Crystallographic Studies for Structure-Based Drug Design

NATALIE JOAN TATUM

CRYSTALLOGRAPHIC STUDIES FOR STRUCTURE-BASED DRUG DESIGN

Natalie Joan Tatum

THESIS SUBMITTED TO

THE DEPARTMENT OF BIOLOGICAL AND BIOMEDICAL SCIENCES

OF THE UNIVERSITY OF DURHAM

FOR A MASTERS DEGREE BY RESEARCH

SEPTEMBER 2011

ABSTRACT

Crystallographic Studies for Structure-Based Drug Design

Natalie J. Tatum

Structure-based drug design is an iterative design cycle reliant upon lead fragments and a known target. The result is two separate avenues of structure-based drug design, explored in this thesis: speculative, with no defined target (in the form of peptoid precursors) and rational, with a characterized target (in the form of EthR inhibitors).

Peptoids are N-substituted glycine molecules, able to mimic peptide structure and function with notable advantages for drug delivery and independent function. Here, the structures of two precursor molecules are presented as part of a wider aim to produce a tool-box of peptoid monomers for a fragment-based approach. The intramolecular interactions made by the two precursors can be extrapolated for the final peptoid.

EthR is a transcriptional repressor from *Mycobacterium tuberculosis*, a part of the activation pathway for the second-line drug ethionamide. It has been shown that inhibition of EthR results in a conformational change which renders the protein inactive, resulting in an increase in the bioactivation of ethionamide and so increased drug efficiency. Inhibitors for EthR have been designed, able to effect this response *in vitro* and *in vivo*.

This thesis details the crystallisation and structural study of peptoid precursors for use as lead fragments or monomers for peptoid click chemistry; and EthR inhibitors for improving the bioactivation and efficacy of ethionamide, with evaluation by molecular docking analysis in *AutoDock Vina*. The results show that shorter ligands, capable of engaging certain residues in the ligand-binding channel, are rated highest and so indicated to be the more effective inhibitors.

The two approaches highlight the diversity of tactics for new therapeutics and the powerful advantage of three-dimensional crystal structures, acquired through X-ray crystallography, to the drug design process.

ACKNOWLEDGEMENTS

Foremost, I would like to thank Dr. Ehmke Pohl for the invaluable, unwavering support and guidance he has given me throughout this year; it would be an understatement to say that this would not have been possible without him.

Secondly, to Dr. Martin Cann, Dr. Gary Sharples, Dr. Victoria Money and my friends and colleagues in CG231 for their advice and for enlightening tea-time conversations on the important questions of life, the universe and everything.

I'd like to thank Dr. Steven Cobb for allowing me to make a little slice of his work my own; and the same to Dr. Alain Baulard and Dr. Nicolas Willand at the Institut Pasteur de Lille and Université de Lille respectively, with added thanks for an interesting and informative trip to their labs in France. They have each been very generous with their time and work and I am exceptionally grateful.

Finally, I would like to take this opportunity to thank my parents for their endless encouragement, their unswerving support, and for never, ever letting me give up. In the end, none of this would have been imaginable, let alone possible, without my mum and dad; I cannot possibly thank them enough.

So this thesis is dedicated to my parents, for their dedication to me.

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Structural characterisation of chemical compounds has a variety of benefits. Foremost, a crystal form can be used to assess the physical attributes of the compound - from bond lengths and angles, to the torsions of individual groups – in the absence of other, potentially structure-altering molecules. Therefore, all intermolecular interactions made are unique to the compound, and can be used in the drug design process to optimise the design for a target or particular purpose.

X-ray crystallography is the most important method available to accurately determine the structure of a compound in three dimensions, and so is a powerful tool in chemical studies and the drug design process. Crystallisation of an active pharmaceutical ingredient is especially important for patenting, particularly for the pharmaceutical companies which commit large sums of money towards research into novel drug compounds. Patents are granted for single crystalline forms – typically the dosage form – and polymorphs, hydrates and salts etc. constitute separate patents (or aspects of a portfolio) which can prove lucrative while also protecting a new development.

In addition, through crystallisation, diffractometry and structure solution, molecular modelling can be used to inform the compound design and make predictions for protein-ligand binding without the need for expending resources on biological assay and co-crystallisation.

This thesis details two aspects of the rational drug design process. Firstly, the creation of lead fragments and sub-units which can be used as a base to develop active drugs: in this case, peptoid monomers. Peptoids, small amino-acid chain mimics, are proven drug alternatives for peptide therapies.[1] The well-defined secondary structures that peptoid oligomers create when linked make for a reliable drug which can have action in its own right, or act as a cellular target or transporter signal.[2,3] Characterising the structure of peptoid monomers in isolation and combination could lead to a database of three-dimensional co-ordinates, which can be used to create lead fragments or signal molecules with known and defined conformations. [4]

The second focus of this research project is a structure-based drug design process: a known target, for which the structure has been determined, is used as a

template to design molecules to affect a particular, desired response. Tuberculosis poses a global world health crisis, and drug resistance is on the rise; new targets for treatment are needed in order to make headway in reducing the rising incidence rate.[5,6]

EthR, a transcriptional regulator in *Mycobacterium tuberculosis*, has been proven to be a viable drug target in treatment of tuberculosis;[7,8] inhibitors of this protein delivered in conjunction with the anti-TB drug ethionamide (ETH) have been shown to reduce the bacterial load as effectively as a three times higher dose of ETH alone.[9] Acquiring the crystal structures of these inhibitors in isolation will not only give insight into the conformation the molecule takes in the absence of the target protein, but also inform the further design process. Optimisation of the inhibitor design requires extensive biophysical assessment if conducted *in vivo* and *in vitro*; molecular docking with software *AutoDock Vina* can use the crystal structure of the inhibitor to predict binding conformations in the protein crystal structure.[10] Docking studies will allow *in silico* examination of the putative protein-ligand complex.

The project itself relies upon the ability to successfully crystallise the compounds for X-ray diffraction analysis; this is not always guaranteed, not least due to the small amounts of each compound provided. This thesis will detail the background of the two avenues of my research project on the structural features of drug and intermediate molecules. The methods used in the endeavour to grow crystals for analysis will be explained, as well as the implications from the data acquired by X-ray diffraction and molecular docking analyses. Finally, conclusions will be drawn from the data which will be used to further the research being conducted in these areas.

CHAPTER 2

BACKGROUND

2.1 | Peptoids

Peptoids are defined as N-substituted glycine molecules which mimic α-amino acid peptides when oligomerised (figure 2.1). The term derives from 'peptidomimetic', which was introduced to indicate that although peptoids are structurally distinct from peptides, they have been shown to mimic peptide biological function. When first developed, they were envisioned as a customisable library of lead compounds for drug discovery.[1] However, they have since been revealed to be useful molecules in their own right.[11]

In practical terms, peptoids can be both drugs and drug delivery agents, vastly outclassing peptide therapies. As drugs, peptides are particularly desirable due to their ability to bind both proteins and DNA effectively. However, their sensitivity to proteases and the limited ability to cross the cell wall of a target are clear stumbling blocks when it comes to using peptides as antimicrobial agents. Peptoids, with their intrinsic resistance to proteolysis and the potential to be designed with greater cell permeability, are a proven alternative.[4]

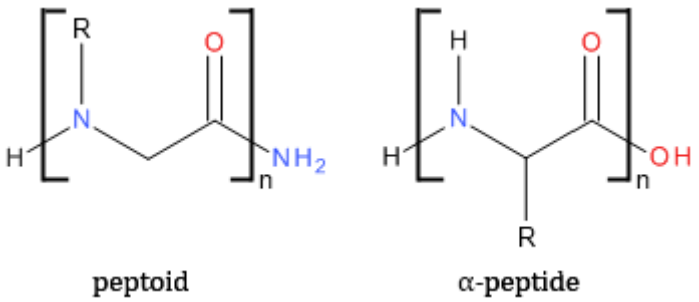


Figure 2.1 | An N-substituted glycine template for peptoids and a classic α peptide.[1]

2.1.1 | Peptoids as transport signals

The ability to cross the cell membrane is of particular importance to using peptides or peptoids as drug molecules. Peptides are largely polar, hydrophilic molecules which are difficult to transport across the hydrophobic, lipophilic cell membrane. However, peptoids analogous to peptide sequences have been shown to be more cell-permeable by a factor of up to 20-fold.[12] This gives peptoids great potential in drug delivery or targeting.

Peptoid sequences have been successfully used as transporter sequences to target drug molecules across cell membranes. For example, the HIV type I Tat protein is an extracellular protein which enters target cells to facilitate viral replication. It is able to do this by an arginine-rich sequence which allows active cellular uptake.[13] The truncated sequence from HIV Tat-1 has been capably used as a drug transporter, as have similar poly-lysine and poly-arginine peptides; it has been put to use recently to deliver a putative tumor repressor protein to hepatoma cells (Chen et al. 2011). Therefore, an analogous peptoid was designed and shown to be resistant to proteases, cheaper and more efficiently produced, and crucially was a better transporter than the original Tat sequence.[3]

Substitution of arginine and lysine residues in an insect antimicrobial peptide with *N*Arginine peptoid monomers (ie. the arginine side chain is attached to the nitrogen rather than the C^{α}) had a reduced effect compared to pure peptoid transport sequences. The peptoid-peptide hybrid was able to cross the plasma membrane and did not show cytotoxic activity; as the substitutions were placed closer to the C-terminal, the hybrid displayed reduced antimicrobial activity.[15]

Furthermore, a library of around 50,000 synthetic peptoids was screened for the ability to bind a specific activation domain of the CREB-binding protein. The hits resulted in synthetic transcription factor mimics which were found to show a 5-fold increase in the reporter gene expression (luciferase in cultured mammalian cells) compared to native transcription factors. The response was also highly selective within the cell, meaning the peptide-peptoid hybrid is specific to the target gene. This is a promising result for peptoid hybrid therapeutics.[16]

2.1.2 | Peptoids with intrinsic drug action

Peptoids can also have drug action in their own right, proving successful in binding crucial, shallow regions on the surfaces of important targets – for example, the vascular endothelial growth factor receptor 2, VEGFR2.[17] Though some clinical successes have been found among peptides, their development is hindered by the various drawbacks to peptides as drugs. Peptoid ligands were found for VEGFR2 through a screening of a peptoid library, and a dimeric peptoid was found to not only interact solely with the receptor, but was an agonist of its function *in vitro*.[2]

In a pre-clinical mouse model, this dimeric peptoid was also seen to have an *in vivo* agonising effect on tumour growth – smaller tumours and a reduced tumour growth rate were observed in comparison to the control animals. Peptoids have also

been seen to inhibit the interaction between HDM2 and p53; over-expression of the former has been linked to aggressive tumours and drug resistance.[18]

Peptoid ligands have also been designed with both agonist and antagonist function against neuropeptide receptors; this has potential applications in hormonal and physiological conditions.[19]

In early 2011, a paper was published showing peptoids as potential 'capture agents' for antibody biomarkers in diseases such as Alzheimer's Disease and multiple sclerosis; the initial studies on mouse models and AD patients (with healthy control patients) have since expanded.[20] Direct peptoid-based therapies are therefore a clinically viable option to current medical conundrums.

2.1.3 | Peptoid secondary structure

The key to all of the aforementioned peptoid uses is the ability to fine-tune the peptoid spatial structure through different N-substituted glycine monomers. Peptoid side chains, placed on the amide nitrogen, are crucial to peptoid function. The folding of the oligomeric peptoid is dictated by these side chains and their effect on the achiral backbone. Without any amide hydrogens in this backbone, there can be no intra-chain hydrogen bond formation. This makes the peptoid chain very flexible, but including monomers with large side chains will restrict the conformations the oligomer is able to adopt; aromatic side chains form intramolecular stacking interactions and electron deficient groups have been observed to stabilise the secondary structure through interactions with the backbone. It follows that, just as in peptides, the primary structure determines the secondary structure adopted, and peptoids have well-defined secondary structures.[21] Folded peptoid conformations can be analogous to α-peptides, even when amino acids are replaced like-for-like with their N-substituted counterparts.[22]

With such well-defined secondary structures derived from the primary sequence, and a history of using peptoid libraries as a starting point in determining drug candidates, well-characterised monomers can be of great use with an eventual 'toolbox' of peptoid monomers with known conformations to be screened for use on a wide range of targets.

2.2 | Inhibitors of EthR, co-drug candidates for ethionamide in TB treatment

Tuberculosis remains a global health crisis. It is anticipated that 2011 will see close to 10 million cases of tuberculosis (TB) worldwide.[23] While global incidence rates per capita are falling year on year by an average of 1%, the global population is increasing by 2% at the same rate; the absolute number of new cases of tuberculosis is in fact rising. This epidemic, which is controlled in Western Europe and the United States, is uncontrollable in sub-Saharan Africa, Eastern Europe and Asia due to high HIV co-infection rates and poor access to effective healthcare. One issue which is common to poor TB control worldwide is the poor efficiency and observance of the Directly Observed and Short-Course therapy (DOTS), with a consequent rise in MDR-TB: multi-drug resistant tuberculosis.[5]

2.2.1 | The life cycle of *Mycobacterium tuberculosis*

TB is caused by *M. tuberculosis* (*Mtb*), a Gram-positive mycobacterium which has an unusual, lipid-rich cell envelope. Mycolic acids are a vital component of the cell well, consisting of a β-hydroxy fatty acid, typically C_{54}-C_{64} atoms long, and an α-branched C_{22}-C_{24} alkyl side chain.[24] Characteristic cyclopropane rings in the chains of the mycolic acid have a structural and functional role, maintaining the integrity of this unique cell wall and protecting the wall and so the mycobacterium from oxidative stress – the primary defence of the mammalian immune system. This gives the bacilli a measure of resistance to the action of the host's innate immune system, essential as the *Mtb* life cycle requires phagocytosis into alveolar macrophages in the host lungs.[25]

Transmission occurs by inhalation of bacilli (or even as few as a single bacillus) in liquid droplets, exhaled by a person with an active TB infection. The droplets are robust enough to remain airborne and transmissible for several hours. It is understood through animal models and studies of human TB progression that the inhaled bacteria are phagocytosed into macrophages on the alveolar epithelial surface; the mycobacterial cell wall fuses with the phagosome membrane, preventing lysosome fusion and attack upon the subsumed pathogen.[26] Therefore, the bacteria are able to replicate in the host cell; the *Mtb* replication cycle takes 15-20 hours, a comparatively slow process when it is considered that the *E. coli* cycle is only 20 minutes; *Mtb* is protected within the macrophages of the mammalian immune system, and so speedy replication is not a priority. The infected macrophages migrate into the alveolar epithelial layer; this induces a localized inflammatory response, recruiting mononuclear cells from neighbouring blood vessels and providing the dividing bacteria with fresh host cells.

A granuloma forms from the macrophages and recruited monocytes and neutrophils. The macrophages begin to differentiate; some fuse into giant multinucleated cells, some into cells filled with liquid droplets ('foamy macrophages'). At 2-3 weeks after initial infection, *Mtb*-specific lymphocytes (of the acquired immune response) arrive at the site of infection and form a mantle around the centre of the granuloma. The rate of bacterial replication has slowed even further by this point, and cells are actively recruited to the vascularised granuloma to allow the infection to persist.[27]

A critical stage is reached when the population of foamy macrophages increases and the vascular structure undergoes necrosis. The granuloma eventually ruptures and collapses, releasing liquid droplets to the bacilli into the patient's lungs; these droplets can then be dispersed in a cough and exhaled into the air. At any given time in an active infection, multiple granulomas may be present in the airways at differential stages of development, with bacteria utilising different metabolic pathways and regulatory systems over the course of the life cycle. Drugs for the treatment of *Mtb* infection preferentially target replicating bacteria, when the bacilli are most vulnerable. However, this means that non-replicative bacilli, either dormant or established prior to drug delivery, will show an innate resistance to the chemotherapy.[28,29] This leads to a treatment of up to two years with the Directly Observed Treatment Short-Course (DOTS) therapy.[30]

2.2.2 | Ethionamide is a second-line drug for TB treatment

The DOTS for TB consists of 8-10 drugs, including the front line antibiotics *isoniazid* and *rifampicin*, as well as second-line drugs such as *pyrazinamide* and *ethionamide*. The second-line drugs are not as efficient, with dose-related side-effects ranging from nausea and abdominal cramps to hepatotoxicity and neuropathies. With patients facing up to two years of such side-effects, it is no surprise that the treatment is often not fully observed and as a result, 30% of patients experience treatment failure.[31]

Although *isoniazid* (INH) was identified as an anti-TB agent in the early 1950s, the mode of action was not identified until much later. Initial observations of the interference of INH with the formation of the cell envelope and mycolic acid synthesis were made by in 1970;[32] two years later, it was shown that INH treatment leads to an immediate inhibition of mycolic acid synthesis.[33] After 24 hours, growth of these drug-treated bacilli had slowed to a complete halt. Viability was completely destroyed within a day, and provided INH treatment continued until this point, inhibition would be

permanent; if INH treatment stopped before cell viability was abolished, mycolic acid production would recover.

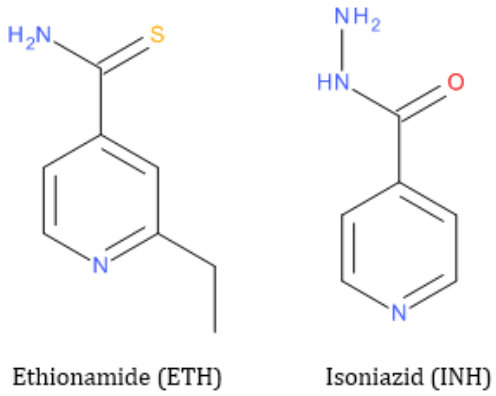


Figure 2.2 | Chemical structures of ETH and INH

The second-line drug *ethionamide* (ETH) is analogous to INH; the chemical structures of both can be seen in figure 2.2. While they have different bioactivation enzymes, controlled by distinct pathways, they share a common mode of action: inhibiting the enoyl-acyl reductase, InhA, of the fatty acid synthase II (FASII) system in *Mtb*.[34] InhA preferentially reduces long-chain 2-trans-enoyl-ACP substrates of 12-14 carbons in length in an NADH-dependent manner. This is essential for fatty acid elongation, and so inhibition prevents the production of mycolic acids for the formation of the mycobacterial outer surface, derailing bacterial replication and causing the cells to lyse.[35]

Though INH and ETH have been in use since 1952 and 1957 respectively as anti- tubercular drugs, their target was not conclusively identified until 1994.[36] The *inhA* gene was identified through INH-resistant clinical isolates of *Mtb* and a year later the protein product InhA was characterised – enzymatically[37] and structurally.[38] It was observed that the two substrates of InhA bound sequentially to the protein, with the NADH taking priority over the enoyl substrate. It was even later still that definitive evidence for the mode of action of INH and ETH was established, as described below.

2.2.3 | Bioactivation of ETH: The story so far, and the possible targets for improvement

The lack of cross-resistance seen between INH and ETH suggested a distinct activation process for ETH, independent of the catalase peroxidise enzyme KatG (responsible for INH-activation). ETH-resistant *M. smegmatis* isolates were used to identify an open reading frame (ORF) which, when disrupted, reversed this resistance. This ORF was

found to code a regulatory protein, termed EthR for its apparent role in ETH resistance and susceptibility. A neighbouring ORF was found to have homology to known mono-oxygenases. When this putative mono-oxygenase was over-expressed in *M. smegmatis*, the mycobacteria became hypersensitive to ETH with a significant inhibition of mycolic acid synthesis with low ETH concentrations in comparison to the wild type. From this it was postulated that the supposed mono-oxygenase was the ETH activator, termed EthA. It was finally identified in 2000 as the protein product of the Mtb H37Rv gene *Rv3854c*. [39]

EthA was confirmed to be a mono-oxygenase, metabolising ETH in a two-step process which is both NADPH and O_2-dependent.[40] The product of the first step is a sulphoxide derivative, ETH-SO; as for the second step, though various metabolites of ETH have been identified[41] and characterised by NMR-based methods,[42] and HPLC,[43] the 'active' and bacteriotoxic form of ETH has never been isolated. Instead, Wang *et al*. in 2007 published their crystal structures and showed the isolation of an ETH-NAD adduct tightly bound to InhA in the normal binding pocket for NADH (figure 2.3a). A pure fraction of this adduct was shown to strongly inhibit native InhA in vitro.[44]

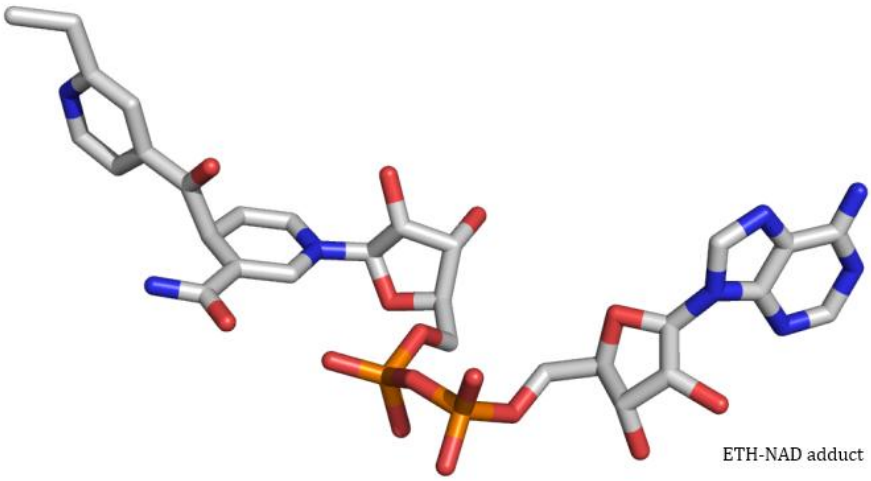


Figure 2.3a | The ETH-NAD adduct, derived from the co-crystal structure of the adduct with InhA (PDB: 2H9I). Image created using PyMol.[44]

What remains to be seen is the structure of the ETH metabolite which forms the covalent bond with NAD^+. Vanelli *et al.* suggested the cytotoxic species as a sulphonic acid derivative of ETH-SO,[40] however the later discovery of the inhibiting ETH-NAD adduct led Wang *et al*. to suggest a mechanism by which ETH^- or $ETH^\cdot$ is produced to react with NAD^+; H_2O then attacks the NH group to create the carboxyl group, as seen in figure 2.3b.

