TECHNISCHE UNIVERSITÄT MÜNCHEN
Fakultät für Chemie
Bayerisches NMR–Zentrum
Lehrstuhl für biomolekulare NMR Spektroskopie

Development and application of NMR and X-ray crystallography in early-stage drug discovery

(Entwicklung und Anwendung von Kernspinresonanz-Spektroskopie und Röntgenstrukturanalyse in früher Wirkstoffentwicklung)

CHARLOTTE ANN SOFTLEY

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzende(r): Prof. Dr. Bernd Reif

Prüfer der Dissertation:
1. Prof. Dr. Michael Sattler
2. Prof. Dr. Kathrin Lang

Die Dissertation wurde am 06.07.20 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 29.07.20 angenommen.

Table of Contents

Abstract

The availability of a huge range of drugs for treating various diseases has become normal in our current daily life. However, the process behind the development of new drugs still holds many challenges. As antibiotic resistance increases, new drugs must continuously be developed, and many illnesses still lack suitable medical cures, particularly neglected tropical diseases and genetic conditions. This thesis covers the structural characterisation and inhibitor screening of two important targets related to infectious diseases: the antibiotic-resistance protein IMP-13 and the trypanosomal PEX14 protein, implicated in the neglected tropical disease, Chagas Disease. The challenges in drug discovery are many-fold, and some of these are highlighted and addressed in this thesis, along with methodological developments to overcome them.

Metallo-β-lactamases lead to antibiotic resistance in a range of bacterial pathogens that both affect humans and act as natural reservoirs in soil and animals. They render current antibiotics of last resort, the carbapenems, useless, whilst being co-expressed with other resistance factors. Fragment-based drug discovery is an efficient process for the rational development of small molecule inhibitors. NMR fragment screens readily identify hits but these ligands are often are not visible in crystallographic structures. To get around this problem, the metallo-β-lactamase target IMP-13 was crystallised with its natural substrates, the carbapenems. Subsequent structure determination highlighted key interactions and the recognition of the ligand by the protein, enabling further work to deduce the mode of binding of the fragments. It also suggested a new screening principle, which has since led to the discovery of fragments that stabilise a specific protein conformation and yielded further crystallographic structures.

PEX14 is a peroxisomal membrane-associated protein with an N-terminal domain that forms protein-protein interactions with the cargo receptor PEX5. PEX5 binds to cargo proteins that are, through the PEX14-PEX5 interactions, translocated into the peroxisomal lumen. PEX14 has been shown to be essential for survival of *Trypanosoma* parasites, such as *T. brucei*, which causes African Sleeping Sickness. Structural studies of PEX14 of *T. cruzi*, which causes Chagas Disease remained elusive, despite the high medical need for Chagas therapy and its promising role in drug discovery. The crystal structure of *T. cruzi* PEX14 N-terminal domain is presented along with solution-state NMR assignments. NMR fragment screening experiments were applied to identify starting points for inhibitor design. However, difficulties in obtaining structural information on fragment-protein complexes using standard crystallographic and NMR techniques led to investigation of the utility of paramagnetic NMR to characterise ligand poses.

A review of the current methods and applications of paramagnetic NMR in drug discovery is presented. Various methods for conjugating paramagnetic lanthanide binding tags were studied and the utility of structural information from pseudocontact shift and paramagnetic relaxation enhancements for structural analysis of PEX14-ligand complexes is assessed. Lanthanide tags were applied to PEX14 and a cysteine mutant and measured to gain restraints for modelling binding poses of fragments.

Water molecules are known to play an important role in binding properties of protein-ligand complexes, as well as the dynamics and folding of proteins themselves, giving valuable information for drug discovery. Whilst high-resolution crystal structures give information on water molecule placement, this is lacking from low-resolution crystal structures, NMR and cryo-EM. A deep learning

algorithm has been trained on high-resolution crystal structures and predict positions of water molecules on NMR and cryo-EM structures. The approach has been validated with experimental structures of IMP-13 and the *T. cruzi* PEX14 N-terminal domain.

The results presented in this thesis provide novel insight and methodology to address current challenges in early stage drug discovery, and are readily applicable to many systems.

Zusammenfassung

Die Verfügbarkeit von Medikamenten zur Behandlung verschiedener Krankheiten ist in westlichen Gesundheitssystemen zur Normalität geworden. Allerdings birgt der Prozess der Medikamentenentwicklung viele Herausforderungen. Da die Antibiotikaresistenz zunimmt, müssen ständig neue Medikamente entwickelt werden, und bei vielen Krankheiten fehlt es noch immer an geeigneten Wirkstoffen, insbesondere bei vernachlässigten Tropenkrankheiten und genetischen Erkrankungen. Diese Arbeit befasst sich mit der strukturellen Charakterisierung und dem Inhibitor-Screening von zwei wichtigen Targets im Zusammenhang mit Infektionskrankheiten: IMP-13, einem Antibiotikaresistenzprotein und PEX14, einem trypanosomalen Protein, das bei der vernachlässigten Tropenkrankheit Chagas-Krankheit eine Rolle spielt. Die Herausforderungen in der Wirkstoffentwicklung sind vielfältig, und einige davon werden in dieser Arbeit hervorgehoben und angesprochen, zusammen mit methodischen Entwicklungen zu ihrer Bewältigung.

Metallo-β-Laktamasen führen zu Antibiotikaresistenzen bei einer Reihe von bakteriellen Krankheitserregern, die sowohl den Menschen betreffen als auch als natürliche Reservoirs in Boden und Tieren wirken. Sie machen die derzeitigen Antibiotika der letzten Instanz, die Carbapeneme, wirkungslos, und sie werden mit anderen Resistenzfaktoren koexprimiert. Die fragmentbasierte Arzneimittelentwicklung ist eine effiziente Methode für die rationale Entwicklung von Inhibitoren. Durch Fragmentscreens mittels Kernspinresonanz Spektroskopie lassen sich Treffer leicht identifizieren, aber diese Liganden sind in Kristallstrukturen oft nicht sichtbar. Um dieses Problem zu umgehen, wurde das Metallo-β-Lactamase IMP-13 mit seinen natürlichen Substraten, den Carbapenemen, kristallisiert. Bei der anschließenden Strukturbestimmung wurden Schlüsselinteraktionen und die Erkennung der Liganden durch das Protein hervorgehoben, so dass weitere Arbeit zur Ableitung des Bindungsmodus der Fragmente möglich waren. Sie schlug auch ein neues Screening-Prinzip vor, das seither zur Entdeckung von Fragmenten geführt hat, die eine bestimmte Proteinkonformation stabilisieren und weitere Kristallstrukturen hervorbrachten.

PEX14 ist ein peroxisomales, membranassoziiertes Protein mit einer N-terminalen Domäne, die Protein-Protein-Interaktionen mit dem Cargo-Rezeptor PEX5 bildet. PEX5 bindet an Cargo-Proteine, die in Abhängigkeit von der PEX14-PEX5-Interaktion in das peroxisomale Lumen transloziert werden. Es hat sich gezeigt, dass PEX14 für das Überleben von Trypanosoma-Parasiten, wie z.B. *T. brucei*, dem Erreger der Afrikanischen Schlafkrankheit, von entscheidender Bedeutung ist. Strukturelle Studien zu PEX14 von *T. cruzi*, dem Erreger der Chagas-Krankheit, blieben trotz des hohen medizinischen Bedarfs für eine Chagas-Therapie und ihrer vielversprechenden Rolle bei der Entdeckung von Medikamenten schwer zugänglich. Die Kristallstruktur der N-terminalen Domäne von PEX14 von *T. cruzi* und eine Zuordnung der NMR Signale werden hier vorgestellt. NMR-Fragmentscreening-Experimente wurden angewandt, um Ausgangspunkte für die Inhibitorenentwicklung zu identifizieren. Schwierigkeiten bei der Gewinnung struktureller Informationen über Fragment-Proteinkomplexe mit Hilfe von oft benutzten Kristallographie- und NMR-Techniken führten zur Untersuchung von paramagnetischen NMR Experimenten zur Charakterisierung der Ligandenpositionen.

Ein Übersichtsartikel beschreibt aktuelle Methoden und Anwendungen von paramagnetischer NMR-Spektroskopie in der Wirkstoffentwicklung. Für die Strukturanalyse von PEX14-Inhibitorkomplexen wurden verschiedene Methoden zur Konjugierung paramagnetischer Lanthanid-Bindungsmarken untersucht und der Nutzen struktureller Informationen aus *Pseudocontact Shifts* und *Paramagnetic*

Relaxation Enhancement bewertet. Lanthanid-Tags wurden auf PEX14 und eine Cysteinmutante davon aufgetragen und gemessen, um strukturelle Information für die Modellierung der Bindung der Fragmente zu erhalten.

Wassermoleküle spielen eine wichtige Rolle bei den Bindungseigenschaften von Protein-Ligand-Komplexen sowie bei der Dynamik und Faltung von Proteinen spielen, was wertvolle Informationen für die Wirkstoffentwicklung liefert. Während hochauflösende Kristallstrukturen Informationen über die Position von Wassermolekülen liefern, fehlt dies bei niedrigauflösenden Kristallstrukturen, NMR und Cryo-EM Strukturen. Ein *deep learning* Algorithmus, der die Position von Wassermolekülen auf NMR- und Kryo-EM-Strukturen voraussagt, wurde an hochaufgelösten Kristallstrukturen trainiert. Der Ansatz wurde mit experimentellen Strukturen von IMP-13 und der *T. cruzi* PEX14 N-terminalen Domäne validiert.

Die in dieser Arbeit vorgestellten Ergebnisse liefern neuartige Erkenntnisse und Methoden zur Bewältigung aktueller Herausforderungen in der frühen Phase der Wirkstoffentwicklung und können auf viele andere Systeme übertragen und angewendet werden.

1. Introduction

In the following section, a description of the process of early-stage drug discovery will be presented, along with a background of the biology of key proteins characterised in the thesis.

Drug development

Drugs have been used by humans as remedies to various ailments since ancient times, mainly using natural products from plants and animals. Many of these traditionally-used plants contain active ingredients that were later extracted and synthesised artificially or optimised for stronger effects. Willow, used for its painkilling and anti-inflammatory properties for more than three and a half millennia, contains salicylate, similar to aspirin. In the 5th Century BC, Hippocrates wrote about a powder that could be extracted from willow that was bitter and had these same properties. More than two millennia later, Reverend Edmond Stone wrote about its antipyretic qualities (Stone 1763) and in 1828, it was first isolated in its pure form in Munich. Salicylic acid was then purified in larger quantities in 1838. This was further developed from this acidic molecule that caused irritation in the stomach into the neutral formulation sodium salicylate by Felix Hoffmann of Bayer, who patented and marketed the drug as aspirin (Desborough and Keeling 2017).

Later, drugs started to be produced synthetically, allowing for production on a large scale, without the need for large quantities of natural-product-producing organisms. This also led to the opportunity to embellish and improve molecules until medicinal chemists were able to design and build their desired molecules from scratch, based on structure; more than half of FDA-approved drugs now have a synthetic origin (Bade et al. 2010). Now, the ability to synthesise a required molecule is no longer the largest hurdle to overcome, with other points in the drug discovery process taking that role, as the methods for finding and developing possible drugs alter over time.

Early-stage drug discovery pipeline

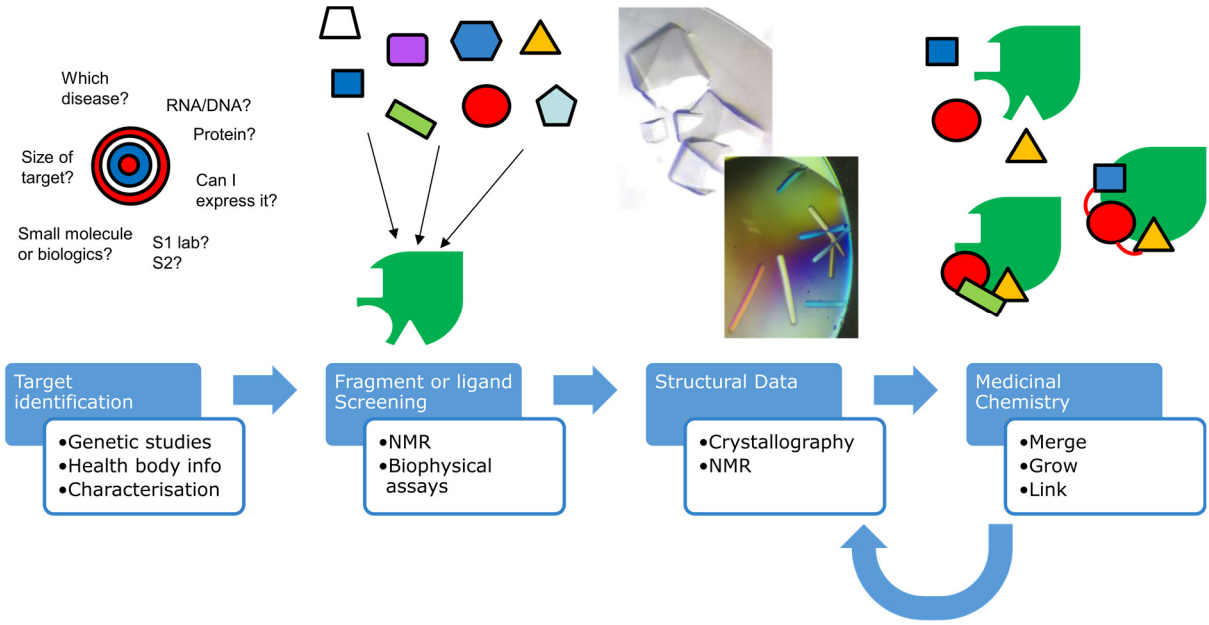

Figure 1.1: *Fragment-based drug design in early stage drug discovery.*

A typical drug discovery pipeline consists of a method of screening for possibilities, followed by determining how it is acting. In some cases, this involves determining binding partners; in others, this requires structure determination, depending on the screening method that is used. Thereafter, a medicinal chemistry programme is normally carried out to further improve binding and pharmacokinetic and pharmacodynamics properties, before iterative testing, characterisation and improvement occurs (Fig. 1.1). Once the molecule is optimised to the point of being a lead molecule, the pharmacological absorption, distribution, metabolism, excretion and toxicological (ADME-TOX) parameters are characterised and further improved, where possible, before entering *in vivo* testing and clinical trials. Here I will focus on the early stages of drug discovery from screening to lead molecule identification.

High-throughput screening

High-throughput screening (HTS) can be either a target- or phenotypic-based approach that is normally automated. It is used to screen vast libraries of drug-like molecules and has increased in prevalence since the 1980s (Chatterjee and Ganguli 1986; Burch and Kyle 1991). These drug-like molecules are generally chosen to have drug-like properties, for example using Lipinski's (Lipinski et al. 2001) rules and variations thereof, and to be chemically diverse.

Some libraries include drugs that have been used for other purposes, or which have been developed and dropped at a later stage in the development process. This method has the advantage that the molecules are already drug-like and so suitable for *in vitro*, cellular or *in vivo* screening, but it also has the disadvantage that many of the molecules are similar as they have been designed for current targets, dominated by proteases and transcription regulators (Cudic and Fields 2009). This means that these libraries are less suitable for finding molecules for new drug target classes, for example protein-protein interface (PPI) interactions, which unlike most enzymes have shallow binding pockets. The library can, however, be added to or designed with this in mind (Valenti et al. 2019), in order to increase the variety of molecules available in the library. The other disadvantage of this is that for relatively large molecules, it is unlikely that they will perfectly fit to the binding site, meaning that many promising molecule classes could be missed in this way. Given its relatively inexpensive nature for companies that already possess such a drug-like HTS library, this is nonetheless a beneficial method, and the fragment-based methods provide an alternative that partially solves this issue.

Phenotypic screening

Phenotypic screening involves testing an array of compounds on cells, tissue cultures or a living organism. These living assays are designed so that the cells or animals exhibit disease-like traits (disease phenotypes). Using a variety of techniques, the researchers can then determine whether the molecules added cause a phenotypic improvement. They can also simultaneously investigate possible side effects in the systems that they use.

Early approaches of drug discovery were focused on phenotypic screening. Having fallen out of fashion as target-based drug discovery gained in popularity, phenotypic screening has now increased again in use, particularly in systems and diseases where a target protein is yet to be discovered (Zheng et al. 2013). Once a molecule is found that has a phenotypic effect, target screening can then be carried out to determine the molecule in the cell or organism that is effecting that change, for example using drop-down experiments. Whilst experiments in mammals, for example mice, can be prohibitively expensive and ethically complex for large-scale screening, other organisms that share much of our

DNA, where relevant mutations with disease phenotypes can be added, can also be used. For example, *C. elegans* and Drosophila can now be used for large-scale screening (Carretero et al. 2017; Bangi 2019), much like typical *in vitro* high-throughput screening methods discussed below. The other value of these models is that their translucency allows real-time tracking, so this can also be used, for example, for biogenesis or developmental phenotypic screening. Of course, drugs designed to combat infections are less suitable for screening in such organisms as the key processes in bacteria and parasites can vary substantially from mammalian and other eukaryotic processes.

The disadvantage of phenotypic screening is that, as the target is not known, it can be challenging to develop and improve the molecule's efficacy and its properties. If a structural improvement is desired, the target must be isolated, the structure of it and the complex with the ligand determined and then the structural improvement can occur.

