. Crowe, et al., New, combined, and reduced dosing treatment protocols cure Trypanosoma cruzi infection in mice, *J Infect Dis* **2014**, *209*, 150-162.
[121] Drugs for Neglected Disease *initiative* DND*i*, *Fexinidazole (HAT)*, https://www.dndi.org/diseases-projects/portfolio/fexinidazole/, accesses date 16. January 2019
[122] W. Raether, H. Seidenath, The activity of fexinidazole (HOE 239) against experimental infections with Trypanosoma cruzi, trichomonads and Entamoeba histolytica, *Ann Trop Med Parasitol* **1983**, *77*, 13-26.
[123] M. T. Bahia, I. M. de Andrade, T. A. Martins, et al., Fexinidazole: a potential new drug candidate for Chagas disease, *PLoS Negl Trop Dis* **2012**, *6*, e1870.
[124] a ClinicalTrials.gov U.S. National Library of Medicine, *Study to Evaluate Fexinidazole Dosing Regimens for the Treatment of Adult Patients With Chagas Disease*, https://www.clinicaltrials.gov/ct2/show/record/NCT02498782, accesses date 16. January 2019; b ClinicalTrials.gov U.S. National Library of Medicine, *Oral Fexinidazole Dosing Regimens for the Treatment of Adults With Chronic Indeterminate Chagas Disease (FEXI12)*, https://clinicaltrials.gov/ct2/show/record/NCT03587766, accesses date 16. January 2019
[125] ClinicalTrials.gov U.S. National Library of Medicine, *Optimization of PCR Technique to Assess Parasitological Response for Patients With Chronic Chagas Disease (PCR)*, https://clinicaltrials.gov/ct2/show/NCT01678599, accesses date 15. January 2019
[126] ClinicalTrials.gov U.S. National Library of Medicine, *BENEFIT: Evaluation of the Use of Antiparasital Drug (Benznidazole) in the Treatment of Chronic Chagas' Disease*, https://clinicaltrials.gov/ct2/show/NCT00123916, accesses date 15. January 2019
[127] ClinicalTrials.gov U.S. National Library of Medicine, *Etiologic Treatment With Benznidazole in Adult Patients With Chronic Chagas Disease. A Randomized Clinical Trial (TRAENA)*, https://clinicaltrials.gov/ct2/show/NCT02386358, accesses date 15. January 2019
[128] J. M. Bustamante, R. L. Tarleton, Potential new clinical therapies for Chagas disease, *Expert Rev Clin Pharmacol* **2014**, *7*, 317-325.
[129] F. A. Botoni, P. A. Poole-Wilson, A. L. Ribeiro, et al., A randomized trial of carvedilol after renin-angiotensin system inhibition in chronic Chagas cardiomyopathy, *Am Heart J* **2007**, *153*, 544.e541-548.
[130] F. R. Quiros, C. A. Morillo, J. P. Casas, et al., CHARITY: Chagas cardiomyopathy bisoprolol intervention study: a randomized double-blind placebo force-titration controlled study with Bisoprolol in patients with chronic heart failure secondary to Chagas cardiomyopathy [NCT00323973], *Trials* **2006**, *7*, 21.
[131] E. Althoff, K. Kontzalis, Novartis Media Relations, Novartis joins the Global Chagas Disease Coalition and also announces first multinational, prospective, randomized study in people with chronic Chagas cardiomyopathy [Press release]. Retrieved from https://www.novartis.com/news/media-releases/novartis-joins-global-chagas-disease-coalition-and-also-announces-first-multinational-prospective-randomized-study-people-chronic-chagas-cardiomyopathy, **15. March 2019**.
[132] N. M. El-Sayed, P. J. Myler, D. C. Bartholomeu, et al., The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease, *Science* **2005**, *309*, 409-415.
[133] F. S. Buckner, C. L. Verlinde, A. C. La Flamme, et al., Efficient technique for screening drugs for activity against Trypanosoma cruzi using parasites expressing beta-galactosidase, *Antimicrob Agents Chemother* **1996**, *40*, 2592-2597.
[134] C. Bot, B. S. Hall, N. Bashir, et al., Trypanocidal activity of aziridinyl nitrobenzamide prodrugs, *Antimicrob Agents Chemother* **2010**, *54*, 4246-4252.

[135] K. V. Hyland, S. H. Asfaw, C. L. Olson, et al., Bioluminescent imaging of Trypanosoma cruzi infection, *Int J Parasitol* **2008**, *38*, 1391-1400.
[136] a F. Zanella, J. B. Lorens, W. Link, High content screening: seeing is believing, *Trends in Biotechnology* **2010**, *28*, 237-245; b P. Brodin, T. Christophe, High-content screening in infectious diseases, *Curr Opin Chem Biol* **2011**, *15*, 534-539; c I. H. Gilbert, Drug discovery for neglected diseases: molecular target-based and phenotypic approaches, *Journal of medicinal chemistry* **2013**, *56*, 7719-7726.
[137] E. Bettiol, M. Samanovic, A. S. Murkin, et al., Identification of three classes of heteroaromatic compounds with activity against intracellular Trypanosoma cruzi by chemical library screening, *PLoS Negl Trop Dis* **2009**, *3*, e384.
[138] a L. A. Jelicks, H. B. Tanowitz, Advances in imaging of animal models of Chagas disease, *Adv Parasitol* **2011**, *75*, 193-208; b G. Andriani, A. D. Chessler, G. Courtemanche, et al., Activity in vivo of anti-Trypanosoma cruzi compounds selected from a high throughput screening, *PLoS Negl Trop Dis* **2011**, *5*, e1298.
[139] L. L. Nohara, C. Lema, J. O. Bader, et al., High-content imaging for automated determination of host-cell infection rate by the intracellular parasite Trypanosoma cruzi, *Parasitol Int* **2010**, *59*, 565-570.
[140] a L. C. Carmody, A. Germain, D. Barker, et al., *Identification of Small-Molecule Inhibitors of Trypansoma cruzi Infection. 2010 Mar 25*, https://www.ncbi.nlm.nih.gov/books/NBK55071/, accesses date 04.03; b J. C. Engel, K. K. Ang, S. Chen, et al., Image-based high-throughput drug screening targeting the intracellular stage of Trypanosoma cruzi, the agent of Chagas' disease, *Antimicrob Agents Chemother* **2010**, *54*, 3326-3334.
[141] a S. Moon, J. L. Siqueira-Neto, C. B. Moraes, et al., An image-based algorithm for precise and accurate high throughput assessment of drug activity against the human parasite Trypanosoma cruzi, *PLoS One* **2014**, *9*, e87188; b R. J. Neitz, S. Chen, F. Supek, et al., Lead identification to clinical candidate selection: drugs for Chagas disease, *J Biomol Screen* **2015**, *20*, 101-111; c I. Pena, M. Pilar Manzano, J. Cantizani, et al., New compound sets identified from high throughput phenotypic screening against three kinetoplastid parasites: an open resource, *Sci Rep* **2015**, *5*, 8771.
[142] a S. Khare, S. L. Roach, S. W. Barnes, et al., Utilizing Chemical Genomics to Identify Cytochrome b as a Novel Drug Target for Chagas Disease, *PLoS Pathog* **2015**, *11*, e1005058; b J. Choi, N. M. El-Sayed, Functional genomics of trypanosomatids, *Parasite Immunol* **2012**, *34*, 72-79; c B. Lomenick, R. Hao, N. Jonai, et al., Target identification using drug affinity responsive target stability (DARTS), *Proc Natl Acad Sci U S A* **2009**, *106*, 21984-21989.
[143] A. R. Vago, L. O. Andrade, A. A. Leite, et al., Genetic characterization of Trypanosoma cruzi directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs, *Am J Pathol* **2000**, *156*, 1805-1809.
[144] S. Brand, E. J. Ko, E. Viayna, et al., Discovery and Optimization of 5-Amino-1,2,3-triazole-4-carboxamide Series against Trypanosoma cruzi, *Journal of medicinal chemistry* **2017**, *60*, 7284-7299.
[145] S. Khare, A. S. Nagle, A. Biggart, et al., Proteasome inhibition for treatment of leishmaniasis, Chagas disease and sleeping sickness, *Nature* **2016**, *537*, 229-233.
[146] G. Alvarez, J. Varela, P. Marquez, et al., Optimization of antitrypanosomatid agents: identification of nonmutagenic drug candidates with in vivo activity, *Journal of medicinal chemistry* **2014**, *57*, 3984-3999.
[147] T. Spangenberg, J. N. Burrows, P. Kowalczyk, et al., The open access malaria box: a drug discovery catalyst for neglected diseases, *PLoS One* **2013**, *8*, e62906.
[148] W. Devine, J. L. Woodring, U. Swaminathan, et al., Protozoan Parasite Growth Inhibitors Discovered by Cross-Screening Yield Potent Scaffolds for Lead Discovery, *Journal of medicinal chemistry* **2015**, *58*, 5522-5537.
[149] a C. Fonseca-Berzal, A. Ibanez-Escribano, F. Reviriego, et al., Antichagasic and trichomonacidal activity of 1-substituted 2-benzyl-5-nitroindazolin-3-ones and 3-alkoxy-2-benzyl-5-nitro-2H-indazoles, *Eur J Med Chem* **2016**, *115*, 295-310; b C. Fonseca-Berzal, A. Ibanez-Escribano, N. Vela, et al., Antichagasic, Leishmanicidal, and Trichomonacidal Activity of 2-Benzyl-5-nitroindazole-Derived Amines, *ChemMedChem* **2018**, Ahead of Print.
[150] P. de Andrade, O. A. Galo, M. R. Carvalho, et al., 1,2,3-Triazole-based analogue of benznidazole displays remarkable activity against Trypanosoma cruzi, *Bioorg Med Chem* **2015**, *23*, 6815-6826.
[151] F. T. Silva, C. H. Franco, D. C. Favaro, et al., Design, synthesis and antitrypanosomal activity of some nitrofurazone 1,2,4-triazolic bioisosteric analogues, *Eur J Med Chem* **2016**, *121*, 553-560.
[152] a D. da Silva Ferreira, V. R. Esperandim, M. P. Toldo, et al., In vivo activity of ursolic and oleanolic acids during the acute phase of Trypanosoma cruzi infection, *Exp Parasitol* **2013**, *134*, 455-459; b W. R. Cunha, E. J. Crevelin, G. M. Arantes, et al., A study of the trypanocidal activity of triterpene acids isolated from Miconia species, *Phytother Res* **2006**, *20*, 474-478.
[153] E. C. Vargas de Oliveira, Z. A. Carneiro, S. de Albuquerque, et al., Development and Evaluation of a Nanoemulsion Containing Ursolic Acid: a Promising Trypanocidal Agent : Nanoemulsion with Ursolic Acid Against T. cruzi, *AAPS PharmSciTech* **2017**, *18*, 2551-2560.
[154] M. N. Soeiro, K. Werbovetz, D. W. Boykin, et al., Novel amidines and analogues as promising agents against intracellular parasites: a systematic review, *Parasitology* **2013**, *140*, 929-951.