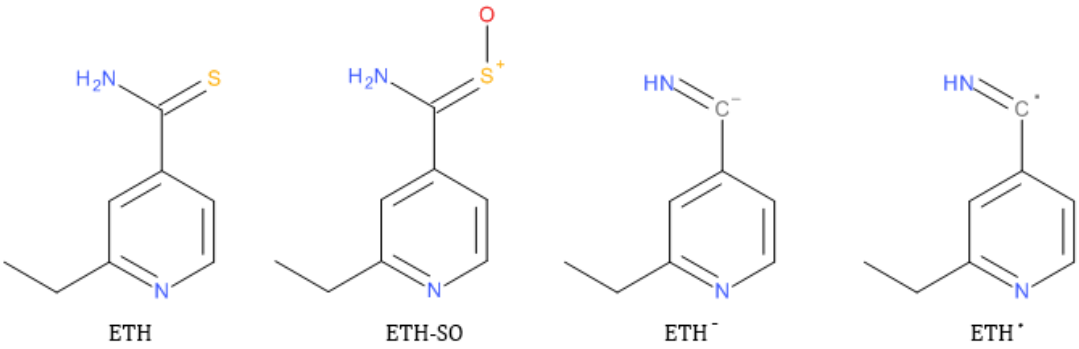


Figure 2.3b | The chemical structures of the potential intermediates of ETH bioactivation.

How this putative anion or radical is formed is unclear: is an intermediate (and as yet unidentified) enzyme required to mediate the covalent bond formation, or does EthA produce this radical as well as facilitate binding to NADH? The presence of quenched metabolites would support the latter supposition, as it would not be possible for 100% of the radical produced at the membrane-bound EthA to react with cytosolic NADH and inhibit InhA. Radical species are exceptionally reactive and would be quenched as quickly as possible to prevent intracellular damage. Added to this is the fact that EthA interacts with NADPH, not NADH, and while EthA certainly catalyses a second step, it remains unclear what this second step is.

2.2.4 | EthR is part of the TetR protein family

As previously stated, EthA expression is regulated by EthR, a protein encoded in a neighbouring open reading frame (figure 2.4) by the gene *ethR.* Under *ethR* over-expression, *ethA* was seen to be repressed (independent of ETH treatment). In those samples treated with ETH, it was observed that the mycobacteria were less sensitive to the treatment. This suggests *ethR* over-expression represses *ethA*, which in turn leads to reduced bioactivation of ETH.[45]

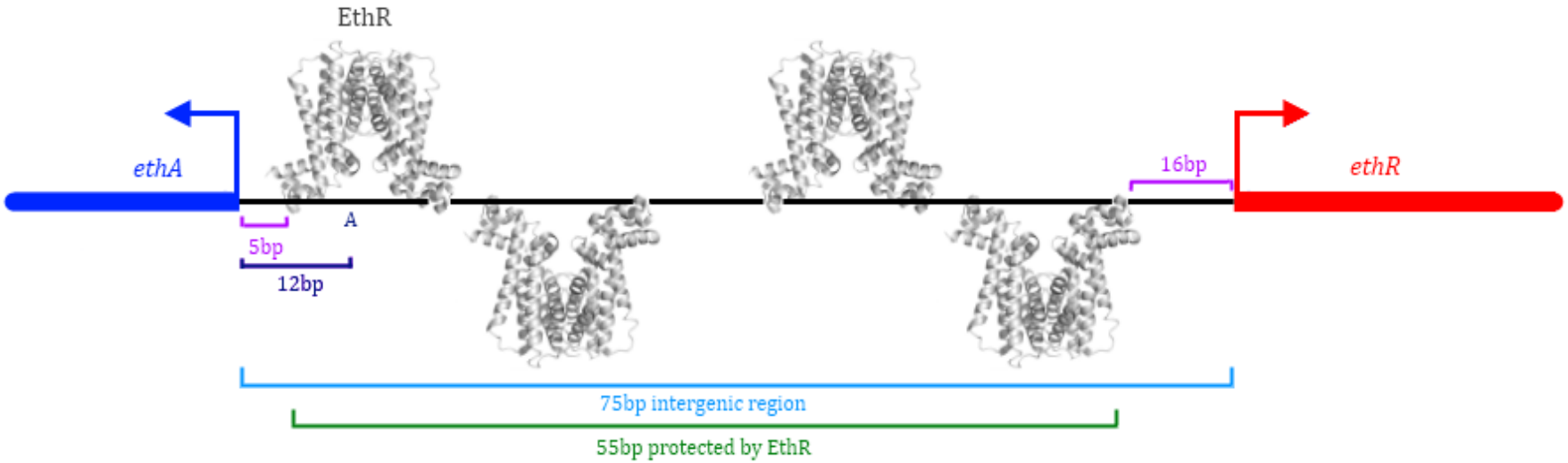


Figure 2.4 | The *ethA* operator is depicted in blue, and the *ethR* operator is depicted in red, beginning at their start codons. These genes are transcribed in opposite directions, as depicted by their respective arrows. The EthR protein, the product of *ethR* expression, protects a 55bp stretch of the 75bp intergenic region. The *ethA* transcription start site is in fact 12bp upstream of the *ethA* start codon, an adenine residue covered by EthR. This, with the unusually short intergenic region, suggests the *ethA* and *ethR* promoters overlap. This would imply – but has yet to be shown conclusively – that *ethR* auto-regulates its expression. The images of EthR (PDB: 1U9N) are of the inactive ligand-bound form, see *figure 2.5*. [45]

Mycobacterium tuberculosis H37Rv gene assignments identified EthR ('Rv3855') as a probable member of the TetR family of transcriptional repressors based on the sequence identity within the family (25%) and the presence of a putative helix-turn-helix motif, a well-known DNA-binding element.[46,47] TetR, the tetracycline expression repressor, recognises and binds a 15bp stretch of the target Tn10 promoter.[48] It was expected that EthR would bind DNA in a similar fashion, and experiments were undertaken to establish how EthR interacts with the *ethA-R* region.

2.2.5 | Determining EthR function

To determine the role of EthR in *Mtb*, the gene was cloned from *M. tuberculosis H37Rv* and expressed in *E. coli*. DNase I protection assays (by which a 323bp fragment containing the *ethA-R* intergenic region are subjected to DNase I digestion in the presence of purified His_6-tagged EthR) revealed that EthR protects 55bp in the 75bp intergenic region. This is almost four times as long as that bound by TetR, indicating EthR binds as a multimer, potentially at separate binding sites.[7]

Surface plasmon resonance was used to determine the stoichiometry of the EthR-DNA interaction. Initially, a 106bp biotinylated double-stranded DNA fragment covering the intergenic region was used as a ligand on the sensor chip. His_6-EthR was then washed onto the surface of the chip at increasing concentrations until the protein was saturated with the available ligand. Using the known masses of the DNA fragment and protein (DNA fragment of 69kDa, and the Mr of His_6-EthR being approximately 25920) the ratio of DNA to protein remaining bound to the chip could be calculated. This revealed eight molecules of His_6-EthR per 106bp DNA molecule. The SPR assay was repeated with a shorter 62bp fragment (which covered the 55bp region identified by DNase I foot-printing) and the same 8:1 stoichiometry was found. In the absence of the operator, EthR was seen to dimerise. All results taken together suggested EthR binds *ethA-R* as four homo-dimers.[7]

This extensive coverage of EthR on *ethA-R* suggests that the protein simply prevents RNA polymerase from binding the operator, repressing the expression of both genes. It follows that EthR auto-regulates its expression, a feature seen in other TetR-family repressors,[49] though no conclusive evidence has been presented to date to confirm this supposition.

2.2.6 | The crystal structure(s) of EthR

The first crystal structures of EthR were solved in 2004.[7,50] The two structures are identical, with an RMSD of 0.634 Å over 179 C^{α} atoms of the protein between residues 2 and 200.[51] These two separate groups working on EthR crystallised the protein in the same crystal form under different crystallisation conditions, indicating that the inhibited conformation is thermodynamically stable and packs very specifically, despite the two structures binding different ligands.

The structures confirmed the suspected similarity to TetR family repressors, including the N-terminal helix-turn-helix DNA-binding domain and the C-terminal ligand-binding domain (figure 2.5). EthR is a functional homodimer with a structure composed of nine α-helices. The dimer associates through the formation of a four helix bundle, which is the association of two helices from each monomer (inset figure 2.5).

Ligand-binding in DNA-binding proteins such as the TetR family of repressors has a profound effect on the conformation of the homodimer. Active homodimeric DNA-binding regulators must contact the major grooves of B-DNA, at a distance of approximately 34 Å.[52] EthR in its ligand-bound conformation was found to have a distance between the DNA-binding domains of 52.49 Å; with a ligand such as HexOc bound, the DNA-binding domains are too far apart to functionally bind DNA.[9] This essentially inactivates EthR.

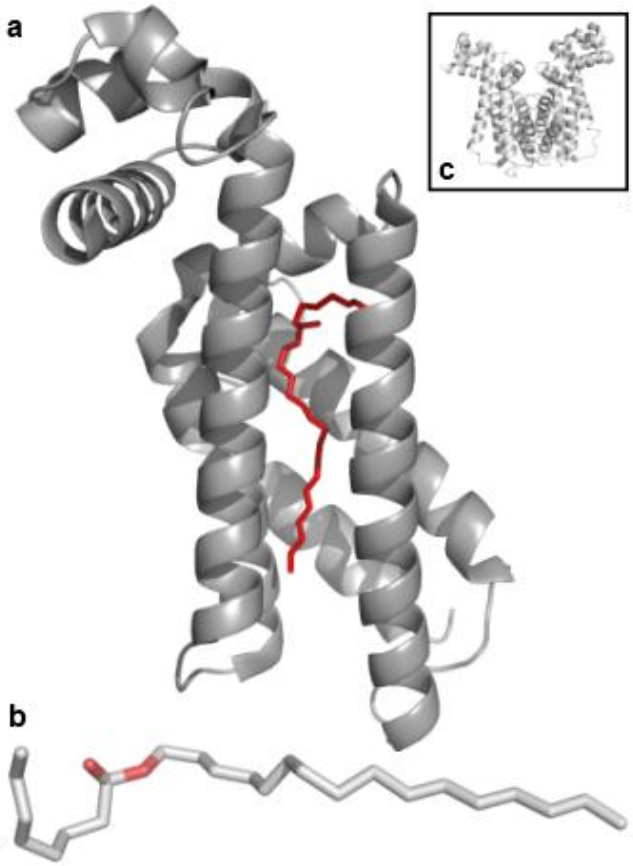


Figure 2.5 | The crystal structure of EthR in a ligand-bound conformation (monomer **a** PDB: 1U9N; dimer **c** PDB: 1U9O). Image **a** shows the DNA-binding domain formed by the helix-turn-helix motif with a third, stabilising helix. Hexadecyl octanoate (**b**) was found within the six-helix ligand-binding domain. [7].

EthR co-purified and co-crystallised with a ligand, hexadecyl octanoate (HexOc) in the ligand-binding domain.[7] It has been postulated that this fortuitous ligand has a role in the *in vivo* regulation of EthR; as a potential by-product in mycolic acid synthesis, long esters binding EthR would provide a feedback system with which to regulate *ethA* expression.

The ligand binding site of EthR is rich in hydrophobic and aromatic residues, and an uncharged polar region is formed by Asn176, Asn179 and Thr149. It was expected that if a moderately lipophilic compound with the ability to hydrogen bond with these residues could bind with EthR, the inactive conformation could be induced.[9]

Figure 2.6 shows how this binding will affect ETH activity. In the presence of inhibited EthR, *ethA* expression increases, causing an increase in EthA protein and ETH bioactivation. The activated ETH binds to NAD^+ to form the inhibitory adduct which prevents InhA from performing its role in mycolic acid synthesis.

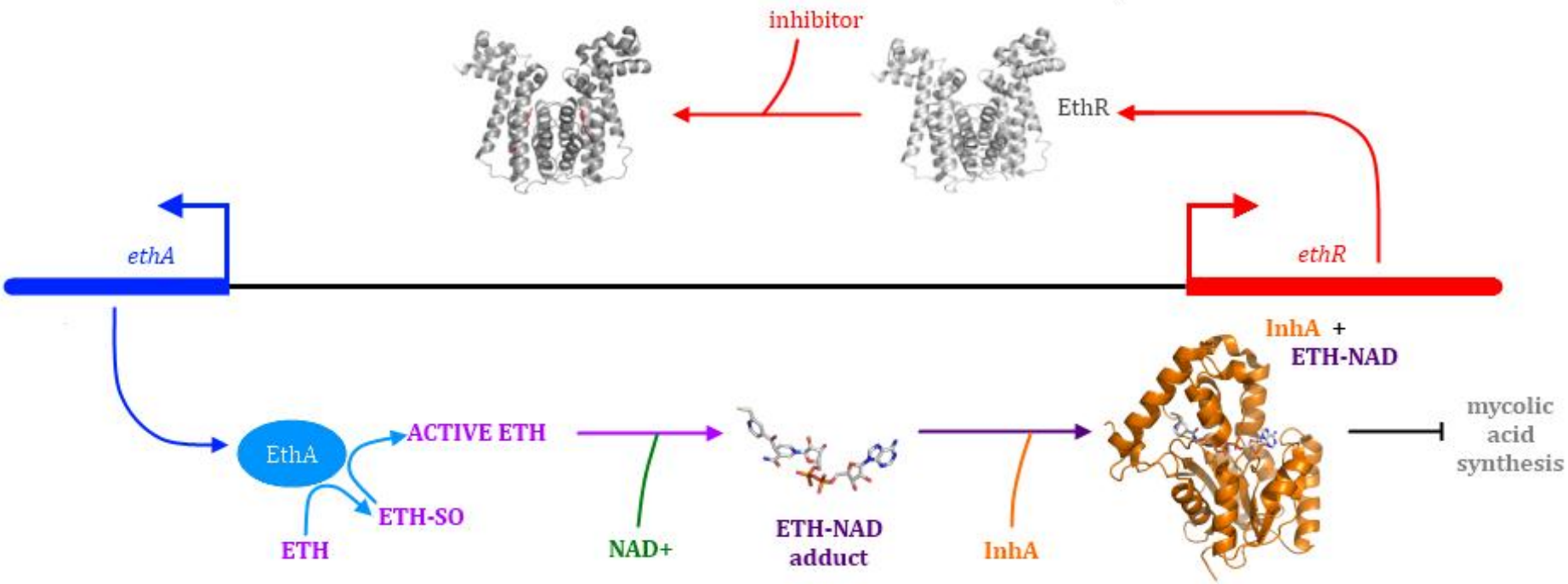


Figure 2.6 | In the presence of an inhibitor, EthR is unable to bind the ethA promoter, and EthA is produced. ETH is converted to an active form, capable of bonding to NAD+ to form the adduct which inhibits InhA. Mycolic acid synthesis is abolished. Images of EthR (PDB: 1U9O) and InhA (PDB: 2H9I), as well as the ETH-NAD adduct derived from the InhA structure, were created using PyMol. [7,9,40,44,45]

2.2.7 | Design and Structure of EthR inhibitors

In order to find suitable ligands for initial trials on EthR inhibition, a pharmacophore was designed to interrogate a virtual library of drug-like compounds. Lipinski *et al*. describe four features which primarily determine the success of a molecule as an effective drug, with a computational method for quantification.[53] These features, which determine good permeation and absorption, are:

- a molecular weight lower than 500
- fewer than 5 hydrogen bond donors
- fewer than 5 hydrogen bond acceptors
- a CLogP < 5 – calculated partition co-efficient, measuring miscibility in water

Lipinski's rules are typically referred to as the 'rule of five', not because there are five rules, but because the relative co-efficients of each feature are multiples of 5. To reflect the desired features, the pharmacophore had a low molecular weight, with two hydrophobic ends separated by a 4-6 Å linker. The linker included groups with the ability to hydrogen bond, designed to interact with the asparagine residues in the channel's polar region.

The library contained 131 compounds which fit this basic pharmacophore. They varied in the nature of the hydrophobic ends (aromatic groups, aliphatic groups, cyclic and heterocyclic); the length and nature of the spacer varied, including triazole and oxadiazole groups; and some ends were substituted with alkyl chains, halogens and electron donating or withdrawing groups. The properties of each of these 131 highly-varied compounds were computed and those with the "best compromise" of features according to Lipinski's rules were taken forward into a functional screen.

The surface plasmon resonance (SPR) assay designed to determine the nature of the EthR-DNA interaction (described in section 2.2.5) was used to determine whether the compounds were capable of preventing EthR-DNA binding. Those that were able to inhibit EthR binding fell into two distinct structural categories. Those compounds possessing a 1,2,4-triazole linker were able to inhibit binding with an average of only 17%; compounds possessing a 1,2,4-oxadiazole linker were able to inhibit binding by over 50%. One of these oxadiazole-containing compounds was BDM14500.[9]

BDM14500, shown below in figure 2.7, was co-crystallised with EthR to determine the position it takes within the ligand-binding channel and if the compound was having the same conformation-induced inhibitory effect on EthR as was shown with HexOc. The crystal structure showed BDM14500 in the ligand-binding channel, as anticipated, with the oxadiazole linker within hydrogen bonding distance of Asn179.

BDM14500 bound to EthR can be seen in figure 2.8, below. Crucially, BDM14500 was able to effect a distance between the DNA-binding domains of the EthR dimer of 50.3 Å, the inhibited form of the repressor.[9]

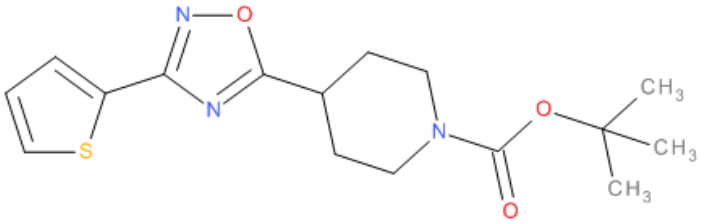


Figure 2.7 | The chemical structure of inhibitor BDM14500

2.2.8 | Inhibitor binding to EthR

With the ability of these drug-like molecules to inhibit EthR proven, analogues of BDM14500 were synthesised and evaluated by the same SPR assay. Of the resulting molecules, only those possessing the piperidyl ring were able to prevent EthR binding DNA. Two of these compounds were particularly effective: BDM31381 and BDM31343. These compounds were taken forward for biological assay, described in section 2.2.10.