Some analyses have found that more first-in-class drugs were discovered using phenotypic screening than target-based screening (Swinney and Anthony 2011), while others have found the opposite (Eder et al. 2014), indicating that both play a significant role in the discovery of new drugs, although the output of FDA-approved drugs found through target-based screening is increasing. Target-based screening, including high-throughput screening and fragment-based screening, produces far more follow-up compounds than found using phenotypic screening.

Fragment-based screening

The libraries of successful or drug-like compounds and natural specimens used (Katz and Baltz 2016), combined with the huge chemical space available leads to many similar compounds being tested and, as the molecules are already drug-sized, this leaves little scope for further alteration without changing large sections of the molecule. It can also be very time and labour-intensive and methods have been developed to improve this.

One such method is fragment-based drug screening (FBS) (Badger 2012; Peng et al. 2016; Erlanson et al. 2019). This involves testing smaller fragments, often containing only one or two functional groups (mass < 250 $g.mol^{-1}$) against the target protein, in order to sample as much of the chemical space as possible in the first rounds of testing (Fig. 1.1). Various methods, X-ray crystallography and NMR prominent among them, are then used to determine which of these fragments bind to the target protein and, when possible, where they bind and whether they alter the binding site (Krimm 2015). This information can then be used to link fragments together that are in similar locations or grow fragments outwards into concave binding pockets or into active sites on the protein to increase binding affinity or improve pharmacokinetic properties.

Fragment-based screening can be carried out using NMR, X-ray crystallography and other biophysical techniques, each with different advantages and disadvantages. Here, NMR techniques are used for screening, both ligand and protein-observed, to determine quickly and in solution whether ligands interact with the target protein.

Fragment- and structure-based design

With a variety of fragment and ligand hits from the screening methods mentioned, the next stage in the drug discovery process is to develop these hits into lead molecules (Krimm 2015; Erlanson et al. 2019). With fragments, there are three main methods to this, often used in conjunction. These are merging, linking and growing of the fragments into larger molecules with better binding properties

(Fig. 1.1). In order to do this efficiently, the relative positions and orientations of binding should be known. This allows, for example, linking elements between two fragments to be designed to the correct length and to optimise possible interactions to the protein regions between the two binding sites. It also allows nearby binding pockets to be located, enabling the fragments to be grown or expanded into them, again maximising positive interactions. Merging requires overlap of binding positions of two or more fragments or ligands. For example, if two fragments overlap in such a way as to share one edge of an aromatic ring each, the two fragments could be merged to form a biaromatic ring system, along with any decorations on the ring systems.

Structure-activity relationships (SAR) also play a key role in determining which route to follow in drug discovery (Hajduk 2006). For a given molecule, exit vectors are determined, which can be optimised. For example, with a phenol fragment, the exit vectors could simply be the para, ortho and meta positions. Each of these would be functionalised with a variety of functional groups and the binding affinity or other measure of activity measured. This leads to an SAR cascade. By systematically approaching this optimisation process, medicinal chemists can more rapidly get to a final molecule by ruling out classes and patterns, rather than testing every permutation in every position.

Protein targets

Two main protein targets, PEX14 and IMP-13 are addressed in this thesis, along with a protein characterised in a collaborative project, ph47^{phox}. The background and disease relevance of these proteins are described below.

Peroxisomal biogenesis in Trypanosoma

Peroxisomes

Peroxisomes are membrane-surrounded organelles found in both humans and the trypanosome parasites. They are located in the cytosol and are the location of many key processes and reactions in the cell. Key reactions such as oxidation of certain biomolecules and lipid biosynthesis occur here (Schrader and Fahimi 2008). Some of these reactions form toxic by-products, for example hydrogen peroxide, so further processing of these molecules must also occur within the peroxisome in order to prevent cell damage and death (Cooper 2020). It is therefore critical that all the required enzymes co-localise in the peroxisomes, and that these reactions do not occur in the cytosol.

Peroxisomes were first isolated by Christian René de Duve (De Duve and Baudhuin 1966), who developed the technique of cell fractionation. As the peroxisomes are denser than other cellular organelles, they could be separated from the rest of the cell contents in this way. They had previously been identified as organelles in 1954 by Rhodin using electron microscopy and were referred to as microbodies. Since their isolation and the analysis of the processes involved, several pathological conditions related to peroxisomes in humans have also been elucidated. The first of these was Zellweger Syndrome, first published by Goldfischer et al. (Goldfischer et al. 1973). This is linked to incorrect or absent lipid metabolism due to a reduced number of peroxisomes; Zellweger Syndrome leads to a huge number of defects, including renal and hepatic failure, musculoskeletal problems and death within six months. Other genetic conditions are also known, many of which cause significant defects and lead to a reduced life expectancy. Infantile Refsum's disease, for example, leads to a life

expectancy of 10-30 years (Schrader and Fahimi 2008), whilst neonatal adrenoleukodystrophy is lethal in infancy (Aubourg et al. 1986).

Trypanosoma

The trypanosomes are single-celled parasitic protozoa that form the genus *Trypanosoma*. Their life cycle is complex and involves multiple organisms, normally with an insect as the vector or reservoir host (Radwanska et al. 2018). A variety of different vertebrates can be affected, including humans. As the disease progression, and form of the parasites, is multi-stage, the symptoms in humans are also seen as an acute and a chronic stage. In the initial period after infection, known as the acute stage, patients are either asymptomatic or suffer from cold- or flu-like symptoms including headache or general fatigue (Pinazo et al. 2017). Some patients experience lesions at the bite site or swelling around the eyes where the faeces of the insect vector enter and cause inflammation (Kennedy and Rodgers 2019). In these cases, the disease is more likely to be caught at the acute stage. After this, if left untreated, the disease enters the chronic stage. During this stage, the parasite is sequestered in the tissues, primarily in the cardiac and digestive muscle (Pérez-Molina and Molina 2018; Kennedy and Rodgers 2019). The chronic stage is again asymptomatic for some time, but this progresses to clinically observable forms. This can include cardiac abnormalities, oesophageal and colonic dilation and problems with the autonomic nervous system. In the end, this can lead to heart failure, haemorrhaging, arrhythmias and ultimately death. The main forms in humans are Human African Trypanosomiasis (HAT), caused by *Trypanosoma brucei,* and Chagas disease, caused by *Trypanosoma cruzi*.

The vector host for *Trypanosoma brucei* is the tsetse fly, a blood-sucking insect found in much of tropical Africa. HAT is therefore found only in sub-Saharan Africa. In 1995, the WHO estimated that 300,000 cases per year were occurring, whilst only 30,000 of these were diagnosed and treated (WHO 2020a). Since then, the numbers of cases, thanks to increased control and information campaigns, dropped to 3796 reported cases in 2014, with under 15,000 total cases estimated. Whilst this is a positive trend, the disease was almost eradicated in the 1960s but, with a lack of interest due to the low number of cases, the numbers of cases gradually increased again. Therefore, despite the lower numbers and better treatment availability, complacency against the *T. brucei* parasite is not an option. A new oral treatment for *T. brucei,* fexinidazole, has recently been recommended for humans following clinical trials (European Medicines Agency 2018). It is hoped that this will replace previous drug regimens that were administered intravenously or by intramuscular injections.

Trypanosoma cruzi is spread by the triatomine bug, as well as by congenital transmission from mother to child, and is found primarily in South America (Pérez-Molina and Molina 2018). However, it is believed that more than 300,000 people in the United States are infected with the disease, with cases also in Europe (Bern and Montgomery 2009). Antiparasitic treatment (Nifurtimox or Benznidazole) are very successful if given at the acute stage of the disease, as soon as possible after infection has occurred. These have treatment times of 60-90 days and common side effects, such as weight loss, insomnia, vomiting and neuropathy, which, given the minimal symptoms of the disease itself at this point, lead to problems with compliance (WHO 2020b). These treatments can be used during the chronic phase, but this is less effective and can only slow or prevent disease progression and not undo damage already caused. According to the WHO, an estimated 6-7 million people worldwide are infected with *T. cruzi*. New drugs with shorter treatment times, fewer side effects and drugs with high efficacy in the chronic stage of the illness are therefore desperately needed.

Glycosomes and protein import

As mentioned above, the reactions carried out in the peroxisome and the importance of co-localisation of enzymes within these pathways make it crucial that protein import into the organelles is targeted and efficient. In the trypanosomal parasites, glycosomes resemble peroxisomes of human cells (Michels et al. 2005). The term peroxisomes is used both for the general classification and for peroxisomes in animals. From here onwards, I will therefore refer to the categories as peroxisomes (all), glycosomes (parasite) and mammalian peroxisomes (as in humans). These are ancestrally related membrane-bound organelles, where many of the key proteins and processes involved are conserved. The biogenesis and proliferation pathways, for example are highly conserved, with homologous proteins involved in protein import and membrane insertion seen, in some cases, across the kingdoms of life (Brown and Baker 2008; Walter and Erdmann 2019). The glyosomes have one key difference from mammalian peroxisomes: the glycosome is the location of a substantial part of the glycolysis pathway in protozoa. The degree to which they rely on this is dependent on the stage of the life cycle of the parasite, as the metabolic pathways alter as they go through the different phases (Haanstra et al. 2016). For example, the trypomastigote stage of *T. brucei*, which is the form that enters the bloodstream from the triatomine bug bite or other skin opening and travels in the bloodstream, is highly mobile and is 100% reliant on ATP production via this mechanism due to the high and conserved oxygen content in the mammalian bloodstream. It therefore has a high glycosome count, whilst the epimastigote stage (the uninfective stage in the digestive system of the insect vector) has a far lower concentration of glycolytic enzymes due to the absence of continuous glucose presence (the insect sugar levels peak after a meal and rapidly decline).

Proteins involved in the biogenesis of peroxisomes are known as peroxins, or PEX, proteins. Currently 37 peroxins are known in total to contribute to processes including targeting, insertion and import of required proteins, and division and proliferation of the peroxisomes (Kalel et al. 2017; Walter and Erdmann 2019). In order to be transported into the peroxisomes, the proteins must first be recognised and targeted. This occurs using peroxisomal targeting sequences (PTS). PTS1, the most abundant PTS, is recognised and bound by the tetratricopeptide repeat (TPR) domain of PEX5 in the cytosol (Fig. 1.2). The tail of PEX has been shown to have a WxxxF motif that binds to the N-terminal domain of PEX14 (Neufeld et al. 2009; Neuhaus et al. 2014). PEX14 is tethered to the peroxisomal membrane, but its N-terminus is soluble and assumed to be facing the cytosolic side of the membrane. On binding of these three partners, the targeted protein is imported through the membrane into the peroxisomal space. A second PTS, PTS2 is targeted in much the same way by PEX7, which in turn targets the PEX14 N-terminal domain for import into the peroxisome (Walter and Erdmann 2019).

In the case of the glycosomes, PEX14 and the membrane protein PEX13 form a “docking” complex in the membrane that aids with import. Molecular and structural details of pore formation and biogenesis (formation and proliferation) of the glycosome have not yet been elucidated.

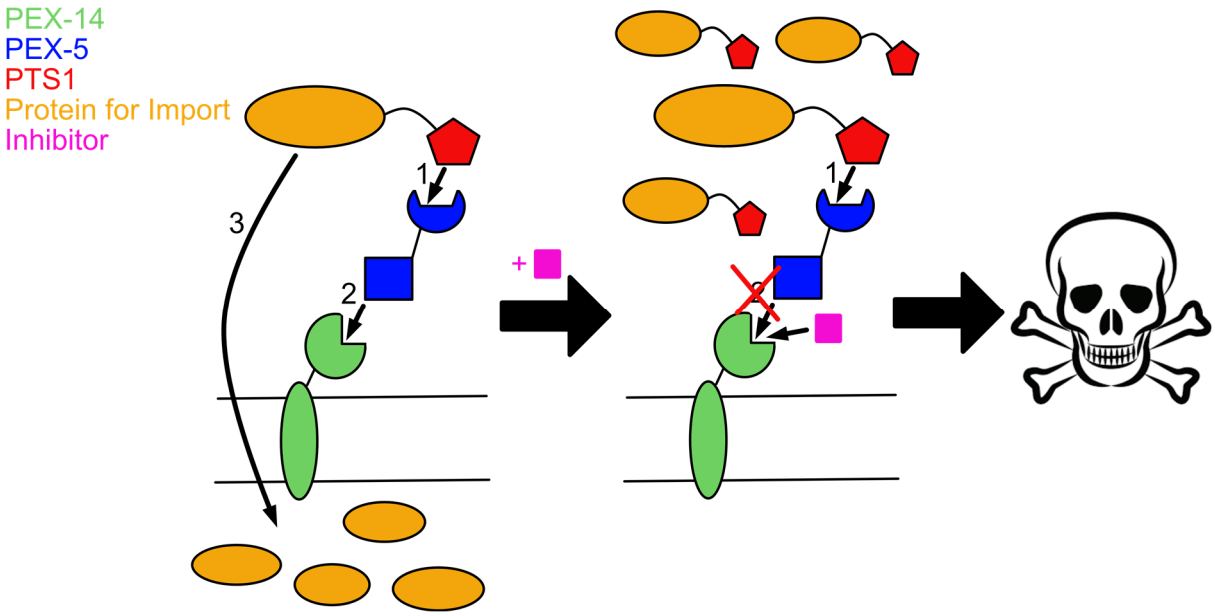

Figure 1.2: *Schematic of PEX14 pathway. An import protein (orange) with a terminal PTS-1 peptide motif (red) binds to the PEX5 (blue) TPR domain. PEX5 then interacts with the PEX14 (green) N-terminal domain, leading to the tagged protein passing the membrane into the glycosome. When an inhibitor (pink) is added to block the PEX14-PEX5 interaction, the proteins do not pass the membrane and therefore build up in the cytoplasm, causing a lack of reactions in the glycosome and runaway reactions in the cytosol, leading to cell death.*

PEX14/PEX5 interface and inhibition

Glycosomes and the reactions that occur within are known to be essential to parasites. Dawidowski et al. showed, using a mixture of fragment and structure-based drug discovery, that small molecule inhibitors of the PEX14/PEX5 interface in *T. brucei* could be designed and synthesised, and that treatment of the bloodstream form of *Trypanosoma brucei brucei* with these small molecules led to glycosomal import defects and death of the parasite (Fig. 1.2) (Dawidowski et al. 2017, 2020). This highlights that disruption of the glycosomal import pathway is a valid target for further drug development campaigns in the trypanosomes.

Due to the sequence similarity of human, *T. cruzi* and *T. brucei* PEX14, care must be taken to ensure that the molecules designed are as selective as possible for the trypanosomes over the human protein, in order to reduce the effect on peroxisomes in the human body. Peroxisomal protein inhibition in humans, given the severity of peroxisomal diseases described above, could cause potentially serious side effects.

Metallo-β-lactamases

The second protein target in the thesis is a metallo-β-lactamase (MBL). MBLs form a class of metalloenzymes that confers antibiotic resistance to bacteria. By cleaving the β-lactam ring on carbapenems, commonly considered the antibiotics of last resort, these critical drugs are rendered useless (Tooke et al. 2019). The protein has been found in every significantly populated continent (Zhao and Hu 2011) and is often co-expressed with other antibiotic resistance genes (Mendes et al. 2004; Ho et al. 2014). Due to an active site with two zinc ions, these enzymes can be very difficult to model *in silico*, increasing the dependence on experimental techniques in drug discovery of this class.

Discovery of molecules to inhibit these enzymes would be highly beneficial as antibiotics that are rendered inactive by them could then be used again in combination therapy with the inhibitors (Tooke et al. 2019). This would have far-reaching benefits as both community and nosocomial outbreaks of antibiotic resistant infections in vulnerable populations could be far more easily brought under control, reducing the mortality of such events.

Imipenemase-13 (IMP-13) is a member of this class of enzymes (Santella et al. 2011) and was previously not structurally characterised. This is a key target approached in this thesis. Due to the broad range of metallo-β-lactamases and their intrinsic similarities and differences, development of inhibitors requires an in-depth knowledge of the structure and function of different members of the class in order to determine which interactions are key when designing drugs and which play only a minor role.

Crystallising IMP-13 with fragments was found to be challenging; the enzyme was instead studied in its apo state and in the presence of its natural substrates, the carbapenem antibiotics, in order to learn more about its mode of binding and dynamics in the different states.

NADPH oxidase 2 complex

In 1933, a process called a respiratory burst was first seen to occur in phagocytes (Baldridge and Gerard 1932), which leads to generation of hydrogen peroxide (Iyer et al. 1961); it was later seen that some isoforms produce superoxide (Babior et al. 1973) and some hydrogen peroxide. In 1987, it was noted that patients with chronic granulomatous disease, who have reduced bactericidal activity in the phagocytes, lack the respiratory burst function (Quie et al. 1967). Over time, the individual components of the complex causing this and the processes involved were elucidated. There are several NADPH oxidases with a variety of cofactors, but we focus on the NADPH Oxidase 2 complex.

The function of this membrane-bound complex is to shuttle an electron from NADPH in the cytosol via Flavin adenine dinucleotide to oxygen on the outside of the membrane. This leads to production of reactive oxygen species (ROS). ROS have several uses in the body, from host defence to regulation of cell growth, but are also known to cause cell damage in many cases, leading to inflammation and some pathological conditions, such as diabetes and cancers (Block and Gorin 2012). Thus, this complex could be a potential drug target for such conditions when targeted to specific tissues.