[155] C. C. Santos, J. R. Lionel, R. B. Peres, et al., In Vitro, In Silico, and In Vivo Analyses of Novel Aromatic Amidines against Trypanosoma cruzi, *Antimicrob Agents Chemother* **2018**, *62*.
[156] a I. Coppens, P. J. Courtoy, The mevalonate pathway in parasitic protozoa and helminths, *Exp Parasitol* **1996**, *82*, 76-85; b V. G. Duschak, Targets and Patented Drugs for Chemotherapy of Chagas Disease in the Last 15 Years-Period, *Recent Pat Antiinfect Drug Discov* **2016**, *11*, 74-173; c W. de Souza, J. C. Rodrigues, Sterol Biosynthesis Pathway as Target for Anti-trypanosomatid Drugs, *Interdiscip Perspect Infect Dis* **2009**, *2009*, 642502.
[157] R. Cosentino, F. Agüero, *Genetic Profiling of the Isoprenoid and Sterol Biosynthesis Pathway Genes of Trypanosoma cruzi*, *Vol. 9*, **2014**.
[158] A. Liendo, G. Visbal, M. M. Piras, et al., Sterol composition and biosynthesis in Trypanosoma cruzi amastigotes, *Mol Biochem Parasitol* **1999**, *104*, 81-91.
[159] G. I. Lepesheva, T. Y. Hargrove, S. Anderson, et al., Structural insights into inhibition of sterol 14alpha-demethylase in the human pathogen Trypanosoma cruzi, *J Biol Chem* **2010**, *285*, 25582-25590.
[160] M. Florin-Christensen, J. Florin-Christensen, C. Garin, et al., Inhibition of Trypanosoma cruzi growth and sterol biosynthesis by lovastatin, *Biochem Biophys Res Commun* **1990**, *166*, 1441-1445.
[161] a E. R. Ferreira, E. Horjales, A. Bonfim-Melo, et al., Unique behavior of Trypanosoma cruzi mevalonate kinase: A conserved glycosomal enzyme involved in host cell invasion and signaling, *Sci. Rep.* **2016**, *6*, 24610; b I. Buhaescu, H. Izzedine, Mevalonate pathway: a review of clinical and therapeutical implications, *Clin Biochem* **2007**, *40*, 575-584; c D. D. Hinson, K. L. Chambliss, M. J. Toth, et al., Post-translational regulation of mevalonate kinase by intermediates of the cholesterol and nonsterol isoprene biosynthetic pathways, *J Lipid Res* **1997**, *38*, 2216-2223.
[162] a M. Ferrer-Casal, C. Li, M. Galizzi, et al., New insights into molecular recognition of 1,1-bisphosphonic acids by farnesyl diphosphate synthase, *Bioorg Med Chem* **2014**, *22*, 398-405; b S. B. Gabelli, J. S. McLellan, A. Montalvetti, et al., Structure and mechanism of the farnesyl diphosphate synthase from Trypanosoma cruzi: implications for drug design, *Proteins* **2006**, *62*, 80-88.
[163] a N. Shang, Q. Li, T. P. Ko, et al., Squalene synthase as a target for Chagas disease therapeutics, *PLoS Pathog* **2014**, *10*, e1004114; b M. Sealey-Cardona, S. Cammerer, S. Jones, et al., Kinetic characterization of squalene synthase from Trypanosoma cruzi: selective inhibition by quinuclidine derivatives, *Antimicrob Agents Chemother* **2007**, *51*, 2123-2129.
[164] G. J. Noguera, L. E. Fabian, E. Lombardo, et al., Studies of 4-arylthiazolylhydrazones derived from 1-indanones as Trypanosoma cruzi squalene epoxidase inhibitors by molecular simulations, *Org Biomol Chem* **2018**, *16*, 8525-8536.
[165] S. Lange, M. Keller, C. Muller, et al., Aminopropylindenes derived from Grundmann's ketone as a novel chemotype of oxidosqualene cyclase inhibitors, *Eur J Med Chem* **2013**, *63*, 758-764.
[166] L. Gros, S. O. Lorente, C. J. Jimenez, et al., Evaluation of azasterols as anti-parasitics, *Journal of medicinal chemistry* **2006**, *49*, 6094-6103.
[167] M. K. Dhar, A. Koul, S. Kaul, Farnesyl pyrophosphate synthase: a key enzyme in isoprenoid biosynthetic pathway and potential molecular target for drug development, *New Biotechnol.* **2013**, *30*, 114-123.
[168] J. M. Rondeau, F. Bitsch, E. Bourgier, et al., Structural basis for the exceptional in vivo efficacy of bisphosphonate drugs, *ChemMedChem* **2006**, *1*, 267-273.
[169] a C. D. Poulter, Farnesyl Diphosphate Synthase. A Paradigm for Understanding Structure and Function Relationships in E-polyprenyl Diphosphate Synthases, *Phytochemistry Reviews* **2006**, *5*, 17-26; b B. M. Lange, T. Rujan, W. Martin, et al., Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes, *Proc Natl Acad Sci U S A* **2000**, *97*, 13172-13177.
[170] a S. Aripirala, S. H. Szajnman, J. Jakoncic, et al., Design, synthesis, calorimetry, and crystallographic analysis of 2-alkylaminoethyl-1,1-bisphosphonates as inhibitors of Trypanosoma cruzi farnesyl diphosphate synthase, *Journal of medicinal chemistry* **2012**, *55*, 6445-6454; b R. Docampo, S. N. Moreno, The acidocalcisome as a target for chemotherapeutic agents in protozoan parasites, *Curr Pharm Des* **2008**, *14*, 882-888.
[171] a L. P. Wright, M. R. Philips, Thematic review series: lipid posttranslational modifications. CAAX modification and membrane targeting of Ras, *J Lipid Res* **2006**, *47*, 883-891; b S. P. Luckman, D. E. Hughes, F. P. Coxon, et al., Nitrogen-containing bisphosphonates inhibit the mevalonate pathway and prevent post-translational prenylation of GTP-binding proteins, including Ras, *J Bone Miner Res* **1998**, *13*, 581-589; c J. E. Fisher, M. J. Rogers, J. M. Halasy, et al., Alendronate mechanism of action: geranylgeraniol, an intermediate in the mevalonate pathway, prevents inhibition of osteoclast formation, bone resorption, and kinase activation in vitro, *Proc Natl Acad Sci U S A* **1999**, *96*, 133-138.
[172] a M. I. Esteva, K. Kettler, C. Maidana, et al., Benzophenone-based farnesyltransferase inhibitors with high activity against Trypanosoma cruzi, *Journal of medicinal chemistry* **2005**, *48*, 7186-7191; b Y. Yang, D. K. Chakravorty, K. M. Merz, Jr., Finding a needle in the haystack: computational modeling of Mg2+ binding in the active site of protein farnesyltransferase, *Biochemistry* **2010**, *49*, 9658-9666.
[173] a K. Yokoyama, J. R. Gillespie, W. C. Van Voorhis, et al., Protein geranylgeranyltransferase-I of Trypanosoma cruzi, *Mol Biochem Parasitol* **2008**, *157*, 32-43; b C. M. Szabo, Y. Matsumura, S. Fukura, et

al., Inhibition of geranylgeranyl diphosphate synthase by bisphosphonates and diphosphates: a potential route to new bone antiresorption and antiparasitic agents, *Journal of medicinal chemistry* **2002**, *45*, 2185-2196.
[174] K. Yokoyama, P. Trobridge, F. S. Buckner, et al., The effects of protein farnesyltransferase inhibitors on trypanosomatids: inhibition of protein farnesylation and cell growth, *Mol Biochem Parasitol* **1998**, *94*, 87-97.
[175] S. Ramkumar, A. Raghunath, S. Raghunath, Statin Therapy: Review of Safety and Potential Side Effects, *Acta Cardiol Sin* **2016**, *32*, 631-639.
[176] F. H. Guedes da Silva, D. Batista, C. F. Da Silva, et al., Successful Aspects of the Co-administration of Sterol 14alpha-Demethylase Inhibitor VFV and Benznidazole in Experimental Mouse Models of Chagas Disease Caused by the Drug-Resistant Strain of Trypanosoma cruzi, *ACS Infect Dis* **2019**.
[177] a M. Sajid, S. A. Robertson, L. S. Brinen, et al., Cruzain : the path from target validation to the clinic, *Adv Exp Med Biol* **2011**, *712*, 100-115; b J. C. Engel, P. S. Doyle, I. Hsieh, et al., Cysteine protease inhibitors cure an experimental Trypanosoma cruzi infection, *J Exp Med* **1998**, *188*, 725-734; c J. Cazzulo, V. Stoka, V. Turk, The major cysteine proteinase of Trypanosoma cruzi: a valid target for chemotherapy of Chagas disease, *Curr Pharm Des* **2001**, *7*, 1143-1156.
[178] a P. S. Doyle, Y. M. Zhou, J. C. Engel, et al., A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection, *Antimicrob Agents Chemother* **2007**, *51*, 3932-3939; b D. A. Otta, F. F. de Araujo, V. B. de Rezende, et al., Identification of Anti-Trypanosoma cruzi Lead Compounds with Putative Immunomodulatory Activity, *Antimicrob Agents Chemother* **2018**, *62*.
[179] Drugs for Neglected Disease initiative DNDi, *K777 (CHAGAS)*, https://www.dndi.org/diseases-projects/portfolio/completed-projects/k777/, accesses date 16. January 2019
[180] a M. P. Hudock, C. E. Sanz-Rodriguez, Y. Song, et al., Inhibition of Trypanosoma cruzi hexokinase by bisphosphonates, *Journal of medicinal chemistry* **2006**, *49*, 215-223; b C. E. Sanz-Rodriguez, J. L. Concepcion, S. Pekerar, et al., Bisphosphonates as inhibitors of Trypanosoma cruzi hexokinase: kinetic and metabolic studies, *J Biol Chem* **2007**, *282*, 12377-12387.
[181] C. L. Verlinde, V. Hannaert, C. Blonski, et al., Glycolysis as a target for the design of new anti-trypanosome drugs, *Drug Resist Updat* **2001**, *4*, 50-65.
[182] R. F. Freitas, I. M. Prokopczyk, A. Zottis, et al., Discovery of novel Trypanosoma cruzi glyceraldehyde-3-phosphate dehydrogenase inhibitors, *Bioorg Med Chem* **2009**, *17*, 2476-2482.
[183] A. A. Zuma, D. P. Cavalcanti, M. C. Maia, et al., Effect of topoisomerase inhibitors and DNA-binding drugs on the cell proliferation and ultrastructure of Trypanosoma cruzi, *Int J Antimicrob Agents* **2011**, *37*, 449-456.
[184] a J. D. Maya, C. O. Salas, B. Aguilera-Venegas, et al., Key proteins in the polyamine-trypanothione pathway as drug targets against Trypanosoma cruzi, *Curr Med Chem* **2014**, *21*, 1757-1771; b K. Vazquez, M. Paulino, C. O. Salas, et al., Trypanothione Reductase: A Target for the Development of Anti- Trypanosoma cruzi Drugs, *Mini Rev Med Chem* **2017**, *17*, 939-946.
[185] a B. R. Miller, 3rd, A. E. Roitberg, Trypanosoma cruzi trans-sialidase as a drug target against Chagas disease (American trypanosomiasis), *Future Med Chem* **2013**, *5*, 1889-1900; b A. F. Nardy, C. G. Freire-de-Lima, A. R. Perez, et al., Role of Trypanosoma cruzi Trans-sialidase on the Escape from Host Immune Surveillance, *Front Microbiol* **2016**, *7*, 348.
[186] C. O. Ogindo, M. H. Khraiwesh, M. George, Jr., et al., Novel drug design for Chagas disease via targeting Trypanosoma cruzi tubulin: Homology modeling and binding pocket prediction on Trypanosoma cruzi tubulin polymerization inhibition by naphthoquinone derivatives, *Bioorg Med Chem* **2016**, *24*, 3849-3855.
[187] G. Benaim, V. Hernandez-Rodriguez, S. Mujica-Gonzalez, et al., In vitro anti-Trypanosoma cruzi activity of dronedarone, a novel amiodarone derivative with an improved safety profile, *Antimicrob Agents Chemother* **2012**, *56*, 3720-3725.
[188] C. L. Bellera, D. E. Balcazar, L. Alberca, et al., Application of computer-aided drug repurposing in the search of new cruzipain inhibitors: discovery of amiodarone and bromocriptine inhibitory effects, *J chem inform model* **2013**, *53*, 2402-2408.
[189] J. A. Urbina, B. Moreno, S. Vierkotter, et al., Trypanosoma cruzi contains major pyrophosphate stores, and its growth in vitro and in vivo is blocked by pyrophosphate analogs, *J Biol Chem* **1999**, *274*, 33609-33615.
[190] F. Lynen, B. W. Agranoff, H. Eggerer, et al., γ,γ-Dimethyl-allyl-pyrophosphat und Geranyl-pyrophosphat, biologische Vorstufen des Squalens Zur Biosynthese der Terpene, VI1), *Angew Chem* **1959**, *71*, 657-663.
[191] Y. X. Ding, X. Ou-Yang, C. H. Shang, et al., Molecular cloning, characterization, and differential expression of a farnesyl-diphosphate synthase gene from the basidiomycetous fungus Ganoderma lucidum, *Biosci Biotechnol Biochem* **2008**, *72*, 1571-1579.
[192] H. C. Rilling, Eukaryotic prenyltransferases, *Methods Enzymol* **1985**, *110*, 145-152.
[193] V. M. Ferriols, R. Yaginuma, M. Adachi, et al., Cloning and characterization of farnesyl pyrophosphate synthase from the highly branched isoprenoid producing diatom Rhizosolenia setigera, *Sci Rep* **2015**, *5*, 10246.
[194] B. T. Sheares, S. S. White, D. T. Molowa, et al., Cloning, analysis, and bacterial expression of human farnesyl pyrophosphate synthetase and its regulation in Hep G2 cells, *Biochemistry* **1989**, *28*, 8129-8135.
[195] A. Montalvetti, B. N. Bailey, M. B. Martin, et al., Bisphosphonates are potent inhibitors of Trypanosoma cruzi farnesyl pyrophosphate synthase, *J. Biol. Chem.* **2001**, *276*, 33930-33937.