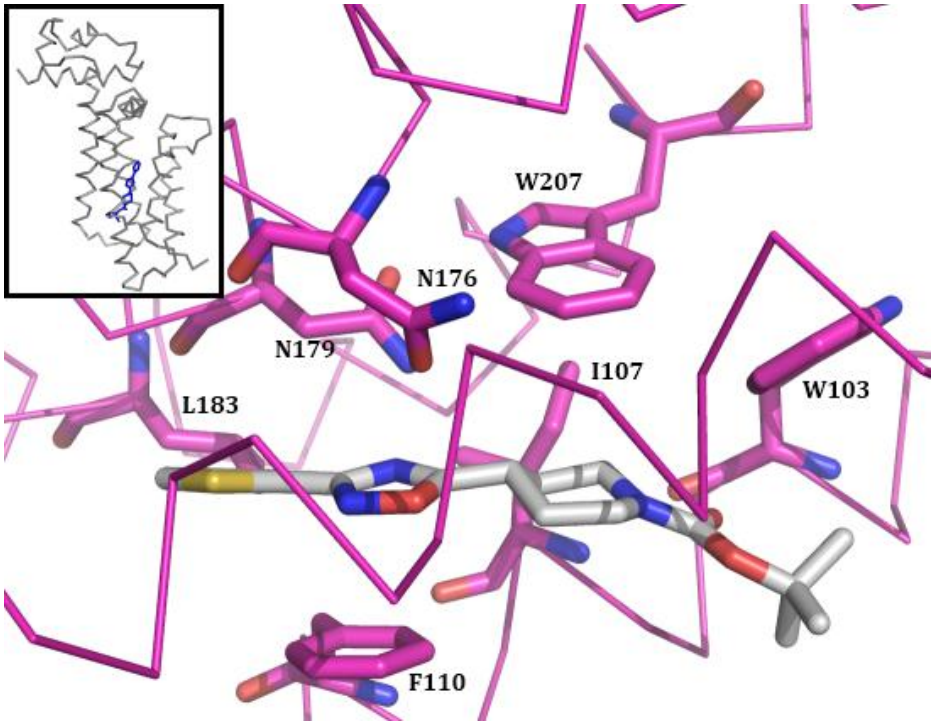


Figure 2.8 BDM14500 co-crystallised with EthR (PDB: 3G1O). The inset shows the molecule (in blue) bound in the ligand-binding channel where HexOc was bound. The orientation of the image has been changed for clarity. The labelled residues represent some of the side groups lining the channel, including the Asn176, Asn179 and Phe110 which surround the oxadiazole linker. The piperidinyl ring is in a pocket formed by Trp207, Trp103, and Ile107. Leu183 is part of the pocket in which the thienyl ring is contained.

The range of flexibility in the ligand-binding site was further tested by developing analogues of BDM31381 (figure 2.9). These analogues were created through *in situ* click chemistry, a method for the quick and reliable generation of new molecules by the assembly of small units.[54] The compound BDM14801, synthesised through the replacement of the thienyl-acetyl group of BDM31381, was found to place itself in the same position in EthR and in a similar orientation to BDM31381, distinct from BDM14500 in that the thienyl-oxadiazole points to the aperture of the channel, rather than to the hydrophobic top, where the acetylazido group of BDM14801 was oriented (PDB: 3O8G). This made BDM14801 an ideal starting point from which to determine the range of flexibility allowed in this hydrophobic pocket.[55] The main product of the tested click chemistry method used was a compound named BDM14950.[55]

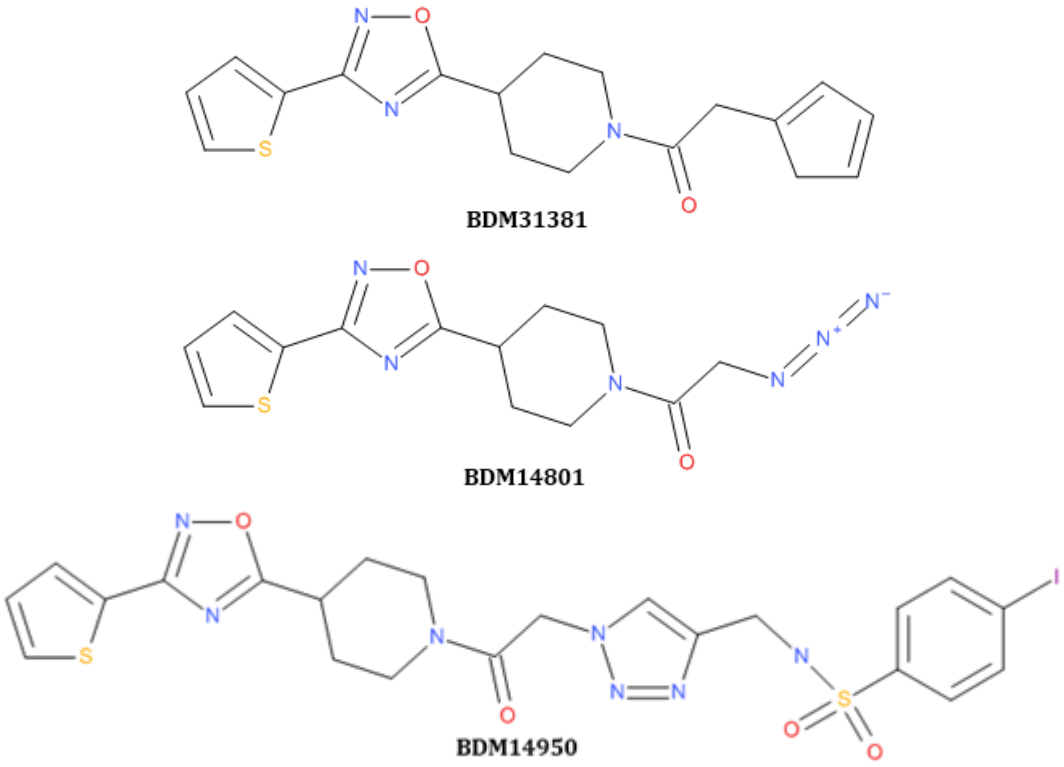


Figure 2.9 | BDM31381 and two analogues. Sixty different aliphatic and aromatic alkynes were chosen for their potential ability to interact with the hydrophobic portion of the EthR ligand-binding channel. They were sub-divided into six groups of 10 in such a way that condensation reactions between the component alkynes would result in combinations with different molecular weights. BDM14801 (125μM) was incubated with each of these alkyne groups and EthR (5μM) for five days at 37°C. The predominant product formed was named BDM14950 and co-crystallised with EthR (PDB: 3O8H).[55]

Willand *et al*. found that the area around the aperture of the ligand-binding channel is structurally conserved, but the upper part of the pocket is able to undergo structural rearrangement upon ligand binding. The ligands were seen to occupy equivalent positions, but for the longer ligand BDM14950, movement of two phenylalanine residues (Phe114 and Phe184) led to the opening of a 'secondary pocket' to accommodate the phenyl-iodide group (figure 2.10). The crystal structure of EthR bound to BDM14950 confirmed the compound's ability to induce the inhibitory conformation, but the variability of the structures to induce this inactive state necessitates further study [55].

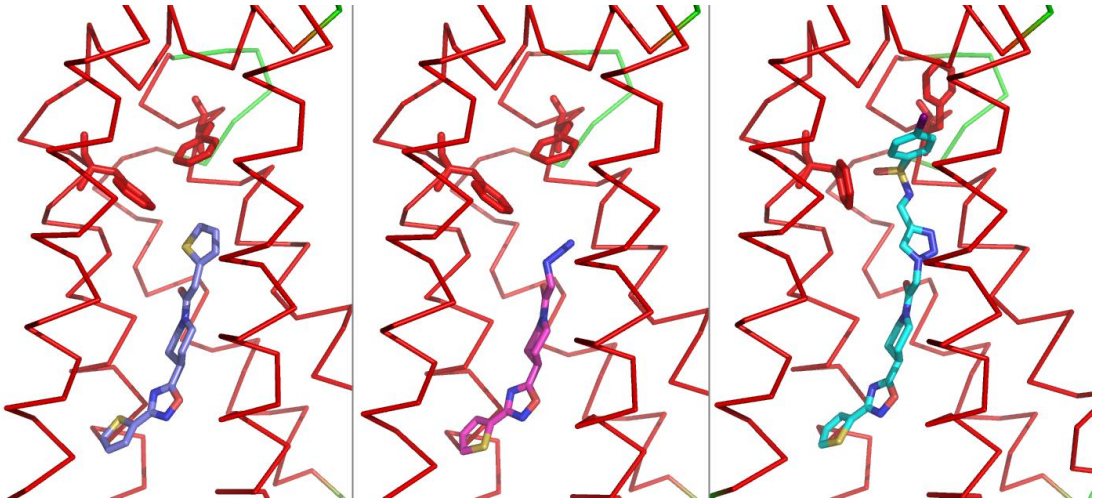


Figure 2.10 BDM31381 (left panel, PDB: 3G1M), BDM14801 (middle panel, PDB: 3O8G) and BDM14950 (right panel, PDB: 3O8H) in the ligand-binding channel of EthR. BDM31381 and BDM14801 the lower portion of the pocket where the residues are structurally conserved, and the phenylalanine residues retain similar positions. However, BDM14950 is longer and to incorporate the phenyl-iodide group, Phe114 and Phe184 rotate to open up this auxiliary pocket. All three compounds share common equivalent positions.

2.2.9 | Structural modifications to the inhibitors

The investigation into the flexibility of the ligand-binding channel was not limited to simply extending the ligand as far as possible. Different terminal groups were introduced to determine if these could increase binding affinity or induce conformational change more readily. One of these modifications was the introduction of CF_3 groups.

Fluorinated groups added to the BDM31381 analogues were shown to have a similar potency to the parent molecule and an increased solubility, but at a cost to the overall stability of the inhibitor. There was a steric limit to the number of fluorine atoms able to fit in this upper portion of the channel, however, and it was suggested that the fluorine additions were an aid to penetrance into the cell and not to the binding

of the protein target [56]. This ability of fluorine to improve cell permeation in drug-like molecules has been seen in other areas of drug design and optimisation.[57,58]

The structural modifications in the ligand-binding channel shown in figure 2.10 are made to accommodate a longer ligand. This requires flexibility of the target protein; here there is a weak, but proven, influence of the ligand upon the protein structure. If this can be taken into account, classical computational docking methods would give some insight into the potential of a drug to bind the target. However, docking simulations have a high computational demand. The speed of the simulation and the time taken can only be reduced with the imposition of limits on the model. These simulations also require three-dimensional co-ordinates of both the ligand and the protein, which can only be derived crystallographically. In this case, there are eight published structures of EthR in ligand-bound conformations (to date, with four as yet unreleased by the PDB, and one EthR mutant structure unreleased) which can be used, provided the ligand structures can be solved separately. If this were possible, crystallisation of the ligands alone would inform the drug design process without the need to co-crystallise with the protein. In cases where this co-crystallisation proves difficult, computational methods could be a boon to the drug design process.

2.2.10 | *In vitro* and *in vivo* testing of EthR Inhibitors

Various biological assays have been used to demonstrate and evaluate the effect of these designed EthR inhibitors *in vitro* and *in vivo*. Once BDM14500 had been crystallised with EthR and shown to inhibit the EthR-DNA interaction, the compound was tested for the ability to prevent *Mtb* H37Rv growth on solid agar plates.

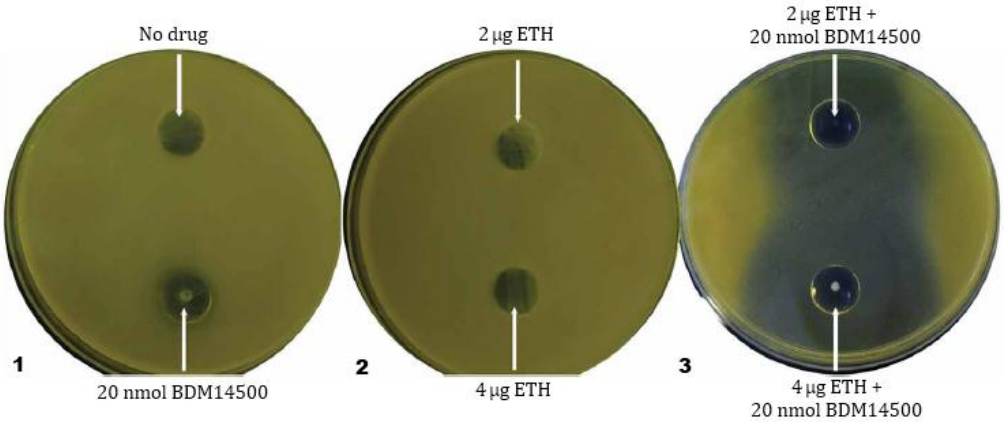


Figure 2.11 | The ability of *Mtb* H37Rv to survive on solid agar was tested in the presence of ETH with and without BDM14500. (Adapted from Willand *et al.*, 2009)[9]

Figure 2.11 shows the radial diffusion experiments performed with ETH and BDM14500. The radial diffusion experiments showed BDM14500 was not bacteriotoxic

alone, and instead had a synergistic effect with ETH. Therefore, BDM14500 analogues – which would also have this effect – were assessed for their pharmacological properties, as potential trial drug candidates.

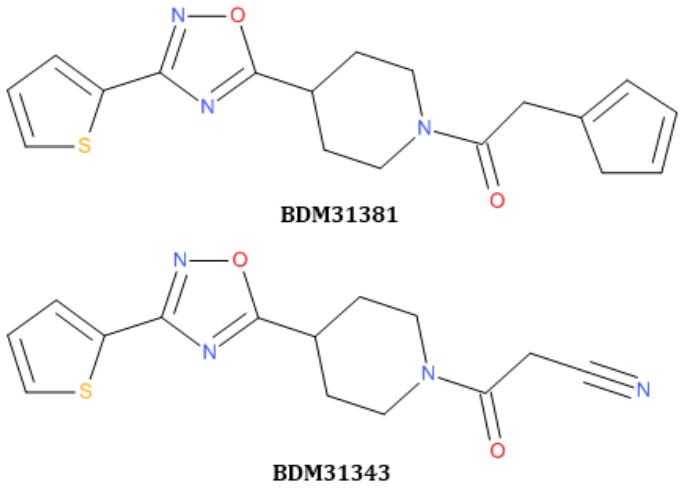


Figure 2.12 | Chemical structures of BDM31381 and BDM31343

SPR assays revealed that BDM31381 was more potent in inhibiting EthR binding DNA, with a half-maximal inhibitory concentration (MIC) of 522 nM (the half-MIC of BDM31343 was determined to be 3.3 µM). As the more potent of the two inhibitors, BDM31381 was used to quantify an effect on *ethA* transcription in *M. bovis* BCG; the level of *ethA* mRNA in cultures treated with BDM31381 were compared with an untreated culture, and 20 nmol BDM31381 was seen to effect a 35-fold increase in *ethA* transcription.

Once the inhibitors were shown not to be cytotoxic up to 30 µg ml^{-1} in human hepatocellular carcinoma cells - a model system for liver cell function - testing moved on to the mouse models. In mouse hepatic microsomes, the *in vitro* plasma half-life of BDM31381 exceeded seven days; for BDM31343 this was only 53 hours. Their respective solubilities in PBS were determined to be within a suitable range for intraperitoneal and oral administration and BDM31343 was seen to have a faster intrinsic clearance.[9]

To determine a basic *in vivo* pharmacokinetic profile of the two inhibitors, male CD-1 mice uninfected with *Mtb* were given a single intraperitoneal administration of either BDM31381 or BDM31343, dissolved in DMSO and at a dosage of 100 mg per kg of body weight. The integrated concentration-time curves show mice were exposed to concentrations of BDM31343 (82 µg h ml^{-1}) exceeding the 7.5 µg ml^{-1} required to boost ethionamide *in vitro*, but not sufficient concentrations of BDM31381 (5 µg h ml^{-1}). Therefore BDM31381 was the most potent inhibitor *in vitro*, but BDM31343

possessed a superior *in vivo* pharmacokinetic profile. Therefore, both inhibitors were taken forward into mice with an active *Mtb* infection.

Male CD-1 mice, as used for the initial pharmacokinetic profile, were infected with *Mtb* and the disease was allowed to progress unhindered for seven days. Mice lack the CD-1 'group 1' subset of the CD-1 family of glycoproteins. These molecules are present on antigen-presenting cells, and the group 1 subset are shown to favourably present foreign lipid antigens and those of mycobacteria; without them, an immune response is prevented. Therefore, CD-1 mice are a good model infection organism for TB and other mycobacterial pathogens.[59]

After seven days, mice were left untreated, treated with ETH only, or treated with ETH in combination with either BDM31343 or BDM31381. Mice treated with ETH/BDM31381 showed only a minor reduction in the bacterial load compared to the mice undergoing ETH-only treatment. However, mice treated with ETH/BDM31343 showed a reduction in bacterial load as efficiently as a three-times-higher dose of ETH alone throughout the three-week treatment; administration of a 9 or 27 mg kg^{-1} body weight dose of ETH was as effective as a 3 or 9 mg kg^{-1} body weight dose of ETH in combination with BDM31343.[9]

This was *in vivo* proof of concept, demonstrating the effect of EthR inhibitors on ETH efficiency. The next step was to optimise the inhibitor design, which can be done through biological assay as well as structural studies. Co-crystal structures of a select few of the compounds have been published; however the lone crystal structures of these inhibitors had not been solved. The inhibitor crystal structures without a protein present can provide a wealth of information in regards to the conformation of the inhibitor, the potential sites for intermolecular interaction, and can be exploited for computational studies which can save time and resources.

2.2.11 | Project Aims

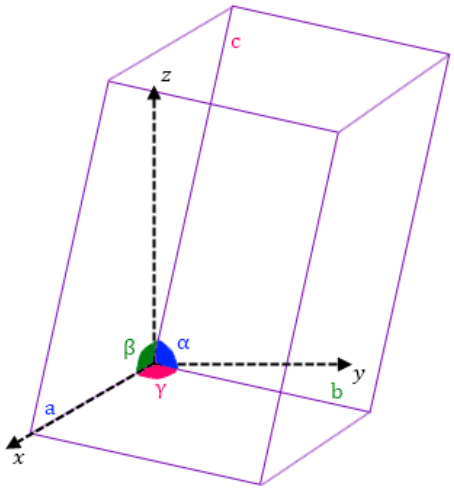
In this project, it was my intention to crystallise a selection of these inhibitors to determine their conformation in three-dimensions to high accuracy, and use this to confidently infer protein-ligand binding information; these inhibitor structures would also be used as a starting point for molecular docking. Due to the limited amounts of inhibitor available, crystallisation would prove challenging; however the information from structure solutions and docking analysis would inform further, more effective inhibitor design.

2.3 | An Introduction to X-ray Crystallography

Crystallography is defined by the International Union of Crystallography as "the branch of science devoted to the study of molecular and crystalline structure and properties" (IUCr). X-ray crystallography is the method by which the pattern of diffraction of X-rays through a crystal is used to determine the molecular structure in three-dimensional space. This section will give a brief, qualitative overview of how the diffraction pattern of a crystal can lead to a structure solution and the underlying physical principles.[61]

2.3.1 | X-ray Diffraction by Crystals

To relate a crystal to its diffraction pattern, various concepts have to be explained. Chapter 3 will describe how crystals are formed; suffice to say at this point that the individual molecules of the crystal align in identical conformations, held by intermolecular interactions, creating a repetitious pattern. In any three dimensional repeating pattern, there is a small section that, if defined and if any symmetry or translation elements are assigned, can be used to describe the entire pattern. In the case of a crystal, this is the *unit cell*.



Triclinic: no unit cell restrictions; $a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$

Monoclinic: $\alpha = \gamma = 90°$

Orthorhombic: $\alpha = \beta = \gamma = 90°$

Tetragonal: $a = b$; $\alpha = \beta = \gamma = 90°$

Trigonal: $a = b$; $\alpha = \beta = 90°$; $\gamma = 120°$

Hexagonal: $a = b$; $\alpha = \beta = 90°$; $\gamma = 120°$

Cubic: $a = b = c$; $\alpha = \beta = \gamma = 90°$

Figure 2.13 | A schematic of a unit cell, showing the three axes (a, b and c) and the three angles (α, β, γ). The classifications on the right denote the geometric qualities of the unit cell, however certain internal symmetries must be satisfied in addition to the axes lengths and angles. For example, a monoclinic cell must have a two-fold rotational symmetry; though trigonal and hexagonal have the same geometric qualities, a trigonal cell has a three-fold rotational symmetry where a hexagonal has a six-fold.

The unit cell may contain one or more, or a fraction of a molecule, depending upon the internal symmetry and how those molecules relate to each other in three-dimensional space. Therefore, a crystal is made of repeating unit cells, defined by three lengths (in Å; 1 Å is equivalent to 10^{-10}m) and three angles (in degrees). There are seven crystal systems based upon the geometry of the unit cell (figure 2.13); the descriptions of

translation and symmetry operations which relate the contents of the unit cell to the entire crystal give 230 crystallographic space groups, each with a distinct notation to define those operations.