The NOX2 complex consists of membrane-bound NOX2 in a heterodimer with $p22^{phox}$, also membrane-bound (Fig. 1.3). These stabilise each other; neither NOX2- nor $p22^{phox}$-deficient patients show significant concentration of either protein in the phagocytes as, when not in dimeric form, these proteins are targeted for degradation (Parkos et al. 1989). $P47^{phox}$ is found in the cytoplasm and binds to $p67^{phox}$, the activator protein, and, in some cases, $p40^{phox}$. In its unphosphorylated form, $p47^{phox}$ does not bind to the membrane-bound dimer. Upon phosphorylation, however, $p47^{phox}$ binds to $p22^{phox}$, bringing $p67^{phox}$ in contact with NOX2 in the membrane, which is necessary for its activation. Rac GTPase also plays a role in activation of the NOX2 complex (Bedard and Krause 2007).

P47^{phox} consists of two SH3 domains, as well as an autoinhibitory region (AIR) (Fig. 1.3A), with a phox region that binds to phospholipid membranes and a proline-rich region that binds to p67^{phox} (Leto et al. 1994). The binding site for p22^{phox} lies in a deep pocket between the two SH3 domains and is rendered inaccessible by the AIR. However, on phosphorylation of the AIR, the binding site then becomes accessible, allowing p22^{phox} to bind, leading to activation of the NOX2 complex by p67^{phox}.

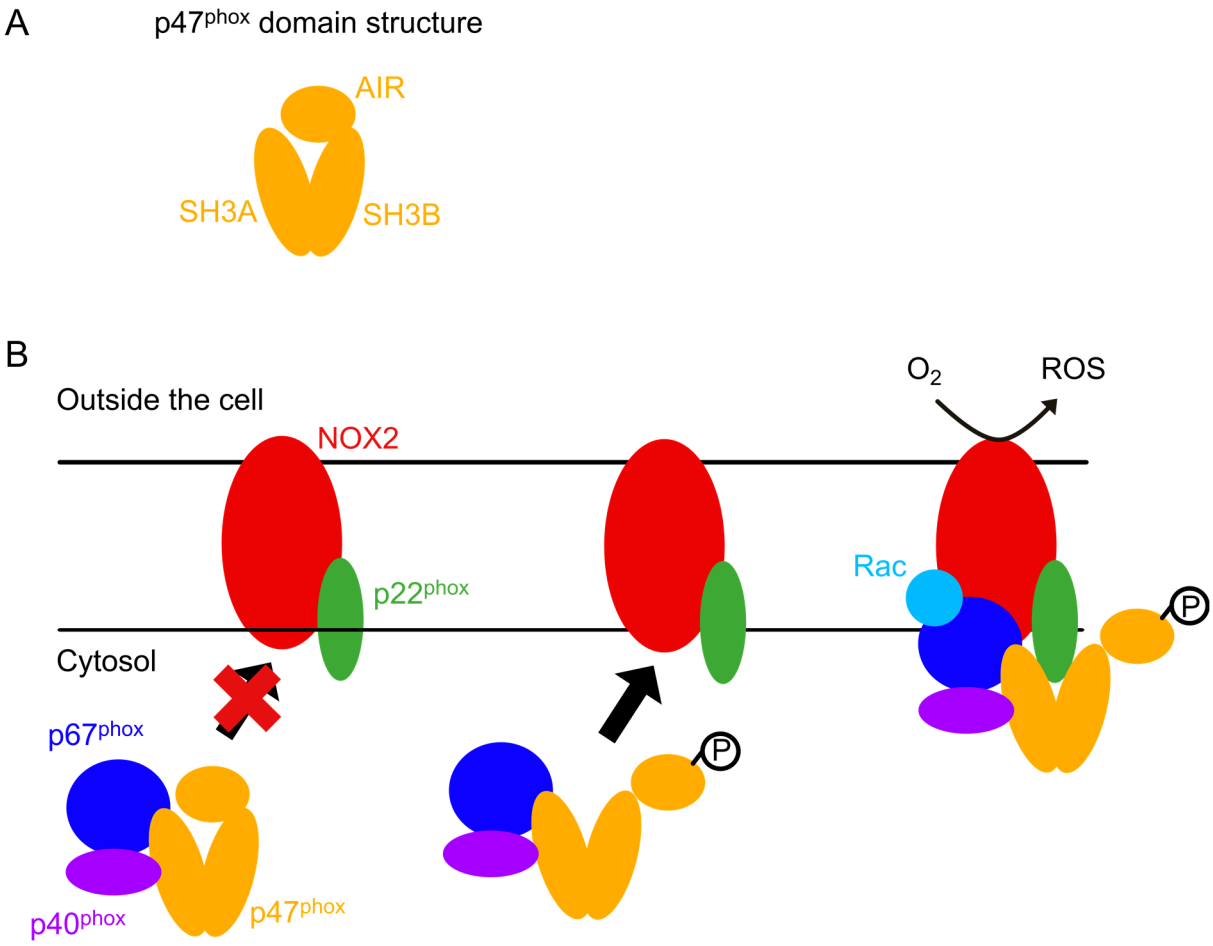

Figure 1.3: *Schematic of NADPH oxidase complex. A. p47phox domain structure. B. p47phox (orange) forms a complex with p67phox (blue) and p40phox (purple). The AIR domain of p47phox is phosphorylated and relocates allowing binding to p22phox (green). This brings p67phox into contact with NOX2 (red), which with Rac (light blue) forms the active complex, leading to production of ROS.*

Fluorescence polarisation assays, along with cell-free activity assays have previously been used to screen for inhibitors of NOX2 by preventing binding of p47^{phox} (Smith et al. 2012). This led to Ebselen (Sakurai et al. 2006) being highlighted as a potential drug for this purpose. Ebselen led to decreased activity of NOX2, as well as preventing localisation of p47^{phox} to the membrane, indicating that it was preventing the binding of p47^{phox} to the NOX2-p22^{phox} dimer as intended. However no structural data was published to show the binding mode or mechanism. This nonetheless highlighted the validity of the use of p47^{phox} as a drug target.

Major challenges in structure-based drug discovery

With the improvement in knowledge of drug mechanisms, screening methodologies and medicinal chemistry, the major limiting factor in the early-stage drug discovery pipeline is structure determination of the complexes. The main methods for structure determination in drug discovery include X-ray crystallography and NMR, while cryo-EM is currently emerging and especially useful for large complexes.

Out of these methods, X-ray crystallography is the oldest and most well-established. A protein is expressed and purified and incubated at high concentrations along with a crystallisation buffer. This is done using the sitting or hanging drop method, allowing a droplet of protein solution to steadily increase in concentration over time, allowing the droplet to reach super-saturation and crystallise (McPherson and Gavira 2014). These crystals grow in a repeating lattice structure, where each repeated block is referred to as a unit cell. The translational and sometimes rotational symmetry of these crystals is a huge benefit in solving the structure of the protein. However, the rigid formation of the crystal lattice can lead to artefacts caused by interactions between molecules within and between unit cells. It also hides any dynamics in the system. The crystalline environment is vastly different from a cellular environment, but the information gleaned from crystals can be incredibly valuable in understanding possible stable forms of the protein. The other key information that can be gained from crystal structures is the position of water molecules in the first hydration shell and other binding or interacting molecules. Water molecules that bind favourably in a specific position due to hydrogen bonding with the surface of the protein can be seen in medium- to high-resolution structures, giving further insight into how binding of a ligand affects this water envelope. This is an effect not seen in NMR or in cryo-EM. The challenge with crystallography for the purpose of studying fragment-protein complexes is that weak fragment binders often require very large concentrations of fragments, leading to crystallisation artefacts, as well as being difficult to crystallise, especially where crystal contacts would be disrupted by binding.

NMR is a commonly-used alternative. NMR has the advantage that dynamics and conformational changes can be studied, both alone in solution and on interaction with ligands or fragments (Ludwig and Guenther 2009). This can be critical in some drug discovery projects, for example where a protein must be stabilised in its inactive conformation in order to prevent its biological activity. Both active and inactive conformations must be studied in this case, as well as the dynamics between the two states. Lower ligand and protein concentrations can be used than in crystallography, leading to fewer artefacts as a result of this. The environment is also more similar to a cellular environment, particularly where phosphate buffers at physiological pH are used. Intermolecular NOEs (Proudfoot et al. 2017), INPHARMA (Sánchez-Pedregal et al. 2005) and transfer-NOESY (Balaram et al. 1972; Ni 1994) are among key NMR techniques used to study ligand-protein interactions, either to detect binding in screening or to gain some information on the mode of interaction. This can provide important information about the mode of binding and the effect this could have on the activity of the protein. However, weak binders can be challenging to work with as a result of their transient interactions, which often leads to a very small bound population. Where only the bound state is measured, for example in the case of intermolecular NOEs, this leads to difficulties in obtaining a signal and long experimental times.

Where direct structural methods are not suitable for the system, indirect experimental or computational approaches can be used to gain insight into binding of ligands. This includes

comparisons of fragments with available substrate or larger ligand complex structures and *in silico* modelling. Larger binding partners (either biomolecules or larger ligands) often bind much more tightly and with less flexibility to the target protein. This can lead to easier crystallisation of the complex due to more favourable interactions, fewer flexible regions or larger occupancy of the bound state. Larger occupancy of the bound state also aids NMR methods such as intermolecular NOE methods. The key interactions and dynamics properties of the system elucidated by either method can then be used to predict fragment or ligand behaviour and to develop further screening methods, as well as instruct medicinal chemists about key interactions within the binding pocket. *In silico* modelling also helps hugely with this, allowing structural biologists to predict the most likely binding poses of the ligand in comparison to the substrate and the protein structure.

Aims of the thesis

The thesis addresses a number of challenges in early-stage drug discovery, applied to important current drug targets. IMP-13, due to difficulty in crystallisation with the available fragments from the original drug screen, was lacking structural information to further the drug discovery project to develop inhibitors. In order to enable rational structure-based drug discovery for IMP-13, crystal structures of IMP-13 in complex with its natural substrates are solved and presented, along with other structural information, such as *in silico* modelling and NMR measurements to gain sufficient insight into the mode of interactions. This will allow the future expansion of the current fragments into hit and lead compounds while an understanding of the mechanism may also lead to the ability to streamline and find new fragments that bind in the active site and are most likely to inhibit the protein.

The thesis also aimed to structurally characterise *T. cruzi* PEX14 and screen for potential fragment hits. NMR and crystallography are both used for characterisation and a variety of NMR methods for screening. Further characterisation of compound hits, in terms of binding parameters and structural features is carried out. Due to the weak binding and optimal single cysteine site, the use of both new and established paramagnetic tags for this protein is tested. In still on-going work the utility of paramagnetic restraints to define ligand binding poses is analysed. The theory and applications of paramagnetism in current drug discovery are reviewed.

The importance of the water envelope of proteins on ligand binding has been exploited by a deep learning algorithm to determine water binding sites on protein structures in the absence of electron density. The algorithm is tested on published structures in the pdb and on presented crystal structures of IMP-13 and PEX14 to determine its applicability in a variety of molecule classes.

The development and application of paramagnetic NMR and a deep learning algorithm to the project's drug targets aims to demonstrate the utility of the methods, whilst highlighting and building on existing knowledge. This work will also provide new structural information and fragment hits for further use in drug discovery in building new ligand molecules to take further towards the clinic.

2. Methods

This chapter summarises the background and broad experimental setup for the data collected for the papers presented in this thesis. Detailed experimental information can be found in the methods sections of the respective publications.

Protein expression, purification and quantification

Proteins were expressed using recombinant expression in *E. coli* cells. Plasmids were transformed into DH5α cells for plasmid amplification and the genetic material isolated. For expression, BL21 cells were used.

For use in crystallography and for ligand-observed NMR methods, protein without isotope labelling was expressed using autoinduction medium (Studier 2005). For protein-observed NMR experiments, ^{15}N-labelled protein was produced with ^{15}N autoinduction medium or, where required, ^{15}N^{13}C-labelled protein was produced using M9 minimal medium, supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose.

After expression, cells were lysed with sonication and affinity chromatography with Ni or Zn NTA beads was used for initial purification. Size-exclusion chromatography was used for a further purification step. Purity and construct size was checked using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and the concentration was measured using a Thermofisher Nanodrop 2000.

All proteins were stored frozen in NMR buffer (HEPES for IMP-13 and phosphate NMR buffer for PEX14) at -80 °C. Fragments were stored at -20 °C at 100 mM in d6-DMSO, while ligands were stored at 50 mM in d6-DMSO, also at -20 °C.

X-ray crystallography

X-ray crystallography is a proven technique in macromolecular structure determination, from proteins to protein-peptide complexes and even nucleic acid structures. It relies on the discovery of many prized, sometimes Nobel-award-winning, scientists (Jaskolski et al. 2014), including the discovery of X-rays by von Laue, and further on the development of Bragg's law to calculate the structure of a crystal lattice. From this, the field developed as Rosalind Franklin measured DNA crystals, allowing Watson and Crick to determine its double helix structure (Watson and Crick 1953).

Currently a large array of software packages is available to solve the required equations, reducing the numerical and algebraic workload of the process for the end user significantly (Kabsch 2010; Adams et al. 2011). However the underlying principles remain the same.

The diffraction pattern (Fig. 2.1) from an X-ray measurement is in inverse space, meaning that the closer together the reflections are, the larger the unit cell is. This means that salts have very largely spaced diffraction patterns, whilst protein crystals lead to a much finer array of spots (Dauter 2017). These reflection spots are very rich in information – their position describes the crystal lattice and the symmetry of the crystal, whilst the intensities hold information on what is within the crystal unit, i.e. the geometry within the unit cell. However some information is not held within: the phasing. This can

be carried out manually using heavy atoms or molecular replacement can be used. This means that phases from previously solved structures that are similar in structure can be used in further structure determinations. The spots must be indexed, then integrated to find their intensity for use in the further calculations. By repeated manual refining of the molecules into the electron density and refinement calculations, the structure slowly improves until completion (Fig. 2.1).

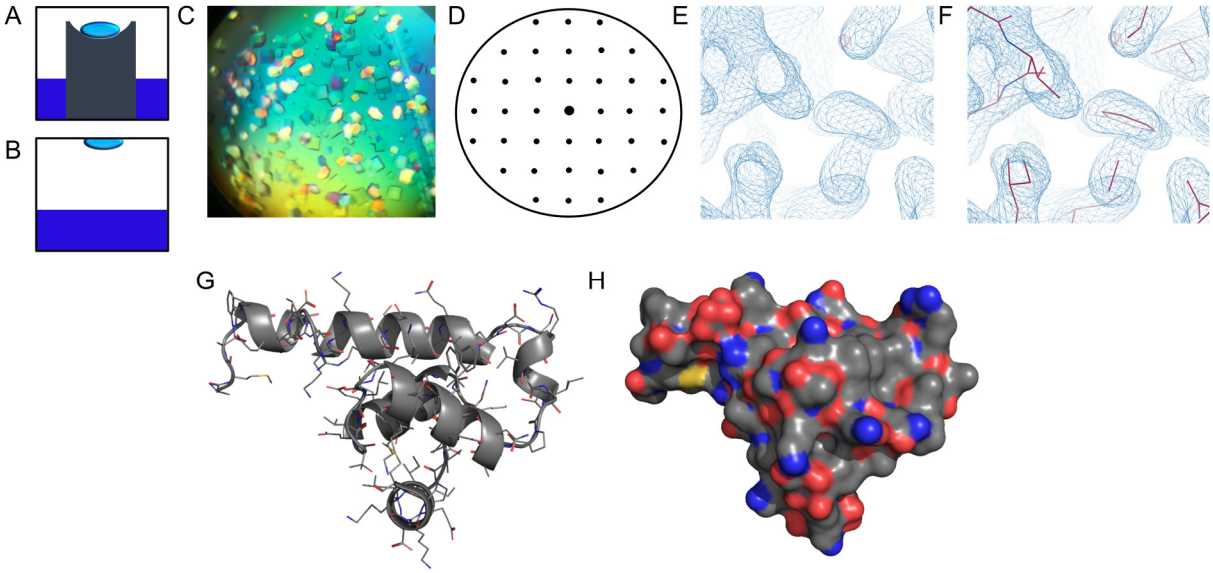

Figure 2.1: *X-ray crystallography pipeline. Using either sitting (A) or hanging (B) drops, small drops of protein and reservoir solution are mixed and left to form crystals (C). These are frozen and taken to a synchrotron, where diffraction patterns are measured (D). The diffraction patterns represent electron density (E), into which protein molecules can be fitted and refined (F). These can then be represented in a variety of ways (G, H).*

Obtaining crystals

The process of obtaining crystals normally involves, after obtaining pure, stable protein, screening under a variety of different conditions to find a condition in which regularly-shaped crystals form, where possible avoiding thin needles due to the difficulty in measuring them with a fixed-width X-ray beam. There are two common procedures for the crystallisation itself – the hanging drop and the sitting drop (McPherson and Gavira 2014).

In the case of the sitting drop (Fig. 2.1A), a reservoir of the screening condition is set up with a small shelf slightly raised from it. On this shelf, a small drop of protein solution is mixed with a small volume of the screening condition, before being sealed and left in the required conditions (normally room temperature or cooled to around 4 °C) for the crystals to grow. As the sitting drop is a mixture of the screening condition and the protein solution, the concentration of the screening solution is higher in the surrounding reservoir that it is in the protein droplet. Therefore, in order to reach equilibrium, water from the droplet evaporates, increasing the concentration of both the ions and the protein in the drop. As the concentration reaches supersaturation, the protein can start to fall out of solution and crystallise.