[196] M. Ferella, Z.-H. Li, B. Andersson, et al., Farnesyl diphosphate synthase localizes to the cytoplasm of Trypanosoma cruzi and T. brucei, *Experimental Parasitology* **2008**, *119*, 308-312.
[197] J. B. Rodriguez, B. N. Falcone, S. H. Szajnman, Approaches for Designing new Potent Inhibitors of Farnesyl Pyrophosphate Synthase, *Expert Opin Drug Discov* **2016**, *11*, 307-320.
[198] K. L. Kavanagh, K. Guo, J. E. Dunford, et al., The molecular mechanism of nitrogen-containing bisphosphonates as antiosteoporosis drugs, *Proc Natl Acad Sci U S A* **2006**, *103*, 7829-7834.
[199] L. C. Tarshis, M. Yan, C. D. Poulter, et al., Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-A resolution, *Biochemistry* **1994**, *33*, 10871-10877.
[200] D. J. Hosfield, Y. Zhang, D. R. Dougan, et al., Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis, *J Biol Chem* **2004**, *279*, 8526-8529.
[201] J. L. Goldstein, M. S. Brown, Regulation of the mevalonate pathway, *Nature* **1990**, *343*, 425-430.
[202] aC. D. Poulter, J. C. Argyle, E. A. Mash, Farnesyl pyrophosphate synthetase. Mechanistic studies of the 1'-4 coupling reaction with 2-fluorogeranyl pyrophosphate, *J Biol Chem* **1978**, *253*, 7227-7233; bL. Sigman, V. M. Sanchez, A. G. Turjanski, Characterization of the farnesyl pyrophosphate synthase of Trypanosoma cruzi by homology modeling and molecular dynamics, *J. Mol. Graphics Modell.* **2006**, *25*, 345-352.
[203] A. Szkopinska, D. Plochocka, Farnesyl diphosphate synthase; regulation of product specificity, *Acta Biochim Pol* **2005**, *52*, 45-55.
[204] A. Montalvetti, A. Fernandez, J. M. Sanders, et al., Farnesyl pyrophosphate synthase is an essential enzyme in Trypanosoma brucei. In vitro RNA interference and in vivo inhibition studies, *J Biol Chem* **2003**, *278*, 17075-17083.
[205] J. Park, M. Zielinski, A. Magder, et al., Human farnesyl pyrophosphate synthase is allosterically inhibited by its own product, *Nat Comm* **2017**, *8*.
[206] L. C. Tarshis, P. J. Proteau, B. A. Kellogg, et al., Regulation of product chain length by isoprenyl diphosphate synthases, *Proc Natl Acad Sci U S A* **1996**, *93*, 15018-15023.
[207] K. Ogura, T. Koyama, Enzymatic Aspects of Isoprenoid Chain Elongation, *Chemical Reviews* **1998**, *98*, 1263-1276.
[208] a K. Wang, S. Ohnuma, Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution, *Trends Biochem Sci* **1999**, *24*, 445-451; b S. Ohnuma, K. Narita, T. Nakazawa, et al., A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product, *J Biol Chem* **1996**, *271*, 30748-30754.
[209] a W. Jahnke, J. M. Rondeau, S. Cotesta, et al., Allosteric non-bisphosphonate FPPS inhibitors identified by fragment-based discovery, *Nat Chem Biol* **2010**, *6*, 660-666; b J. W. Schmidberger, R. Schnell, G. Schneider, Structural characterization of substrate and inhibitor binding to farnesyl pyrophosphate synthase from Pseudomonas aeruginosa, *Acta Crystallogr D Biol Crystallogr* **2015**, *71*, 721-731; c J. E. Gisselberg, Z. Herrera, L. M. Orchard, et al., Specific Inhibition of the Bifunctional Farnesyl/Geranylgeranyl Diphosphate Synthase in Malaria Parasites via a New Small-Molecule Binding Site, *Cell Chem Biol* **2017**.
[210] J. Mao, Y. G. Gao, S. Odeh, et al., Crystallization and preliminary x-ray diffraction study of the farnesyl diphosphate synthase from Trypanosoma brucei, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *D60*, 1863-1866.
[211] C. H. Huang, S. B. Gabelli, E. Oldfield, et al., Binding of nitrogen-containing bisphosphonates (N-BPs) to the Trypanosoma cruzi farnesyl diphosphate synthase homodimer, *Proteins* **2010**, *78*, 888-899.
[212] M. B. Martin, J. S. Grimley, J. C. Lewis, et al., Bisphosphonates inhibit the growth of Trypanosoma brucei, Trypanosoma cruzi, Leishmania donovani, Toxoplasma gondii, and Plasmodium falciparum: a potential route to chemotherapy, *Journal of medicinal chemistry* **2001**, *44*, 909-916.
[213] a N. B. Watts, Bisphosphonate treatment of osteoporosis, *Clin Geriatr Med* **2003**, *19*, 395-414; b G. A. Rodan, A. A. Reszka, Osteoporosis and bisphosphonates, *J Bone Joint Surg Am* **2003**, *85-A Suppl 3*, 8-12; c A. A. Reszka, G. A. Rodan, Nitrogen-containing bisphosphonate mechanism of action, *Mini Rev Med Chem* **2004**, *4*, 711-719; d K. Miller, R. Erez, E. Segal, et al., Targeting bone metastases with a bispecific anticancer and antiangiogenic polymer-alendronate-taxane conjugate, *Angew Chem Int Ed Engl* **2009**, *48*, 2949-2954.
[214] a H. Fleisch, R. G. Russell, F. Straumann, Effect of pyrophosphate on hydroxyapatite and its implications in calcium homeostasis, *Nature* **1966**, *212*, 901-903; b H. Fleisch, R. G. Russell, M. D. Francis, Diphosphonates inhibit hydroxyapatite dissolution in vitro and bone resorption in tissue culture and in vivo, *Science* **1969**, *165*, 1262-1264; c M. D. Francis, R. G. Russell, H. Fleisch, Diphosphonates inhibit formation of calcium phosphate crystals in vitro and pathological calcification in vivo, *Science* **1969**, *165*, 1264-1266.
[215] a E. van Beek, E. Pieterman, L. Cohen, et al., Farnesyl pyrophosphate synthase is the molecular target of nitrogen-containing bisphosphonates, *Biochem Biophys Res Commun* **1999**, *264*, 108-111; b P. P. Lehenkari, M. Kellinsalmi, J. P. Napankangas, et al., Further insight into mechanism of action of clodronate: inhibition of mitochondrial ADP/ATP translocase by a nonhydrolyzable, adenine-containing metabolite, *Mol Pharmacol* **2002**, *61*, 1255-1262.
[216] M. J. Rogers, New insights into the molecular mechanisms of action of bisphosphonates, *Curr Pharm Des* **2003**, *9*, 2643-2658.

[217] T. H. Cromartie, K. J. Fisher, Method of controling platns by inhibition of farnesy pyrophosphate synthase, *US Patent No 5 756 423* **1998**.
[218] V. D. Ding, B. T. Sheares, J. D. Bergstrom, et al., Purification and characterization of recombinant human farnesyl diphosphate synthase expressed in Escherichia coli, *Biochem J* **1991**, *275 ( Pt 1)*, 61-65.
[219] a T. H. Cromartie, K. J. Fisher, J. N. Grossman, The discovery of a novel site of action for herbicidal bisphosphonates, *Pestic Biochem Phys* **1999**, *63*, 114-126; b D. Fernandez, R. Ramis, J. Ortega-Castro, et al., New insights into human farnesyl pyrophosphate synthase inhibition by second-generation bisphosphonate drugs, *J Comput Aided Mol Des* **2017**.
[220] L. Widler, K. A. Jaeggi, M. Glatt, et al., Highly potent geminal bisphosphonates. From pamidronate disodium (Aredia) to zoledronic acid (Zometa), *Journal of medicinal chemistry* **2002**, *45*, 3721-3738.
[221] J. D. Bergstrom, R. G. Bostedor, P. J. Masarachia, et al., Alendronate is a specific, nanomolar inhibitor of farnesyl diphosphate synthase, *Arch Biochem Biophys* **2000**, *373*, 231-241.
[222] J. E. Frampton, C. M. Perry, Ibandronate: a review of its use in the management of postmenopausal osteoporosis, *Drugs* **2008**, *68*, 2683-2707.
[223] C. Crandall, Risedronate: a clinical review, *Arch Intern Med* **2001**, *161*, 353-360.
[224] M. Takeuchi, S. Sakamoto, K. Kawamuki, et al., Studies on novel bone resorption inhibitors. II. Synthesis and pharmacological activities of fused aza-heteroarylbisphosphonate derivatives, *Chem Pharm Bull (Tokyo)* **1998**, *46*, 1703-1709.
[225] a J. E. Dunford, K. Thompson, F. P. Coxon, et al., Structure-activity relationships for inhibition of farnesyl diphosphate synthase in vitro and inhibition of bone resorption in vivo by nitrogen-containing bisphosphonates, *J Pharmacol Exp Ther* **2001**, *296*, 235-242; b J. E. Dunford, A. A. Kwaasi, M. J. Rogers, et al., Structure-activity relationships among the nitrogen containing bisphosphonates in clinical use and other analogues: time-dependent inhibition of human farnesyl pyrophosphate synthase, *Journal of medicinal chemistry* **2008**, *51*, 2187-2195.
[226] W. Jahnke, C. Henry, An in vitro assay to measure targeted drug delivery to bone mineral, *ChemMedChem* **2010**, *5*, 770-776.
[227] D. E. Hughes, K. R. Wright, H. L. Uy, et al., Bisphosphonates promote apoptosis in murine osteoclasts in vitro and in vivo, *J Bone Miner Res* **1995**, *10*, 1478-1487.
[228] a H. Monkkonen, S. Auriola, P. Lehenkari, et al., A new endogenous ATP analog (ApppI) inhibits the mitochondrial adenine nucleotide translocase (ANT) and is responsible for the apoptosis induced by nitrogen-containing bisphosphonates, *Br J Pharmacol* **2006**, *147*, 437-445; b J. Raikkonen, M. Taskinen, J. E. Dunford, et al., Correlation between time-dependent inhibition of human farnesyl pyrophosphate synthase and blockade of mevalonate pathway by nitrogen-containing bisphosphonates in cultured cells, *Biochem Biophys Res Commun* **2011**, *407*, 663-667.
[229] L. Widler, W. Jahnke, J. R. Green, The chemistry of bisphosphonates: from antiscaling agents to clinical therapeutics, *Anti-Cancer Agents Med. Chem.* **2012**, *12*, 95-101.
[230] A. L. Marzinzik, R. Amstutz, G. Bold, et al., Discovery of Novel Allosteric Non-Bisphosphonate Inhibitors of Farnesyl Pyrophosphate Synthase by Integrated Lead Finding, *ChemMedChem* **2015**, *10*, 1884-1891.
[231] a J. Mao, S. Mukherjee, Y. Zhang, et al., Solid-state NMR, crystallographic, and computational investigation of bisphosphonates and farnesyl diphosphate synthase-bisphosphonate complexes, *J Am Chem Soc* **2006**, *128*, 14485-14497; b T. Yokoyama, M. Mizuguchi, A. Ostermann, et al., Protonation State and Hydration of Bisphosphonate Bound to Farnesyl Pyrophosphate Synthase, *Journal of medicinal chemistry* **2015**, *58*, 7549-7556.
[232] M. B. Martin, W. Arnold, H. T. Heath, 3rd, et al., Nitrogen-containing bisphosphonates as carbocation transition state analogs for isoprenoid biosynthesis, *Biochem Biophys Res Commun* **1999**, *263*, 754-758.
[233] J. F. Glickman, A. Schmid, Farnesyl pyrophosphate synthase: real-time kinetics and inhibition by nitrogen-containing bisphosphonates in a scintillation assay, *Assay Drug Dev Technol* **2007**, *5*, 205-214.
[234] J. Park, Y. S. Lin, Y. S. Tsantrizos, et al., Structure of human farnesyl pyrophosphate synthase in complex with an aminopyridine bisphosphonate and two molecules of inorganic phosphate, *Acta Crystallogr F Struct Biol Commun* **2014**, *70*, 299-304.
[235] S. H. Szajnman, B. N. Bailey, R. Docampo, et al., Bisphosphonates derived from fatty acids are potent growth inhibitors of Trypanosoma cruzi, *Bioorg Med Chem Lett* **2001**, *11*, 789-792.
[236] B. Bouzahzah, L. A. Jelicks, S. A. Morris, et al., Risedronate in the treatment of Murine Chagas' disease, *Parasitol Res* **2005**, *96*, 184-187.
[237] R. Docampo, S. N. Moreno, Acidocalcisomes, *Cell Calcium* **2011**, *50*, 113-119.
[238] a L. R. Garzoni, A. Caldera, M. d. N. L. Meirelles, et al., Selective in vitro effects of the farnesyl pyrophosphate synthase inhibitor risedronate on Trypanosoma cruzi, *Int. J. Antimicrob. Agents* **2004**, *23*, 273-285; b L. R. Garzoni, M. C. Waghabi, M. M. Baptista, et al., Antiparasitic activity of risedronate in a murine model of acute Chagas' disease, *Int J Antimicrob Agents* **2004**, *23*, 286-290; c S. H. Szajnman, A. Montalvetti, Y. Wang, et al., Bisphosphonates derived from fatty acids are potent inhibitors of Trypanosoma cruzi farnesyl pyrophosphate synthase, *Bioorg Med Chem Lett* **2003**, *13*, 3231-3235.
[239] I. C. Cuevas, P. Rohloff, D. O. Sanchez, et al., Characterization of farnesylated protein tyrosine phosphatase TcPRL-1 from Trypanosoma cruzi, *Eukaryot Cell* **2005**, *4*, 1550-1561.