2.3.2 | The crystal lattice

All of the descriptions so far have been in real space, however to relate the crystal to the diffraction events, mathematical constructs are used. The first of these is the crystal lattice.

If each unit cell in the crystal was to be represented as a point rather than a constructed box containing atoms, then a three-dimensional lattice of points would be formed. The spacing between these points depends upon the geometry of the unit cell – a longer unit cell will show a more widely-spaced lattice. Lines can be drawn through the points in two dimensions, and in three, these become lattice planes (figure 2.14).

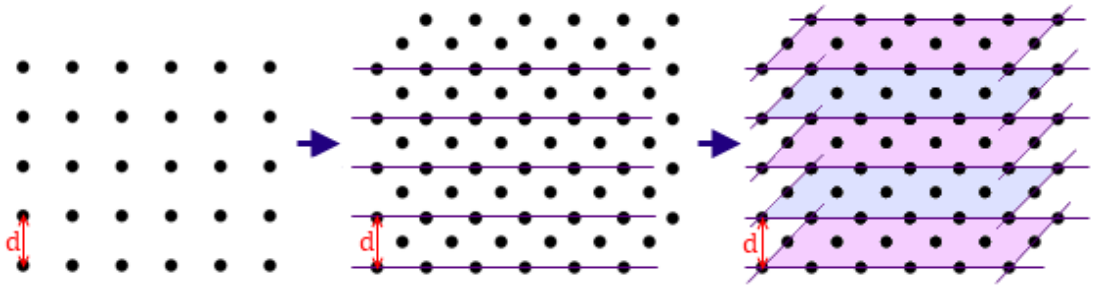


Figure 2.14 | Lattice points to lattice planes.

A *reciprocal lattice* is derived from the dimensions of the 'direct' crystal lattice, and it is the reciprocal lattice that is seen in a diffraction pattern. The reciprocal lattice is measured in distances of $Å^{-1}$ and degrees and as such, long spaces between lattice points become small spaces between reciprocal lattice points; wide distances of d on the lattice become small distances of d^{-1}. The reciprocal lattice is put into use to describe a diffraction event through Bragg's Law.

$$n\lambda = 2d \sin\theta$$

Equation 1 Bragg's Law states that constructive interference of waves will occur when the path difference between two incident waves is equal to the wavelength multiplied by an integer: $n\lambda = 2d \sin\theta$.

Constructive interference is a requirement for measurable diffraction; diffraction by a single atom of a molecule is too weak, and so the constructive interference of diffracted waves by identically positioned atoms in the crystal magnifies the signal. Therefore, diffraction is described as reflection by lattice planes.

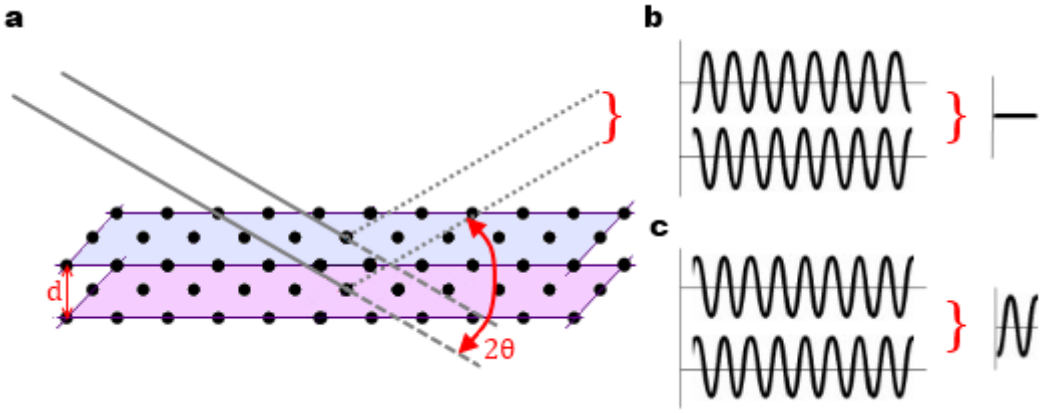


Figure 2.15 (**a**) shows the principle behind Bragg's Law. Incident waves diffract from identical points on different planes. (**b**) depicts two waves out of phase, with the same frequency and wavelength; their amplitudes sum to zero. (**c**) shows the opposite: two waves in with the same phase and frequency, which combine to create a wave with an amplitude equal to the sum of the amplitudes of the constituent waves.

2.3.3 | Miller Indices

To define planes in a crystal and describe complicated crystal shapes, Miller Indices are used. They were introduced as a notation in the early 1800s, but developed and published in 1839 by WH Miller, then the Chair of Mineralogy at Cambridge University.[62] To assign the indices, axes are chosen so that the axes x, y and z align respectively with the unit cell axes a, b, and c. The Miller indices - h, k and l – are always integers (figure 2.16). Planes can be defined by Miller Indices by $1/ha$, $1/kb$ and $1/lc$, thus defining the reciprocal lattice.

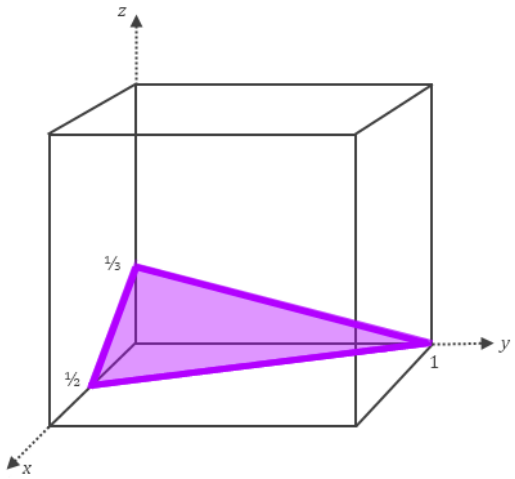


Figure 2.16 | Where the plane intersects the unit cell axes, fractional co-ordinates are assigned, such that the plane that intersects x at a/h, y at b/k and z at c/l. Therefore, the plane is defined as h, k and l, or **hkl**.

Here, axis x is intersected by the plane at ½, axis y at 1 and axis z at ⅓.

Therefore, this plane is defined as 213.

2.3.4 | Combining Waves to Give An Image of Electron Density: The Phase Problem

When diffracted waves intercept the detector, they are recorded as intensities; the intensity is proportional to the square of the amplitude of the wave (figure 2.16), but the phases of those waves are lost.

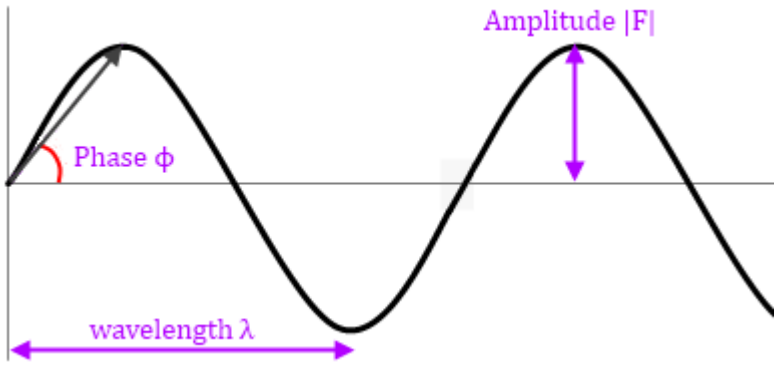


Figure 2.17 | The geometry of a periodic wave.

Euler's formula establishes the relationship between trigonometric functions (as used to describe periodic waves) and complex exponential functions (as used to calculate electron density). This formula can be seen as equation 2, below.

$$e^{i\Phi} = \cos\Phi + i\sin\Phi$$

Equation 2 | Euler's formula, where 'e' is the exponential function, 'i' represents an imaginary number, and 'ϕ' is the angle given in radians.

To relate diffracted waves to the electron density and so solve the crystallographic structure, a Fourier transformation is used (equation 3). Euler's formula is used to encompass the cosine and sine functions of the diffracted waves into the notation using *2πi*. However, it is not as straight-forward as simply solving the resulting equation; we must take into account the structure factors. Structure factors describe the amplitudes and phase of each spot on the diffraction pattern, but the phases have been lost in the course of our diffraction experiment – this is the *phase problem*.

$$F(hkl) = \int_{cell} \rho\,(xyz) \cdot \exp\,[2\pi i\,(hx, ky, lz)]\,dV$$

Equation 3 | Structure factors (red) are related to the electron density at each point in the cell (green) by integration. Reciprocal space and real space must be reconciled using a complex term (blue).

Without the phases, only the observed amplitudes can be used. As F(hkl) is a complex number, with an angle and length, we can consider the phases as the angle ϕ_{hkl} and the length (or amplitude) as |F(hkl)|; we arrive at equation 4 for the electron density.

$$\rho(xyz) = \frac{1}{V}\sum |F(hkl)| \exp[i\phi(hkl)] . \exp[-2\pi i(hx, ky, lz)]$$

Equation 4 | The amplitudes (red) and phases (grey) of all diffracted waves are summed (and adjusted to relate real and reciprocal space, blue) to determine the electron density (green).

However, the electron density still cannot be calculated without the phase information (the grey term in equation 3). Therefore, structure solution requires solving the 'Phase Problem'.

2.3.5 | Solving Structures in Chemical Crystallography

There are various methods in chemical crystallography to solve small molecule structures, involving different approaches to the phase problem. The first and oldest of these methods is known as 'trial and error' and was particularly useful in the early days of X-ray crystallography where structure solutions were attempted upon simple structures with only a few parameters.

The *trial and error* method begins with assigning a chosen reflection the structure factor F(000); this is then used as a reference point to calculate the places of the other atoms. A trial structure can then be tested for accuracy by calculating an R factor (equation 5).

$$R = \frac{\Sigma [|F_o| - |F_c|]}{\Sigma |F_o|}$$

Equation 5 | The reliability index for trial structures, comparing calculated structure factors (F_c) and structure factors derived from observed intensities (F_o). The R factor is minimised during solution to an agreement between the observed and calculated structure factors.

The lower the value of R, then the more similar the calculated and experimental structure factors are, making the trial structure more likely to be correct; when this is the case, the trial structure can then undergo refinement. If the R value shows the structure factors differ greatly, then the process is taken back to the beginning and a different F(hkl) position is used. This method can be powerful where there are many constraints upon the crystal packing; for example, high symmetry space groups where atoms must occupy special positions to satisfy all symmetry operations. However, this is a very inefficient method of structure solution, when the origin could lie at one of

many positions of F(hkl) within the crystal, particularly in low symmetry space groups.[63]

A second method, suggested in 1934, is the Patterson method.[64] Until this point, *trial and error* had been the only way to solve chemical crystal structures. The Patterson method relies upon the presence of a notably heavy atom within the structure, and takes advantage of known inter-atomic distances. The Patterson function is a Fourier series which does not require phase information (equation 6).

$$P(x,y,z) = \frac{1}{V}\sum_{hkl}|F(hkl)|^2\cos[2\pi i(hx+ky+lz)]$$

Equation 6 | The Patterson function, whereby P(x, y, z) will have a peak which corresponds to each inter-atomic vector in the structure.[64]

$$\mathbf{r}_i = x_i\mathbf{a} + y_i\mathbf{b} + z_i\mathbf{c}$$

Equation 7 | Defining the position of an atom in a Patterson function.

If each position is defined as $\mathbf{r}_i$ (according to equation 7), then the Patterson function gives peaks $\mathbf{r}_i - \mathbf{r}_j$, where i is 1 and j is the number of atoms in the unit cell. The total weight of the peak will be the sum of the two atomic numbers of the two atoms forming the vector. The highest peak will be the vector between the heaviest atom and itself at an equivalent position; as a vector will be created between every atom in a cell, a unit cell containing n atoms will give n^{n-1} Patterson peaks. By picking vectors carefully and using the atomic numbers of constituent atoms carefully, atoms can be assigned to each vector and so each electron density peak; heavy atoms make this particularly easy, as the vectors are distinguishable from those between lighter atoms, allowing the heavy atom to become a reference point. This is a powerful method, still in use and development after nearly seventy years, however it is not applicable to this project, as none of the compounds in question contained sufficiently heavy atoms to make this method a viable option.[64]

The most widely utilised method of structure solution today is direct methods.[65] Direct methods are an *ab initio* approach requiring two assumptions; first, that the electron density will only be positive where atoms are present (thus zero elsewhere) and second, that the structure consists of point-like atoms. With these assumptions, normalised structure factors calculated from the observed amplitudes can be used in the place of the unaltered observed values; these are |E| values rather than |F|. Only high |E| values are used to contribute to the electron density map,

ensuring that the most positive areas are included, and it logically follows that these strong positive areas of electron density are point-like atoms.[66]

Next, a limited number of phases are assigned to those Bragg reflections which are *probably equal*, known as 'triplets'.[65] The probability of these phase relationships giving a positive electron density is calculated. Only the best set of phases (ie. the most correct, positive set) is used to calculate the electron density map, with |E(hkl)| taking the place of |F(hkl)| in equation 3. Once atoms are assigned to the resulting electron peaks, the phases undergo refinement to arrive at the final structure.[63,67]

2.3.6 | Refining Structures in Chemical Crystallography

Refinement is the adjustment and optimisation of the parameters in the crystal model, defined by the atomic co-ordinates obtained in the structure solution. As the sine and cosine functions of the Fourier series for calculating the electron density are not linear, crystallographic refinement makes use of non-linear least squares regression. This method depends on over-determinancy in the model, and small molecule structures generally require a data-to-parameter ratio of 10:1 or more, meaning 10 Bragg reflections for every parameter. Parameters for refinement include (but are not necessarily limited to, depending on the nature of the model) the atomic coordinates (x, y, z) and atomic displacement parameters (ADPs are described below). The refinement indicators are two R-factor statistics, R1 and wR2.[68]

Least squares (in a non-linear form) is derived from the Taylor series and differs slightly depending upon whether the parameters are being refined against F or F^2. SHELXL refines against F^2, utilising equation 8. Refinement against F^2 has advantages over refinement against F, which can result in mathematical problems with weak reflections or intensities with negative measurements. Refinement against F^2 is not subject to these issues, and so reduces the possibility of the refinement stalling in a local minimum. With F^2, all data can be used with appropriate weighting; the weighting is determined by experimentally-derived standard deviations.

The atomic positions and thermal factors for each atom in the structure are derived from the initial model, giving F_{calc} (F_c).

$$\frac{\delta(F_o^2 - F_c^2)}{\delta x} = -4 \Sigma (F_o^2 - F_c^2) Fc \frac{\delta F_c}{\delta x}$$

Equation 8 | The least squares refinement equation used by SHELXL.[69]

F_o represents the experimentally observed structure factors, while F_c represents the calculated structure factors from the model and subsequent refinement. While F_o

carries random errors and systematic experimental errors, F_c carries no errors. What follows is an iterative refinement of F_c from F_o with corrections, until they converge to a global minimum. Theoretical R-factors represent the difference between the observed and calculated structure factors; as the refinement progresses, the R factors should fall. A good R1 factor for a small molecule in chemical crystallography is 5% or below.

$$R = \frac{\Sigma\,(|F_o| - |F_c|)}{\Sigma|F_o|}$$

Equation 9 | The R1 factor is the clearest and most reliable of the two refinement indicators.[70]

$$wR2 = \sqrt{\frac{\Sigma w(F_o^2 - F_c^2)^2}{\Sigma w(F_o^2)^2}}$$

Equation 10 | wR2 uses squared factors, increasing the value by up to 3x that of R1. This value is the one actually minimised by least squares, but R1 is more conventionally used to compare refinements as it is less dependent on weighting and is easier to understand.[69]

The wR2 factor is typically greater value than R1 as it uses squared factors. It is also subject to a user-adaptable weighting scheme and so R1 is considered a more reliable refinement indicator.[70]

As the refined model begins with values determined from the structure solution, the starting data must be good, with atomic positions correct to within 0.5 Å. A poor starting model is just one of the issues which can cause refinement to fail; a poor understanding of how to handle the data can lead to the same failure. For example, data from poor quality crystals with low resolution will not have the accuracy to determine the atomic co-ordinates accurately. A chemical understanding is required to recognise disorder, whether in the structure of in the solvent, as well as correctly impose any restraints or constraints on the model to improve the refinement. Similarly, including spurious atoms can incorrectly inflate the anisotropic displacement parameters – ADPs.[69]

ADPs, often represented as 50% probability ellipsoids, represent the dynamic disorder, where vibration causes movement, and static disorder, where slightly differing atomic positions between unit cells cause the electron density to show a shape other than perfectly spherical. Extended ADPs may indicate disorder or the omission of a neighbouring atom in the structure.[71]

While the refinement is built upon a straight-forward and clearly defined principle, a great deal of user understanding and interpretation is required to ensure a successful result. This requires various standards to achieve a good data set to a high

resolution (for small molecules, ideally better than 1 Å), with a well-defined starting model, for example a high data-to-parameter ratio (ten or greater) and diffraction data collected at lower temperatures to minimise thermal motion and radiation damage.

CHAPTER 3

METHODS

3.1 | Introduction to the crystallisation of small organic compounds

The benefits of crystallisation in structural studies have been described in the previous chapter; in practice, crystallisation of organic compounds can prove time-consuming and technically demanding. This section will introduce the kinetic and thermodynamic principles behind crystal formation as well as describe the methods frequently used in crystallisation to generate single crystals for X-ray diffraction studies.

Solution methods, by which the solid compound for study is crystallised from a solvent, are the most flexible methods when attempting to grow good quality crystals. The wide selection of potential solvents, alone or in combination, make for a dynamic set of possibilities to choose from. The aim is to bring a concentrated homogenous solution to the point of supersaturation, where it can then relax into the metastable zone (figure 3.1). This increases the likelihood of nucleation and the formation of a single, ordered crystal.[72]

Crystal formation is a two stage process. First, nucleation must occur, followed by sustainable crystal growth. Nucleation is a stochastic, localised phase change from solution to solid, in which a small number of identically oriented molecules associate to form a regularly ordered cluster.[73] This event can occur once in a solution, or multiple times depending upon various factors including (but not limited to) the concentration of the solution, the temperature, and the degree of manual disturbance to the solution. Once the initial nucleation point has been reached, molecules of the compound which are in the correct, identical orientation can be adsorbed to the surface of the critical nucleus; thus a crystal is formed and grows. This allows the system to relax to a thermodynamic equilibrium, and the main growth phase of the crystal takes place in the metastable zone (figure 3.1).

If crystals grow too quickly, there is a risk of constituent molecules misaligning, leading to disordered crystals or in more severe cases, multiple crystals growing together. This would render them useless for analysis by X-ray diffraction, as single ordered crystals are essential. Alternatively, an under-saturated solution can lead to nucleation sites disassociating, preventing crystal growth. Intermolecular interactions between molecules of the compound will support the crystal packing structure and stabilise the formation and growth process. In organic molecules, the most common interactions are hydrogen bonds, electrostatic interactions and hydrophobic

interactions; however, when present, metal co-ordination and solvent-mediated contacts are another strong determinant of crystal packing [74]. Therefore, the occurrence of crystal formation and the rate of crystal growth are unpredictable.

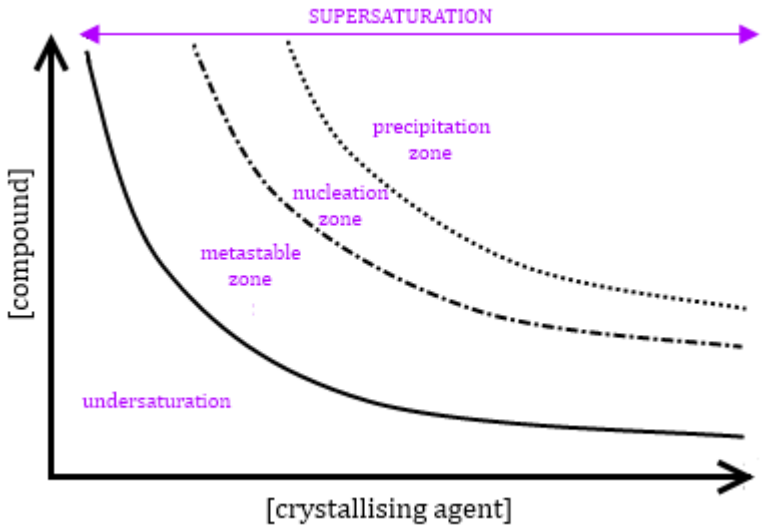


Figure 3.1 | Phase diagram of crystallisation with a solvent in which the compound is moderately soluble. The relative concentrations of the compound and crystallising agent (solvent) determine solubility. If supersaturation can be reached beyond the solubility curve (solid black line), nucleation and initial crystal growth will occur in the nucleation zone. As growth continues, the concentration of compound decreases and the main growth phase continues in the metastable zone. With increased solubility of the compound in the solvent, the solubility curve moves to the right, making supersaturation harder to achieve.