The principle of the hanging drop method (Fig. 2.1B) is much the same – vapour diffusion leads to supersaturation of the droplet and therefore crystallisation of the protein. However the setup of the drops themselves is different. Instead of sitting on a shelf or support above the reservoir, they are

suspended from a cover slip above the reservoir. In practice, this means that a small volume of the protein, mixed with a small amount of the given reservoir solution, are placed on a thin cover sheet that is then upturned onto the plate, so that the drop is hanging from it. The surface tension and friction of the drop therefore prevent it from dropping. This does however mean that the drop size, if too large, can lead to loss of the sample. As these are normally prepared manually, whilst sitting drop plates can be prepared robotically, this is less suitable for large-scale screening campaigns. However, some proteins or complexes are seen to crystallise better in a hanging drop set-up.

In order to reduce dynamics in the protein system further and to minimise damage caused by the heating effect of the X-ray beams, the crystals, once fished from their droplets, have a cryo-protectant added to them, if not already present in the crystallising condition, and are then frozen in liquid nitrogen. The crystals are cooled during measurement by a cryogen stream.

Crystallisation with ligands

It is often the case that in macromolecular structure determination, the required or desired structure consists of a complex with a peptide or small molecule ligand. There are multiple ways to obtain such structures, namely soaking or co-crystallisation (Rodrigues et al. 2018).

Soaking is the addition of the peptide or small molecule to the pre-formed apo crystal structure. In an apo crystal, the unit cell of the protein is already in a stable state, with crystal contacts maintaining the interactions between protein molecules, both within and between unit cells (Müller 2017). The protein molecules, thanks to their lack of tessellation properties have channels between them within the crystal lattice, where solvents and other small molecules can pass. While some solvent molecules and ions maintain a relatively favoured position in different unit cells within this lattice when cryogen-cooled due to particularly strong interactions, at room temperature the crystals still maintain some dynamic character, particularly in the solvent channels. This opens the possibility of introducing small molecules into the lattice once the crystals are formed. Ligand, fragment or peptide solutions are then added to the crystal and left to incubate for a period of time before freezing to allow the molecule to diffuse through the protein. This approach requires a reliable condition for production of apo crystals and the binding site must be accessible via the solvent channels in the crystal. It also is essential that the ligand does not cause a large conformational change in the protein that would disrupt the protein crystal, leading to either cracking of the crystal or a loss of diffraction.

Co-crystallisation on the other hand involves the addition of the additional molecule before the crystals are grown. Typically, the ligand or peptide is added in large excess to the protein solution, the mixture is incubated for a time, and in some cases further concentrated, and then crystal plates are set up. This allows for the formation of a complex, which then together forms crystals maintaining this complex arrangement. This leads to the formation of crystals of the complex even when the binding site in the apo symmetry state is obstructed by other molecules in the unit cell, although can lead to a lower ligand occupancy (Müller 2017). The disadvantage of this method is that it works mainly with ligands and peptides that have a high affinity for the protein; the complex formation must be more stable than the formation of the crystal contacts in the apo state for the position of the ligand to be maintained during crystal formation. It is therefore not as useful as soaking in the case of fragment crystallisation, but in the case of known substrates or very strong binders can work very well.

Experimental outline and software

Screening for crystallisation conditions was performed using commercially available buffer sets in a sitting-drop vapour diffusion setup. For co-crystallisation screens, compounds were added as either powder or DMSO solution.

Diffraction data reported in this thesis were taken at the ESRF (Grenoble, France) at the ID30b beamline and at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland) on the X06SA beamline. The data was indexed using XDS (Kabsch 2010) and scaled using XScale or Aimless. The initial phases were obtained using molecular replacement carried out using molrep or Phaser, with TbPEX-14 (PDB ID: 5L87) (Dawidowski et al. 2017) as the search model for TcPEX14 and PDB 1DD6 for the IMP-13. For additional IMP-13 structures, apo IMP-13 was used as the search model. Manual rebuilding using electron density maps was carried out in Coot (Emsley et al. 2010). Elbow was used for obtaining restraints for small molecules in the crystallisation conditions. Further refinements were carried out using Phenix Refine (Adams et al. 2011) or Refmac (Winn et al. 2003) with proSMART (Nicholls et al. 2012) used to generate additional restraints. Five percent of the reflections were used for cross-validation analysis. The final model of TcPEX14 was deposited in the PDB under code **6ZFW**, which will be released on publication of the Zaucha et al. article. Final models of the apo and complex structures of IMP-13 were deposited under accession numbers **6R79** and **6R78** for apo and **6R73**, **6RZR**, **6RZS** and **6S0H** for meropenem, imipenem, ertapenem and doripenem bound forms, respectively in the Protein Data Bank.

NMR

Basic principles

Nuclear Magnetic Resonance (NMR) experiments have become a key source of information in structural biology in recent years. It relies on the principle of 'spin'. Each sub-atomic particle has a spin value. When these are paired (i.e. both the number of neutrons and the number of protons are even), the nucleus has spin 0. If either or both of these numbers are odd, the nucleus has an non-zero spin value, I (Keeler and Wothers 2013).

The value of I can be any half-integer (…-1,$-\frac{1}{2}$, 0 , $\frac{1}{2}$, 1… etc), but in biomolecular NMR, we normally focus on spin $\frac{1}{2}$ nuclei, for example ^{15}N, ^{1}H, ^{13}C and ^{19}F. I will therefore focus on this situation. For a nucleus with spin I within a magnetic field, the energy levels are split into $2I + 1$ possible values, meaning that the spin $\frac{1}{2}$ nuclei have two spin combinations (Fig. 2.2). In the absence of a magnetic field, these states have the same energy and so a statistical Boltzmann distribution leads to identical populations in each spin combination. In the presence of a magnetic field however, these combinations have distinct energy levels. The difference between these energy levels is equivalent to the Larmor frequency, $\Delta E = \gamma \hbar B_0$, where γ is the gyromagnetic ratio, an intrinsic property of nuclei, $\hbar$ is the reduced Planck's constant and B_0 is the external magnetic field strength. This means that the higher the magnetic field strength of the spectrometer, the larger the difference in energy. The spin states in which the spins partially align with the magnetic field are marginally lower in energy, leading to a slightly larger occupancy of this spin state.

By inducing a change of spin alignment using a radio pulse of frequency $\omega_0 = -\Delta E = -\gamma\hbar B_0$, spins are flipped from the lower energy level to the higher level. The decay of the spins from this high energy state back to the lower energy state leads to a measurable signal that can be transformed into a spectrum, where slight differences in chemical environments are measured as small differences in frequencies arriving at the detector.

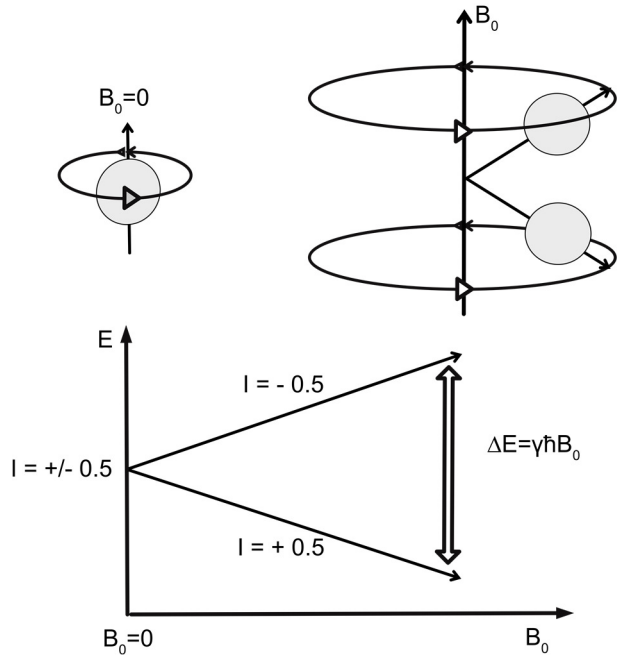

Figure 2.2: *NMR basic principles. A nucleus of spin half has one energy level in zero magnetic field. In the presence of a magnetic field, this splits into parallel and antiparallel states with different energies with an energy difference of $\gamma\hbar B_0$. The larger the field, therefore, the larger the energy difference.*

The Fourier transform (FT) of these data can be shown as a spectrum, depicting each environment of the chosen nucleus type as a single peak. The disadvantage of this now-simple method is that as target molecules increase in size, the number of peaks increases hugely, leading to significant overlap of peaks and difficulties in interpretation.

Multi-dimensional protein NMR

The development of more advanced pulse programs and improvement in isotope labelling (Otting and Wüthrich 1990; Ikura et al. 1990; Pardi 1992; Sattler et al. 1999; Lacabanne et al. 2018) have led to more subtle and specific measurements being possible and has increased the maximum protein size for such studies. Two-dimensional spectra in which magnetisation is transferred via bonds between pairs of nuclei, for example between ^{1}H and ^{15}N (Bodenhausen and Ruben 1980), enabled significant reductions in overlap of peaks in spectra of large molecules and led to the uprise of biological macromolecule studies.

Further progress in computing power and spectrometer hardware and pulse programmes led to higher-dimensional spectra (Oschkinat et al. 1988), once more reducing overlap between signals and increasing the capability of studies of larger molecules. The coupling of magnetisation in this way (through-bond or through-space) can be used to study the structure of biological macromolecules, including RNA and proteins, in solution (Barbieri et al. 2004). The benefit of this in comparison to X-

ray crystallography is that the conditions of the samples are much closer to the cellular environment than those used in crystallography. In order to assign each resonance in a ^{1}H, ^{15}N 2D correlation experiment, for example, 3D experiments such as HNCACB and CBCACONH spectra are used (Grzesiek and Bax 1992). These involve transfer of magnetisation along the chain of the molecule in the order referred to, so from amide H to amide N, further to the Cα and then to the Cβ in the case of HNCACB. CBCACONH spectra link the Cβ and Cα of the previous amino acid ($n-1$) with the N and H of the current amino acid (n). By overlapping the two spectra, this enables 'walking' along the chain of the molecule, which allows the individual residues to be assigned due to additional intrinsic properties of different amino acid types and knowledge of the amino acid sequence.

For TcPEX14 backbone assignment ^{1}H,^{15}N 2D correlation experiments, 3D ^{15}N-edited spectra, HNCA, HNCACB and CBCACONH experiments were recorded on uniformly ^{15}N,^{13}C-labelled samples on Bruker Avance III 800 MHz spectrometer (^{1}H frequency 800 MHz) equipped with a 5 mm TCI cryoprobe. Samples at 600 µM were prepared in NMR phosphate buffer supplemented with 10% D_2O. Spectra were processed using NMRpipe (Delaglio et al. 1995) and nmrDraw and assignments made using CCPN Analysis (Vranken et al. 2005). Amide assignments are deposited in the BMRB under deposition number **50345** and will be released upon publication.

Backbone assignment experiments of IMP-13 are described in the article Structure and Molecular Recognition Mechanism of IMP-13 Metallo-β-lactamase (Softley et al. 2020b).

Heteronuclear NOE

Heteronuclear NOE (hetNOE) experiments are used to study the flexibility of different regions of a protein (Neuhaus and van Mierlo 1992; Farrow et al. 1994). The {^{1}H}-^{15}N heteronuclear NOE is commonly used to study the protein backbone flexibility by recording two $^{1}H^{15}N$ correlation spectra, normally in an interleaved fashion, with one including proton saturation {^{1}H} and the other not. The comparison of the intensity of these two spectra gives the NOE value: a measure of the motion of the N-H bond vectors within the molecule. In areas where the protein is more flexible than other regions, the N-H bond will undergo more motion, leading to a decreased NOE value in comparison to less flexible regions:

$$NOE = \frac{I_{sat}}{I_{equil}},$$

where I_{sat} is the intensity of a peak in the spectrum measured with proton saturation and I_{equil} is the intensity in the spectrum without.

Paramagnetic NMR

One NMR method investigated in detail in this thesis is the use of paramagnetism in drug discovery (Fig. 2.3). A protein is tagged and the paramagnetic effects, namely pseudocontact shifts, residual dipolar couplings and paramagnetic relaxation enhancements are measured on the protein and, in the case of drug discovery, on the fragment or ligand. As these effects provide distance- and/or angle-dependent restraints, they are very informative and can be used for structure calculations or guided docking. This could be particularly useful in drug discovery due to the difficulties involved at the structure determination stage and the fact that, unlike chemical shift perturbation studies, the effects are not affected by allosteric changes.

Paramagnetism in NMR and its various applications in drug discovery are discussed in detail in the article Paramagnetic NMR in Drug Discovery (Softley et al. 2020a), with a summary on page 33.

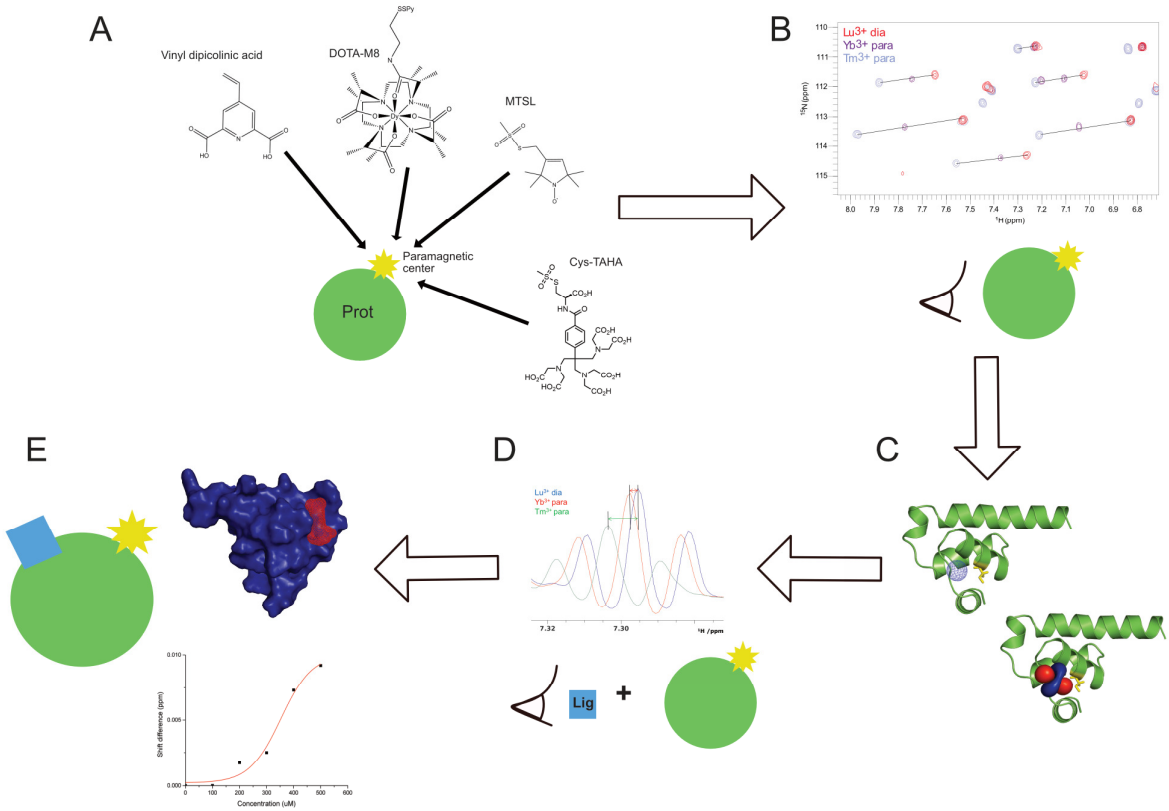

Figure 2.3: *Use of paramagnetism in fragment placement. A protein is tagged with a paramagnetic label (A), either with a metal (eg vinyl dipicolinic acid, DOTA-M8 or TAHA) or with a nitroxyl spin label (eg MTSL). Protein-observed spectra are measured, assigned and analysed (B, PCS shown). The metal position and tensor are then back-calculated (C). Ligand spectra can also be measured using ligand-observed techniques (D). The proportion bound may need taking into consideration, and the position of the ligand on the surface of the protein can then be calculated or docked (E).*

In order to utilise the effects of paramagnetism in protein NMR, a paramagnetic centre must be inserted using a paramagnetic tag, with the exception of the case where a natural metal binding site is present. A variety of tags are available, each with different properties, some of which are depicted in Fig. 2.3A. The tags shown here are single-point binding tags that bind to Cys residues. In the article Paramagnetic NMR in Drug Discovery, other tag types are also discussed. For the PEX14 protein target, however, single-point tags were deemed the most suitable due to the protein's small size. Other tags, for example peptide tags or CLaNP-5, a two-site binding site, were not suitable as they would require mutation of multiple residues, potentially causing large disruption to the structure of such a small protein.

Paramagnetic tagging, NMR experiments and data analysis

These experiments, described here in more detail, are part of still-ongoing work which is intended for publication in the near future.

TcPEX14 constructs were tagged using DOTA-M8 (Häussinger et al. 2009) and vinyl-dipicolinic acid (Su et al. 2008), as well as the tag presented in the article by Denis et al. to introduce a paramagnetic centre into the protein. In all cases, samples containing BME were first incubated overnight at 4 °C

with TCEP, as BME has been found to maintain its interaction with the Cys residue even after attempted removal with buffer exchange columns or concentrate-dilute cycles. This interaction is also shown in the presented crystal structure of TcPEX14. After displacement with TCEP, this was buffer exchanged with a desalting column or concentrate-dilute cycles to buffers without reducing agents. In all cases, buffer exchange and removal of excess occurred with gravity columns PD-10 or NAP-5 (GE Healthcare) or with concentrate-dilute cycles.