[240] H. M. Berman, J. Westbrook, Z. Feng, et al., The Protein Data Bank, *Nucleic Acids Res* **2000**, *28*, 235-242.
[241] H. G. Lu, L. Zhong, W. de Souza, et al., Ca2+ content and expression of an acidocalcisomal calcium pump are elevated in intracellular forms of Trypanosoma cruzi, *Mol Cell Biol* **1998**, *18*, 2309-2323.
[242] D. A. Scott, R. Docampo, J. A. Dvorak, et al., In situ compositional analysis of acidocalcisomes in Trypanosoma cruzi, *J Biol Chem* **1997**, *272*, 28020-28029.
[243] A. S. Kesselheim, J. Avorn, The most transformative drugs of the past 25 years: a survey of physicians, *Nat Rev Drug Discov* **2013**, *12*, 425-431.
[244] a C. Orozco, N. M. Maalouf, Safety of bisphosphonates, *Rheum Dis Clin North Am* **2012**, *38*, 681-705; b S. Paulo, A. M. Abrantes, M. Laranjo, et al., Bisphosphonate-related osteonecrosis of the jaw: specificities, *Oncol Rev* **2014**, *8*, 254.
[245] P. Merino, L. Maiuolo, I. Delso, et al., Chemical approaches to inhibitors of isoprenoid biosynthesis: targeting farnesyl and geranylgeranyl pyrophosphate synthases, *Rsc Advances* **2017**, *7*, 10947-10967.
[246] a R. Docampo, D. A. Scott, A. E. Vercesi, et al., Intracellular Ca2+ storage in acidocalcisomes of Trypanosoma cruzi, *Biochem J* **1995**, *310 ( Pt 3)*, 1005-1012; b D. A. Scott, W. de Souza, M. Benchimol, et al., Presence of a plant-like proton-pumping pyrophosphatase in acidocalcisomes of Trypanosoma cruzi, *J Biol Chem* **1998**, *273*, 22151-22158.
[247] C. O. Rodrigues, D. A. Scott, R. Docampo, Characterization of a vacuolar pyrophosphatase in Trypanosoma brucei and its localization to acidocalcisomes, *Mol Cell Biol* **1999**, *19*, 7712-7723.
[248] A. E. Vercesi, C. O. Rodrigues, R. Catisti, et al., Presence of a Na(+)/H(+) exchanger in acidocalcisomes of Leishmania donovani and their alkalization by anti-leishmanial drugs, *FEBS Lett* **2000**, *473*, 203-206.
[249] S. Aripirala, D. Gonzalez-Pacanowska, E. Oldfield, et al., Structural and thermodynamic basis of the inhibition of Leishmania major farnesyl diphosphate synthase by nitrogen-containing bisphosphonates, *Acta Crystallogr D Biol Crystallogr* **2014**, *70*, 802-810.
[250] S. Luo, F. A. Ruiz, S. N. Moreno, The acidocalcisome Ca2+-ATPase (TgA1) of Toxoplasma gondii is required for polyphosphate storage, intracellular calcium homeostasis and virulence, *Mol Microbiol* **2005**, *55*, 1034-1045.
[251] Y. M. Drozdowicz, M. Shaw, M. Nishi, et al., Isolation and characterization of TgVP1, a type I vacuolar H+-translocating pyrophosphatase from Toxoplasma gondii. The dynamics of its subcellular localization and the cellular effects of a diphosphonate inhibitor, *J Biol Chem* **2003**, *278*, 1075-1085.
[252] V. Yardley, A. A. Khan, M. B. Martin, et al., In vivo activities of farnesyl pyrophosphate synthase inhibitors against Leishmania donovani and Toxoplasma gondii, *Antimicrob Agents Chemother* **2002**, *46*, 929-931.
[253] D. Mukkamala, J. H. No, L. M. Cass, et al., Bisphosphonate inhibition of a Plasmodium farnesyl diphosphate synthase and a general method for predicting cell-based activity from enzyme data, *Journal of medicinal chemistry* **2008**, *51*, 7827-7833.
[254] a L. Sinigaglia, M. Varenna, S. Casari, Pharmacokinetic profile of bisphosphonates in the treatment of metabolic bone disorders, *Clin Cases Miner Bone Metab* **2007**, *4*, 30-36; b S. C. Cremers, G. Pillai, S. E. Papapoulos, Pharmacokinetics/pharmacodynamics of bisphosphonates: use for optimisation of intermittent therapy for osteoporosis, *Clin Pharmacokinet* **2005**, *44*, 551-570; c R. G. Russell, N. B. Watts, F. H. Ebetino, et al., Mechanisms of action of bisphosphonates: similarities and differences and their potential influence on clinical efficacy, *Osteoporos Int* **2008**, *19*, 733-759; d W. Jahnke, G. Bold, A. L. Marzinzik, et al., A General Strategy for Targeting Drugs to Bone, *Angew Chem Int Ed Engl* **2015**, *54*, 14575-14579; e H. M. Weiss, U. Pfaar, A. Schweitzer, et al., Biodistribution and plasma protein binding of zoledronic acid, *Drug Metab Dispos* **2008**, *36*, 2043-2049; f J. H. Lin, Bisphosphonates: a review of their pharmacokinetic properties, *Bone* **1996**, *18*, 75-85.
[255] J. Kotz, Going to bis-school, *Science-Business eXchange* **2010**, *3*, 1030-1030.
[256] H. C. Chen, N. Joalland, J. S. Bridgeman, et al., Synergistic targeting of breast cancer stem-like cells by human gammadelta T cells and CD8(+) T cells, *Immunol Cell Biol* **2017**, *95*, 620-629.
[257] S. Pelleieux, C. Picard, L. Lamarre-Theroux, et al., Isoprenoids and tau pathology in sporadic Alzheimer's disease, *Neurobiol Aging* **2018**, *65*, 132-139.
[258] X.-L. Xu, W.-L. Gou, A.-Y. Wang, et al., Basic research and clinical applications of bisphosphonates in bone disease: what have we learned over the last 40 years?, *J. Transl. Med.* **2013**, *11*, 303/301-303/308, 308 pp.
[259] M. S. Marma, Z. Xia, C. Stewart, et al., Synthesis and biological evaluation of alpha-halogenated bisphosphonate and phosphonocarboxylate analogues of risedronate, *Journal of medicinal chemistry* **2007**, *50*, 5967-5975.
[260] J. Liu, W. Liu, H. Ge, et al., Syntheses and characterization of non-bisphosphonate quinoline derivatives as new FPPS inhibitors, *Biochim Biophys Acta* **2014**, *1840*, 1051-1062.
[261] S. H. Szajnman, E. L. Ravaschino, R. Docampo, et al., Synthesis and biological evaluation of 1-amino-1,1-bisphosphonates derived from fatty acids against Trypanosoma cruzi targeting farnesyl pyrophosphate synthase, *Bioorg Med Chem Lett* **2005**, *15*, 4685-4690.
[262] S. H. Szajnman, V. S. Rosso, L. Malayil, et al., 1-(Fluoroalkylidene)-1,1-bisphosphonic acids are potent and selective inhibitors of the enzymatic activity of Toxoplasma gondii farnesyl pyrophosphate synthase, *Org Biomol Chem* **2012**, *10*, 1424-1433.

[263] a V. S. Rosso, S. H. Szajnman, L. Malayil, et al., Synthesis and biological evaluation of new 2-alkylaminoethyl-1,1-bisphosphonic acids against Trypanosoma cruzi and Toxoplasma gondii targeting farnesyl diphosphate synthase, *Bioorg Med Chem* **2011**, *19*, 2211-2217; b S. H. Szajnman, G. E. Garcia Linares, Z. H. Li, et al., Synthesis and biological evaluation of 2-alkylaminoethyl-1,1-bisphosphonic acids against Trypanosoma cruzi and Toxoplasma gondii targeting farnesyl diphosphate synthase, *Bioorg Med Chem* **2008**, *16*, 3283-3290.
[264] C. A. Rodrigues-Poveda, D. Gonzalez-Pacanowska, S. H. Szajnman, et al., 2-alkylaminoethyl-1,1-bisphosphonic acids are potent inhibitors of the enzymatic activity of Trypanosoma cruzi squalene synthase, *Antimicrob Agents Chemother* **2012**, *56*, 4483-4486.
[265] M. Recher, A. P. Barboza, Z.-H. Li, et al., Design, synthesis and biological evaluation of sulfur-containing 1,1-bisphosphonic acids as antiparasitic agents, *Eur. J. Med. Chem.* **2013**, *60*, 431-440.
[266] T. Galaka, M. Ferrer Casal, M. Storey, et al., Antiparasitic Activity of Sulfur- and Fluorine-Containing Bisphosphonates against Trypanosomatids and Apicomplexan Parasites, *Molecules* **2017**, *22*, 82/81-82/22.
[267] E. Oldfield, Y. Zhang, F. Yin, Bisphosphonate compounds and mehtods with enhanced potency for multiple targets inlcluding FPPS, GGPPS, and DPPS, *US Patent 8 012 949 B2* **2011**.
[268] P. Vachal, J. J. Hale, Z. Lu, et al., Synthesis and study of alendronate derivatives as potential prodrugs of alendronate sodium for the treatment of low bone density and osteoporosis, *Journal of medicinal chemistry* **2006**, *49*, 3060-3063.
[269] a B. Demoro, F. Caruso, M. Rossi, et al., Risedronate metal complexes potentially active against Chagas disease, *J Inorg Biochem* **2010**, *104*, 1252-1258; b B. Demoro, F. Caruso, M. Rossi, et al., Bisphosphonate metal complexes as selective inhibitors of Trypanosoma cruzi farnesyl diphosphate synthase, *Dalton Trans* **2012**, *41*, 6468-6476.
[270] R. A. Sanchez-Delgado, K. Lazardi, L. Rincon, et al., Toward a novel metal-based chemotherapy against tropical diseases. 1. Enhancement of the efficacy of clotrimazole against Trypanosoma cruzi by complexation to ruthenium in RuCl2(clotrimazole)2, *Journal of medicinal chemistry* **1993**, *36*, 2041-2043.
[271] V. M. de Oliveira, L. T. Mendes, D. J. Almeida, et al., New treatments for Chagas disease and the relationship between chagasic patients and cancers, *Cancer Res. J.* **2014**, *2*, 11-29.
[272] B. Demoro, S. Rostan, M. Moncada, et al., Ibandronate metal complexes: solution behavior and antiparasitic activity, *J Biol Inorg Chem* **2018**.
[273] J. J. Irwin, B. K. Shoichet, ZINC--a free database of commercially available compounds for virtual screening, *Journal of chemical information and modeling* **2005**, *45*, 177-182.
[274] L. Oliveira, J. Araújo, D. Costa Júnior, et al., *Virtual Screening for the Selection of New Candidates to Trypanosoma cruzi Farnesyl Pyrophosphate Synthase Inhibitors, Vol. 29*, **2018**.
[275] S. Lindert, W. Zhu, Y. L. Liu, et al., Farnesyl diphosphate synthase inhibitors from in silico screening, *Chem Biol Drug Des* **2013**, *81*, 742-748.
[276] J. W. De Schutter, J. Park, C. Y. Leung, et al., Multistage screening reveals chameleon ligands of the human farnesyl pyrophosphate synthase: implications to drug discovery for neurodegenerative diseases, *Journal of medicinal chemistry* **2014**, *57*, 5764-5776.
[277] J. P. Changeux, Feedback Control Mechanism of Biosynthetic L-Threonine Deaminase by L-Isoleucine, *Cold Spring Harbor Symposia on Quantitative Biology* **1961**, *26*, 313-&.
[278] R. Amstutz, G. Bold, S. Cotesta, et al., Quinolines as inhibiotors of farnesyl pyrophosphate synthase, **2009**.
[279] S. Cotesta, J. F. Glickman, W. Jahnke, et al., Salicilic acid derivatives being farnesyl pyrophosphate synthase activity inhibitors, *WO 2010/043584 A1* **2010**.
[280] S. Sun, C. E. McKenna, Farnesyl pyrophosphate synthase modulators: a patent review (2006 - 2010), *Expert Opin Ther Pat* **2011**, *21*, 1433-1451.
[281] J. Gao, X. Chu, Y. Qiu, et al., Discovery of potent inhibitor for farnesyl pyrophosphate synthase in the mevalonate pathway, *Chem Commun (Camb)* **2010**, *46*, 5340-5342.
[282] J. Gao, J. Liu, Y. Qiu, et al., Multi-target-directed design, syntheses, and characterization of fluorescent bisphosphonate derivatives as multifunctional enzyme inhibitors in mevalonate pathway, *Biochim Biophys Acta* **2013**, *1830*, 3635-3642.
[283] Y.-L. Liu, R. Cao, Y. Wang, et al., Farnesyl Diphosphate Synthase Inhibitors With Unique Ligand-Binding Geometries, *ACS Med. Chem. Lett.* **2015**, *6*, 349-354.
[284] C. Y. Leung, J. Park, J. W. De Schutter, et al., Thienopyrimidine bisphosphonate (ThPBP) inhibitors of the human farnesyl pyrophosphate synthase: optimization and characterization of the mode of inhibition, *Journal of medicinal chemistry* **2013**, *56*, 7939-7950.
[285] D. Gritzalis, J. Park, W. Chiu, et al., Probing the molecular and structural elements of ligands binding to the active site versus an allosteric pocket of the human farnesyl pyrophosphate synthase, *Bioorg Med Chem Lett* **2015**, *25*, 1117-1123.
[286] Y.-L. Liu, S. Lindert, W. Zhu, et al., Taxodione and arenarone inhibit farnesyl diphosphate synthase by binding to the isopentenyl diphosphate site, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E2530-E2539.
[287] M. Scrima, G. Lauro, M. Grimaldi, et al., Structural evidence of N6-isopentenyladenosine as a new ligand of farnesyl pyrophosphate synthase, *Journal of medicinal chemistry* **2014**, *57*, 7798-7803.