The degree of solubility of the compound in the chosen solvent is important to crystal formation, dictating how quickly a saturation point can be reached with a given amount of material. Generally, a compound will be most soluble in a solvent which is chemically similar; if functional groups or polar moieties of the compound are known, this information can be used to choose a solvent with similar characteristics. The best solvent for growing single crystals of suitable size for x-ray diffraction studies is one in which the compound for study is moderately soluble. If the compound is exceptionally soluble in a solvent, any micro-crystals formed at nucleation will re-dissolve, preventing crystal growth; if the compound is barely soluble, there will not be enough dissolved material to form crystals at all. For this reason, the first step taken in attempting to re-crystallise my samples was always an initial solubility screen against a library of solvents.[75]

3.1.1 | Method: Initial Solubility Screens

After initial observations on the nature of the sample were made under a microscope (such as colour, habit, if the sample was crystalline or not, if any crystals could be seen in the sample from which predictions on the eventual crystal shape could be made), solubility screens were carried out on a standard microscope slide, with up to 5 µl of solvent and a small amount of the compound for study.

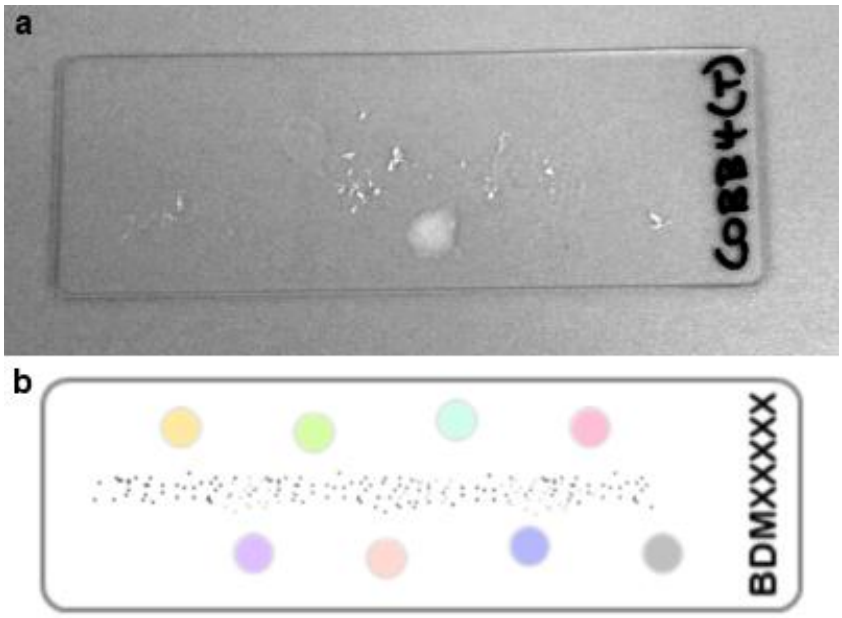


Figure 3.2 | When performing a solubility screen, the slide is labelled to keep track of the results (seen in the photograph **a**, and the schematic **b**). The coloured circles denote different solvent drops with the powdered compound distributed through the slide. I developed a convention by which my first, far left solvent drop would be water, with the subsequent drops of organic solvents (beginning with acetonitrile and toluene, ending in DMF and DMSO) increasing in boiling point.

Each solvent was added to the slide using a 2-200 µl pipette, and the powdered compound was added to the solvent drop using a clean needle. Observations under the microscope on the relative solubility of the compound in each solvent were made, including any additional observations: for example, if the powder broke apart in any specific manner. For example, if the powder appeared to 'splinter' apart, this could indicate the presence of needle-shaped crystalline material in the powder, and this suggests crystallisation may occur in this form.

3.1.2 | Theory: Crystallisation by Evaporation

The evaporation of the crystallisation solvent brings the system into supersaturation in the metastable zone, where nucleation can occur. However, since evaporation is an

open system, there are many conditions affecting the experiment. This makes evaporation difficult to reproduce. For the organic solvents in the library with comparatively high boiling points (such as DMF, DMSO and NMP, show in figure 3.5), initial evaporation experiments were carried out on the solubility screen slide.

If an evaporation experiment cannot be performed on the screen slide, or the experiment bears repeating, it is preferable to do so in a watch glass or glass 9-well plate, or a suitable glass vial (specific protocols used for acquiring my crystals are described in sections 3.1.6 – 3.1.11). The glass should be clean, with an aperture wide enough to facilitate removing small crystals; the vial can be stored at an angle to encourage crystals to grow on the side of the vial to make crystals easier to harvest. The vial can be left uncapped, or a perforated cap can be used to slow the rate of evaporation of volatile solvents.

The most undesirable result in an evaporation experiment is for the solvent/compound mixture to evaporate completely without crystallisation. Such failed crystallisation could occur because the solution was under-saturated and did not reach a nucleation point, or due to a misjudged solvent-to-compound ratio. Where there is enough material available, varying this ratio may bring the system to supersaturation; however, with limited material (ie. less than 8-10 mg), experimenting with different conditions can potentially waste material on an unsuitable solvent. Therefore, if under-saturation due to the high solubility of the compound in the chosen solvent is suspected, an alternative solvent is the preferable choice.

Crystallisation may fail because of poor and transient intermolecular bonds at the nucleation stage. While a nucleation point may have been reached, it is possible growth is prevented because the molecules of the compound were simply unable to remain in an identical conformation long enough to find a stable and sustainable growth phase. Altering the temperature or the speed of evaporation is the most effective way to encourage sustainable crystal growth. As such, evaporation experiments (and diffusion, described below) were often repeated in a refrigerator at 4°C.

When evaporation techniques failed, as they did for the protonated peptoid precursor form and BDM41325, diffusion techniques in sealed vessels were my next step. With a sealed vessel, more control is gained over the conditions of the crystallisation experiment.

3.1.3 | Theory and Method: Crystallisation by Liquid Diffusion

Evaporation techniques can fail for a number of reasons, not least that the solubility of the compound in a chosen solvent is too high and supersaturation cannot be reached. To reduce this, anti-solvents can be used. By definition, anti-solvents are solvents in which the compound is not (or is barely) soluble; such anti-solvents would have been identified in the initial solubility screen for each compound. Using solvents in this combinatorial fashion alters the solubility of the compound, lowering the supersaturation point and theoretically making nucleation and sustainable crystal growth easier to achieve.

Liquid diffusion involves creating an interface (figure 3.3) between two solvents; one in which the compound is soluble, one in which the compound is not.[76]

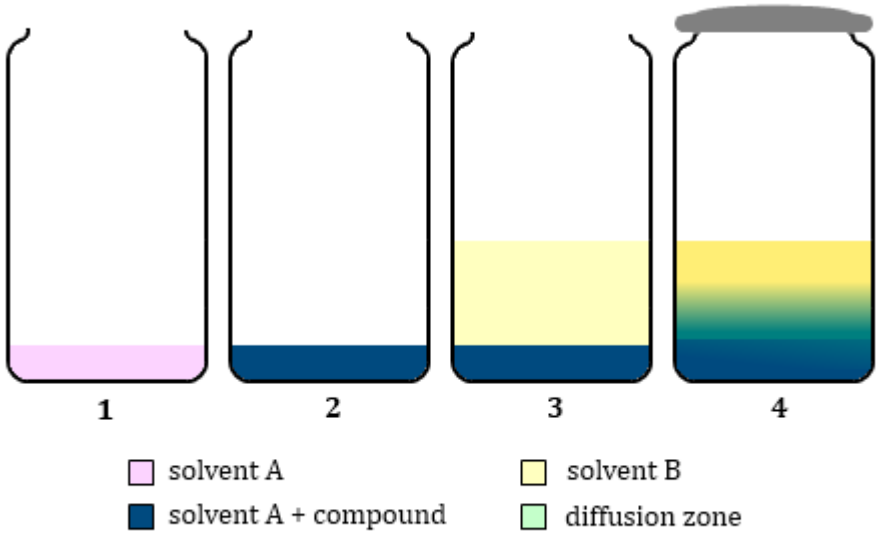


Figure 3.3 | The compound (~0.2 mg) is dissolved in a small volume (≤100 µL) of solvent A (**1** and **2**), in which it is moderately soluble. The anti-solvent, 'solvent B' (**3**) should be added to the vial in such a way as not to mix with the solvent-compound mixture. Solvent to anti-solvent ratios of 1:2, 1:3 or 1:4 were typically used. This can be achieved by adding the anti-solvent dropwise down the side of the glass vial. As solvent B will be less dense than solvent A, layers should form. Seal the vial and leave undisturbed in a stable temperature environment for a few hours to a day or two. If successful, crystals will form in the concentration gradient of the diffusion zone formed by the interface of the two solvents (**4**).

The key to liquid diffusion is the difference in solubility between the two solvents. Solvent A should provide moderate solubility for the compound for study, just as would be used in an evaporation experiment; if liquid diffusion is being used because no solvent has been found to fit this criteria, solvent A must be one in which the compound is soluble, and a stronger anti-solvent (solvent B) can be used. While the two solvents should be miscible, they should have distinct densities. The aim is to create a diffusion

zone (seen above in **4** as a green band between the two solvents) where the two solvents are in contact and have a limited interaction; if they are too similar in their densities and volatility, the two solvents will mix too thoroughly. An ideal ratio of solvent A to solvent B is around 1:4, but this method is versatile. If the compound is readily soluble in solvent A, more anti-solvent will be required; if the initial volumes do not yield crystals, it is quite straight-forward to remove the cap and add more anti-solvent, taking care not to disturb the solutions too much and cause over-mixing. This was of particular advantage to me, having a finite supply of material to work with – the solution can be cooled, or even heated and then allowed to cool to room temperature, in an effort to encourage nucleation. As evaporation of the compound-solvent mixture is not allowed to occur, none of the compound is lost and the experiment can be altered and repeated easily.

Even with appropriate solvent choices, this method can (and does) fail. For example, if the system is not closed – if the vial is not completely sealed – the volatile anti-solvent will evaporate and will not aid in crystallisation. If the experiment is stored somewhere with a fluctuating temperature, or which is not still and free from movement or tremors, a nucleation point will be more difficult to reach. Crystals will take up to a few days to form and grow, even at low temperatures where energy and movement of the molecules is restricted. Manual disturbance should be avoided so a stable, dry and ideally temperature-controlled storage place should be chosen, and the experiment should not be removed to be checked frequently – this method is slow enough that daily observation should be sufficient.

In the event that the only anti-solvent found for the compound is denser and less volatile than the solvent in which the compound is soluble, there must be an alteration to this method. For example, if the compound was soluble in acetonitrile, but not soluble in water, the water fraction can be layered on the bottom of the vial, and a mixture of acetonitrile and the compound can be carefully added to make the top layer. This would make it impossible to add more anti-solvent should crystals not form, but does further demonstrate the usefulness and versatility of anti-solvent/solvent mixtures.

3.1.4| Theory and Method: Crystallisation by Vapour Diffusion

Vapour diffusion is a technique typically used for the growth of single crystals of peptides and proteins, using similar principles to liquid diffusion. However, where liquid diffusion works best with high compound concentrations, vapour diffusion can handle comparatively smaller amounts of sample (ie. ~0.2 mg) and multiple experiments can be run in the same environment at one time. This gives a greater through-put to the experiment and can yield multiple crystals. Vapour diffusion is also slower, which should result in single, more ordered crystals for diffraction experiments. Figure 3.4 shows a basic schematic of a vapour diffusion experiment.

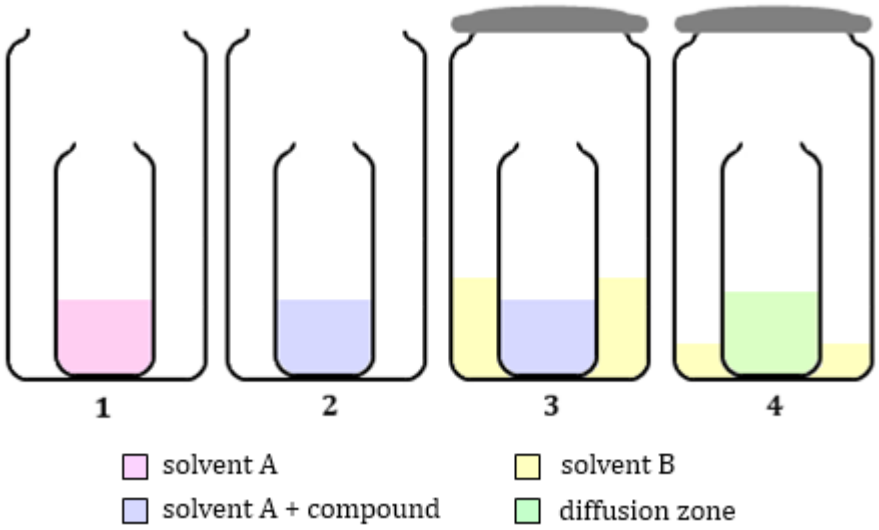


Figure 3.4 | Vapour diffusion involves two vessels, one placed inside the other, where the volatile anti-solvent ('solvent B' in **3** and **4**) is allowed to diffuse into the solvent-compound mixture. First, a small amount of compound is dissolved into a solvent (≤100 μL) in which it is moderately soluble, just as in evaporation and liquid diffusion experiments (**2**) in a small vessel. This vessel is placed inside a larger vessel, uncapped, filled with a volume of the anti-solvent (**3**) typically at a 1:3 or 1:4 ratio. This larger vessel is then capped. The volatile anti-solvent diffuses into the smaller vessel, reducing the overall solubility and inducing supersaturation.

As with liquid diffusion and evaporation, solvent choice is key, and solvents suitable for liquid diffusion will work for vapour diffusion in a similar way. The anti-solvent ought to be more volatile than the solvent, however in this case the degree of difference is not as crucial as with liquid diffusion. As long as the anti-solvent will diffuse into the solvent-compound mixture, this technique can be successful (meaning the anti-solvent must be the more volatile of the two, otherwise a volatile solvent will diffuse into the less volatile anti-solvent, leaving a decreasing solvent-to-compound ratio: this is essentially evaporation in a closed system). Whichever technique is used, sustained

nucleation events are serendipitous, and any method utilised is simply an attempt to make nucleation more likely.

3.1.5 | Method: Crystallisation of an anionic peptoid precursor

Crystalline material of an **anionic peptoid precursor** was provided for analysis, grown by slow evaporation of acetonitrile at room temperature in a glass vial. Long, thin and fragile single crystals were formed on the sides of the glass.

3.1.6 | Method: Crystallisation of a protonated peptoid precursor

Initial attempts were made to grow crystals of the 5 mg of protonated precursor compound (a white, crystalline powder) by evaporation. The compound was found to be soluble in ethanol, acetonitrile, DMF and DMSO, and the compound would re-crystallise at room temperature and at 4°C readily from both DMF and DMSO. However, nucleation was rapid and abundant; crystals grown by evaporation from these solvents were invariably long, thin and fragile – too small for analysis (ie. thinner than 10 μm in two dimensions).

Liquid diffusion techniques were used in an effort to control this rapid crystal formation with anti-solvents; the protonated peptoid precursor was observed to be insoluble in water and toluene. No crystals formed from the DMSO or DMF liquid diffusions with either of these anti-solvents after two weeks and so the solutions were allowed to evaporate; crystals of the similar poor quality were formed and under closer inspection, what appeared to be long single crystals were in fact multiple crystals growing side-by-side in succession.

Evaporation of 100% ethanol did not yield crystals, nor did evaporation of acetonitrile. Liquid diffusion experiments of ethanol/water and acetonitrile/water at 1:2 ratios were set up with the remainder of the sample but none were successful in forming crystals. The caps sealing both liquid diffusion experiments were removed and the solutions were mixed and heated, then finally re-sealed. The cooling of the mixed solutions did not lead to sustained nucleation, and so the caps were removed and evaporation was allowed as a last effort to re-crystallise the molecule. Evaporation of the 50% acetonitrile solution did not yield crystals.

However, crystals were formed by the evaporation of the 200 μl 50% ethanol solution, in two distinct habits: needles and blocks. A small (200 μm x 100 μm x 100 μm) clear, light, block-shaped crystal was used for X-ray diffraction analysis.

3.1.7 | Ongoing Crystallisation Attempts of Peptoid Precursor Molecules

Two more peptoid-related molecules (of unknown chemical structure) are under investigation, but as yet no crystals of a sufficient quality for X-ray diffraction analysis have been achieved. Compound A is exceptionally soluble in DMF and quite soluble in DMSO, readily crystallising from the evaporation of both, and also soluble in acetonitrile and ethanol. Crystals formed from the evaporation of DMF and DMSO were thin and exceptionally fragile; though crystals derived from DMSO evaporation were marginally less fragile, they weren't robust enough to use for analysis. Evaporation of both acetonitrile and ethanol at 25%, 50% and 75% did not result in crystal formation. Similar results were seen for Compound B.

Anti-solvents suitable for use on both samples included water, toluene and hexane. The liquid diffusion technique used with the protonated peptoid precursor was repeated for Compounds A and B without success – evaporation occurred without any re-crystallisation. Liquid diffusion and vapour diffusion experiments (of DMF/water, DMSO/water and ethanol/hexane) have not yet been able to slow the rapid crystal growth to yield larger and more robust crystals. Therefore, despite all efforts, no suitable crystals have been achieved from these two peptoid-related molecules.

3.1.8 | Method: Crystallisation of BDM31343

Crystallisation had been successful from a 20 µl volume of 100% dioxane, by seeding this solution in a glass well plate with small crystals formed in the initial solubility screen. However, these crystals did not diffract under X-ray analysis to a sufficient resolution (ie. less than 1 Å). Re-crystallisation was also attempted in NMP (figure 3.5, below).

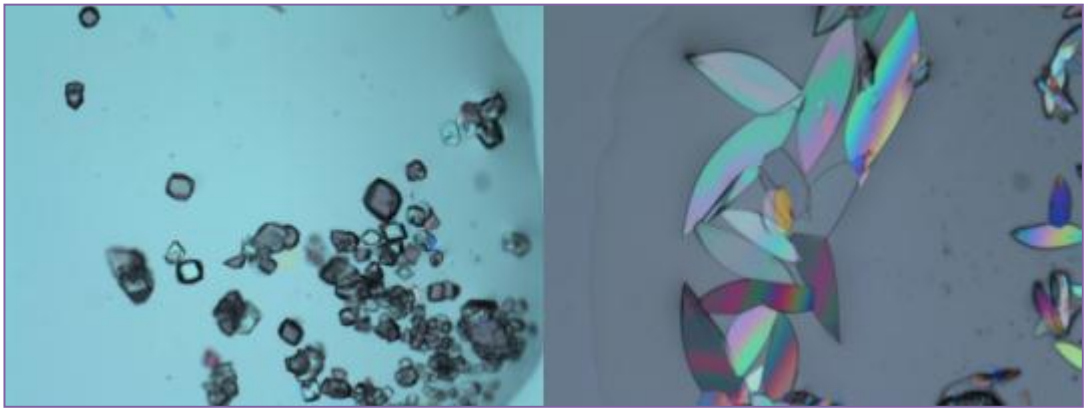


Figure 3.5 | Crystals of BDM31343 grown by evaporation of NMP (N-methyl pyrrolidone) at room temperature. Crystals formed in a drop on the initial solvent screen (left) were too small to use for x-ray analysis and were instead used to seed a drop in a glass well-plate (right). These crystals are thin (estimated at ~10 µm) with poorly defined edges and are not suitable for diffraction experiments.

As seen with dioxane, the NMP initial solubility screen also resulted in the formation of small crystals; the crystals formed in 20 µl of 100% NMP in a glass well plate were of poor quality (figure 3.5). BDM31343 was observed to be soluble in DMSO, and crystals were grown by the evaporation of 20 µl 100% DMSO at room temperature in a glass well plate.

3.1.9 | Method: Crystallisation of BDM41420

As described above, the initial solubility screens involve spreading some of the compound across a slide. In the case of BDM41420, some of the powdered sample appeared to contain single crystals. Before commencing the solubility screen, four of these apparently-single, grey and block-shaped crystals were harvested.

In initial solubility screens, BDM41420 was seen to be soluble in polar, organic solvents such as ethanol, acetonitrile, DMF and DMSO. No crystallisation occurred on this screen slide, however, and so evaporation experiments were performed in glass well plates using 20 µl of DMF and DMSO at room temperature. These experiments were not performed on ethanol or acetonitrile because at room temperature, these solvents evaporate rapidly; with a limited amount of compound, it was preferable to use the solvents with the higher boiling points, allowing as much time as possible for nucleation and crystal growth to occur.