For paramagnetic tagging with DOTA-M8, a four-fold excess of tag was added and gently mixed and the mixture incubated for a reaction time of 4 hours at 25 °C. Excess tag was then removed before measurement.

For paramagnetic tagging with vinyl dipicolinic acid, the protein was buffer exchanged into PEX14 HEPES NMR buffer. A ten-fold excess of vinyl dipicolinic acid was added and incubated overnight at 25 ^{o}C. The excess was removed, lanthanide metal ions (lutetium – diamagnetic, ytterbium and thulium – paramagnetic) in the form of chloride salts were added and any excess of ions then removed before measurement.

^{1}H,^{15}N 2D correlation spectra were measured to check tagging was successful and temperature series were measured where necessary to aid with assignment of paramagnetic peaks from the diamagnetic reference. 2D TOCSY experiments using Dipsi2 cross-polarisation (Rucker and Shaka 1989) were carried out on the protein-ligand complex at 200 µM TcPEX14 and 360 µM fragment concentration to determine ligand peak assignments, as the peaks shifted significantly on protein binding. Tensors were calculated using FANTEN (Rinaldelli et al. 2015) and Paramagpy (William Orton et al. 2020) and visualised using Pymol (Schrödinger LLC). PCS-guided docking was carried out using Haddock (Schmitz and Bonvin 2011; Van Zundert et al. 2016).

Drug discovery

A variety of methods for drug discovery were covered briefly in the introduction; here I will focus on techniques used for early-stage drug discovery in the projects described.

Fragment screening methods

Fragment screening can be carried out using NMR, X-ray crystallography and other biophysical techniques. In our lab, we use mainly NMR techniques, both ligand and protein-observed. Ligand-observed spectra have the advantage that no labelled protein is required, which aids the ease and cost of sample preparation for large-scale screening campaigns. Protein-observed is often more useful for smaller proteins, due to the relaxation-induced size limitations in NMR. Ligand-observed spectra include relaxation-editing (Hajduk et al. 1997), Saturation Transfer Difference (STD) (Mayer and Meyer 1999), WaterLOGSY (Dalvit et al. 2001) and simple 1D ^{19}F spectra (Ludwig and Guenther 2009).

Relaxation-edited experiments are based on the principle that upon binding to a protein, the effective size of a ligand increases dramatically, leading to a slower tumbling rate and faster relaxation compared to when the ligand is measured alone. STD involves pulsing specific resonances within the protein (avoiding hitting the ligand signals; usually the methyl region is used), in order to saturate the protein. Due to efficient cross-relaxation spin diffusion within proteins, this leads to the entire protein being magnetised, along with any ligands that are bound, which are then visible in the spectra. Ligands

that are not bound to the protein are not magnetised, meaning that their signals are not visible in the spectrum. WaterLOGSY shows differential positive and negative NOEs to the ligand depending on the route of NOE transfer, in turn dependant on whether the ligand is bound or not. The water is magnetised and this can transfer to surrounding molecules either directly or indirectly. In the case of the ligand, this means that it can be transferred directly to the ligand, which due to the ligand's small size results in a negative NOE or from the protein (to which the magnetisation has already been transferred from the water), which would lead to a positive NOE, due to the protein's large size. ^{19}F spectra can be used to look for shifts or changes in peak intensity; the advantage of ^{19}F screening is that due to the absence of ^{19}F signals from the protein, only the ligand is visible and the background is low, meaning that large mixtures of compounds can be tested at once and the spectra are easy to analyse. ^{19}F is of particular relevance in drug discovery as many drug-like molecules contain fluorine or other halogen atoms.

Protein-observed spectra, such as 2D correlation experiments (HSQC, HMQC) (Bodenhausen and Ruben 1980) tend to be used for screening smaller proteins and as an orthogonal method after testing with ligand-observed signals. It has the disadvantage that isotopically labelled protein is required, but the added advantage that, if the protein spectrum is already assigned, a chemical shift perturbation (CSP) analysis (Williamson 2013) can be carried out. This analysis highlights which amino acid signals have moved, giving an idea of the areas of the protein most affected. This can give an idea of the binding site, but CSP effects can also be induced allosterically so this is not always accurate.

Fluorine Library Screening, Novartis

The CF_3 library screened consists of mixtures of compounds, designed to have clearly separated peaks and no cross-reactivity within each mixture. Initial screening was carried out on these compound mixtures, with validation being carried out using the single compounds.

A ^{19}F CPMG sequence was run twice on each sample with delays of 80 and 400ms and spectra were analysed using TopSpin NMR. Hits were then tested individually in the same manner. Hits observed in both mixture and individual screening were further screened by protein-observed NMR using ^{1}H,^{15}N 2D correlation experiments.

Helmholtz Zentrum München fragment library and compound testing

The Helmholtz Zentrum fragment library was screened against ^{15}N-labelled TcPEX14 using SOFAST ^{1}H,^{15}N HMQC (Schanda and Brutscher 2005) spectra. Samples consisted of 190 µM protein with mixtures of 5 fragments. The DMSO signal was used as the lock signal and a reference spectrum of protein with 8µl DMSO was also measured. Hits were determined by manual inspection of the spectra using Topspin and deconvolution spectra of individual fragments in mixtures deemed to be hits were run at 1 mM fragment, supplemented with 10% D_2O.

Compounds from collaborators were tested with ^{1}H, ^{15}N 2D correlation spectra at 1:1, 3:1 or 1 mM compound, depending on the expected affinity and stage of development in the drug discovery process. The protein concentration was 100-200 µM for these spectra and a DMSO blank was measured for each set of tests.

Biophysical and biochemical assays

In order to characterise fragment or ligand hits from screening or optimisation rounds, further assays are required. These can be *in vivo* or *in vitro* and either direct or competitive binding assays. Two key

techniques used are AlphaScreen™ (Ullman et al. 1994; Bosse et al. 2002) and fluorescence polarisation (FP) (Lea and Simeonov 2011) assays.

AlphaScreen™

AlphaScreen™ (PerkinElmer) is an assay technique that uses the short lifetime of the singlet state oxygen molecules to distinguish between binding and non-binding events (Bosse et al. 2002). Acceptor and donors bead are each bound to one of the two partners, the protein and its binder (Fig. 2.4), in our case PEX14 protein and the PEX5 WxxxF motif peptide.

On laser excitation at 680 nm, the donor bead converts normal triplet oxygen into the excited singlet state. This is easily quenched on interactions with other molecules. However, if the acceptor bead is in close proximity to the donor bead as the excitation takes place, the singlet oxygen can react with a thioxene derivative present in the acceptor bead, leading to chemiluminescence at 370 nm. This then further stimulates other fluorophores in the acceptor bead leading to emission of light at 520 nm (Fig. 2.4).

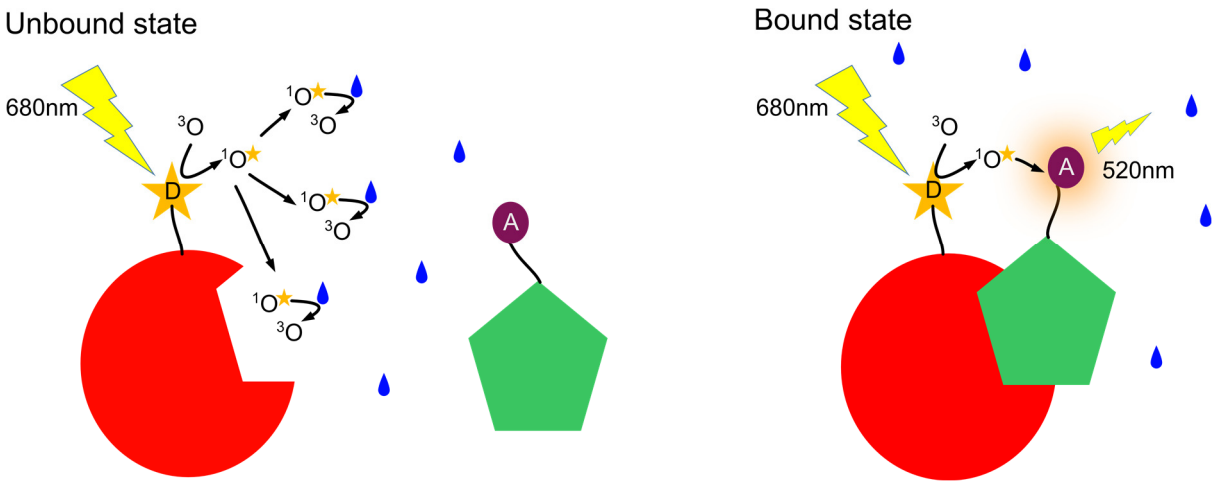

Figure 2.4: *AlphaScreen™ schematic. Donar beads are excited by a laser at 680nm. Normal triplet oxygen is converted into the excited singlet state. This is quickly quenched on interactions with other molecules, for example water. In the bound state, the donor and acceptor beads are close so the singlet oxygen can travel far enough to reach the accepter bead, triggering a reaction that leads to fluorescence at 520nm, which is measured.*

Therefore, in the absence of an inhibitor, the beads emit light upon laser excitation, which can then be detected. When an inhibitor is added, the protein and peptide binding is obstructed, so a lower proportion of the beads are in close contact with each other. This leads to less light being emitted and so a lower intensity is recorded. As more inhibitor is added, more of the protein-peptide complexes dissociate, leading to a concentration-dependant change in the light emitted, which can be measured, plotted and fitted to a binding curve to determine the EC50 of this assay.

Fluorescence Polarisation

Originally described theoretically by Perrin in 1926 (Perrin 1926), fluorescence polarisation (FP) rests on the fact that polarisation of light emitted from a fluorophore is dependent on the rotation time, and therefore the size, of the molecule or any complexes that it forms (Lea and Simeonov 2011).

Polarised light is beamed onto the experimental mixture. Light is then emitted from the fluorophore in a mixture of parallel and perpendicularly polarised light (Fig. 2.5). The parallel light intensity ($I_{\parallel}$) and the perpendicular light intensity ($I_{\perp}$) can then be detected and used to calculate the FP value:

$$FP = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}$$

This method has the advantage that it is independent of the concentration of fluorophores in the mixture.

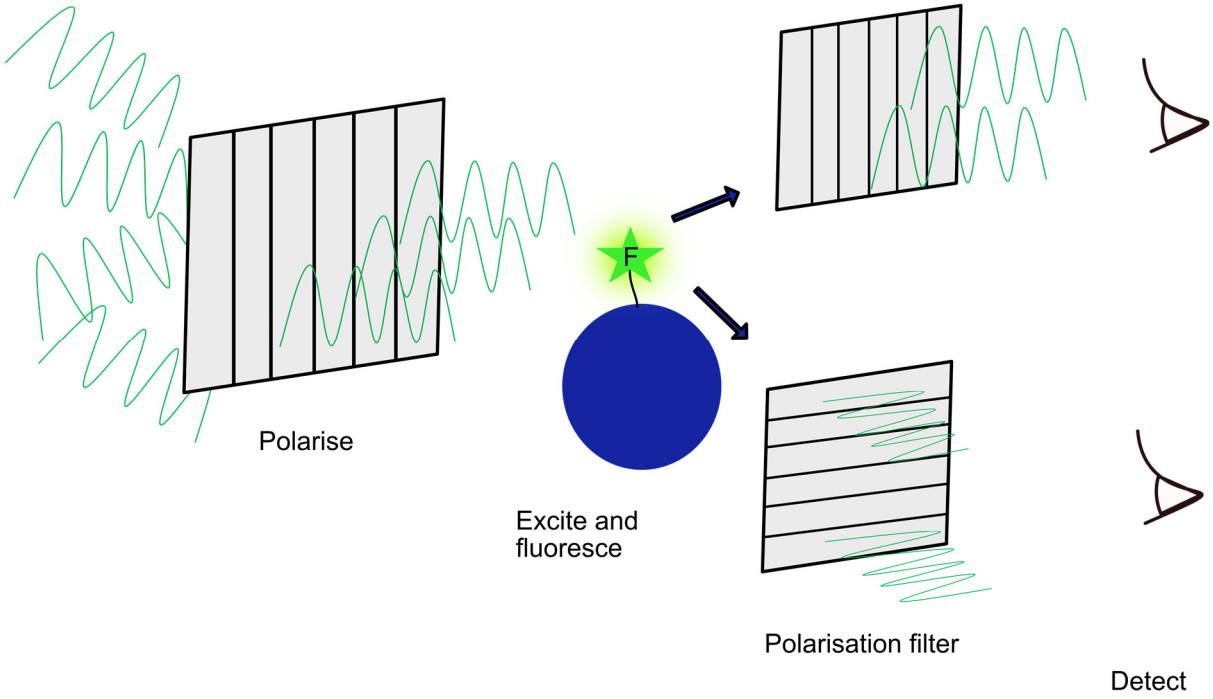

Figure 2.5: *Fluorescence polarisation schematic. Polarised light excites a fluorophore that emits light that has different polarisation properties depending on the tumbling speed of the molecule(s) it is bound to. Polarisation filters are used to separate the emitted light into parallel and perpendicular contributions, which are detected and used to calculate the fluorescence polarisation value, FP.*

In drug screening, this is used to detect complex formation and inhibition. In our case, the peptide binder is bound to a fluorophore, the protein partner is also present and we make a concentration series of steadily increasing ligand concentration. In the absence of ligand, the peptide and its fluorophore bind to the protein, causing a specific FP readout. In the presence of binding ligand, the protein-peptide interface is disrupted, leading to a lower proportion of the fluorophore in the bound state. The average rotation time of the fluorophore is then significantly decreased due to the decrease in the effective weight of the complex, leading to an altered polarisation state which can be detected. This can be used to form a titration curve, allowing estimation of the IC50 from this data.

One significant disadvantage of this method is that many ligands fluoresce themselves and so would increase the amount of fluorescence and alter the proportions of parallel and perpendicular light. It is important to note this and to check after the measurements whether the overall fluorescence is increasing with increased ligand concentration and to account for this during the analysis.

Surface Plasmon Resonance

Surface plasmon resonance (SPR) is a method for monitoring molecular binding events and their kinetics. For direct binding studies, one binding partner is tethered to a film in a manner that minimally affects the structure and function. The other binding partner, or potential binding partner, is passed over the surface of the film, allowing interactions between the two to occur (Patching 2014; Nguyen et al. 2015).

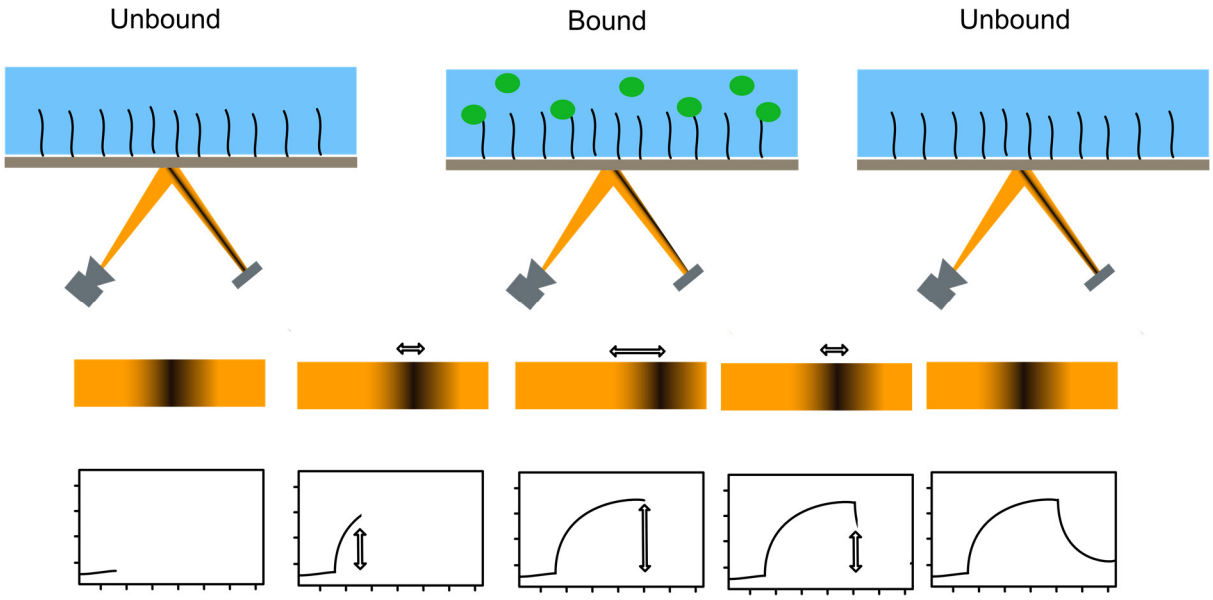

Figure 2.6: Surface Plasmon Resonance schematic. One of the two binding partners is immobilised on a metal chip. Light is shone through a prism at the underside of the chip and the reflected light detected. An absorption band is present at a certain angle of incidence. The other binding partner is passed over the surface of the chip. As the binding alters, the position of the absorption band alters and is measured. When the soluble molecules are removed from the flow-through again, the proportion bound decreases again (for reversible binding), causing the absorption band to shift back towards its original position. This allows the measurement of kinetics as well as binding.

The principle behind the technique is dependent on oscillations of free electrons (or surface plasmons) at the interface of two materials. In this case, a light is shone through a prism and is reflected off the metal film, to which the molecules are tethered. At a fixed angle of incidence, the light excites the surface plasmons, leading to resonance, and an absorption band in the reflected light. By studying the intensity of different regions of the reflected beam, as well as the position of the absorption band over time, kinetics of binding events can be studied (Fig. 2.6). This can be measured with variation of concentrations in the flow-through solution to allow both on- and off-rates of binding to be calculated.