[288] R. Macarron, M. N. Banks, D. Bojanic, et al., Impact of high-throughput screening in biomedical research, *Nat Rev Drug Discov* **2011**, *10*, 188-195.
[289] a J. J. Irwin, D. Duan, H. Torosyan, et al., An Aggregation Advisor for Ligand Discovery, *Journal of medicinal chemistry* **2015**, *58*, 7076-7087; b S. L. McGovern, E. Caselli, N. Grigorieff, et al., A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening, *Journal of medicinal chemistry* **2002**, *45*, 1712-1722.
[290] A. Barker, J. G. Kettle, T. Nowak, et al., Expanding medicinal chemistry space, *Drug Discov Today* **2013**, *18*, 298-304.
[291] P. C. Ray, M. Kiczun, M. Huggett, et al., Fragment library design, synthesis and expansion: nurturing a synthesis and training platform, *Drug Discov Today* **2017**, *22*, 43-56.
[292] B. Lamoree, R. E. Hubbard, Current perspectives in fragment-based lead discovery (FBLD), *Essays Biochem* **2017**, *61*, 453-464.
[293] aM. M. Hann, A. R. Leach, G. Harper, Molecular complexity and its impact on the probability of finding leads for drug discovery, *J Chem Inf Comput Sci* **2001**, *41*, 856-864; bD. A. Erlanson, S. W. Fesik, R. E. Hubbard, et al., Twenty years on: the impact of fragments on drug discovery, *Nat Rev Drug Discov* **2016**; cR. E. Hubbard, B. Davis, I. Chen, et al., The SeeDs approach: integrating fragments into drug discovery, *Curr Top Med Chem* **2007**, *7*, 1568-1581.
[294] G. G. Ferenczy, G. M. Keseru, How are fragments optimized? A retrospective analysis of 145 fragment optimizations, *Journal of medicinal chemistry* **2013**, *56*, 2478-2486.
[295] a N. Baurin, F. Aboul-Ela, X. Barril, et al., Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets, *J Chem Inf Comput Sci* **2004**, *44*, 2157-2166; b G. M. Keseru, D. A. Erlanson, G. G. Ferenczy, et al., Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia, *Journal of medicinal chemistry* **2016**, *59*, 8189-8206.
[296] C. J. Radoux, T. S. Olsson, W. R. Pitt, et al., Identifying Interactions that Determine Fragment Binding at Protein Hotspots, *Journal of medicinal chemistry* **2016**, *59*, 4314-4325.
[297] a I. J. Chen, R. E. Hubbard, Lessons for fragment library design: analysis of output from multiple screening campaigns, *J Comput Aided Mol Des* **2009**, *23*, 603-620; b P. J. Hajduk, J. R. Huth, S. W. Fesik, Druggability indices for protein targets derived from NMR-based screening data, *Journal of medicinal chemistry* **2005**, *48*, 2518-2525.
[298] R. F. Ludlow, M. L. Verdonk, H. K. Saini, et al., Detection of secondary binding sites in proteins using fragment screening, *Proc Natl Acad Sci U S A* **2015**, *112*, 15910-15915.
[299] a J. M. Ostrem, U. Peters, M. L. Sos, et al., K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions, *Nature* **2013**, *503*, 548-551; b Q. Sun, J. P. Burke, J. Phan, et al., Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation, *Angew Chem Int Ed Engl* **2012**, *51*, 6140-6143.
[300] a C. W. Murray, D. C. Rees, The rise of fragment-based drug discovery, *Nat Chem* **2009**, *1*, 187-192; b J. P. Renaud, C. W. Chung, U. H. Danielson, et al., Biophysics in drug discovery: impact, challenges and opportunities, *Nat Rev Drug Discov* **2016**, *15*, 679-698.
[301] a P. D. Leeson, S. A. St-Gallay, The influence of the 'organizational factor' on compound quality in drug discovery, *Nat Rev Drug Discov* **2011**, *10*, 749-765; b M. M. Hann, Molecular obesity, potency and other addictions in drug discovery, *Medchemcomm* **2011**, *2*, 349-355; cR. J. Young, in *Tactics in Contemporary Drug Design* (Ed.: N. A. Meanwell), Springer Berlin Heidelberg, Berlin, Heidelberg, **2015**, pp. 1-68.
[302] S. B. Shuker, P. J. Hajduk, R. P. Meadows, et al., Discovering high-affinity ligands for proteins: SAR by NMR, *Science* **1996**, *274*, 1531-1534.
[303] P. J. Hajduk, J. Greer, A decade of fragment-based drug design: strategic advances and lessons learned, *Nat Rev Drug Discov* **2007**, *6*, 211-219.
[304] V. L. Nienaber, P. L. Richardson, V. Klighofer, et al., Discovering novel ligands for macromolecules using X-ray crystallographic screening, *Nat Biotechnol* **2000**, *18*, 1105-1108.
[305] M. Congreve, R. Carr, C. Murray, et al., A 'rule of three' for fragment-based lead discovery?, *Drug Discov Today* **2003**, *8*, 876-877.
[306] D. A. Erlanson, J. A. Wells, A. C. Braisted, Tethering: fragment-based drug discovery, *Annu Rev Biophys Biomol Struct* **2004**, *33*, 199-223.
[307] C. N. Johnson, D. A. Erlanson, C. W. Murray, et al., Fragment-to-Lead Medicinal Chemistry Publications in 2015, *Journal of medicinal chemistry* **2016**.
[308] D. A. Erlanson, *Fragments in the clinic: 2018 edition*, https://practicalfragments.blogspot.com/2018/10/fragments-in-clinic-2018-edition.html, accesses date 10. April 2019
[309] a T. Sugiki, K. Furuita, T. Fujiwara, et al., Current NMR Techniques for Structure-Based Drug Discovery, *Molecules* **2018**, *23*; b A. D. Gossert, W. Jahnke, NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules, *Prog Nucl Magn Reson Spectrosc* **2016**, *97*, 82-125.

[310] a D. G. Myszka, R. L. Rich, Implementing surface plasmon resonance biosensors in drug discovery, *Pharm Sci Technolo Today* **2000**, *3*, 310-317; b A. M. Giannetti, From experimental design to validated hits a comprehensive walk-through of fragment lead identification using surface plasmon resonance, *Methods Enzymol* **2011**, *493*, 169-218.
[311] a F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability, *Nature Protocols* **2007**, *2*, 2212-2221; b D. A. Erlanson, W. Jahnke, R. Mannhold, et al., *Fragment-based Drug Discovery: Lessons and Outlook*, Wiley, **2016**.
[312] E. H. Mashalidis, P. Sledz, S. Lang, et al., A three-stage biophysical screening cascade for fragment-based drug discovery, *Nat Protoc* **2013**, *8*, 2309-2324.
[313] D. S. Chan, A. J. Whitehouse, A. G. Coyne, et al., Mass spectrometry for fragment screening, *Essays Biochem* **2017**, *61*, 465-473.
[314] R. E. Hubbard, J. B. Murray, in *Methods in Enzymology, Vol. Volume 493* (Ed.: C. K. Lawrence), Academic Press, **2011**, pp. 509-531.
[315] a J. Schiebel, N. Radeva, S. G. Krimmer, et al., Six Biophysical Screening Methods Miss a Large Proportion of Crystallographically Discovered Fragment Hits: A Case Study, *ACS Chem Biol* **2016**, *11*, 1693-1701; b J. Schiebel, N. Radeva, H. Koster, et al., One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists, *ChemMedChem* **2015**, *10*, 1511-1521.
[316] a P. S. Kutchukian, A. M. Wassermann, M. K. Lindvall, et al., Large scale meta-analysis of fragment-based screening campaigns: privileged fragments and complementary technologies, *J Biomol Screen* **2015**, *20*, 588-596; b J. Wielens, S. J. Headey, D. I. Rhodes, et al., Parallel screening of low molecular weight fragment libraries: do differences in methodology affect hit identification?, *J Biomol Screen* **2013**, *18*, 147-159; c E. Meiby, H. Simmonite, L. le Strat, et al., Fragment screening by weak affinity chromatography: comparison with established techniques for screening against HSP90, *Anal Chem* **2013**, *85*, 6756-6766.
[317] P. Baldi, R. Nasr, When is chemical similarity significant? The statistical distribution of chemical similarity scores and its extreme values, *J Chem Inf model* **2010**, *50*, 1205-1222.
[318] a M. J. Harner, A. O. Frank, S. W. Fesik, Fragment-based drug discovery using NMR spectroscopy, *J Biomol NMR* **2013**, *56*, 65-75; b W. F. Lau, J. M. Withka, D. Hepworth, et al., Design of a multi-purpose fragment screening library using molecular complexity and orthogonal diversity metrics, *J Comput Aided Mol Des* **2011**, *25*, 621-636.
[319] C. M. Dobson, Chemical space and biology, *Nature* **2004**, *432*, 824-828.
[320] aT. J. Ritchie, S. J. Macdonald, R. J. Young, et al., The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types, *Drug Discov Today* **2011**, *16*, 164-171; bA. P. Hill, R. J. Young, Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity, *Drug Discov Today* **2010**, *15*, 648-655; cT. J. Ritchie, S. J. Macdonald, The impact of aromatic ring count on compound developability--are too many aromatic rings a liability in drug design?, *Drug Discov Today* **2009**, *14*, 1011-1020.
[321] a F. Lovering, J. Bikker, C. Humblet, Escape from flatland: increasing saturation as an approach to improving clinical success, *Journal of medicinal chemistry* **2009**, *52*, 6752-6756; b Y. Yang, O. Engkvist, A. Llinas, et al., Beyond size, ionization state, and lipophilicity: influence of molecular topology on absorption, distribution, metabolism, excretion, and toxicity for druglike compounds, *Journal of medicinal chemistry* **2012**, *55*, 3667-3677.
[322] B. Over, S. Wetzel, C. Grutter, et al., Natural-product-derived fragments for fragment-based ligand discovery, *Nat Chem* **2013**, *5*, 21-28.
[323] A. Vulpetti, C. Dalvit, Design and generation of highly diverse fluorinated fragment libraries and their efficient screening with improved (19) F NMR methodology, *ChemMedChem* **2013**, *8*, 2057-2069.
[324] M. O'Reilly, A. Cleasby, T. G. Davies, et al., Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design, *Drug Discovery Today* **2019**.
[325] C. Dalvit, D. Caronni, N. Mongelli, et al., NMR-based quality control approach for the identification of false positives and false negatives in high throughput screening, *Curr Drug Discov Technol* **2006**, *3*, 115-124.
[326] a R. F. Bruns, I. A. Watson, Rules for identifying potentially reactive or promiscuous compounds, *Journal of medicinal chemistry* **2012**, *55*, 9763-9772; b R. Brenk, A. Schipani, D. James, et al., Lessons learnt from assembling screening libraries for drug discovery for neglected diseases, *ChemMedChem* **2008**, *3*, 435-444; c G. M. Rishton, Nonleadlikeness and leadlikeness in biochemical screening, *Drug Discov Today* **2003**, *8*, 86-96.
[327] J. B. Baell, 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*, 2719-2740.
[328] C. Lipinski, A. Hopkins, Navigating chemical space for biology and medicine, *Nature* **2004**, *432*, 855-861.
[329] A. H. Lipkus, Q. Yuan, K. A. Lucas, et al., Structural diversity of organic chemistry. A scaffold analysis of the CAS Registry, *J Org Chem* **2008**, *73*, 4443-4451.
[330] R. D. Taylor, M. MacCoss, A. D. Lawson, Rings in drugs, *Journal of medicinal chemistry* **2014**, *57*, 5845-5859.

[331] W. R. Pitt, D. M. Parry, B. G. Perry, et al., Heteroaromatic rings of the future, *Journal of medicinal chemistry* **2009**, *52*, 2952-2963.
[332] J. Fejzo, C. A. Lepre, J. W. Peng, et al., The SHAPES strategy: an NMR-based approach for lead generation in drug discovery, *Chem Biol* **1999**, *6*, 755-769.
[333] A. Schuffenhauer, S. Ruedisser, A. L. Marzinzik, et al., Library design for fragment based screening, *Curr Top Med Chem* **2005**, *5*, 751-762.
[334] X. Wang, Y. Choe, C. S. Craik, et al., Design and synthesis of novel inhibitors of gelatinase B, *Bioorg Med Chem Lett* **2002**, *12*, 2201-2204.
[335] D. J. Maly, I. C. Choong, J. A. Ellman, Combinatorial target-guided ligand assembly: identification of potent subtype-selective c-Src inhibitors, *Proc Natl Acad Sci U S A* **2000**, *97*, 2419-2424.
[336] O. B. Cox, T. Krojer, P. Collins, et al., A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain, *Chem Sci* **2016**, *7*, 2322-2330.
[337] a C. Boesch, Nobel Prizes for nuclear magnetic resonance: 2003 and historical perspectives, *J Magn Reson Imaging* **2004**, *20*, 177-179; b Nobel Prizes for Physics for 1943 and 1944: Prof. Otto Stern, *Nature* **1944**, *154*, 634-634.
[338] a R. Frisch, O. Stern, Über die magnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. I, *Zeitschrift für Physik* **1933**, *85*, 4-16; b I. Estermann, O. Stern, Über die magnetische Ablenkung von Wasserstoffmolekülen und das magnetische Moment des Protons. II, *Zeitschrift für Physik* **1933**, *85*, 17-24; c I. Estermann, R. Frisch, O. Stern, Magnetic Moment of the Proton, *Nature* **1933**, *132*, 169-170; d W. Gerlach, O. Stern, *Der experimentelle Nachweis der Richtungsquantelung im Magnetfeld, Vol. 9*, **1922**.
[339] a I. I. Rabi, S. Millman, P. Kusch, et al., The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments. The Magnetic Moments ofLi63,Li73andF199, *Physical Review* **1939**, *55*, 526-535; b J. M. B. Kellogg, I. I. Rabi, N. F. Ramsey, et al., The Magnetic Moments of the Proton and the Deuteron. The Radiofrequency Spectrum ofH2in Various Magnetic Fields, *Physical Review* **1939**, *56*, 728-743.
[340] a F. Bloch, W. W. Hansen, M. Packard, Nuclear Induction, *Physical Review* **1946**, *69*, 127-127; b E. M. Purcell, H. C. Torrey, R. V. Pound, Resonance Absorption by Nuclear Magnetic Moments in a Solid, *Physical Review* **1946**, *69*, 37-38.
[341] H. Friebolin, *Ein- und zweidimensionale NMR-Spektroskopie. Eine Einführung.*, 5. ed., Wiley-VCH, **2013**.
[342] E. M. Purcell, in *Pysics 1942-1962* (Ed.: S. Stuart), Elsevier, Amsterdam, **1964**, pp. 219-231.
[343] R. R. Ernst, W. A. Anderson, Application of Fourier Transform Spectroscopy to Magnetic Resonance, *Review of Scientific Instruments* **1966**, *37*, 93-102.
[344] a W. Aue, E. Bartholdi, R. R. Ernst, Two-dimensional spectroscopy. Application to nuclear magnetic resonance, *J Chem Phys* **1976**, *64*, 2229-2246; b W. P. Aue, P. Bachmann, A. Wokaun, et al., Sensitivity of two-dimensional NMR spectroscopy, *J Magn Reson(1969)* **1978**, *29*, 523-533; c R. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of NMR in one and two dimensions*, Clarendron Press, Oxford, **1987**.
[345] K. Wuthrich, NMR studies of structure and function of biological macromolecules (Nobel Lecture), *J Biomol NMR* **2003**, *27*, 13-39.
[346] a C. A. Lepre, Practical aspects of NMR-based fragment screening, *Methods Enzymol* **2011**, *493*, 219-239; b A. D. Gossert, C. Henry, M. J. Blommers, et al., Time efficient detection of protein-ligand interactions with the polarization optimized PO-WaterLOGSY NMR experiment, *J Biomol NMR* **2009**, *43*, 211-217.
[347] A. Ciulli, Biophysical screening for the discovery of small-molecule ligands, *Methods Mol Biol* **2013**, *1008*, 357-388.
[348] B. J. Stockman, C. Dalvit, NMR screening techniques in drug discovery and drug design, *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *41*, 187-231.
[349] M. Mayer, B. Meyer, Characterization of ligand binding by saturation transfer difference NMR spectroscopy, *Angew Chem-Int Ed* **1999**, *38*, 1784-1788.
[350] A. W. Overhauser, Polarization of Nuclei in Metals, *Physical Review* **1953**, *92*, 411-415.
[351] J. Robson-Tull, Biophysical screening in fragment-based drug design: a brief overview, *Bioscience Horizons: The International Journal of Student Research* **2018**, *11*.
[352] P. J. Hajduk, E. T. Olejniczak, S. W. Fesik, One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules, *J Am Chem Soc* **1997**, *119*, 12257-12261.
[353] a C. Dalvit, P. Pevarello, M. Tato, et al., Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water, *J Biomol NMR* **2000**, *18*, 65-68; b C. Dalvit, G. Fogliatto, A. Stewart, et al., WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability, *J Biomol NMR* **2001**, *21*, 349-359.
[354] R. S. Norton, E. W. Leung, I. R. Chandrashekaran, et al., Applications of (19)F-NMR in Fragment-Based Drug Discovery, *Molecules* **2016**, *21*, 860.
[355] a H. Y. Carr, E. M. Purcell, Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, *Physical Review* **1954**, *94*, 630-638; b S. Meiboom, D. Gill, Modified Spin-Echo Method for Measuring Nuclear Relaxation Times, *Review of Scientific Instruments* **1958**, *29*, 688-691.
[356] W. Jahnke, P. Floersheim, C. Ostermeier, et al., NMR reporter screening for the detection of high-affinity ligands, *Angew Chem Int Ed Engl* **2002**, *41*, 3420-3423.