No crystals were formed from the evaporation of 100% DMSO, however needle-shaped, clear crystals were formed through the evaporation of 20 µl 100% DMF at room temperature in a glass well plate. When screening these crystals on the diffractometer, they proved too small for diffraction analysis. The four crystals harvested from the original sample were screened; three diffracted poorly and the final crystal was not, in fact, single. However, the crystal was successfully cleaved to around half its size, and was subsequently used for diffraction study.

3.1.10 | Method: Crystallisation of BDM41907

BDM41907 was a white powder which appeared as a collection of small needles. These clusters were found to be soluble in ethanol, acetonitrile, DMF, DMSO. A thin, clear, needle-shaped crystal of this compound was grown in 20 µl 100% DMF in a glass well plate at room temperature.

3.1.11 | Method: Crystallisation of BDM41325

BDM41325 showed a similar solubility profile to the compounds crystallised before, seen to be soluble in ethanol, acetonitrile, DMF and DMSO, yet insoluble in water and

toluene. Evaporation from all solvents with the exception of DMF was unsuccessful. In a glass well plate, evaporation of 20 µl 100% DMF yielded small clusters of needle-shaped crystals. In an attempt to slow nucleation, I set up a liquid diffusion of DMF and water at a ratio of 1:4, at room temperature. The same clusters were observed, though the clusters were larger than had been grown through evaporation. At that point, untried solvent combinations (including DMF/toluene, DMSO/water and DMSO/toluene) were used in an attempt to grow single crystals in isolation; this proved fruitless. Therefore, a thin paper taper was used to harvest the crystal clusters grown in the DMF/water liquid diffusion. The cluster was moved into a drop of inert oil on a glass slide (discussed later in section 3.2) and a large enough single crystal fragment was successfully separated from the crystal cluster for analysis by X-ray diffraction (figure 3.7 in section 3.2).

3.1.12 | Ongoing Crystallisation Attempts of EthR Inhibitors

At the beginning of this project, eight inhibitors were supplied, only four of which crystallised sufficiently to take forward to X-ray diffraction studies. The four remaining inhibitors are at various stages of investigation.

BDM41906 and BDM41281 are the least studied so far. Initial solubility screens on both showed solubility in acetonitrile, DMF, DMSO and ethanol, and insolubility in water. Evaporation trials of BDM41281 in these solvents have not led to any re-crystallisation. BDM41906 was moderately soluble in DMF, but so far all evaporation and liquid diffusion experiments have led to precipitation.

The initial solubility screen on BDM41425 resulted in irregularly-shaped small crystals forming from the evaporation of 100% DMF. However, this could not be repeated in a glass well plate with varying amounts of compound. A liquid diffusion experiment of BDM41425 dissolved in DMF with toluene as an anti-solvent (a 1:3 ratio) resulted in the growth of tiny star-shaped clusters at room temperature. This was repeated in a vapour diffusion experiment in an attempt to slow the growth of crystals, and though crystals formed, they were no larger and arranged in the same clusters. Temperature variations have had no effect on the formation and growth of these crystals, and unlike BDM41325, these crystals are too small to separate one for analysis.

BDM41231 has undergone the most investigation of the four remaining inhibitors. The initial solubility screen showed a slightly different solubility profile, as it was soluble in toluene as well as ethanol, acetonitrile, DMF and DMSO. BDM41231 demonstrated a tight meta-stable zone for nucleation and crystal growth.

On the initial screen, thin crystalline ‘rafts’ formed on the surface of the DMF drop; disturbance of the drop with a needle caused these to aggregate. This was repeated in 1 µl and 3 µl drops of DMF on a separate clean slide (these volumes are too small to warrant using a glass well plate), but the same observation was made. However, in 20 µl DMF in a glass well plate, clear needles formed after a period of 90 mins, but so rapidly that they began to adhere to each other and were impossible to separate, even when transferred to inert oil. Some of these aggregated needles were transferred to a new 20 µl DMF drop and re-dissolved; evaporation occurred without re-crystallisation and the original effect could not be repeated. BDM41231 dissolved in a 20 µl DMSO drop in a glass well plate led to precipitation in under an hour. With decreasing amounts of compound, re-crystallisation did occur, however in the first instance these crystals were thin, thread-like and aggregated almost like fibres, and with a further decrease in compound concentration, similar crystalline rafts as seen with DMF formed on the top of the drop. Liquid diffusions of ethanol/water and DMF/water were set up at 1:3 ratios at both room temperature and 4°C; no re-crystallisation occurred. A DMSO/water vapour diffusion (1:4) at room temperature led to the formation of small clusters which were tightly inter-grown and impossible to separate. This final round of liquid and vapour diffusions consumed the last of the available sample.

3.2 | X-ray Diffraction Experiments

3.2.1 | Choosing a Crystal

The most apparent indicator of quality is shape; a good crystal should have well-defined edges and not be curved or have 'satellite' crystals which cannot be removed from the surface. Crystals typically transmit polarised light, with three specific exceptions: tetragonal and hexagonal crystals if viewed along the *c* axis, and cubic crystals. However, other features of the crystal (such as the crystal shape) would indicate they were something other than amorphous if transmission did not occur. As the crystals are rotated under the polariser, the transmission should be even across the whole crystals and at some but not all crystal orientations. Non-crystalline material will not display this behaviour. Under cross-polarisation, crystals which transmit polarised light will be extinguished; this can be used to help identify if the examined crystal is singular as extinction should be sharp and complete. The size of crystal necessary for diffraction depends largely upon the equipment being used. For example, the protonated peptoid precursor crystal used for analysis had dimensions 200 μm x 100 μm x 100 μm, but structure determination was still possible.

3.2.2 | Harvesting a Crystal

Once a crystal (or in some cases, crystals) had been chosen, attempts would be made to remove them intact from the experimental solvent using a thin paper taper, and transfer them to a drop of inert oil on a glass slide with as much care as possible. The inert oil is more viscous than the experimental solvent, which can aid in manipulating the crystal, as well as securing the crystal once frozen in the mounting process. If clusters of crystals were grown which could be separated, or if smaller crystals were adhered to the main crystal which could potentially be removed, these crystals would also be harvested, and using the needle (or the taper, which is softer and so not as likely to cause destructive damage to the desired crystal) attempts would be made to separate the crystals. In the case of BDM41325, this led to a single crystal being removed from a large cluster.

Once crystals have been moved and provided they still hold up to the same standard once immersed in oil, the next step is to screen these crystals on the diffractometer.[77]

3.2.3 | Screening Crystals

Crystal screening will reveal the practical diffraction quality of a crystal; crystals which appear to be good quality under a microscope can diffract poorly, and less attractive crystals can prove strong once tested.

The first step is to mount the crystal on the glass fibre on a pin. This is easiest when the pin is already secured in the goniometer head. Before attempting to pick up a crystal, the tip of the pin would be roughly centred on the diffractometer, making use of the attached camera. This would mean that once the crystal had been mounted on the goniometer head, it would be in a good starting position as soon as the head was secured. The aim is to use a minimal amount of oil when picking the crystal up on the tip of the fibre, enough to keep the crystal secure once cooled by the nitrogen-flow device, but not so much as to risk interfering with the diffraction.

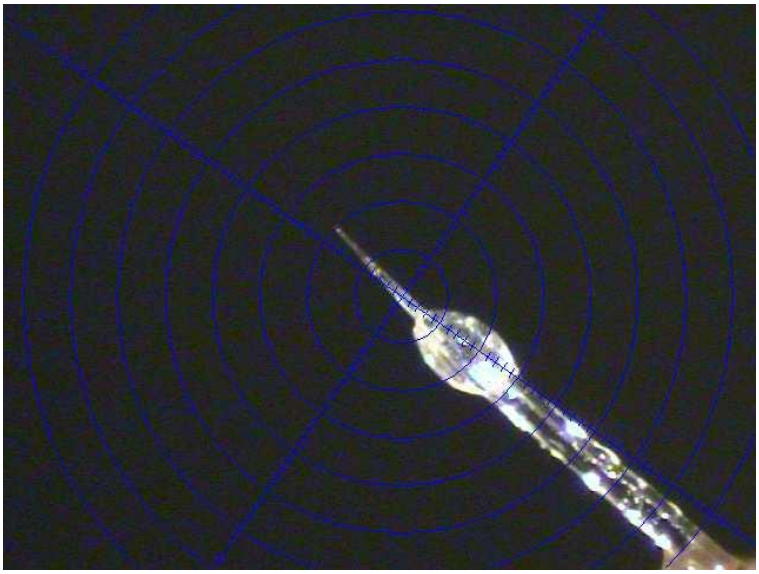


Figure 3.6 | The needle-shaped crystal of BDM41325 used for analysis, mounted in inert oil on a glass fibre on the goniometer head.
Size: 300 µm x 20 µm x 20 µm.

Once mounted, I would centre the crystal – this is not finding the centre of the crystal, but finding the centre of *gravity* of the crystal, so rotation will not cause dislodging or slipping, and the crystal will remain in the X-ray beam at all times. A camera is set up to aid centring, as can be seen in figure 3.7, above. Mounting needle-shaped crystals is something of a challenge, particularly as the needles are difficult to pick up on the fibre in the first place, but also because they are narrow and must be centred carefully. The crystal is rotated through *phi* while at a 60° *omega* angle, using small pins on the goniometer head to alter the angle of the fibre. Once the centre has been determined, it

is useful to move to a wider *omega* angle on the other side of the beam to ensure that centre of gravity remains steady and the crystal isn't potentially going to be lost during the collection.

To test a crystal, single frames of a narrow width (typically 0.5° per frame) were taken, with *phi* at varying degrees (0°, 60°, 90°, 150°). Also, this is the ideal time to find the appropriate exposure time, should the crystal prove to diffract well. A good crystal will give intense reflections which do not exceed the detector limits, with well-defined edges; smeared reflections will be difficult (if not impossible) to integrate reliably and doubled reflections are an indication of the crystal not being single.

If a crystal diffracts poorly, even with exposure times of up to a minute, then it is not worth wasting the time and energy to collect potentially unusable data; this happened frequently with my compounds, and in many cases there were other crystals to choose from until it was necessary to re-crystallise again.

If a crystal diffracts well, the first step is to determine the orientation matrix of the crystal. This is the process by which a limited number of frames are collected in order to determine a unit cell, or more accurately, determine the matrix function which relates the real-space unit cell to the reciprocal lattice distances. Enough data must be collected at different crystal orientations to gather as much information as possible from intense, well-formed reflections; most data collection software has a matrix strategy to collect the sufficient data.

3.2.4 | Data Collection

Attaining the unit cell dimensions has a three-fold role in the process of solving a structure. Firstly, a unit cell determination allows the crystallographer to make a decision about whether the compound is what is expected, using the *18 Å^3 Rule*. This 'rule' is a fast way of determining if the volume of the cell matches the expected compound, based on each non-hydrogen atom having a volume of 18 Å^3; it is a rough estimate, but can immediately flag any problems – salt rather than compound, for example. The second role of unit cell determination is to see whether the crystal diffracts well at a variety of *phi* and *omega* angles, and if the reflections remain as sharp and intense as the initial test. Finally, the unit cell determination leads to a determination of the crystal system, which informs how much data it is necessary to collect to achieve a complete data set. For example, more data would need to be collected on a monoclinic crystal than an orthorhombic crystal, which has a greater symmetry. If there is any doubt (such as a β angle very close to 90°, which could indicate orthorhombic character or simply a monoclinic cell with an incidentally ~90°

angle) then collecting with an assumption of a lower symmetry is safest, avoiding incomplete and unsolvable data.

Data for all compounds were collected on a Bruker MicroStar diffractometer equipped with a copper rotating anode X-ray source, Helios mirror monochromator system and a Platinum 135 CCD area detector. A combination of *omega* and *phi* scans were used at varying 2θ angles. The 2θ ranges for each data collection can be found in the relevant crystallographic data tables in the Appendix. The diffractometer is connected to a computer with software for designing, controlling, saving and processing the data collection.[78]

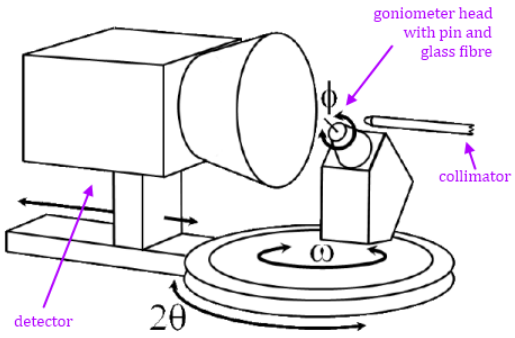


Figure 3.7 | The collimator controls the width of the beam, and is perfectly aligned with respect to the beam from the source and the goniometer. The 2θ angle is determined by the position of the detector; the ϕ angle is the rotation of the pin around 360°; the ω angle is the rotation of the goniometer head. The nitrogen flow device for cooling crystals throughout collection is not shown for the clarity of the image.

Data collection strategy software was used for designing the diffraction experiment. The strategy uses the unit cell determination to suggest the scans necessary to achieve a high completeness and reasonable redundancy, with an expected duration for the collection based upon the exposure times for each frame. With higher symmetry in the crystal, there is a small portion of unique data necessary to cover the entire crystal structure, with a greater number of equivalent reflections. Again, if there is any doubt, a lower symmetry class should be chosen so that all of the unique data is collected (completeness) and then as much data as time allows; repeated collection of equivalent reflections is termed 'redundancy'. The data collection strategy is particularly useful after a stock set of scans has been entered or even performed, as it will 'fill in the blanks' and select scans from a virtual bank of possible scans to increase completeness and redundancy. If time is an important factor in the data collection, a target can be set to optimise the scans chosen; data collection is a trade-off between completeness and time.

The anionic peptoid precursor, BDM41420 and BDM41907 were collected using a manually chosen data collection strategy, which was sufficient for completeness of the data. For the protonated peptoid precursor, BDM31343 and BDM41325, the data collection strategy was utilised to design an optimum data collection within a set time-

frame. Further details regarding the diffraction experiments on the six crystals can be found in the Appendix.

3.2.5 | Data Processing

Once collected, data were merged, integrated and scaled using SAINT;[78] absorption corrections were performed using SADABS, conducted automatically by SAINT. The R_{int} statistic is a measure of the internal consistency of the unprocessed data. The programme XPREP[78] is used to examine the systematic absences in the reflection data to determine the space group of the crystal. An R_{sym} value provided for the suggested space groups can inform the decision should there be any doubt. Finally, XPREP writes the files required for structure solution by SHELXS-97.[69]

All structures were solved using direct methods by SHELXS-97, and refined by least squares methods against F^2 using SHELXL-97 through the Olex2 interface.[79] For each structure, all non-hydrogen atoms were refined with anisotropic displacement parameters and all hydrogen atoms were unambiguously identified and refined isotropically on calculated positions using a riding model. Further crystallographic solution and refinement details on each of the six crystal structures can be found in the Appendix.

3.3 | *AutoDock Vina*

3.3.1 | Theory: Molecular Docking

AutoDock Vina (often referred to as simply '*Vina*') is a molecular docking software which predicts binding affinities and the bound conformations of ligands with receptor molecules - typically small organic molecules to proteins.[10]

Molecular docking is a computational approach to protein-ligand binding when experimental data has not yet been acquired. In all computational methods, there is an inevitable trade-off between exhaustive accuracy and speed; the algorithm utilised by *Vina* is an iterative process from a random starting seed position of the ligand, which seeks to maximise the accuracy of the determined conformation and binding affinity with a manageable computational time.

Docking in this form is a statistical method utilising biomolecular force fields.[80] The binding affinity of ligand to protein is determined by structural and energetic compatibility; the biomolecular force fields are comprised of a large number of energy terms for covalent and non-covalent interactions. The force field calculations assume a vacuum, but use empirical data to give appropriate, consistent values for the force field parameters. For example, structural data (such as crystal structures) give molecular geometry and bond lengths and torsions to inform movement; spectroscopic data give molecular vibrations and force constants, whereas thermodynamic data give the partition co-efficient. Experimental data from small molecules calibrate the docking calculations to ensure high-probability results with reliable accuracy.[81]

3.3.2 | Using *Vina*

Certain assumptions are made in order to decrease the time and detail of the computations necessary to determine the result. First, *Vina* assumes the protonation of both the ligand and protein remain the same between the bound and unbound states. Second, *Vina* assumes the receptor remains rigid, with constant and unchanging bond lengths, bond angles and torsion angles; with the preparation of particular input files, however, *Vina* is capable of taking some flexible residues into account. Finally, *Vina* requires the user to specify the rotatable bonds of the ligand.

Vina has various defaults built in to the computational process, which can be specified. For example, *Vina* will start at a random seed position if one is not specified; unless included in the instruction configuration file, a maximum of nine binding modes will be reported with the difference between the first (and best) and the ninth (with the lowest binding affinity) being no more than 3 kcal/mol; the exhaustiveness of the

search is given an arbitrary value on a linear scale with time, with the default value of '8' – increasing this exhaustiveness will increase the computational time but will increase the accuracy of the results.

The result of the *Vina* run is a number of binding modes with 3D co-ordinates, given with the binding affinity of the ligand and protein for each mode. In addition, the RMSD of each mode successive binding mode is compared to the best result and reported. Two RMSD values are given, one is the 'lower bound' or 'lb' and the other is the 'upper bound' or 'ub'. The upper bound RMSD is defined in the *Vina* manual as the statistical match of each atom in one binding mode with itself in the top binding mode, ignoring any symmetry. The lower bound RMSD utilises *rmsd'* to give an RMSD where each atom in one binding mode is matched to the closest atom of the same element type. Below is an example of the statistical read-out from *Vina* (figure 3.8).

```
Detected 2 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 127522912
Performing search ... done.
Refining results ... done.

mode |   affinity | dist from best mode
     | (kcal/mol) | rmsd l.b.| rmsd u.b.
-----+------------+----------+----------
   1        -10.6      0.000      0.000
   2         -9.2      1.929      9.779
   3         -8.9      3.484     11.192
   4         -8.4      1.634      2.427
   5         -8.3      1.960      9.867
   6         -8.1      3.523      5.323
   7         -8.1      3.401     10.265
   8         -8.0      3.729      6.006
Writing output ... done.
```

Figure 3.8 | The log file of a *Vina* run, detailing the random seed position used and the eight binding modes determined with their relative RMSD values and binding affinities. Details of this run are described in section 3.3.3, where a ligand was removed from a co-crystal structure to test *Vina*.

The two RMSD values inform differently, as a rigid symmetrical ligand such as a benzene ring could be rotated 180°; the RMSD/ub would give a significantly higher result than the RMSD/lb, where the upper bound takes into account the movement of specific atoms, whereas the lower bound would only take into account the nearest atom of the same type.

Finally, *Vina* generates the co-ordinates of the calculated binding mode in relation to the protein, and saves these in the same format as the input file to be viewed. *Vina* uses a unique type of input file. The standard PDB format does not carry

information which *Vina* requires (hydrogen atoms, charges and atom assignments), and so and extended format, PDBQT, is used instead. These files can be prepared in the *AutoDock* visualiser *AutoDock Tools* (ADT) from the PDB files. This interface is also used to determine the search space or 'grid box'.

3.3.3 | Testing *Vina*

In order to understand how best to utilise *Vina* for docking the inhibitor crystal structures to EthR as well as determine just how accurate *Vina* was, the following test model was used: the published co-crystal structure (PDB: 31GM) of EthR with BDM31381.[99] Using ADT, the ligand was separated from the protein structure into separate PDB files. Following this, the ligand file was prepared by assigning the rotatable bonds of the molecule and subsequently written out in PDBQT format; the protein file was prepared by deleting the water molecules (which *Vina* cannot take into account), adding polar hydrogens to the structure, assigning the charges and the atom names and finally writing out the unbound EthR structure for docking. The intention was to use different search spaces volumes to assess two things: (a) how closely *Vina* could determine the experimental binding mode (by comparing the resulting modes with the experimental 3G1M structure) and (b) how long a run upon a search space can be expected to take.

The result of this exploratory test showed that *Vina* was most accurate with an increasingly exhaustive search of the *entire* protein structure (with a box volume of 126,720 $Å^3$). The default value of exhaustiveness is 8, which was steadily increased to 40; with no other alterations to the programme configuration, this took only twenty minutes. The runs were conducted on a personal laptop running Windows 7, equipped with 4GB RAM and a 2.2GHz dual-core processor. The log file of the initial test can be seen as figure 3.8, above. Therefore, it was decided that *Vina* would perform exhaustive runs on the whole protein search space with each ligand - and this model form of 3G1M - with an empty ligand-binding site in a ligand-binding conformation.