Competition assays can also be carried out (Huber et al. 2015) using this method by altering the protocol, for example adding multiple molecules to the flow-through solution.

Thermal Shift Assay

Thermal shift assays (TSA) are a way of measuring the thermal stability of a protein or complex and characterising protein ligand interactions (Huynh and Partch 2015). A dye molecule, for example SYPRO orange, is added to a protein solution. The hydrophobic properties of the dye leads to it binding to hydrophobic regions on the protein, causing measureable fluorescence of the dye. The protein-dye mixture is subjected to a temperature gradient over time; as the temperature increases, more of the protein denatures, leading to more exposed hydrophobic areas that would normally be buried within the protein. This leads to more dye binding and higher fluorescence. Above a certain temperature, the protein starts to aggregate, reducing the hydrophobic surface again, leading to a drop in fluorescence at very high temperatures. The melting temperature is taken as the turning point, or half-way point, of the initial increase in fluorescence as the protein denatures. By repeating the experiments with different conditions in the protein solutions, including different buffers or binding partners, the effect of those environments on the stability of the protein can be measured.

3. Publications

This is a publication-based thesis. The work has been published in international peer-reviewed journal articles, which are summarised in this section.

3.1 Structure and Molecular Recognition Mechanism of IMP-13 Metallo-β-Lactamase

The article, Structure and Molecular Recognition Mechanism of IMP-13 Metallo-β-Lactamase (Softley et al. 2020b) *was published in Antibiotic Agents and Chemotherapy and is found at http://aac.asm.org/lookup/doi/10.1128/AAC.00123-20 . The author of this thesis, Charlotte Softley, along with Mark Bostock and Krzysztof Zak, contributed equally to the original work and writing of this article. Charlotte Softley carried out molecular biology and protein expression and purification, crystallisation, data acquisition and analysis and preparation of the manuscript. In the following summary, citations of this paper are not included.*

Antibiotic resistance is a growing threat worldwide, leading to many treatable infections becoming more lethal as the resistance spreads into larger populations (Tooke et al. 2019). While many different antibiotic resistance mechanisms occur in bacteria, the WHO has formulated a list of critical pathogens for development of new antibiotics. On this list, all three categories of the most critical pathogens include carbapenem resistance (Tacconelli et al. 2017). Carbapenems are considered to be the drugs of last resort in the treatment of antibiotic resistant infections, but metallo-β-lactamases are rendering even these last-resort drugs useless (Mendes et al. 2004). MBLs are a family of metalloenzymes, including IMP-13 (Santella et al. 2011), that cleave the β-lactam ring of the carbapenems, preventing binding of the antibiotic to its intended target, the penicillin-binding proteins (Hong et al. 2015; Bush 2018).

The article “Structure and Molecular Recognition Mechanism of IMP-13 Metallo-β-Lactamase” presents a detailed structure and dynamics study of IMP-13 in the presence and absence of its substrates, the carbapenems. The interactions between IMP-13 and four carbapenems were studied by X-ray crystallography. Two apo structures were also presented, one with a critical flexible L1 loop in the open form, where the apical tryptophan residue was situated far from the active site, pointing into solution, and one with the flexible L1 loop in the closed form, positioned over the active site, with the tryptophan forming a tunnel between the binding site and free solution. This highlighted the presence of flexibility and dynamics of this loop, which were subsequently tested by NMR and molecular dynamics (MD) simulations. Both MD and NMR confirmed this hypothesis of the flexible loop, with the apo state having considerably more dynamic activity of the loop than in the presence of the antibiotic substrates.

Furthermore the modes of binding to the antibiotics were studied. The carbapenem antibiotics have a carbapenem core motif consisting of decorated β-lactam and pyrroline ring moieties, as well as a tail region that varies more widely between different antibiotics in the class. Structural analysis indicates that the binding mode is highly conserved within the core motif, whilst the binding modes of the tails varied significantly as a result of the differing functional groups and orientations. Critical interactions include hydrophobic interactions of the rings and sulphur with the tryptophan and valines of L1, the interactions of the core’s carboxylic acid moieties to the zinc ions and to Lys 161 and the interactions of the N in the core to the Zn ions. The co-localisation of these key interactions in the core lends itself

to unspecific binding to different carbapenems, allowing the same enzyme to destroy an entire class of antibiotics, despite differing tail moieties.

Comparison of this with existing structures of penicillin binding protein 3 (PBP) (Han et al. 2010) indicate that the tail regions are more critical to the binding to PBP, leading to different degrees of selectivity to different carbapenems. This could be key for drug discovery as the most important interactions differ in many ways between the carbapenems' target molecules and the enzymes that degrade them. If differences can be exploited and similarities avoided, this could lead to development of further classes of molecules that are active against PBP without being destroyed by the MBL enzymes.

3.2 Paramagnetic NMR in Drug Discovery

The article, Paramagnetic NMR in Drug Discovery (Softley et al. 2020a) *was published in the Journal of Biomolecular NMR and is found at https://doi.org/10.1007/s10858-020-00322-0. The author of this thesis, Charlotte Softley, and Mark Bostock contributed equally to this article. In the following summary, citations of this paper are not included.*

The process of structure-based drug discovery is a multi-step iterative process involving screening, structure determination of the complex, characterisation of the interaction with biophysical methods, medicinal chemistry to grow or link these fragments and ligands and then a repeat of the whole process (Krimm 2015; Erlanson et al. 2019). Many methods are available for all steps of this pathway, but in many cases, it is the structure determination of the complex that is the rate-limiting step. This is because weak-affinity ligands and fragments can often not be crystallised and NMR methods such as intermolecular NOEs (Proudfoot et al. 2017), whilst available, are often challenging and time consuming.

Paramagnetism is a property of molecules where an unpaired electron is present. It causes magnetic properties in the presence of a magnetic field. As NMR uses strong static magnetic fields, paramagnetic samples show distinct characteristic alterations in the spectra. This includes shifts in the peaks of a protein ^{1}H, ^{15}N 2D correlation experiments (pseudo-contact shifts, PCS), an increase in relaxation and therefore a decrease in intensity of the signals (paramagnetic relaxation enhancement, PRE) and an alteration in the dipolar coupling of the signals (residual dipolar coupling, RDC) (Bertini et al. 2002a; Clore and Iwahara 2009). All three of these effects are distance and/or angle dependent. This means that by adding a paramagnetic centre, the extent of the PCS, PRE and RDC effects on each amino acid or ligand atom gives structural information on the molecule or complex. Whilst not yet widely used for this purpose, this can be exploited in drug discovery.

The article 'Paramagnetic NMR in Drug Discovery' reviews the current methodology and applications in this area, from different methods available to introduce a paramagnetic spin label or metal ion, a brief description of the theory behind each paramagnetic effect and considerations that must be taken into account to literature examples of how these have been used and exploited in the drug discovery world. This includes works involving ^{1}H, ^{15}N, ^{13}C and ^{19}F and for PCS covers the entire range of interactions from slow exchange of strong binders through intermediate exchange to fast exchange of weaker binders. All three of these scenarios are found in drug discovery at different stages, making it critical for such methods to be applicable to all cases in order to speed up the drug discovery pipeline

using these methods. This comprehensive study provides clarity and examples for other scientists in the field to apply, as well as perspectives for the future.

Figures are made, where relevant, using experimental data from TcPEX14, a peroxisomal protein in the *T. cruzi* parasite that causes Chagas Disease and is a key drug target in the fight against it. These data were collected as part of a drug discovery project on this target. Other figures are schematic depictions of the process.

3.3 Deep Learning Model Predicts Water Interaction Sites on the Surface of Proteins using Limited-Resolution Data

The article, Deep learning model predicts water interaction sites on the surface of proteins using limited-resolution data, *was submitted to ChemComm on 24 June 2020. The author of this thesis, Charlotte Softley, and Jan Zaucha contributed equally to the original work and writing of this article. In the following summary, citations of this paper are not included. The structural analysis of the outputs of the algorithms, the crystal structure refinement TcPEX14 and structural biology and drug discovery discussion were carried out by Charlotte Softley, whilst design and development of the algorithm from the training set was carried out by Jan Zaucha.*

Interactions of water with proteins are known to have an effect on folding (Rhee et al. 2004), as well as binding of other proteins, ligands and nucleic acids (Hong and Kim 2016; Brini et al. 2017). Understanding where these water molecules are and how they interact, however, is less well studied. A prediction algorithm that could predict where waters on the surface of the protein lie could be highly beneficial in understanding the processes involved in binding, thereby significantly aiding in drug discovery efforts, where an understanding of factors affecting binding pockets and water networks is highly beneficial.

The article 'Deep learning model predicts water interaction sites on the surface of proteins using limited-resolution data' by Jan Zaucha, Charlotte Softley et al. presents a water prediction algorithm based on a deep learning artificial intelligence model. A training set from the PDB is taken to allow the algorithm to learn from positions where water molecules are found in a wide subset of structures. From this, 2.8 million water molecules, each interacting with at least two protein atoms, were taken as the positive dataset and negative samples were chosen with a random sampling of a mesh around the protein, where the point chosen had no water molecule within 1.4 Å. An equal number of negative and positive training points were chosen.

These positions were then analysed using information on their nearest neighbours. Up to 50 of the closest atoms, up to a maximum of 7.5 Å were included. Vectors from the position to each of these nearest neighbours, along with information on the character, for example hydrogens, aliphatic carbons, and the B factors of the atoms. These were passed through six different neural network architectures and analysed to choose the architecture that gives the highest classification accuracy. This architecture was then used.

For evaluation, three PDB pairs of homologous proteins with different resolutions and methods were used. All structurally homologous proteins had been excluded from the training set to ensure that any information provided by the algorithm was from prediction and not just repetition of a protein in the

training set. The algorithm was run on the lower-resolution x-ray structures, and on cryo-EM and NMR structures, where waters are either limited or completely absent, and the results compared to the higher-resolution structures.

A new crystal structure of the N-terminal domain of TcPEX14 (PDB: 6ZFW) is also presented. After testing of the algorithm on the three PDB pairs as described above, the TcPEX14 structure, with all of its water molecules removed, was provided to the algorithm for water prediction. The resulting predicted waters were then clustered and the results compared to TcPEX14, showing a recall of 0.27. This was then compared with the X-ray crystallography modules Coot Find Waters (Emsley et al. 2010) and Phenix Update Waters (Adams et al. 2011), both of which use the electron density map to locate water molecules, at a variety of resolution cut-offs. It was found that with a resolution >2.6 Å, a maximum of six water molecules were predicted with Coot and no water molecules at all with Phenix. At a resolution of <2.5 Å, the Phenix algorithm has a higher recall than hotWater, but Coot was not able to reach the same level of recall even at a resolution of 2 Å.

The algorithm was run on a further structure, added to the PDB since the algorithm was developed. This was a crystal structure of IMP-13 (PDB: 6RZR) (Softley et al. 2020b), which has a binding pocket with a loop forming a tunnel-like structure over it. The original structure had an antibiotic, imipenem, bound. We found that the waters predicted using hotWater on the backbone structure of the pdb predicted water locations in the same place as the oxygen atoms of the antibiotic. On studying the other structures, we found that this was also the case with glycerol molecules in the PEX14 structure.

3.4 Introducing the CSP Analyzer: a Novel Machine Learning-based Application for Automated Analysis of two-dimensional NMR spectra in NMR Fragment-based Screening

The article, Introducing the CSP Analyzer: a Novel Machine Learning-based Application for Automated Analysis of two-dimensional NMR spectra in NMR Fragment-based Screening (Fino et al. 2020), *was published in the Computational and Structural Biotechnology Journal and is found at https://doi.org/10.1016/j.csbj.2020.02.015. The author of this thesis, Charlotte Softley, contributed to this work in the form of contribution to the design, testing of the functionality and user interface, protein expression and purification, NMR screening experiments and manual analysis to produce training sets. In the following summary, citations of this paper are not included.*

A time-demanding initial task for NMR spectroscopists on drug screening projects is the screening and subsequent data analysis of the protein target. There are many ways to do this, including both protein- and ligand-observed methods (Ludwig and Guenther 2009; Gossert and Jahnke 2016). Typically the output for such a screen is hundreds to thousands of NMR spectra of samples containing the protein target with each of the members of the fragment or ligand library to be screened. For 2D protein-observed spectra, the analysis for these datasets involves manual inspection of the spectra to determine whether signals have shifted on addition of the ligand. Whilst software is available for automatic or computer-aided processing of 1D datasets (Peng et al. 2016; East et al. 2019), this was lacking for 2D spectra.

As part of the study “Introducing the CSP Analyzer: a Novel Machine Learning-based Application for Automated Analysis of two-dimensional NMR spectra in NMR Fragment-based Screening” by R. Fino et al, an application, with front end written by Roberto Fino and back end written by Ryan Byrne, was produced to combat this need in NMR spectroscopy. Using training sets from drug discovery projects within the Helmholtz Zentrum München, a machine learning algorithm was designed and trained to automatically determine which spectra (and therefore which fragments or ligands) were deemed active and which were not, combined with the ability to manually check these assignments and label the spectra as active or inactive where these results were in conflict. This was then tested on new proteins to determine relevant statistical parameters.

The software allows the user to upload hundreds of spectra of the same protein for analysis at once. Using the topspin peak picking, it shows schematically the test spectra in one panel, displayed one by one, each overlaid with the reference spectrum. It then automatically categorises them as hits or not hits and shows this as a diagram and numerically. The user has the option to scroll through each of the spectra and add manual labels where relevant. The software can be used to display just those labelled as hits or just those labelled as inactive for manual inspection. The number of peaks in each spectrum is also indicated so that errors in spectral data or processing can easily be identified. These data can then be printed to a spreadsheet for further analysis.

The back end was designed using a machine learning algorithm. Initially a set of descriptors is defined, including the structural similarity index, which is a measure of similarity, including differences in noise and new or disappearing peaks. A paired-distance approach is then taken to further describe the similarities of the spectrum. The Hu moment estimation (Hu 1962) is used to compare invariant properties of the two spectra and as an additional metric, the Jensen-Shannon entropy difference between reference and test spectra is taken. This variety of metrics then undergoes dimensionality reduction and passed to a support vector classifier. In the training set, each spectrum is pre-labelled as active or inactive to allow categorisation of the test set.

3.5 The Photocatalysed Thiol-ene reaction: A New Tag to Yield Fast, Selective and Irreversible Paramagnetic Tagging of Proteins

The article, The Photocatalysed Thiol-ene reaction: A New Tag to Yield Fast, Selective and Irreversible Paramagnetic Tagging of Proteins (Denis et al. 2020), *was published in ChemPhysChem and is found at https://doi.org/10.1002/cphc.202000071. The author of this thesis, Charlotte Softley, contributed to this work in the form of testing this tag, along with those designed and synthesised previously, on T. cruzi PEX14, along with calculating tensors where relevant using FANTEN. Some of these tags were found to have multiple conformations and, as such, were not reported in the final publication, due to the added complication of tensor calculation and flexibility not being conducive for the intended purpose. Precursors such as the dipicolinic acid-based tag, vinyl dipicolinic acid, were also tested on PEX14 to prove that the cysteine residue present was accessible enough to be successfully tagged.*

Paramagnetic tagging of proteins is a method being increasingly used in drug discovery and structural biology (Bertini et al. 2002b; Pintacuda et al. 2007). The addition of a paramagnetic centre causes changes to NMR spectra including PCS, RDC and PRE. Key properties of paramagnetic tags for use for these purposes include rigidity of the tag, sometimes with the ability to chelate or interact with an

extra residue on the protein for added rigidity, strong chelation of the metal ion, a stable bond between protein and tag (usually via a cysteine or lysine residue) and the chosen metal in the tag (Joss and Häussinger 2019). With regards to the metal, lanthanides or paramagnetic transition metals are usually chosen.

In the article "The Photocatalysed Thiol-ene reaction: A New Tag to Yield Fast, Selective and Irreversible Paramagnetic Tagging of Proteins" by Maxime Denis et al., a newly designed and synthesised tag is presented. This tag is bound using a photo-catalysed thiolene reaction to cysteine residues on the surface of the protein. The thiolene reaction renders a stable thioether bond between the sulphur of the cysteine thiol and the double-bonded carbon of the tag. This reaction is irreversible, meaning that should reducing agents be necessary to stabilise other parts of the protein, they can be re-added after tagging without disrupting the ligation of the tag.

The reaction itself takes place under UV radiation, speeding up the tagging process significantly. This is helpful for proteins that are unstable at room temperature overnight, as is necessary for some existing paramagnetic tagging protocols. Instead the reaction can run to completion within a few hours, with the added benefit that the measurement can be initiated on the same day as the tagging occurs.