[357] a P. Schanda, E. Kupce, B. Brutscher, SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds, *J Biomol NMR* **2005**, *33*, 199-211; b P. Schanda, B. Brutscher, Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds, *J Am Chem Soc* **2005**, *127*, 8014-8015; c L. Mueller, Alternate HMQC experiments for recording HN and HC-correlation spectra in proteins at high throughput, *J Biomol NMR* **2008**, *42*, 129-137.
[358] F. S. Midani, M. L. Wynn, S. Schnell, The importance of accurately correcting for the natural abundance of stable isotopes, *Anal Biochem* **2017**, *520*, 27-43.
[359] M. P. Williamson, Using chemical shift perturbation to characterise ligand binding, *P Nucl Mag Reson Spec* **2013**, *73*, 1-16.
[360] a L. Fielding, NMR methods for the determination of protein-ligand dissociation constants, *P Nucl Magn Reson Spec* **2007**, *51*, 219-242; b L. Fielding, NMR methods for the determination of protein-ligand dissociation constants, *Curr Top Med Chem* **2003**, *3*, 39-53.
[361] W. Jahnke, Perspectives of biomolecular NMR in drug discovery: the blessing and curse of versatility, *J Biomol NMR* **2007**, *39*, 87-90.
[362] T. L. Hwang, A. J. Shaka, Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients, *J Magn Reson, Series A* **1995**, *112*, 275-279.
[363] a F. L. Hunefeld, *Die Chemismus in der Thierischen Organization*, Leipzig, **1840**; b A. Mcpherson, A Brief-History of Protein Crystal-Growth, *J Crystal Growth* **1991**, *110*, 1-10.
[364] A. McPherson, J. A. Gavira, Introduction to protein crystallization, *Acta Crystallogr F Struct Biol Commun* **2014**, *70*, 2-20.
[365] International Union of Crystallography, *Novel Prize winners associated with crystallography*, https://www.iucr.org/people/nobel-prize, accesses date 18th of March 2019
[366] J. B. Sumner, The Isolation and Crystallization of the Enzyme Urease, *J Biol Chem* **1926**, *69*, 435-441.
[367] a J. H. Northrop, Crystalline Pepsin : I. Isolation and Tests of Purity, *J Gen Physiol* **1930**, *13*, 739-766; b J. H. Northrop, Crystalline Pepsin : Ii. General Properties and Experimental Methods, *J Gen Physiol* **1930**, *13*, 767-780.
[368] W. M. Stanley, Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus, *Science* **1935**, *81*, 644-645.
[369] K. L. Manchester, The crystallization of enzymes and virus proteins: laying to rest the colloidal concept of living systems, *Endeavour* **2004**, *28*, 25-29.
[370] W. C. Röntgen, Ueber eine neue Art von Strahlen (Vorläufige Mitteilung). *Aus den Sizungsberichten der Würzburger Physik.-medic. Gesellschaft,* **Würzburg 1895**.
[371] M. Eckert, Disputed discovery: the beginnings of X-ray diffraction in crystals in 1912 and its repercussions, *Acta Crystallogr A* **2012**, *68*, 30-39.
[372] Y. Shi, A glimpse of structural biology through X-ray crystallography, *Cell* **2014**, *159*, 995-1014.
[373] A. L. Carvalho, J. Trincao, M. J. Romao, X-ray crystallography in drug discovery, *Methods Mol Biol* **2009**, *572*, 31-56.
[374] W. H. Bragg, W. L. Bragg, The Reflection of X-rays by Crystals, *Proc. R. Soc. London, Ser. A* **1913**, *88*, 428-438.
[375] J. C. Kendrew, G. Bodo, H. M. Dintzis, et al., A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis, *Nature* **1958**, *181*, 662-666.
[376] S. de Chadarevian, John Kendrew and myoglobin: Protein structure determination in the 1950s, *Protein Sci* **2018**, *27*, 1136-1143.
[377] a S. W. Wukovitz, T. O. Yeates, Why protein crystals favour some space-groups over others, *Nat Struct Biol* **1995**, *2*, 1062-1067; b T. Hahn, Editor, *International Tables for Crystallography, Volume A: Space-Group Symmetry; Fifth Edition*, Kluwer Academic Publishers, **2002**.
[378] D. E. Danley, Crystallization to obtain protein-ligand complexes for structure-aided drug design, *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 569-575.
[379] D. W. Green, V. M. Ingram, M. F. Perutz, et al., The structure of haemoglobin - IV. Sign determination by the isomorphous replacement method, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1954**, *225*, 287-307.
[380] N. Isaacs, A history of experimental phasing in macromolecular crystallography, *Acta Crystallogr D Struct Biol* **2016**, *72*, 293-295.
[381] A. M. Hassell, G. An, R. K. Bledsoe, et al., Crystallization of protein-ligand complexes, *Acta Crystallogr D Biol Crystallogr* **2007**, *63*, 72-79.
[382] H. Hope, Cryocrystallography of biological macromolecules: a generally applicable method, *Acta Crystallogr B* **1988**, *44 ( Pt 1)*, 22-26.
[383] J. R. Helliwell, *Macromolecular crystallography with synchrotron radiation*, Cambridge, **1992**.
[384] a J. Schiebel, S. G. Krimmer, K. Rower, et al., High-Throughput Crystallography: Reliable and Efficient Identification of Fragment Hits, *Structure* **2016**, *24*, 1398-1409; b A. Joachimiak, High-throughput crystallography for structural genomics, *Curr Opin Struct Biol* **2009**, *19*, 573-584.

[385] M. J. Hartshorn, C. W. Murray, A. Cleasby, et al., Fragment-based lead discovery using X-ray crystallography, *Journal of medicinal chemistry* **2005**, *48*, 403-413.
[386] a D. Stock, O. Perisic, J. Lowe, Robotic nanolitre protein crystallisation at the MRC Laboratory of Molecular Biology, *Prog Biophys Mol Biol* **2005**, *88*, 311-327; b F. Cipriani, M. Rower, C. Landret, et al., CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films, *Acta Crystallogr D Biol Crystallogr* **2012**, *68*, 1393-1399.
[387] a D. Nurizzo, M. W. Bowler, H. Caserotto, et al., RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments, *Acta Crystallogr D Struct Biol* **2016**, *72*, 966-975; b U. Zander, G. Hoffmann, I. Cornaciu, et al., Automated harvesting and processing of protein crystals through laser photoablation, *Acta Crystallogr D Struct Biol* **2016**, *72*, 454-466.
[388] a D. de Sanctis, A. Beteva, H. Caserotto, et al., ID29: a high-intensity highly automated ESRF beamline for macromolecular crystallography experiments exploiting anomalous scattering, *J Synchrotron Radiat* **2012**, *19*, 455-461; b M. W. Bowler, D. Nurizzo, R. Barrett, et al., MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules, *J Synchrotron Radiat* **2015**, *22*, 1540-1547.
[389] C. Broennimann, E. F. Eikenberry, B. Henrich, et al., The PILATUS 1M detector, *J Synchrotron Radiat* **2006**, *13*, 120-130.
[390] a A. A. McCarthy, R. Barrett, A. Beteva, et al., ID30B - a versatile beamline for macromolecular crystallography experiments at the ESRF, *J Synchrotron Radiat* **2018**, *25*, 1249-1260; b O. Svensson, S. Malbet-Monaco, A. Popov, et al., Fully automatic characterization and data collection from crystals of biological macromolecules, *Acta Crystallogr D Biol Crystallogr* **2015**, *71*, 1757-1767; c G. Winter, K. E. McAuley, Automated data collection for macromolecular crystallography, *Methods* **2011**, *55*, 81-93.
[391] a A. Sharff, C. Vonrhein, O. Smart, et al., Pipedream, version 1.2.1. Global Phasing Ltd, **2011**, *Cambridge, United Kingdom*; b T. G. Battye, L. Kontogiannis, O. Johnson, et al., iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM, *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 271-281; c W. Kabsch, Xds, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125-132; d K. M. Sparta, M. Krug, U. Heinemann, et al., Xdsapp2.0, *J Appl Crystallogr.* **2016**, *49*, 1085-1092; e D. G. Waterman, G. Winter, R. J. Gildea, et al., Diffraction-geometry refinement in the DIALS framework, *Acta Crystallogr D Struct Biol* **2016**, *72*, 558-575.
[392] a T. Krojer, R. Talon, N. Pearce, et al., The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination, *Acta Crystallogr D Struct Biol* **2017**, *73*, 267-278; b W. T. Mooij, M. J. Hartshorn, I. J. Tickle, et al., Automated protein-ligand crystallography for structure-based drug design, *ChemMedChem* **2006**, *1*, 827-838.
[393] a N. Pearce, A. R. Bradley, P. Collins, et al., A Multi-Crystal Method for Extracting Obscured Signal from Crystallographic Electron Density, *bioRxiv* **2016**; b N. M. Pearce, T. Krojer, A. R. Bradley, et al., A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density, *Nat Commun* **2017**, *8*, 15123.
[394] N. Echols, N. Morshed, P. V. Afonine, et al., Automated identification of elemental ions in macromolecular crystal structures, *Acta Crystallogr D Biol Crystallogr* **2014**, *70*, 1104-1114.
[395] a H. R. Powell, X-ray data processing, *Biosci Rep* **2017**, *37*, BSR20170227; b A. T. Brunger, Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature* **1992**, *355*, 472-475.
[396] M. C. Deller, B. Rupp, Models of protein-ligand crystal structures: trust, but verify, *J Comput Aided Mol Des* **2015**, *29*, 817-836.
[397] D. E. Scott, A. G. Coyne, S. A. Hudson, et al., Fragment-based approaches in drug discovery and chemical biology, *Biochemistry* **2012**, *51*, 4990-5003.
[398] N. M. Pearce, T. Krojer, F. von Delft, Proper modelling of ligand binding requires an ensemble of bound and unbound states, *Acta Crystallogr D Struct Biol* **2017**, *73*, 256-266.
[399] a P. J. McIntyre, P. M. Collins, L. Vrzal, et al., Characterization of Three Druggable Hot-Spots in the Aurora-A/TPX2 Interaction Using Biochemical, Biophysical, and Fragment-Based Approaches, *ACS Chem Biol* **2017**, *12*, 2906-2914; b D. A. Keedy, Z. B. Hill, J. T. Biel, et al., An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering, *Elife* **2018**, *7*, e36307.
[400] Diamond light source, *Fragment Screening - XChem*, www.diamond.ac.uk/Instruments/Mx/Fragment-Screening.html, accesses date 10. of October 2017
[401] U. Mueller, M. Ühlein, J. Linnik, et al., *Fragment-screening at the BESSY II MX-beamlines of the Helmholtz-Zentrum Berlin, Vol. 70*, **2014**.
[402] a J. Jancarik, S. H. Kim, Sparse-Matrix Sampling - a Screening Method for Crystallization of Proteins, *Journal of Applied Crystallography* **1991**, *24*, 409-411; b T. S. Walter, J. M. Diprose, C. J. Mayo, et al., A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization, *Acta Crystallogr D Biol Crystallogr* **2005**, *61*, 651-657.
[403] T. Bergfors, *Protein Crystallization: Techniques, Strategies, and Tips*, International University Line, La Jolla California USA, **1999**.