3.3.4 | Method: *Vina* applied to EthR inhibitors

Based upon observations made by Willand *et al.*,[55] where the movement of two phenylalanine residues in the EthR structure were seen to be rotatable and as such, could open up an auxillary binding pocket for longer ligands, *Vina* was instructed to compute binding with these two residues as flexible in the ligand-binding channel. This required preparation of three input files in ADT: the ligand file (one for each of the four

inhibitor crystal structures), the torsions of the flexible Phe114 and Phe184, and the rigid protein without these two residues.

Normally, *Vina* considers the protein as rigid and the ligand as flexible in user-specified rotatable bonds. However, by including flexible residues, the user is able to select rotatable bonds in the side chains (Phe114 and Phe184 in this case) while the rest of the protein remains rigid. However, there is a computational cost and an increase in time required. 3G1M was used as the protein model because the co-crystal structure contained a ligand of a similar size with the same base chemical structure as the ligand crystal structures determined. The following configuration file was used for each run:

```
receptor = 3g1mrig.pdbqt     center_x = 26
flex = 3g1mflex.pdbqt        center_y = 74
ligand = XXX.pdbqt           center_z = 7
out = outputXXXflex.pdbqt    size_x = 68
log = logXXXflex.txt         size_y = 48
exhaustiveness = 40          size_z = 36
num_modes = 4
```

Figure 3.9 | Configuration file for *AutoDock Vina* runs.

The *receptor*, *flex* and *ligand* fields specify the source files (figure 3.9, above), with the *out* field giving a file name for the co-ordinates of the binding modes and the *log* field giving a file name for a log file to be written; the *center* and *size* fields define the search space; the *exhaustiveness* field, as described above, determines how long and detailed the iterative process should be. Each of the four docking runs (one for each compound) took roughly 45 minutes. The results of these docking runs are summarised for each compound in chapter four, with more detailed discussion in chapter six.

CHAPTER 4

RESULTS & DISCUSSION: PEPTOIDS

4.1 | Anionic Peptoid Precursor

The anionic precursor is an intermediate product of the target peptoid formation.[82] The crystal structure of this anionic peptoid precursor revealed it be a sodium salt (figure 4.1).

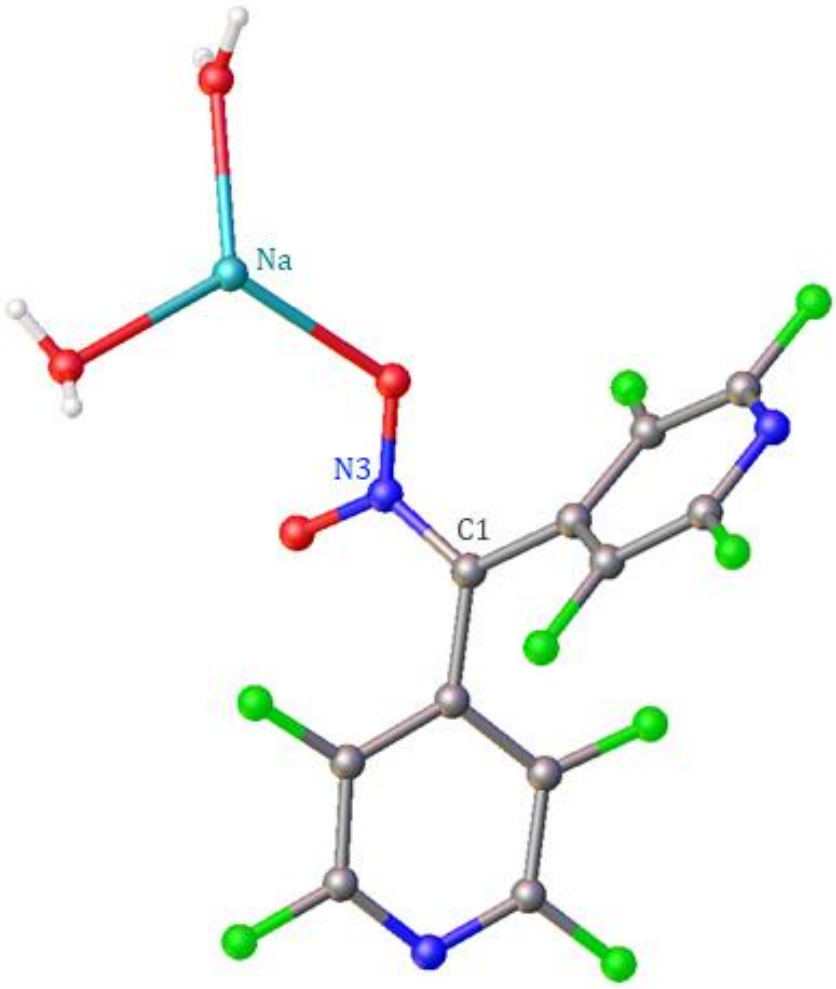


Figure 4.1 | Asymmetric unit of the anionic peptoid precursor. All bond lengths and angles determined were as expected.

For both the anionic and protonated precursor form, the full crystallographic details (including bond lengths and angles, and figures showing the 50% probability anisotropic ellipsoids of constituent atoms) can be found in the appendix as sections A-I and A-II. All images were created in Olex2.[79]

4.1.1 | Structural Analysis

The central C1 atom is trigonal, with the C1-N3 bond displaying double bond character - evidenced by a bond length of 1.338(13) Å. The sodium ion co-ordinates two oxygen atoms in the asymmetric unit, however in the wider crystal structure, the sodium ion is trigonal bipyramidal (figure 4.2), with the five co-ordinate sodium ion holding two anions and three bridging water molecules in place.

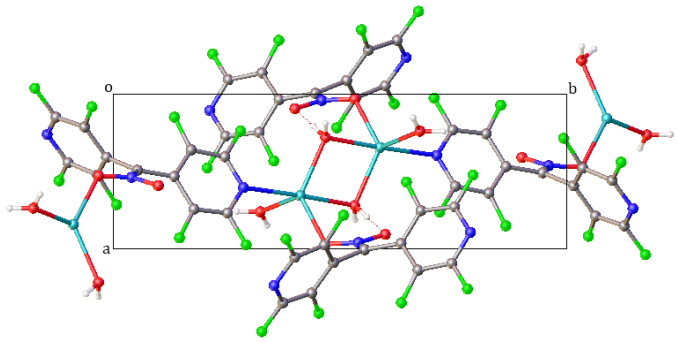


Figure 4.2 | Cell-centered image of the anionic peptoid precursor, viewed along the *c* axis. A clearer image of the crystal structure itself is given in figure 4.4, overleaf.

However, the angle formed by the axial atoms O3-Na-N2 is 177.4°, which is distorted from the ideal 180° (figure 4.3). Similarly, the equatorial positions are not at 90° to these axial positions. Specifically, the angle O3-Na-O3' is reduced to 73.4°, but the angle formed by O3-Na-O4 is widened to 118.9°, due to sodium from another molecule co-coordinating with the shared oxygen atoms. From this we can surmise that hydrogen bond interactions as well as the sodium co-ordination determine the crystal packing structure, as can be seen in figures 4.2 and 4.4.

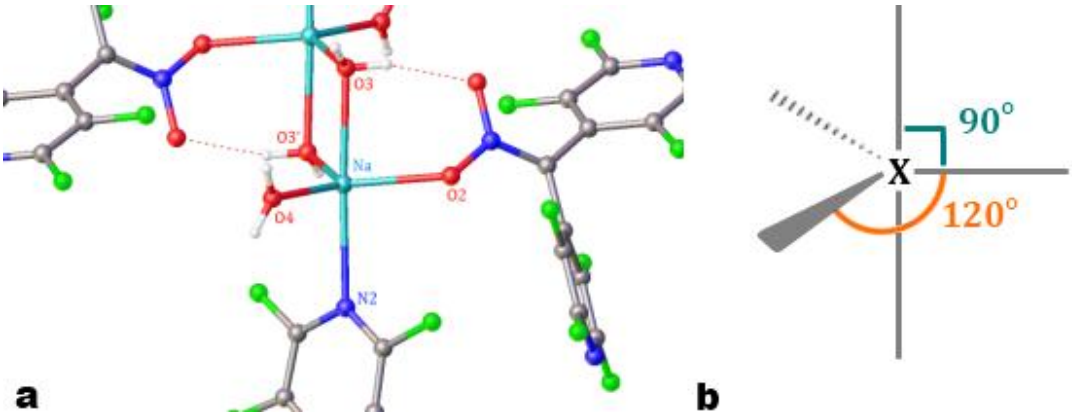


Figure 4.3 | The co-ordination around the sodium ion (**a**) is not perfectly trigonal bipyramidal – angles shown in **b** are the expected trigonal bipyramidal arrangement about central atom/ion **X**.

There is an observed delocalization of electrons across the perfluoropyridyl rings, evidenced by the equivalency in the carbon-carbon bond lengths of the heterocycles.

The NO_2 group is involved in hydrogen bonding, and there is evidence to support the two nitrogen-oxygen bonds being equivalent. A typical nitrogen-oxygen single bond has a distance of approximately 1.40 Å, and a double bond is typically around 1.21 Å; the N3-O1 and N3-O2 bonds are of almost identical length - 1.282(11) Å and 1.288(11) Å respectively, forming an O1-N3-O2 angle of 120.7°. This potentially gives a more localised charge to the NO_2 group, increasing the potential for electrostatic bonding.

4.1.2 | Crystal Packing Interactions

Expanded from the asymmetric unit, the crystal structure of the anionic peptoid precursor is demonstrated below. The structure itself is held together by the co-ordination around sodium ions and hydrogen bonds formed by oxygen atoms (figure 4.4, overleaf).

There are two key hydrogen bonds repeated in the packing structure, shown as dotted lines in figure 4.5 below; the second Na-N2 ionic bond is shown for clarity. The O1-H3a (1.968 Å) and N1-H4b (2.227 Å) bonds stabilize the co-ordination around the sodium ion and maintain the positions of the perfluorinated pyridyl rings.

Though potential stacking interactions were anticipated between the perfluoropyridyl groups,[83] the conformation of the precursor does not support it. The planes of the two rings are at a 95.2° angle with respect to each other, and the co-ordination of the sodium ion in the structure precludes π-π interactions. However, this does not prevent the potential for this kind of non-polar interaction in a covalently linked sequence of peptoid or peptoid-based molecules.

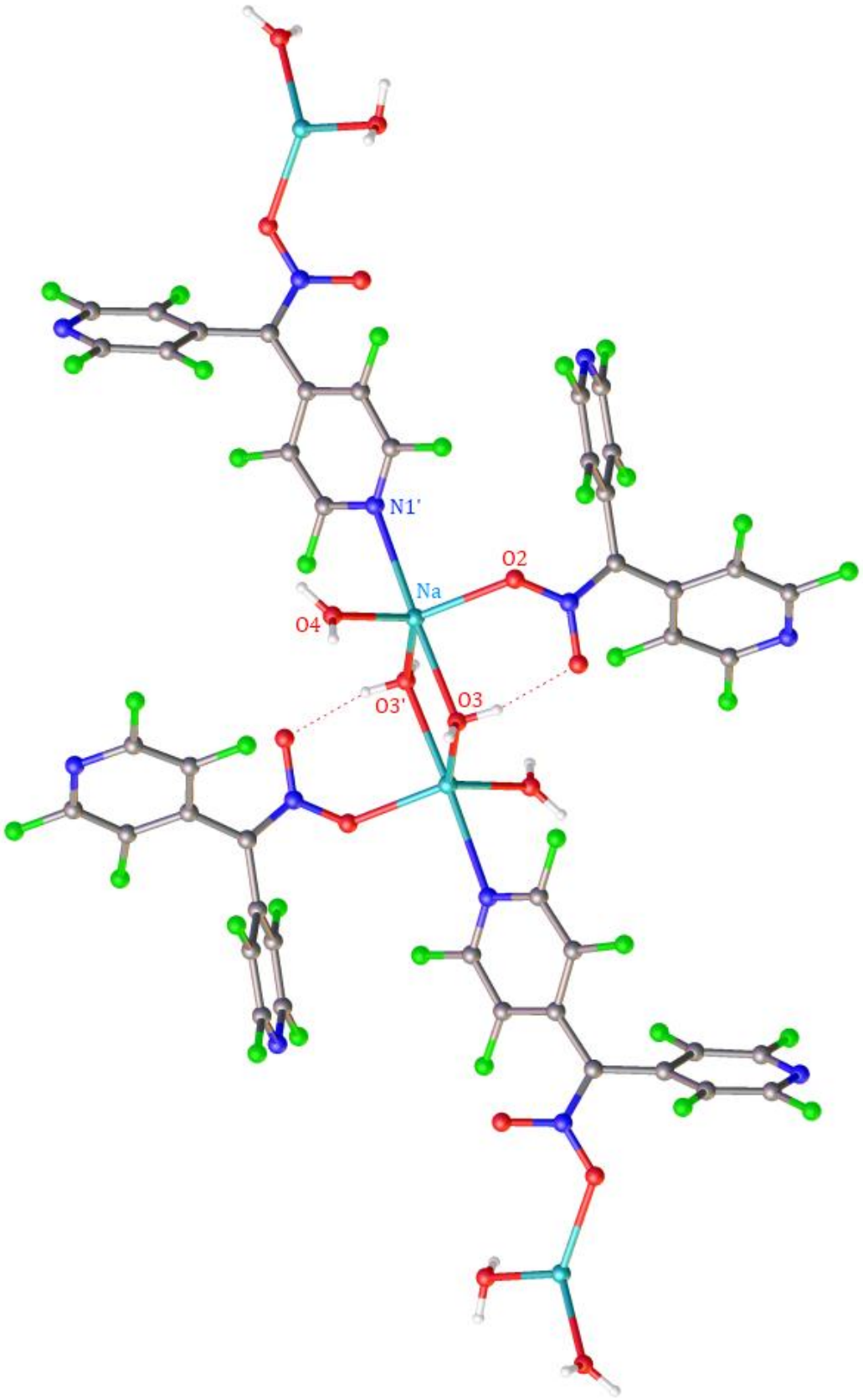


Figure 4.4 | Crystal structure of the anionic peptoid precursor.

4.2 | Protonated Peptoid Precursor

The precursor was prepared[82] with the intention of isolating the reduced product. The asymmetric unit of the crystal structure was determined to be the protonated form of the peptoid precursor.

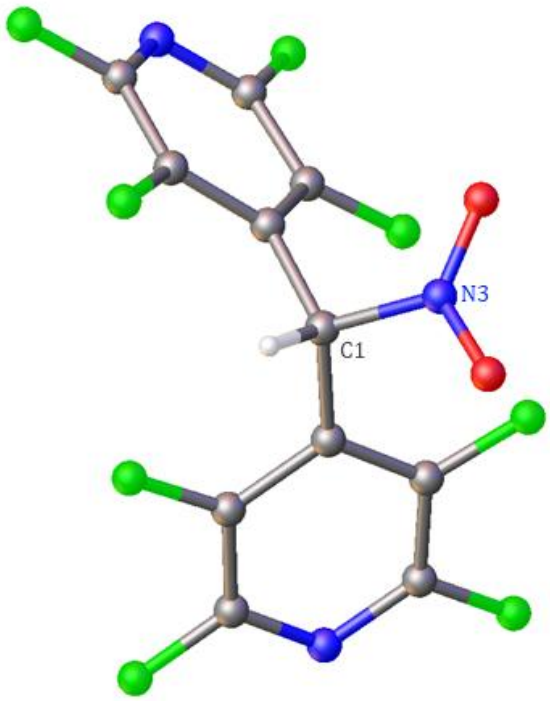


Figure 4.5| Asymmetric unit of the protonated peptoid precursor

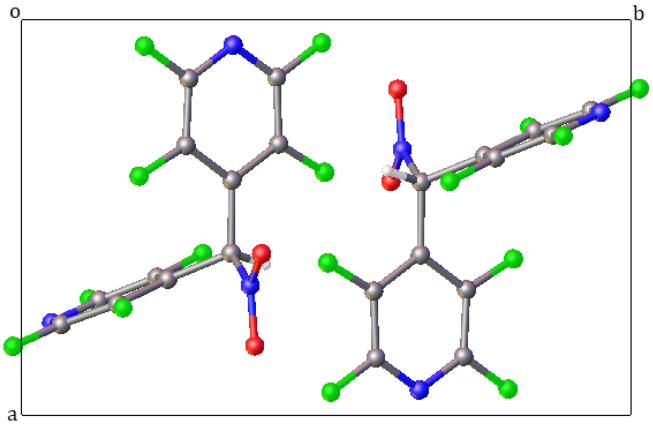


Figure 4.6 | Cell-centered image of the protonated peptoid precursor molecule. The crystallographic inversion centre is evident in the centre of the cell.

4.2.1 | Structural Analysis

Unlike the anionic precursor form, the protonated peptoid precursor shows a typical tetrahedral formation about the central C1 atom due to the addition of hydrogen. The C1-N3 bond shows a carbon-nitrogen single bond distance of 1.487(4) Å, longer than the corresponding bond length in the anionic form. The new arrangement around C1 is less restrained by this single bond and as a result, the planar perfluoropyridyl rings are oriented at an angle of 100.3°, greater than in the anionic form (95.2°). The same

delocalization of electrons over the heterocyclic rings, as seen in the anionic form, can be observed in here.

The NO_2 group, in contrast to the anionic form, does not show equivalent nitrogen-oxygen bonds. Though the angle formed is similar (O1-N3-O2 angle of 123.9°) the two bonds are distinctly different. The N3-O1 bond, with a distance of 1.200(4) Å, and N3-O2 bond, which is 1.246(4) Å, are equivalent; the bonds differ only slightly from the expected 1.21 Å length for the N-O bonds of an NO_2 group.

4.2.2 | Crystal Packing Interactions

As with the anionic form, there are no π-π interactions of the ring systems, and instead of packing through sodium co-ordination, hydrogen bonds are formed between molecules through the hydrogen atom (figure 4.7).

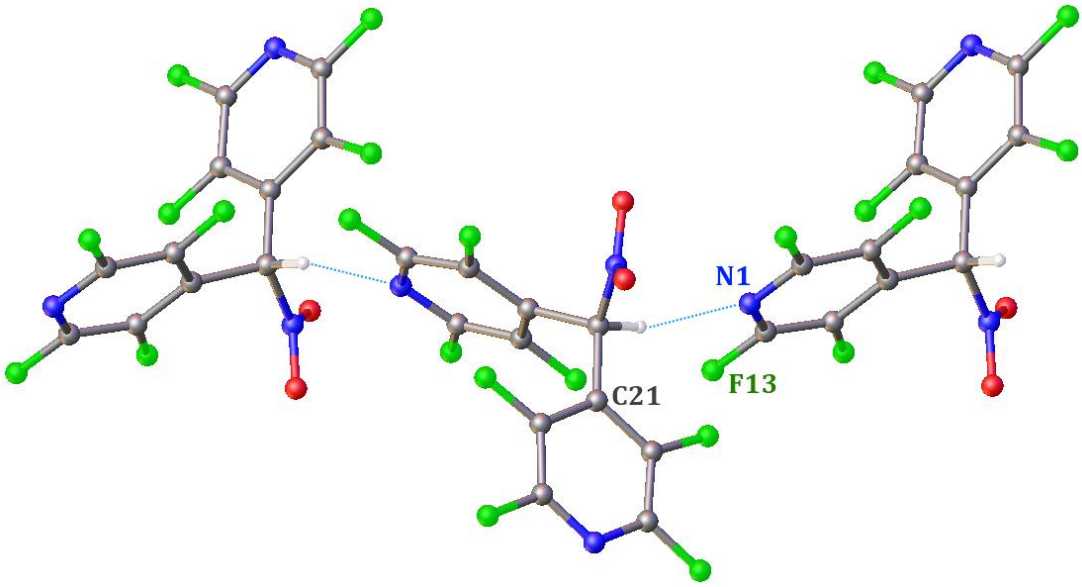


Figure 4.7 | Hydrogen bonding stabilizes crystal packing; H1···N1' has a distance of 2.467 Å.

Also present is evidence of C-F···π interactions. The F13 atom is 3.259 Å from the plane of the C21 ring of the adjacent molecule (figure 4.7). Due to the delocalisation of negative charge across the perfluorinated ring, the centre of the ring is relatively electropositive; F13 is electronegative. Interactions of this nature have been shown to be widespread amongst fluoro-organic crystal structures.[84] Therefore, the only interactions maintaining this packing structure are the N1···H1 hydrogen bonds and the F13···π electrostatic attraction.