3.6 Developing Inhibitors of the p47phox−p22phox Protein−Protein Interaction by Fragment-Based Drug Discovery

The article Developing Inhibitors of the $p47^{phox}$–$p22^{phox}$ Protein−Protein Interaction by Fragment-Based Drug Discovery (Solbak et al. 2020), *was published in the Journal of Medicinal Chemistry and is found at https://dx.doi.org/10.1021/acs.jmedchem.9b01492. The author of this thesis, Charlotte Softley, contributed to this work in the form of validation screens of the hits from FP and TSA using NMR and contributions to the analysis of the tryptophan region.*

The production of reactive oxygen species by NADPH oxidase 2 causes a wide variety of disease states and damage within the cells, including inflammation, cancer, brain damage after traumatic injury and diabetes (Bedard and Krause 2007; Block and Gorin 2012). This complex consists of a membrane-bound heterodimer of NOX2 and $p22^{phox}$, along with three cytosolic proteins – $p40^{phox}$, $p47^{phox}$ and $p67^{phox}$ – and Rac GTPase, a G protein (Groemping and Rittinger 2005; Bedard and Krause 2007). The interaction between $p47^{phox}$ and $p22^{phox}$ has been shown to be a key interaction in the formation of the active complex and a compound that inhibits this interaction, Ebselen, has been published (Sakurai et al. 2006; Smith et al. 2012), although not characterised in-depth.

The $p47^{phox}$ protein is known to have two conformations: it has two domains, SH3A and SH3B which form one compact elongated form where the two domains are in close contact along their length and one 'open' form where the domains are separated (Groemping et al. 2003). On binding the $p22^{phox}$ peptide, $p47^{phox}$ adopts the more compact form.

The study "Developing Inhibitors of the $p47^{phox}$–$p22^{phox}$ Protein−Protein Interaction by Fragment-Based Drug Discovery" by Sara Solbak et al. shows a fragment-based drug design pathway, from screening using fluorescence polarisation and thermal shift assays to further characterisation and binding studies using NMR and surface plasmon residence studies.

The validation screens of the hits from FP and TSA using NMR showed that all of the tryptophans (and therefore the tryptophans of both SH3 domain regions) were being affected by the addition of two of the fragments, while only a subset of the tryptophans were affected by the other fragments, indicating that only one domain was being affected at a time.

SAR was carried out on these two fragments and, after optimisations, a PEG-based linger was used between two molecules to form a dimer of each of the resulting fragments. This led to a large increase in affinity, with the best of these shown to have a K_i of 20 μM, comparable to that of the $p22^{phox}$ peptide.

4. Conclusions

Structural biology in the drug discovery process still faces numerous challenges. This thesis presents methodological solutions to several of these issues, along with applications to drug targets with relevance in infectious diseases.

The antibiotic resistance protein IMP-13 was structurally characterised using X-ray crystallography in complex with several antibiotic substrates and in the apo state. The protein dynamics were also elucidated with the help of NMR. This allowed subsequent development of a novel screening approach, which has since widened the variety of fragment hits, including several that were able to be crystallised.

The crystal structure of TcPEX14 was elucidated and its solution NMR assignment obtained. It was screened using ligand-observed CPMG and fluorine experiments in collaboration with Novartis, as well as with protein-observed NMR. As these structural information on these fragments could be obtained by neither X-ray crystallography nor intermolecular NOE experiments, the use of paramagnetic NMR was explored. PEX14 was tagged with several lanthanide-binding tags, including a novel tag. Existing theory and applications of paramagnetic relaxation enhancement and pseudocontact shifts in drug discovery were extensively reviewed.

An algorithm for predicting water interaction sites on the surface of a protein using deep learning was developed and tested on crystal structures presented in this thesis, as well as on structures from the pdb. It was found that this was able to replicate key parameters found in the training set, such as likely distance from protein surface and number of proximal protein atoms, as well as providing better recall than standard algorithms using electron density at low resolutions.

A software method for automatic analysis of 2D NMR screening data was also developed, trained and tested. Furthermore, a collaborative project on $p47^{phox}$ highlighted an additional use of NMR CSP data and led to the development of a compound series that bound simultaneously to two domains of the target protein, indicating the strength of multi-faceted approaches to challenging targets.

In summary, whilst there are a huge number of problems and challenges throughout the drug discovery process, some caused by the characteristics of the biomolecular target and others by the intrinsic flaws in the techniques applied, by utilising new methodology or orthogonal methods in new ways, these can slowly be broken down and combatted in order to improve the drug development toolbox available. Whether combining *in silico* and experimental techniques, or chemical alterations of the target with standard structural techniques, novel integrative approaches can open doors to new knowledge and expertise, leading to faster development of new molecules to enhance or inhibit biomolecular activity.

5. Abbreviations

AIR	Autoinhibitory region
Cryo-EM	Cryo-electron microscopy
CSP	Chemical shift perturbation
DMSO	Dimethylsulphoxide
EC50	Half-maximal effective concentration
ESRF	European synchrotron radiation facility
FBS	Fragment-based screening
FDA	US Food and Drug Administration
FP	Fluorescence polarisation
GTP	Guanosine triphosphate
HAT	Human African Trypanosomiasis
hetNOE	Heteronuclear nuclear overhauser effect
HMQC	Heteronuclear multi-quantum correlation
HSQC	Heteronuclear single quantum correlation
HTS	High-throughput screening
IMP	Imipenemase
INPHARMA	Interligand NOEs for pharmacophore mapping
MBL	Metallo-β-lactamase
MD	Molecular Dynamics
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NOX2	NADPH oxidase 2
NTA	Nitrilotriacetic acid
PBP	Penicillin Binding Protein
PCS	Pseudocontact shifts
PDB	Protein data bank
PEG	Poly-ethylene glycol
PEX	Peroxin protein
Phox	Phagocytic oxidase
PPI	Protein-protein interface
PRE	Paramagnetic relaxation enhancement
PTS	Peroxisomal targeting sequence
RDC	Residual dipolar couplings
RNA	Ribonucleic acid
ROS	Reactive oxygen species
SAR	Structure-activity relationship
SAXS	Small angle X-ray scattering
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
STD	Saturation transfer difference
T. brucei	*Trypanosoma brucei*
T. cruzi	*Trypanosoma cruzi*
TPR	Tetratricopeptide repeat
TSA	Thermal shift assay
WaterLOGSY	Water-ligand observed via gradient spectroscopy
WHO	World Health Organisation

6. Bibliography

Adams PD, Afonine P V, Bunkóczi G, et al (2011) The Phenix software for automated determination of macromolecular structures. Methods 55:94–106. https://doi.org/10.1016/j.ymeth.2011.07.005

Aubourg P, Scotto J, Rocchiccioli F, et al (1986) Neonatal adrenoleukodystrophy. J Neurol Neurosurg Psychiatry 49:77–86. https://doi.org/10.1136/jnnp.49.1.77

Babior BM, Kipnes RS, Curnutte JT (1973) Biological defense mechanisms. The production by leukocytes of superoxide, a potential bactericidal agent. J Clin Invest 52:741–744. https://doi.org/10.1172/JCI107236

Bade R, Chan HF, Reynisson J (2010) Characteristics of known drug space. Natural products, their derivatives and synthetic drugs. Eur J Med Chem 45:5646–5652. https://doi.org/10.1016/j.ejmech.2010.09.018

Badger J (2012) Crystallographic fragment screening. Methods Mol Biol 841:161–177. https://doi.org/10.1007/978-1-61779-520-6_7

Balaram P, Bothner-By AA, Dadok J (1972) Negative nuclear Overhauser effects as probes of macromolecular structure. J Am Chem Soc 94:4015–4017. https://doi.org/10.1021/ja00766a063

Baldridge CW, Gerard RW (1932) The extra respiration of phagocytosis. Am J Physiol Content 103:235–236. https://doi.org/10.1152/ajplegacy.1932.103.1.235

Bangi E (2019) A drosophila based cancer drug discovery framework. Adv Exp Med Biol 1167:237–248. https://doi.org/10.1007/978-3-030-23629-8_14

Barbieri R, Luchinat C, Parigi G (2004) Backbone-only protein solution structures with a combination of classical and paramagnetism-based constraints: A method that can be scaled to large molecules. ChemPhysChem 5:797–806. https://doi.org/10.1002/cphc.200301058

Bedard K, Krause KH (2007) The NOX family of ROS-generating NADPH oxidases: Physiology and pathophysiology. Physiol Rev 87:245–313. https://doi.org/10.1152/physrev.00044.2005

Bern C, Montgomery SP (2009) An Estimate of the Burden of Chagas Disease in the United States. Clin Infect Dis 49:e52–e54. https://doi.org/10.1086/605091

Bertini I, Luchinat C, Parigi G (2002a) Magnetic susceptibility in paramagnetic NMR. Prog Nucl Magn Reson Spectrosc 40:249–273. https://doi.org/10.1016/S0079-6565(02)00002-X

Bertini I, Luchinat C, Parigi G (2002b) Paramagnetic constraints: An aid for quick solution structure determination of paramagnetic metalloproteins. Concepts Magn Reson Part A 14:259–286. https://doi.org/10.1002/cmr.10027

Block K, Gorin Y (2012) Aiding and abetting roles of NOX oxidases in cellular transformation. Nat Rev Cancer 12:627–637. https://doi.org/10.1038/nrc3339

Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem Phys Lett 69:185–189. https://doi.org/10.1016/0009-2614(80)80041-8

Bosse R, Illy C, Chelsky D (2002) Clinical Screening Amplified Luminescent Proximity Homogenous Assay

Brini E, Fennell CJ, Fernandez-Serra M, et al (2017) How Water's Properties Are Encoded in Its Molecular Structure and Energies. Chem Rev 117:12385–12414.

https://doi.org/10.1021/acs.chemrev.7b00259

Brown L-A, Baker A (2008) Shuttles and cycles: transport of proteins into the peroxisome matrix. Mol Membr Biol 25:363–375. https://doi.org/10.1080/09687680802130583

Burch RM, Kyle DJ (1991) Mass Receptor Screening for New Drugs. Pharm Res 8:141–147. https://doi.org/10.1023/A:1015819431855

Bush K (2018) Past and present perspectives on β-lactamases. Antimicrob Agents Chemother 62:e01076-18. https://doi.org/10.1128/AAC.01076-18

Carretero M, Solis GM, Petrascheck M (2017) C. elegans as Model for Drug Discovery. Curr Top Med Chem 17:2067–2076. https://doi.org/10.2174/1568026617666170131114401

Chatterjee S, Ganguli BN (1986) Improved screening for β-Lactam antibiotics a sensitive, high-throughput assay using dd-carboxypeptidase and a novel chromophore-labeled substrate. J Antibiot (Tokyo) 39:53–57. https://doi.org/10.7164/antibiotics.39.53

Clore MG, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev 109:4108–4139. https://doi.org/10.1021/cr900033p

Cooper J (2020) Peroxisome. In: Encycl. Br. https://www.britannica.com/science/peroxisome. Accessed 20 May 2020

Cudic M, Fields G (2009) Extracellular Proteases as Targets for Drug Development. Curr Protein Pept Sci 10:297–307. https://doi.org/10.2174/138920309788922207

Dalvit C, Fogliatto GP, Stewart A, et al (2001) WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. J Biomol NMR 21:349–359. https://doi.org/10.1023/A:1013302231549

Dauter Z (2017) Collection of X-ray diffraction data from macromolecular crystals. Methods Mol Biol 1607:165–184. https://doi.org/10.1007/978-1-4939-7000-1_7

Dawidowski M, Emmanouilidis L, Kalel VC, et al (2017) Inhibitors of PEX14 disrupt protein import into glycosomes and kill Trypanosoma parasites. Science (80-) 355:1416–1420. https://doi.org/10.1126/science.aal1807

Dawidowski M, Kalel VC, Napolitano V, et al (2020) Structure-Activity Relationship in Pyrazolo[4,3-c]pyridines, First Inhibitors of PEX14-PEX5 Protein-Protein Interaction with Trypanocidal Activity. J Med Chem 63:847–879. https://doi.org/10.1021/acs.jmedchem.9b01876

De Duve C, Baudhuin P (1966) Peroxisomes (microbodies and related particles). Physiol Rev 46:323–357. https://doi.org/10.1152/physrev.1966.46.2.323

Delaglio F, Grzesiek S, Vuister GW, et al (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293. https://doi.org/10.1007/BF00197809

Denis M, Softley CA, Giuntini S, et al (2020) The photo-catalysed thiol-ene reaction: a new tag to yield fast, selective and irreversible paramagnetic tagging of proteins. ChemPhysChem 21:863–869. https://doi.org/10.1002/cphc.202000071

Desborough MJR, Keeling DM (2017) The aspirin story - from willow to wonder drug. Br J Haematol 177:674–683. https://doi.org/10.1111/bjh.14520

East KW, Leith A, Ragavendran A, et al (2019) NMRdock: Lightweight and Modular NMR Processing. BioRxiv. https://doi.org/10.1101/679688

Eder J, Sedrani R, Wiesmann C (2014) The discovery of first-in-class drugs: Origins and evolution. Nat Rev Drug Discov 13:577–587. https://doi.org/10.1038/nrd4336

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493

Erlanson DA, Davis BJ, Jahnke W (2019) Fragment-Based Drug Discovery: Advancing Fragments in the Absence of Crystal Structures. Cell Chem Biol 26:9–15. https://doi.org/10.1016/j.chembiol.2018.10.001

European Medicines Agency (2018) CHMP recommends first oral-only treatment for sleeping sickness

Farrow NA, Muhandiram R, Pascal SM, et al (1994) Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. Biochemistry 33:5984–6003. https://doi.org/10.1021/bi00185a040

Fino R, Byrne R, Softley CA, et al (2020) Introducing the CSP Analyzer: A novel Machine Learning-based application for automated analysis of two-dimensional NMR spectra in NMR fragment-based screening. Comput Struct Biotechnol J 18:603–611. https://doi.org/10.1016/j.csbj.2020.02.015

Goldfischer S, Moore CL, Johnson AB, et al (1973) Peroxisomal and Mitochondrial Defects in the Cerebro-Hepato-Renal Syndrome. Science (80-) 182:62–64. https://doi.org/10.1126/science.182.4107.62

Gossert AD, Jahnke W (2016) NMR in drug discovery: A practical guide to identification and validation of ligands interacting with biological macromolecules. Prog Nucl Magn Reson Spectrosc 97:82–125. https://doi.org/10.1016/j.pnmrs.2016.09.001

Groemping Y, Lapouge K, Smerdon SJ, Rittinger K (2003) Molecular basis of phosphorylation-induced activation of the NADPH oxidase. Cell 113:343–355. https://doi.org/10.1016/S0092-8674(03)00314-3

Groemping Y, Rittinger K (2005) Activation and assembly of the NADPH oxidase: A structural perspective. Biochem J 386:401–416. https://doi.org/10.1042/BJ20041835

Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. J Magn Reson 96:432–440. https://doi.org/10.1016/0022-2364(92)90099-S

Haanstra JR, González-Marcano EB, Gualdrón-López M, Michels PAM (2016) Biogenesis, maintenance and dynamics of glycosomes in trypanosomatid parasites. Biochim Biophys Acta - Mol Cell Res 1863:1038–1048. https://doi.org/10.1016/j.bbamcr.2015.09.015

Hajduk PJ (2006) SAR by NMR: Putting the pieces together. Mol Interv 6:266–272. https://doi.org/10.1124/mi.6.5.8

Hajduk PJ, Olejniczak ET, Fesik SW (1997) One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules. J Am Chem Soc 119:12257–12261. https://doi.org/10.1021/ja9715962

Han S, Zaniewski RP, Marr ES, et al (2010) Structural basis for effectiveness of siderophore- conjugated monocarbams against clinically relevant strains of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A 107:22002–22007. https://doi.org/10.1073/pnas.1013092107

Häussinger D, Huang JR, Grzesiek S (2009) DOTA-M8: An extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy. J Am Chem Soc 131:14761–14767. https://doi.org/10.1021/ja903233w

Ho PL, Lo WU, Chan J, et al (2014) pIMP-PH114 carrying blaIMP-4in a Klebsiella pneumoniae strain is closely related to other multidrug-resistant IncA/C2 plasmids. Curr Microbiol 68:227–232. https://doi.org/10.1007/s00284-013-0471-x

Hong DJ, Bae IK, Jang IH, et al (2015) Epidemiology and characteristics of metallo-ß-lactamase-producing Pseudomonas aeruginosa. Infect Chemother 47:81–97. https://doi.org/10.3947/ic.2015.47.2.81

Hong S, Kim D (2016) Interaction between bound water molecules and local protein structures: A statistical analysis of the hydrogen bond structures around bound water molecules. Proteins Struct Funct Bioinforma 84:43–51. https://doi.org/10.1002/prot.24953

Hu MK (1962) Visual Pattern Recognition by Moment Invariants. IRE Trans Inf Theory 8:179–187. https://doi.org/10.1109/TIT.1962.1057692

Huber W, Sinopoli A, Kohler J, et al (2015) Elucidation of direct competition and allosteric modulation of small-molecular-weight protein ligands using surface plasmon resonance methods. J Mol Recognit 28:480–491. https://doi.org/10.1002/jmr.2465

Huynh K, Partch CL (2015) Analysis of protein stability and ligand interactions by thermal shift assay. Curr Protoc protein Sci 79:28.9.1-28.9.14. https://doi.org/10.1002/0471140864.ps2809s79

Ikura M, Krinks M, Torchia DA, Bax A (1990) An efficient NMR approach for obtaining sequence-specific resonance assignments of larger proteins based on multiple isotopic labeling. FEBS Lett 266:155–158. https://doi.org/10.1016/0014-5793(90)81528-V

Iyer GYN, Islam MF, Quastel JH (1961) Biochemical aspects of phagocytosis. Nature 192:535–541. https://doi.org/10.1038/192535a0

Jaskolski M, Dauter Z, Wlodawer A (2014) A brief history of macromolecular crystallography, illustrated by a family tree and its Nobel fruits. FEBS J 281:3985–4009. https://doi.org/10.1111/febs.12796