[404] A. D'Arcy, A. Mac Sweeney, A. Haber, Using natural seeding material to generate nucleation in protein crystallization experiments, *Acta Crystallogr D Biol Crystallogr* **2003**, *59*, 1343-1346.
[405] a G. C. Ireton, B. L. Stoddard, Microseed matrix screening to improve crystals of yeast cytosine deaminase, *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 601-605; b A. D'Arcy, T. Bergfors, S. W. Cowan-Jacob, et al., Microseed matrix screening for optimization in protein crystallization: what have we learned?, *Acta Crystallogr F Struct Biol Commun* **2014**, *70*, 1117-1126.
[406] J. R. Luft, G. T. DeTitta, A method to produce microseed stock for use in the crystallization of biological macromolecules, *Acta Crystallogr D Biol Crystallogr* **1999**, *55*, 988-993.
[407] a B. W. Matthews, Solvent content of protein crystals, *J Mol Biol* **1968**, *33*, 491-497; b B. W. Matthews, Determination of molecular weight from protein crystals, *J Mol Biol* **1974**, *82*, 513-526.
[408] L. Z. Vilenchik, J. P. Griffith, N. St Clair, et al., Protein crystals as novel microporous materials, *J. Am. Chem. Soc.* **1998**, *120*, 4290-4294.
[409] I. Muller, Guidelines for the successful generation of protein-ligand complex crystals, *Acta Crystallogr D Struct Biol* **2017**, *73*, 79-92.
[410] N. Fuller, L. Spadola, S. Cowen, et al., An improved model for fragment-based lead generation at AstraZeneca, *Drug Discov Today* **2016**, *21*, 1272-1283.
[411] D. A. Erlanson, B. J. Davis, W. Jahnke, Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures, *Cell Chem Biol* **2019**, *26*, 9-15.
[412] a G. C. Terstappen, C. Schlupen, R. Raggiaschi, et al., Target deconvolution strategies in drug discovery, *Nat Rev Drug Discov* **2007**, *6*, 891-903; b M. Congreve, G. Chessari, D. Tisi, et al., Recent developments in fragment-based drug discovery, *Journal of medicinal chemistry* **2008**, *51*, 3661-3680.
[413] W. P. Jencks, On the attribution and additivity of binding energies, *Proc Natl Acad Sci U S A* **1981**, *78*, 4046-4050.
[414] J. D. Chodera, D. L. Mobley, Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design, *Annu Rev Biophys* **2013**, *42*, 121-142.
[415] R. L. M. van Montfort, P. Workman, Structure-based drug design: aiming for a perfect fit, *Essays Biochem* **2017**, *61*, 431-437.
[416] a D. C. Rees, M. Congreve, C. W. Murray, et al., Fragment-based lead discovery, *Nat Rev Drug Discov* **2004**, *3*, 660-672; b D. A. Erlanson, R. S. McDowell, T. O'Brien, Fragment-based drug discovery, *Journal of medicinal chemistry* **2004**, *47*, 3463-3482.
[417] A. Wlodawer, Rational approach to AIDS drug design through structural biology, *Annu Rev Med* **2002**, *53*, 595-614.
[418] M. Kaur, R. K. Rawal, G. Rath, et al., Structure Based Drug Design: Clinically Relevant HIV-1 Integrase Inhibitors, *Curr Top Med Chem* **2018**, *18*, 2664-2680.
[419] Y. Liu, J. Zhang, W. Xu, Recent progress in rational drug design of neuraminidase inhibitors, *Curr Med Chem* **2007**, *14*, 2872-2891.
[420] O. Ichihara, J. Barker, R. J. Law, et al., Compound Design by Fragment-Linking, *Mol Inform* **2011**, *30*, 298-306.
[421] A. J. Souers, J. D. Leverson, E. R. Boghaert, et al., ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets, *Nat Med* **2013**, *19*, 202-208.
[422] a P. J. Hajduk, SAR by NMR: putting the pieces together, *Mol Interv* **2006**, *6*, 266-272; b M. Nazare, H. Matter, D. W. Will, et al., Fragment deconstruction of small, potent factor Xa inhibitors: exploring the superadditivity energetics of fragment linking in protein-ligand complexes, *Angew Chem Int Ed Engl* **2012**, *51*, 905-911.
[423] a G. Bollag, J. Tsai, J. Zhang, et al., Vemurafenib: the first drug approved for BRAF-mutant cancer, *Nat Rev Drug Discov* **2012**, *11*, 873-886; b J. Tsai, J. T. Lee, W. Wang, et al., Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity, *Proc Natl Acad Sci U S A* **2008**, *105*, 3041-3046.
[424] S. Howard, V. Berdini, J. A. Boulstridge, et al., Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity, *Journal of medicinal chemistry* **2009**, *52*, 379-388.
[425] S. Roughley, L. Wright, P. Brough, et al., Hsp90 inhibitors and drugs from fragment and virtual screening, *Top Curr Chem* **2012**, *317*, 61-82.
[426] R. E. Hubbard, Fragment approaches in structure-based drug discovery, *J Synchrotron Radiat* **2008**, *15*, 227-230.
[427] A. J. Massey, J. Schoepfer, P. A. Brough, et al., Preclinical antitumor activity of the orally available heat shock protein 90 inhibitor NVP-BEP800, *Mol Cancer Ther* **2010**, *9*, 906-919.
[428] S. M. Bertrand, N. Ancellin, B. Beaufils, et al., The Discovery of in Vivo Active Mitochondrial Branched-Chain Aminotransferase (BCATm) Inhibitors by Hybridizing Fragment and HTS Hits, *Journal of medicinal chemistry* **2015**, *58*, 7140-7163.
[429] a G. Klebe, Applying thermodynamic profiling in lead finding and optimization, *Nat Rev Drug Discov* **2015**, *14*, 95-110; b J. E. Ladbury, G. Klebe, E. Freire, Adding calorimetric data to decision making in lead discovery: a hot tip, *Nat Rev Drug Discov* **2010**, *9*, 23-27.

[430] a A. L. Hopkins, C. R. Groom, A. Alex, Ligand efficiency: a useful metric for lead selection, *Drug Discov Today* **2004**, *9*, 430-431; b A. L. Hopkins, G. M. Keseru, P. D. Leeson, et al., The role of ligand efficiency metrics in drug discovery, *Nat Rev Drug Discov* **2014**, *13*, 105-121.
[431] M. Majewski, S. Ruiz-Carmona, X. Barril, Dynamic Undocking: A Novel Method for Structure-Based Drug Discovery, *Methods Mol Biol* **2018**, *1824*, 195-215.
[432] S. Shaabani, C. G. Neochoritis, A. Twarda-Clapa, et al., Scaffold hopping via ANCHOR.QUERY: beta-lactams as potent p53-MDM2 antagonists(dagger), *Medchemcomm* **2017**, *8*, 1046-1052.
[433] a D. Kozakov, L. E. Grove, D. R. Hall, et al., The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins, *Nat Protoc* **2015**, *10*, 733-755; b C. H. Ngan, T. Bohnuud, S. E. Mottarella, et al., FTMAP: extended protein mapping with user-selected probe molecules, *Nucleic Acids Res* **2012**, *40*, W271-275.
[434] D. Alvarez-Garcia, X. Barril, Molecular simulations with solvent competition quantify water displaceability and provide accurate interaction maps of protein binding sites, *Journal of medicinal chemistry* **2014**, *57*, 8530-8539.
[435] *Hit generation Keymical Fragments*, http://www.edelris.com/solutions-services/hit-generation, accesses date 29. April 2009
[436] Enamine, *Enamine Golden Fragment Library*, https://enamine.net/download/FL/Enamine_Golden_Fragment_Library.pdf, accesses date 29. April 2019
[437] G. W. Bemis, M. A. Murcko, The properties of known drugs. 1. Molecular frameworks, *Journal of medicinal chemistry* **1996**, *39*, 2887-2893.
[438] Gasteiger E., Hoogland C., Gattiker A., et al., in *The Proteomics Protocols Handbook* (Ed.: J. M. Walker), Humana Press, Totowa, NJ, **2005**, pp. 571-607.
[439] P. Kraft, A. Bergamaschi, C. Broennimann, et al., Performance of single-photon-counting PILATUS detector modules, *J Synchrotron Radiat* **2009**, *16*, 368-375.
[440] P. R. Evans, G. N. Murshudov, How good are my data and what is the resolution?, *Acta Crystallogr D Biol Crystallogr* **2013**, *69*, 1204-1214.
[441] C. Vonrhein, C. Flensburg, P. Keller, et al., Data processing and analysis with the autoPROC toolbox, *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 293-302.
[442] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, et al., Phaser crystallographic software, *J Appl Crystallogr* **2007**, *40*, 658-674.
[443] M. Wojdyr, R. Keegan, G. Winter, et al., *DIMPLE - a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands, Vol. 69*, **2013**.
[444] aP. Emsley, B. Lohkamp, W. G. Scott, et al., Features and development of Coot, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 486-501; bP. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 2126-2132.
[445] O. S. Smart, T. O. Womack, C. Flensburg, et al., Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER, *Acta Crystallogr D Biol Crystallogr* **2012**, *68*, 368-380.
[446] G. J. Kleywegt, A. T. Brunger, Checking your imagination: applications of the free R value, *Structure* **1996**, *4*, 897-904.
[447] J. M. Grimes, D. R. Hall, A. W. Ashton, et al., Where is crystallography going?, *Acta Crystallogr D Struct Biol* **2018**, *74*, 152-166.
[448] P. M. Collins, J. T. Ng, R. Talon, et al., Gentle, fast and effective crystal soaking by acoustic dispensing, *Acta Crystallogr D Struct Biol* **2017**, *73*, 246-255.
[449] J. T. Ng, C. Dekker, M. Kroemer, et al., Using textons to rank crystallization droplets by the likely presence of crystals, *Acta Crystallogr D Biol Crystallogr* **2014**, *70*, 2702-2718.
[450] G. Winter, xia2: an expert system for macromolecular crystallography data reduction, *Journal of Applied Crystallography* **2010**, *43*, 186-190.
[451] P. Evans, Scaling and assessment of data quality, *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 72-82.
[452] G. N. Murshudov, P. Skubak, A. A. Lebedev, et al., REFMAC5 for the refinement of macromolecular crystal structures, *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 355-367.
[453] M. D. Winn, C. C. Ballard, K. D. Cowtan, et al., Overview of the CCP4 suite and current developments, *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 235-242.
[454] S. Delageniere, P. Brenchereau, L. Launer, et al., ISPyB: an information management system for synchrotron macromolecular crystallography, *Bioinformatics* **2011**, *27*, 3186-3192.
[455] P. D. Adams, P. V. Afonine, G. Bunkoczi, et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213-221.
[456] F. Long, R. A. Nicholls, P. Emsley, et al., AceDRG: a stereochemical description generator for ligands, *Acta Crystallogr D Struct Biol* **2017**, *73*, 112-122.
[457] O. Smart, T. O. Womack, A. Sharff, et al., Grade, version 1.2.9. Cambridge, United Kingdom: Global Phasing Ltd, **2011**, *Cambridge, United Kingdom*.
[458] N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation, *Acta Crystallogr D Biol Crystallogr* **2009**, *65*, 1074-1080.