Joss D, Häussinger D (2019) Design and applications of lanthanide chelating tags for pseudocontact shift NMR spectroscopy with biomacromolecules. Prog Nucl Magn Reson Spectrosc 114–115:284–312. https://doi.org/10.1016/j.pnmrs.2019.08.002

Kabsch W (2010) XDS. Acta Crystallogr Sect D Biol Crystallogr 66:125–132. https://doi.org/10.1107/S0907444909047337

Kalel VC, Emmanouilidis L, Dawidowski M, et al (2017) Inhibitors of glycosomal protein import provide new leads against trypanosomiasis. Microb Cell 4:229–232. https://doi.org/10.15698/mic2017.07.581

Katz L, Baltz RH (2016) Natural product discovery: past, present, and future. J Ind Microbiol Biotechnol 43:155–176. https://doi.org/10.1007/s10295-015-1723-5

Keeler J, Wothers P (2013) Chemical Structure and Reactivity: An Integrated Approach. OUP Oxford

Kennedy PGE, Rodgers J (2019) Clinical and neuropathogenetic aspects of human African trypanosomiasis. Front Immunol 10:39. https://doi.org/10.3389/fimmu.2019.00039

Krimm I (2015) Le criblage de fragments: Une voie prometteuse pour la conception de médicaments. médecine/sciences 31:197–202. https://doi.org/10.1051/medsci/20153102017

Lacabanne D, Meier BH, Böckmann A (2018) Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. J Biomol NMR 71:141–150. https://doi.org/10.1007/s10858-017-0156-z

Lea WA, Simeonov A (2011) Fluorescence polarization assays in small molecule screening. Expert Opin Drug Discov 6:17–32. https://doi.org/10.1517/17460441.2011.537322

Leto TL, Adams AG, De Mendez I (1994) Assembly of the phagocyte NADPH oxidase: Binding of Src homology 3 domains to proline-rich targets. Proc Natl Acad Sci U S A 91:10650–10654. https://doi.org/10.1073/pnas.91.22.10650

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development q settings. Adv Drug Deliv Rev 46:3–26. https://doi.org/10.1016/s0169-409x(00)00129-0

Ludwig C, Guenther UL (2009) Ligand based NMR methods for drug discovery. Front Biosci 14:4565–4574. https://doi.org/10.2741/3549

Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew Chemie Int Ed 38:1784–1788. https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q

McPherson A, Gavira JA (2014) Introduction to protein crystallization. Acta Crystallogr Sect F Struct Biol Commun 70:2–20. https://doi.org/10.1107/S2053230X13033141

Mendes RE, Toleman MA, Ribeiro J, et al (2004) Integron carrying a novel metallo-β-lactamase gene, bla IMP-16, and a fused form of aminoglycoside-resistant gene aac(6')-30/aac(6')-Ib': Report from the SENTRY Antimicrobial Surveillance Program. Antimicrob Agents Chemother 48:4693–4702. https://doi.org/10.1128/AAC.48.12.4693-4702.2004

Michels PAM, Moyersoen J, Krazy H, et al (2005) Peroxisomes, glyoxysomes and glycosomes. Mol Membr Biol 22:133–145. https://doi.org/10.1080/09687860400024186

Müller I (2017) Guidelines for the successful generation of protein-ligand complex crystals. Acta Crystallogr Sect D Struct Biol 73:79–92. https://doi.org/10.1107/S2059798316020271

Neufeld C, Filipp F V, Simon B, et al (2009) Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. EMBO J 28:745–754. https://doi.org/10.1038/emboj.2009.7

Neuhaus A, Kooshapur H, Wolf J, et al (2014) A novel Pex14 protein-interacting site of human Pex5 is critical for matrix protein import into peroxisomes. J Biol Chem 289:437–448. https://doi.org/10.1074/jbc.M113.499707

Neuhaus D, van Mierlo CPM (1992) Measurement of heteronuclear NOE enhancements in biological macromolecules. A convenient pulse sequence for use with aqueous solutions. J Magn Reson 100:221–228. https://doi.org/10.1016/0022-2364(92)90382-H

Nguyen HH, Park J, Kang S, Kim M (2015) Surface plasmon resonance: A versatile technique for biosensor applications. Sensors (Switzerland) 15:10481–10510. https://doi.org/10.3390/s150510481

Ni F (1994) Recent developments in transferred NOE methods. Prog Nucl Magn Reson Spectrosc 26:517–606. https://doi.org/10.1016/0079-6565(94)90000-0

Nicholls RA, Long F, Murshudov GN (2012) Low-resolution refinement tools in REFMAC5. Acta Crystallogr Sect D Biol Crystallogr 68:404–417. https://doi.org/10.1107/S090744491105606X

Oschkinat H, Griesinger C, Kraulis PJ, et al (1988) Three-dimensional NMR spectroscopy of a protein in solution. Nature 332:374–376. https://doi.org/10.1038/332374a0

Otting G, Wüthrich K (1990) Heteronuclear filters in two-dimensional [1H, 1H]-NMR spectroscopy: Combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions. Q Rev Biophys 23:39–96. https://doi.org/10.1017/S0033583500005412

Pardi A (1992) Isotope labelling for NMR studies of biomolecules. Curr Opin Struct Biol 2:832–835. https://doi.org/10.1016/0959-440X(92)90107-I

Parkos CA, Dinauer MC, Jesaitis AJ, et al (1989) Absence of both the 91kD and 22kD subunits of human neutrophil cytochrome b in two genetic forms of chronic granulomatous disease. Blood 73:1416–1420. https://doi.org/10.1182/blood.v73.6.1416.bloodjournal7361416

Patching SG (2014) Surface plasmon resonance spectroscopy for characterisation of membrane protein-ligand interactions and its potential for drug discovery. Biochim Biophys Acta - Biomembr 1838:43–55. https://doi.org/10.1016/j.bbamem.2013.04.028

Peng C, Frommlet A, Perez M, et al (2016) Fast and Efficient Fragment-Based Lead Generation by Fully Automated Processing and Analysis of Ligand-Observed NMR Binding Data. J Med Chem 59:3303–3310. https://doi.org/10.1021/acs.jmedchem.6b00019

Pérez-Molina JA, Molina I (2018) Chagas disease. Lancet 391:82–94. https://doi.org/10.1016/S0140-6736(17)31612-4

Perrin F (1926) Polarization of light of fluorescence, average life of molecules in the excited state. J Phys Radium

Pinazo M-J, Aldasoro E, Calvo-Cano A, et al (2017) Blood and Tissue Protozoa. In: Infectious Diseases, 4th edn. Elsevier, pp 1751-1762.e1

Pintacuda G, John M, Su X-C, Otting G (2007) NMR structure determination of protein-ligand complexes by lanthanide labeling. Acc Chem Res 40:206–212. https://doi.org/10.1021/ar050087z

Proudfoot A, Bussiere DE, Lingel A (2017) High-confidence protein-ligand complex modeling by NMR-guided docking enables early hit optimization. J Am Chem Soc 139:17824–17833. https://doi.org/10.1021/jacs.7b07171

Quie PG, White JG, Holmes B, Good RA (1967) In vitro bactericidal capacity of human polymorphonuclear leukocytes: diminished activity in chronic granulomatous disease of childhood. J Clin Invest 46:668–679. https://doi.org/10.1172/JCI105568

Radwanska M, Vereecke N, Deleeuw V, et al (2018) Salivarian trypanosomosis: A review of parasites involved, their global distribution and their interaction with the innate and adaptive mammalian host immune system. Front Immunol 9:2253. https://doi.org/10.3389/fimmu.2018.02253

Rhee YM, Sorin EJ, Jayachandran G, et al (2004) Simulations of the role of water in the protein-folding

mechanism. Proc Natl Acad Sci U S A 101:6456–6461. https://doi.org/10.1073/pnas.0307898101

Rinaldelli M, Carlon A, Ravera E, et al (2015) FANTEN: A new web-based interface for the analysis of magnetic anisotropy-induced NMR data. J Biomol NMR 61:21–34. https://doi.org/10.1007/s10858-014-9877-4

Rodrigues M, Baptista B, Lopes JA, Sarraguça MC (2018) Pharmaceutical cocrystallization techniques. Advances and challenges. Int J Pharm 547:404–420. https://doi.org/10.1016/j.ijpharm.2018.06.024

Rucker SP, Shaka AJ (1989) Broadband homonuclear cross polarization in 2D N.M.R. using DIPSI-2. Mol Phys 68:509–517. https://doi.org/10.1080/00268978900102331

Sakurai T, Kanayama M, Shibata T, et al (2006) Ebselen, a seleno-organic antioxidant, as an electrophile. Chem Res Toxicol 19:1196–1204. https://doi.org/10.1021/tx0601105

Sánchez-Pedregal VM, Reese M, Meiler J, et al (2005) The INPHARMA method: Protein-mediated interligand NOEs for pharmacophore mapping. Angew Chemie - Int Ed 44:4172–4175. https://doi.org/10.1002/anie.200500503

Santella G, Docquier JD, Gutkind G, et al (2011) Purification and biochemical characterization of IMP-13 metallo-β-lactamase. Antimicrob Agents Chemother 55:399–401. https://doi.org/10.1128/AAC.00421-10

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 34:93–158. https://doi.org/10.1016/S0079-6565(98)00025-9

Schanda P, Brutscher B (2005) 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 127:8014–8015. https://doi.org/10.1021/ja051306e

Schmitz C, Bonvin AMJJ (2011) Protein-protein HADDocking using exclusively pseudocontact shifts. J Biomol NMR 50:263–266. https://doi.org/10.1007/s10858-011-9514-4

Schrader M, Fahimi HD (2008) The peroxisome: Still a mysterious organelle. Histochem Cell Biol 129:421–440. https://doi.org/10.1007/s00418-008-0396-9

Schrödinger LLC The PyMOL Molecular Graphics System, Version 1.8

Smith SME, Min J, Ganesh T, et al (2012) Ebselen and congeners inhibit NADPH oxidase 2-dependent superoxide generation by interrupting the binding of regulatory subunits. Chem Biol 19:752–763. https://doi.org/10.1016/j.chembiol.2012.04.015

Softley CA, Bostock MJ, Popowicz GM, Sattler M (2020a) Paramagnetic NMR in drug discovery. J Biomol NMR 74:287–309. https://doi.org/10.1007/s10858-020-00322-0

Softley CA, Zak KM, Bostock MJ, et al (2020b) Structure and molecular recognition mechanism of IMP-13 metallo-β-lactamase. Antimicrob Agents Chemother 64:e00123-20. https://doi.org/https://doi.org/10.1128/AAC.00123-20

Solbak SMØ, Zang J, Narayanan D, et al (2020) Developing Inhibitors of the p47phox-p22phox Protein-Protein Interaction by Fragment-Based Drug Discovery. J Med Chem 63:1156–1177. https://doi.org/10.1021/acs.jmedchem.9b01492

Stone E (1763) XXXII. An account of the success of the bark of the willow in the cure of agues. In a

letter to the Right Honourable George Earl of Macclesfield, President of R. S. from the Rev. Mr. Edward Stone, of Chipping-Norton in Oxfordshire. Philos Trans R Soc London 53:195–200. https://doi.org/10.1098/rstl.1763.0033

Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. Protein Expr Purif 41:207–234. https://doi.org/10.1016/j.pep.2005.01.016

Su XC, Man B, Beeren S, et al (2008) A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy. J Am Chem Soc 130:10486–10487. https://doi.org/10.1021/ja803741f

Swinney DC, Anthony J (2011) How were new medicines discovered? Nat Rev Drug Discov 10:507–519. https://doi.org/10.1038/nrd3480

Tacconelli E, Carrara E, Savoldi A, et al (2017) Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics

Tooke CL, Hinchliffe P, Bragginton EC, et al (2019) β-Lactamases and β-Lactamase Inhibitors in the 21st Century. J Mol Biol 431:3472–3500. https://doi.org/10.1016/j.jmb.2019.04.002

Ullman EF, Kirakossian H, Singh S, et al (1994) Luminescent oxygen channeling immunoassay: Measurement of particle binding kinetics by chemiluminescence. Proc Natl Acad Sci U S A 91:5426–5430. https://doi.org/10.1073/pnas.91.12.5426

Valenti D, Neves JF, Cantrelle FX, et al (2019) Set-up and screening of a fragment library targeting the 14-3-3 protein interface. Medchemcomm 10:1796–1802. https://doi.org/10.1039/c9md00215d

Van Zundert GCP, Rodrigues JPGLM, Trellet M, et al (2016) The HADDOCK2.2 Web Server: User-friendly integrative modeling of biomolecular complexes. J Mol Biol 428:720–725. https://doi.org/10.1016/j.jmb.2015.09.014

Vranken WF, Boucher W, Stevens TJ, et al (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. Proteins Struct Funct Genet 59:687–696. https://doi.org/10.1002/prot.20449

Walter T, Erdmann R (2019) Current Advances in Protein Import into Peroxisomes. Protein J 38:351–362. https://doi.org/10.1007/s10930-019-09835-6

Watson JD, Crick FH (1953) The structure of DNA. Cold Spring Harb Symp Quant Biol 18:123–131. https://doi.org/10.1101/SQB.1953.018.01.020

WHO (2020a) Trypanosomiasis, human African (sleeping sickness). In: World Heal. Organ. Fact Sheets. https://www.who.int/en/news-room/fact-sheets/detail/trypanosomiasis-human-african-(sleeping-sickness). Accessed 22 May 2020

WHO (2020b) Chagas disease (also known as American trypanosomiasis). In: World Heal. Organ. Fact Sheets. https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis). Accessed 22 May 2020

William Orton H, Huber T, Otting G (2020) Paramagpy: software for fitting magnetic susceptibility tensors using paramagnetic effects measured in NMR spectra. Magn Reson 1:1–12. https://doi.org/10.5194/mr-1-1-2020

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc 73:1–16. https://doi.org/10.1016/j.pnmrs.2013.02.001

Winn MD, Murshudov GN, Papiz MZ (2003) Macromolecular TLS Refinement in REFMAC at Moderate Resolutions. Methods Enzymol 374:300–321. https://doi.org/10.1016/S0076-6879(03)74014-2

Zhao WH, Hu ZQ (2011) IMP-type metallo-β-lactamases in Gram-negative bacilli: Distribution, phylogeny, and association with integrons. Crit Rev Microbiol 37:214–226. https://doi.org/10.3109/1040841X.2011.559944

Zheng W, Thorne N, McKew JC (2013) Phenotypic screens as a renewed approach for drug discovery. Drug Discov Today 18:1067–1073. https://doi.org/10.1016/j.drudis.2013.07.001

7. Acknowledgements

Firstly, I would like to thank my supervisor, and project head Prof. Michael Sattler, for the opportunity and support over the past four years, and for upholding and cultivating a sense of scientific excellence and collaborative discovery within the group.

Secondly, to Dr. Grzegorz Popowicz, with whom I worked closely for the duration, my heartfelt thanks. The support and guidance along the way was invaluable and your enthusiasm for new and exciting projects is infectious and keeps us always on our toes.

My wonderful AEGIS co-workers, collaborators and friends, the constant scientific and personal exchange livened up many a day, and the schools and secondments together were a highlight of my doctoral studies. To the PIs and to Eva, thank you for the organisation and stimulating questions and input to the projects and for making AEGIS such a success.

I was lucky enough to have an amazing working environment at HMGU. To all of you, and to those at TUM, thank you for your generosity with knowledge and time; with care and support. Particular thanks go to Mark Bostock, Krzysztof Zak and Michael Ostertag, and to Astrid for making life in the lab so easy with her diligence and ever-present smile.

And last, but of course not least, to my friends and family, both in Germany and abroad, who cannot all be named, for their undying love and support, and simply for being there. In particular, to my parents, for 26 years of love, advice and encouragement; to Bex, for the laughs, for reminding me what is important in life, and for the many pictures of Max; to Abi, Alice and Joy, for motivation and support, and of course to Adam, for spiriting me away to the mountains when I needed a break, I could not have done it without you all.

8. Appendices

Appendix Details	Type
1 Structure and Molecular Recognition Mechanism of IMP-13 Metallo-β-Lactamase C. A. Softley, M. J. Bostock, K. M. Zak et al. Antimicrobial Agents and Chemotherapy https://doi.org/10.1128/AAC.00123-20	Publication
2 Paramagnetic NMR in Drug Discovery C. A. Softley, M. J. Bostock et al. Journal of Biomolecular NMR https://doi.org/10.1007/s10858-020-00322-0	Publication
3 Deep Learning Model Predicts Water Interaction Sites on the Surface of Proteins using Limited-Resolution Data J. Zaucha, C. A. Softley et al. RSC Chemical Communications (Submitted)	Publication (submitted)
4 Introducing the CSP Analyzer: a Novel Machine Learning-based Application for Automated Analysis of two-dimensional NMR spectra in NMR Fragment-based Screening R. Fino, R. Byrne et al. Computation and Structural Biotechnology Journal https://doi.org/10.1016/j.csbj.2020.02.015	Publication
5 The Photocatalysed Thiol-ene reaction: A New Tag to Yield Fast, Selective and Irreversible Paramagnetic Tagging of Proteins M. Denis et al. ChemPhysChem https://doi.org/10.1002/cphc.202000071	Publication
6 Developing Inhibitors of the p47phox−p22phox Protein−Protein Interaction by Fragment-Based Drug Discovery S. M. Ø. Solbak et al. Journal of Medicinal Chemistry https://doi.org/10.1021/acs.jmedchem.9b01492	Publication