[459] M. F. Incardona, G. P. Bourenkov, K. Levik, et al., EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis, *J Synchrotron Radiat* **2009**, *16*, 872-879.
[460] S. Monaco, E. Gordon, M. W. Bowler, et al., Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF, *J Appl Crystallogr* **2013**, *46*, 804-810.
[461] O. Smart, T. O. Womack, A. Sharff, et al., RHOFIT, version 1.2.4. Cambridge, United Kingdom: Global Phasing Ltd, **2014**, *Cambridge, United Kingdom.*
[462] D. R. Koes, A. Domling, C. J. Camacho, AnchorQuery: Rapid online virtual screening for small-molecule protein-protein interaction inhibitors, *Protein Sci* **2018**, *27*, 229-232.
[463] G. M. Sastry, M. Adzhigirey, T. Day, et al., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, *J Comput Aided Mol Des* **2013**, *27*, 221-234.
[464] a R. A. Friesner, R. B. Murphy, M. P. Repasky, et al., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes, *Journal of medicinal chemistry* **2006**, *49*, 6177-6196; b R. A. Friesner, J. L. Banks, R. B. Murphy, et al., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *Journal of medicinal chemistry* **2004**, *47*, 1739-1749; c T. A. Halgren, R. B. Murphy, R. A. Friesner, et al., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *Journal of medicinal chemistry* **2004**, *47*, 1750-1759.
[465] J. P. Ebejer, G. M. Morris, C. M. Deane, Freely available conformer generation methods: how good are they?, *Journal of chemical information and modeling* **2012**, *52*, 1146-1158.
[466] D. A. Case, T. A. Darden, T. E. Cheatham, et al., Amber 10. University of California., **2008**.
[467] R. U. Kumar, K. H. V. Reddy, B. S. P. A. Kumar, et al., Metal free amination of 2-chloroazoles in aqueous medium, *Tetrahedron Letters* **2016**, *57*, 637-640.
[468] aA. F. Abdel-Magid, S. J. Mehrman, A review on the use of sodium triacetoxyborohydride in the reductive amination of ketones and aldehydes, *Organic Process Research & Development* **2006**, *10*, 971-1031; bA. F. Abdel-Magid, K. G. Carson, B. D. Harris, et al., Reductive Amination of Aldehydes and Ketones with Sodium Triacetoxyborohydride. Studies on Direct and Indirect Reductive Amination Procedures(1), *J Org Chem* **1996**, *61*, 3849-3862.
[469] V. U. Jeankumar, R. S. Reshma, R. Vats, et al., Engineering another class of anti-tubercular lead: Hit to lead optimization of an intriguing class of gyrase ATPase inhibitors, *Eur J Med Chem* **2016**, *122*, 216-231.
[470] F. W. Studier, Protein production by auto-induction in high density shaking cultures, *Protein Expr Purif* **2005**, *41*, 207-234.
[471] A. D'Arcy, F. Villard, M. Marsh, An automated microseed matrix-screening method for protein crystallization, *Acta Crystallogr D Biol Crystallogr* **2007**, *63*, 550-554.
[472] M. Mueller, M. Wang, C. Schulze-Briese, Optimal fine phi-slicing for single-photon-counting pixel detectors, *Acta Crystallogr D Biol Crystallogr* **2012**, *68*, 42-56.
[473] L. Messina, J. A. Gavira, S. Pernagallo, et al., Identification and characterization of a bacterial hyaluronidase and its production in recombinant form, *FEBS Lett* **2016**, *590*, 2180-2189.
[474] Z. W. Chen, C. Vignaud, A. Jaafar, et al., High resolution crystal structure of rat long chain hydroxy acid oxidase in complex with the inhibitor 4-carboxy-5-[(4-chlorophenyl)sulfanyl]-1, 2, 3-thiadiazole. Implications for inhibitor specificity and drug design, *Biochimie* **2012**, *94*, 1172-1179.
[475] C. Browning, M. M. Shneider, V. D. Bowman, et al., Phage pierces the host cell membrane with the iron-loaded spike, *Structure* **2012**, *20*, 326-339.
[476] J. A. Cummings, A. A. Fedorov, C. Xu, et al., Annotating enzymes of uncertain function: the deacylation of D-amino acids by members of the amidohydrolase superfamily, *Biochemistry* **2009**, *48*, 6469-6481.
[477] E. Thomassen, G. Gielen, M. Schutz, et al., The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold, *J Mol Biol* **2003**, *331*, 361-373.
[478] S. B. Needleman, C. D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, *J Mol Biol* **1970**, *48*, 443-453.
[479] a H. Ashkenazy, S. Abadi, E. Martz, et al., ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, *Nucleic Acids Res* **2016**, *44*, W344-350; b M. Landau, I. Mayrose, Y. Rosenberg, et al., ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures, *Nucleic Acids Res* **2005**, *33*, W299-302.
[480] F. Jeanmougin, J. D. Thompson, M. Gouy, et al., Multiple sequence alignment with Clustal X, *Trends Biochem Sci* **1998**, *23*, 403-405.
[481] K. Stierand, M. Rarey, Drawing the PDB: Protein-Ligand Complexes in Two Dimensions, *ACS Med Chem Lett* **2010**, *1*, 540-545.
[482] F. R. Ehrmann, J. Stojko, A. Metz, et al., Soaking suggests "alternative facts": Only co-crystallization discloses major ligand-induced interface rearrangements of a homodimeric tRNA-binding protein indicating a novel mode-of-inhibition, *PloS one* **2017**, *12*, e0175723-e0175723.
[483] A. C. Anderson, The process of structure-based drug design, *Chem Biol* **2003**, *10*, 787-797.
[484] H. Jhoti, G. Williams, D. C. Rees, et al., The 'rule of three' for fragment-based drug discovery: where are we now?, *Nat Rev Drug Discov* **2013**, *12*, 644-645.

[485] a I. Boscaro-Clarke, L. Cuddon, Enamine supplies DSI poised fragment and analogue libraries to Diamond Light Source XChem facility and SGC Oxford screening efforts [Press release]. Retrieved from https://enamine.net/public/press-release/Enamine-Diamond-Light-Source-SGC-PR-January-09th-2018.pdf, **9. January 2018**; b Enamine, *Real Compound Libraries*, https://enamine.net/library-synthesis/real-compounds/real-compound-libraries, accesses date 07. August 2019

[486] A. Domling, W. Wang, K. Wang, Chemistry and biology of multicomponent reactions, *Chem Rev* **2012**, *112*, 3083-3135.

[487] E. Kroon, J. O. Schulze, E. Suss, et al., Discovery of a Potent Allosteric Kinase Modulator by Combining Computational and Synthetic Methods, *Angew Chem Int Ed Engl* **2015**, *54*, 13933-13936.

[488] S. Shaaban, B. F. Abdel-Wahab, Groebke-Blackburn-Bienayme multicomponent reaction: emerging chemistry for drug discovery, *Mol Divers* **2016**, *20*, 233-254.

[489] a I. Ugi, R. Meyr, U. Fetzer, et al., Versuche mit isonitrilen, *Angewandte Chemie-International Edition* **1959**, *71*, 386-386; b C. G. Neochoritis, T. Zhao, A. Domling, Tetrazoles via Multicomponent Reactions, *Chem Rev* **2019**, *119*, 1970-2042.

[490] M. Keenan, P. W. Alexander, J. H. Chaplin, et al., Selection and optimization of hits from a high-throughput phenotypic screen against Trypanosoma cruzi, *Future Med Chem* **2013**, *5*, 1733-1752.

[491] M. Keenan, J. H. Chaplin, A new era for chagas disease drug discovery?, *Prog Med Chem* **2015**, *54*, 185-230.

[492] M. N. Drwal, G. Bret, C. Perez, et al., Structural Insights on Fragment Binding Mode Conservation, *Journal of medicinal chemistry* **2018**, *61*, 5963-5973.

[493] A. V. Onufriev, E. Alexov, Protonation and pK changes in protein-ligand binding, *Q Rev Biophys* **2013**, *46*, 181-209.

[494] M. Sanchez-Sanchez, G. Rivera, E. A. Garcia, et al., Therapeutic Targets for the Development of Anti-Trypanosoma Cruzi Drugs: A Brief Review, *Mini-Reviews in Organic Chemistry* **2016**, *13*, 227-243.

[495] F. Sievers, A. Wilm, D. Dineen, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol* **2011**, *7*, 539.

[496] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, *Nucleic Acids Research* **2014**, *42*, W320-W324.

# Acknowledgements

This work would not have been possible without the support, guidance and help I received from many people during my time as a PhD student. Thanks to all of you!

In particular, I would like to thank:
My supervisor, Dr. Wolfgang Jahnke, Novartis Pharma AG, Basel, Switzerland, for his support and guidance and for giving me the great opportunity to work on such a versatile, interdisciplinary and challenging project. I will never forget his call on a late Monday evening in January 2016 when he offered me the PhD position. Dr. Andreas Marzinzik, Novartis Pharma AG, Basel, Switzerland, for his support when starting medicinal chemistry and for all the motivation and many encouraging words. Thank you very much for the time you both invested in our regular meetings over the past three and a half years.
Prof. Dr. Gerhard Klebe, Philipps-Universität Marburg, Germany, for being my academic supervisor and for his constant interest in my PhD project throughout the last three and a half years.
Prof. Dr. Carsten Culmsee and Prof. Dr. Moritz Bünemannn, Philipps-Universität Marburg, Germany, for agreeing to be members of my thesis committee and for taking their time to evaluate my thesis.
My PhD colleague Lena Münzker, who shared the last three and a half years with me. Thank you for sitting in the same boat. It was worth priceless that at least one of us stayed motivated. Lena, ich danke dir!

This project has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Grant Agreement No. 675555, Accelerated Early staGe drug discovery (AEGIS) and under the infrastructure for NMR, EM and X-rays for translational Research (iNEXT) grant agreement ID 653706 (project number 2847).

I am very grateful for the AEGIS Marie Skłodowska Curie Fellowship and the chance to be part of such a multidisciplinary and multicultural scientific network. I thank all the collaborators within the AEGIS consortium, especially the ESRs Francesca Magari, Giulia Opassi, Markella Konstantinidou and Ryan Byrne, the project instructors Prof. Helena Danielson, Uppsala University, Sweden, and Prof. Dr. Alexander Dömling, University Groningen, the Netherlands, as well as the coordinator Prof. Dr. Michael Sattler and the project manager Dr. Eva Schlosser, Helmholtz Zentrum Munich, Germany.

I thank my collaborators at the Diamond Light source in Oxford, UK, especially Elliot Nelson and Prof. Frank von Delft, as well as Dr. Carien Dekker (Novartis) for initiating the great cooperation.

The results obtained had breakthrough character for my work. In addition, I thank my collaborators at the HTX lab in Grenoble, France, especially Dr. Damien Clavel, Dr. Irina Cornaciu and Dr. Jose Marquez, who made the 2nd FBS by X-ray crystallography campaign possible. Many thanks to the beamline staff at the European Synchrotron Facility, the Diamond Light Source and the Swiss Light Source for their outstanding work and support.

I would like to thank all the former and present members of the Chemical Biology and Therapeutics (CBT) group and the Global Discovery Chemistry (GDC), who in various ways contributed to my personal and professional time at Novartis. I had the great pleasure to work with or alongside amazing colleagues who have been a source of good advice. I would especially like to mention:
Dr. Alvar Gossert, Dr. Felix Freuler, Simon Haenni, Valentina Tundo, Cecile Delmas and Julia Klopp for plasmid design, training in protein expression and/or purification and invaluable help in the biolabs.
Rene Hemmig and Daniela Scherer-Becker for assistance in acquiring HPLC results.
Dr. Alvar Gossert, Dr. Lukasz Skora, Andrea Wiget and Chrystèlle Henry for sharing their extensive knowledge in NMR spectroscopy.
Dr. Jean-Michel Rondeau for passing on his knowledge on crystallography and Sylvie Lehmann, Aude Izaac and Frederic Villard for helping me to set up my first crystallization experiments.
Dr. Markus Krömer and Dr. Christian Schleberger for their thorough help with software and scripting related to X-ray data.
Dr. Rainer Wilcken for his ideas for fragment follow-up and modelling.
Dr. Christian Bergsdorf for his help to get started with SPR experiments.
Dr. Sebastién Ripoche for the great support in the chemistry lab and sharring his work bench when running a reaction overnight.
David Papale and Adriano Marra for being there whenever my computer failed to do what I wanted.
Sarah Barbet and Lydia Hartleben for assistance in dealing with bureaucracy and any paper work within Novartis and at the Philipps-Universität Marburg, respectively.
Thanks to all my proof readers whose helpful suggestions increased readability and reduced ambiguity.

Lukasz, Matthias, Sara, Katharina, Claudia, Anna, Ole, it was so great that you were there! Thanks for staying late in the lab and for all the fun and good conversations we had. If only we had all been there at the same time!

Very special thanks goes to my family and friends for their continuous support despite the many kilometres in between. Thanks to their tolerance, understanding and love, this journey became possible. Finally, I thank my husband Björn for all his love, the strength to put up with me and to always stand by my side.

# Curriculum Vitae

Aus Gründen des Persönlichkeitsschutzes wird von der elektronischen Veröffentlichung des Lebenslaufes abgesehen.

# Conference participation

06/2018 **Oral communication**

5th Novalix conference: Biophysics in Drug discovery 2018, 13th – 15th June 2018, Boston, US,

- „ Novel active site and allosteric site binders identified for FPPS of *T. cruzi*“

11/2017 **Oral communication**

Revolutions in Structural Biology: Celebrating the 100th Anniversary of Sir John Kendrew, 16th – 17th Nov 2017, EMBL Heidelberg, Germany,

- „Targeting *T. cruzi* FPPS by fragment based drug discovery“

11/2017 **Poster presentation**

Revolutions in Structural Biology: Celebrating the 100th Anniversary of Sir John Kendrew, 16th – 17th Nov 2017EMBL Heidelberg, Germany,

- „Targeting *T. cruzi* FPPS by fragment based drug discovery“
- Petrick, J., Muenzker, L., Klebe, G., Marzinzik, K., Jahnke, W.

---

03/2019 **Poster**

7th RSC-BMCS Fragment-based Drug Discovery meeting, 24th – 26th March 2019, Cambridge, United Kingdom

- „ Fragment-based discovery of novel active and allosteric site binders of *T. brucei* farnesyl pyrophosphate synthase“
- Muenzker, L., Petrick, J., Klebe, G., Marzinzik, K., Jahnke, W.

07/2018 **Poster**

3rd Integrative Structural Biology School, Institut Pasteur, 16th – 21th July 2018, Paris, France

- „ Fragment-based discovery of novel active and allosteric site binders of *T. brucei* farnesyl pyrophosphate synthase“
- Muenzker, L., Petrick, J., Klebe, G., Marzinzik, K., Jahnke, W.

09/2017 **Poster**

9th International Conference on Structural Biology, 8th – 20th September 2017, Zurich, Switzerland

- „Targeting *Trypanosoma brucei* FPPS by fragment-based drug discovery“
- Muenzker, L., Petrick, J., Klebe, G., Marzinzik, K., Jahnke, W.

# Erklärung

Ich versichere, dass ich meine Dissertation

**„ Targeting farnesyl pyrophosphate synthase**
**of *Trypanosoma cruzi* by fragment-based lead discovery “**

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

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

Marburg, den _____ _____

Joy Petrick