This implies that the crystal packing is not tightly held, particularly in comparison to the anionic precursor form, giving some explanation as to the formation of two different crystal habits (section 3.1.7). Without determining the structure of the

alternative, needle-shaped crystal habit, it is impossible to say whether hydrogen bonding was occurring between different atoms, however it could be suggested that the alternative habit arises from hydrogen bonds between H1 and N2, O1 or O2. It is unfortunate that these needle-shaped crystals were too small to analyse, for the sake of a complete picture of the potential packing arrangements.

The crystal packing interactions of the monomeric precursor forms gives an insight into the interactions which could form in a multimer, dictating secondary structures.

4.3 | Implications for Use as Peptoids

To transition into potential use as peptoids, the precursor molecule (specifically the protonated form, as the anion would not be appropriate for a peptoid-based peptide mimic as suggested in chapter two) would have to undergo reduction (figure 4.8).

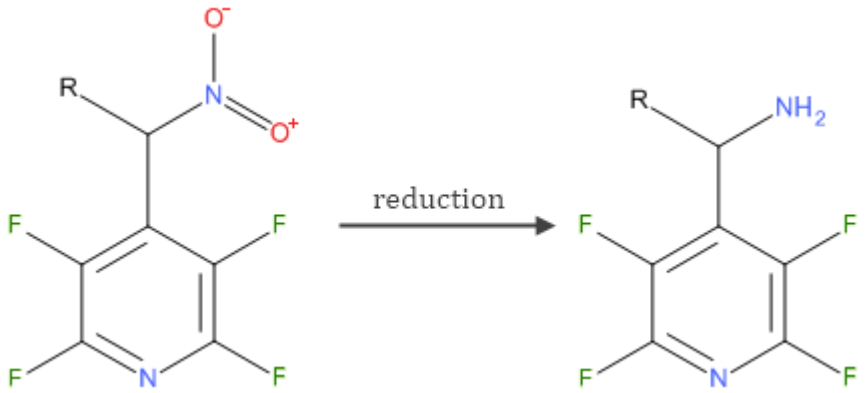


Figure 4.8 | Scheme for synthetic preparation of a peptoid from a precursor.[82]

Although the NO_2 group is present in the precursor form, it would be reduced to an NH_2 group in the peptoid; this would be condensed to then create a linkable peptoid for oligomerization (figure 4.9). Therefore, the hydrogen bonding potential of this group, found in the anionic form, is not available here.

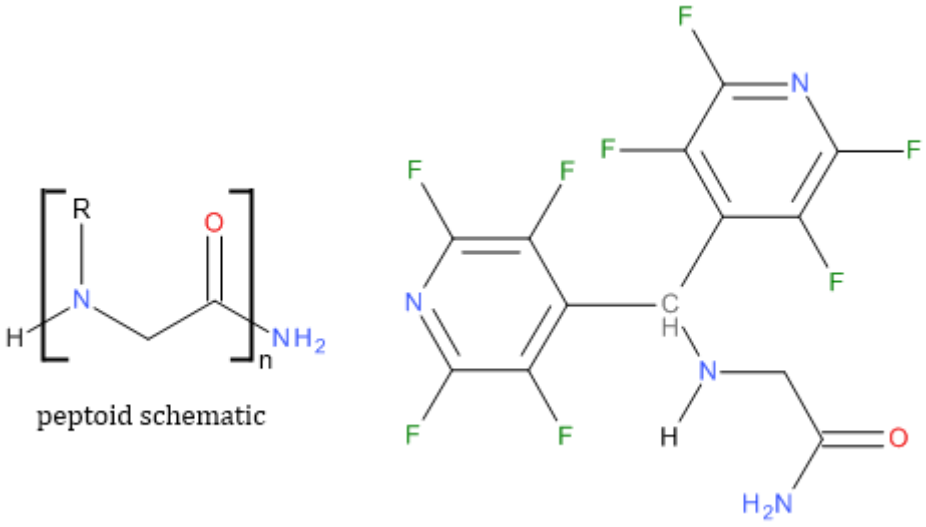


Figure 4.9 | Given the schematic for a peptoid, left, the potential peptoid formed from the protonated precursor would have a chemical structure such as that on the right.

As these two perfluorinated rings are bulky, the backbone and secondary structure of any oligomeric peptoid would be highly distorted from a typical α-helix. However, the nitrogen atoms of the perfluorinated pyridyl rings are still available for hydrogen bonding, as would be the hydrogen of the N-linkage.

CHAPTER 5

RESULTS AND DISCUSSION: ETHR INHIBITORS

5.1 | BDM31343

The crystal structure of the first inhibitor compound was determined, with the contents of the asymmetric unit – one molecule of the inhibitor – shown as figure 5.1.

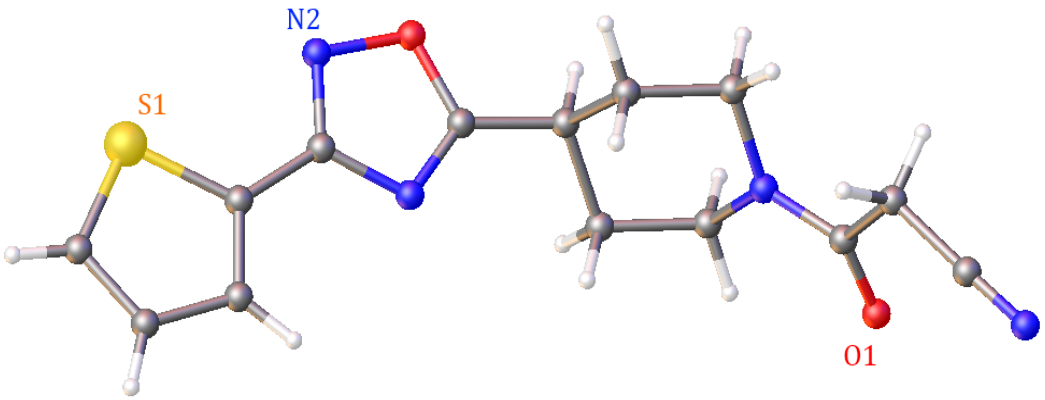


Figure 5.1 | Asymmetric unit contents of BDM31343. All bond lengths and angles were as expected.

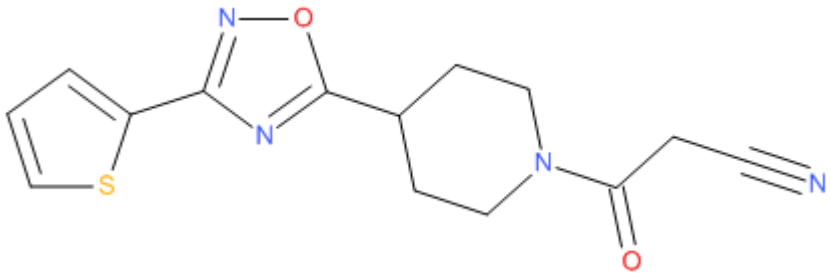


Figure 5.2 | Chemical structure of BDM31343.

It is of note that the oxadiazole ring takes an orientation distinct from the chemical structure given in Willand et al., 2009, which can be seen as figure 5.2, above. This results in an S1···N2 distance of 3.086 Å, a potential source of non-bonding interactions between the ligand and the protein. A similar intramolecular interaction was suggested by Flipo et al., 2011 based on the co-crystal structure of EthR and BDM31381, in which a sulphur in the second thiophene ring was within 3.5 Å of the carboxyl oxygen atom, however BDM31343 does not share that structural group. Interestingly, the published

Full crystallographic details for all four inhibitor structures (including bond lengths and angles, and figures showing the 50% probability anisotropic ellipsoids of constituent atoms) can be found in the appendix as sections A-III to A-VI. All images were created in Olex2.[79]

structure of BDM31381 with EthR (PDB: 3G1M) shows the inhibitor as drawn in the chemical structure, with S1 and N2 in a *trans* conformation. The 3G1M crystal structure has a resolution of 1.7 Å, which should be sufficient to accurately identify the atom positions of the bound ligand BDM31343. The structure shows electron density around the thiophene ring which would support the orientation observed in the BDM31343 crystal structure rather than the one presented in the refined co-crystal structure, implying an incorrectly modeled ligand in the published EthR-BDM31343 complex.

The thiophene ring of BDM31341 and the oxadiazole ring are planar (RMSDs of each ring are 0.003 Å and 0.001 Å respectively). The angle between these ring planes is 5.6°, meaning the compound crystallizes with these two rings nearly, but not quite, planar to each other. This conformation would be supported by an intramolecular interaction between S1 and N2. Though the oxadiazole ring is capable of both polar and possible stacking intermolecular interactions to stabilize the crystal packing, only the O1 atom makes any connections within a suitable distance to this effect.

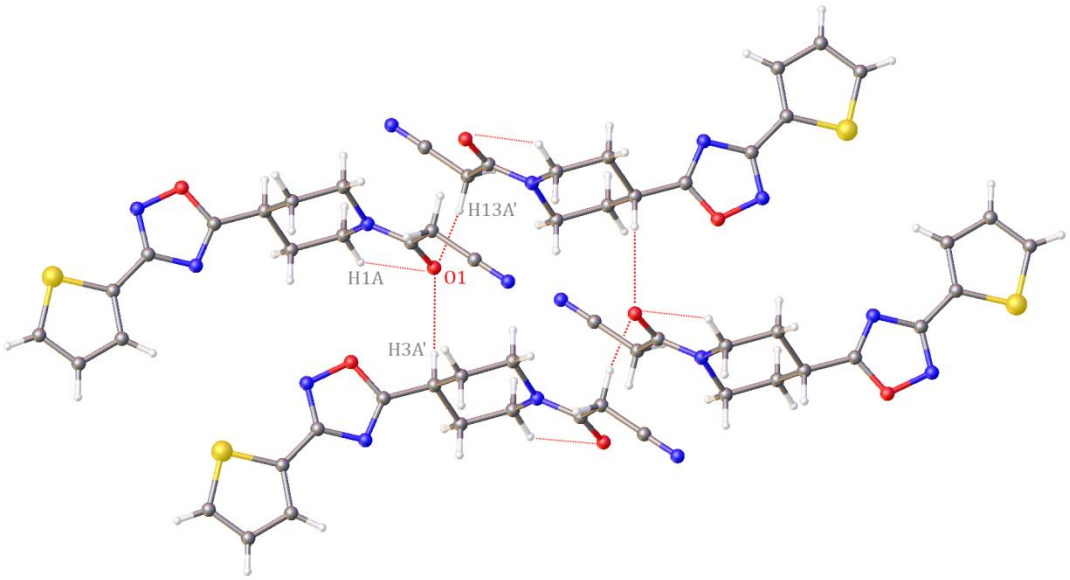


Figure 5.3 | Crystal packing of BDM31343. Interactions between atoms are shown in red.

Atom O1 of the crystal structure is within the expected 4 Å van Der Waals distance[74] of two hydrogen atoms, one positioned in a molecule above (taking the furthest left molecule as the example, this would be H13A) and the other in a molecule below (H3A). The orientation this enforces upon the rest of the molecule precludes any π-π stacking interactions involving either the oxadiazole, piperidyl or thiazole rings. However, O1 also forms an intramolecular interaction with H1A (at a distance of 2.379 Å) which would be responsible for ensuring the nitrile 'tail' of the molecule is held at a specific angle.

The interactions present in this crystal structure would suggest that the ability of BDM31343 to form similar interactions with the ligand binding channel of EthR would be largely dependent of the position and orientation of O1.

5.2 | BDM41420

BDM41420 shares a similar core structure to BDM31343, with a piperidyl ring and carboxyl group. However, in place of an oxadiazole linker, there is an oxazole – this structure has no N3 atom. Furthermore, the sulphur is no longer present and instead a phenyl ring takes the place of the thienyl found in BDM31343; in the place of the nitrile is a trifluorocarbon group, which infers more possibility for intermolecular bonding.

BDM41420 crystallised with two distinct molecules in the asymmetric unit, unlike the previous crystal structures where the Z' – the parameter which is used to define the number of molecules in the asymmetric unit – is equal to 1. These two molecules are overlaid in figure 5.4, below.

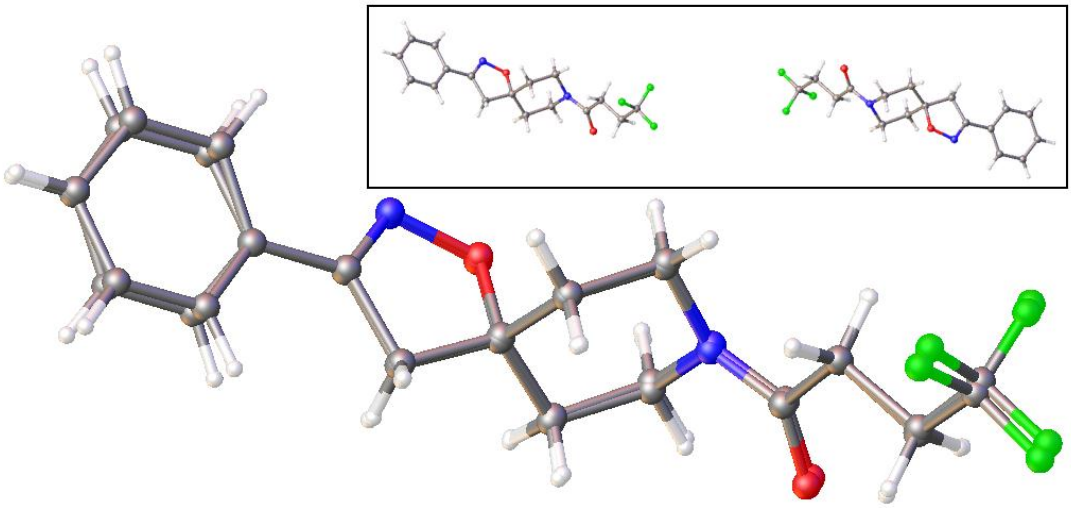


Figure 5.4 | Least squares superimposition of the two molecules of BDM41420 in the asymmetric unit (inset). There is a non-crystallographic centre of inversion in the asymmetric unit.

The two molecules BDM41420 match, with an inversion of one molecule, with an RMSD of 0.217 Å. This shows that the two molecules are chemically identical, yet structurally distinct enough to be considered different. What has been unintentionally captured is an indication of a flexibility of the molecule, even though crystallography is typically technique which produces static, averaged snapshots of a crystal structure.

The overlay reveals movement primarily in the phenyl ring (which is planar, within 0.004 Å) and around the tetrahedral CF_3 group. The C3 atom has a mildly distorted tetrahedral geometry in both molecules as it is part of the oxazole ring, so that O2-C3-C2 and O2-C3-C4 angles are 107.7(9)° and 107.2(9)° respectively, and the C2-C3-C4 angle is 110.8(9)°; this also leads to the oxazole no longer being planar, as seen in BDM31343.

The intramolecular interaction of O1 with H1A is seen in both molecule *1* (at a distance of 2.354 Å) and molecule *2* (O3···H22A with a distance of 2.338 Å). The

interaction occurs on the alternate side of the piperidyl ring in molecule *2*, indicating the piperidyl-oxazole ring has inverted; this is the cause of the difference between the two molecules in the packing structure.

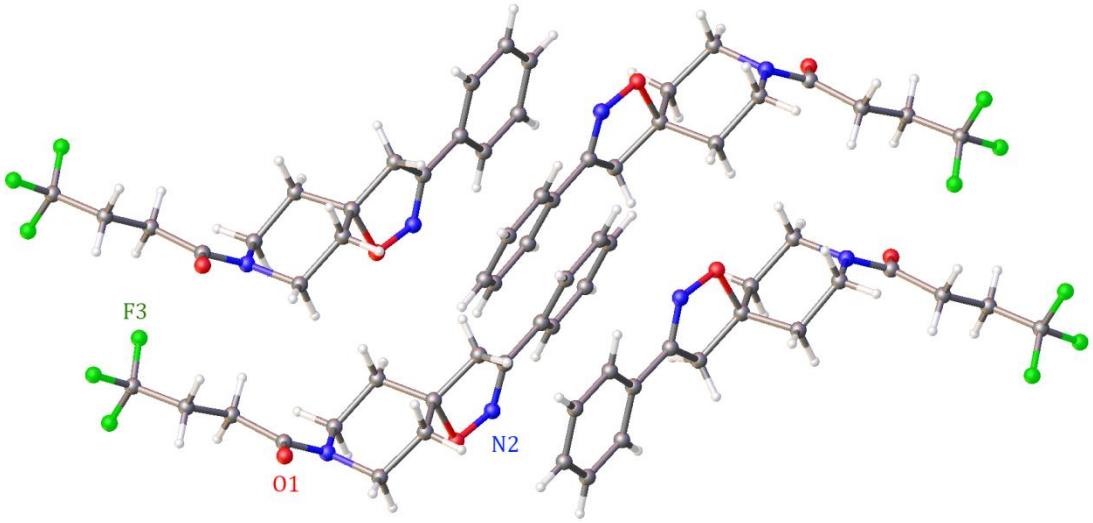


Figure 5.5 | Example of crystal packing in BDM41420. Note that while the orientation of the molecules in the figure suggests π-π stacking, this is not the case.

The crystal packing structure is also supported by hydrogen bonds made by N2, O1 and F3 with hydrogens (H23A, H15B and H16B respectively) in molecule *2* – with the equivalent atoms in alternate molecules; fluorine hydrogen bonds are a strong determinant of crystal packing in fluorine-substituted rings.[85] However, the structure shows no evidence of π-π stacking, just as observed in BDM31343. Overall, this implies more opportunity for protein-ligand binding, even with the reduction of flexibility caused by specific arrangement around the C3 atom.

The presence of two molecules in the asymmetric unit (Z' = 1) is an important feature of the crystal packing. An Z' >1 can be a result of various factors including chirality or an increased ability to make hydrogen bonds, but the underlying laws governing crystal packing are not well understood. A study of the Cambridge Crystallographic Database quotes the frequency of a Z' > 1 as being 8.8% across the CSD as a whole, and 11.5% for organic structures.[86] Of these structures, the most common space group is $P2_1/c$, which accounts for almost 27% of the CSD structures with Z' =2;[87] the separate molecules are typically observed to avoid the crystallographic glide plane and to pack in such a way as to minimize the packing surface area, usually with a non-crystallographic inversion centre.[88] This inversion centre can be seen in the figure 5.4.

5.3 | BDM41907

The structure of BDM41907 revealed it to have the same core structure as BDM31343, with some of the modifications seen in BDM41420. This structure possesses an oxadiazole linker, a trifluorocarbon group and the essential piperidyl ring with a carboxyl group attached at N1. However, unlike the two previous molecules, BDM41907 possesses a methyl-sulphoxide group linked at C11.

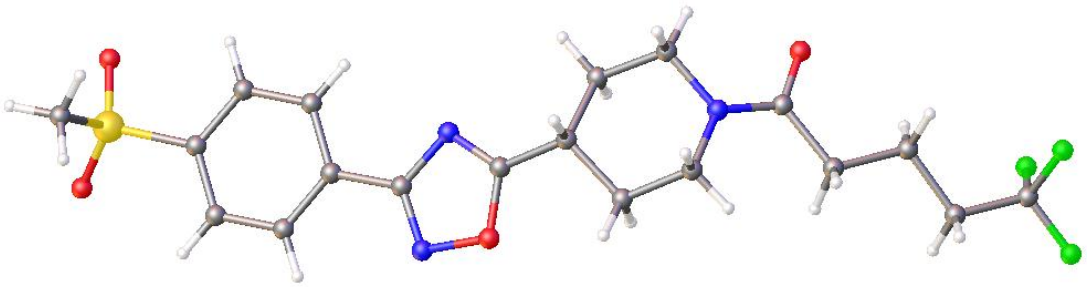


Figure 5.6 | Asymmetric unit (one molecule) of BDM41907

The oxadiazole ring is planar (RMSD of 0.003 Å), as is the phenyl ring (RMSD of 0.006 Å). The two rings are not perfectly planar with respect to each other, being at an angle of 7.4° - however this reflects a similar arrangement observed in BDM31343, suggesting that the lowest energy conformation in the crystal packing structure benefits from the planarity of these rings, despite them forming no π-π stacking interactions.

BDM41907 contains a longer chain between the carboxyl group and the triflurocarbon group than seen in BDM41420 by one CH_2. This longer molecule can potentially occupy a larger portion of the ligand-binding channel, forming more intermolecular interactions and prospective tighter binding effect; molecular docking analysis will reveal if this is the case (section 5.5).

Though sulphur is present in this structure, it is not in a position capable of forming intramolecular interactions with any nitrogen or oxygen present in the molecule. Figure 5.7, overleaf, indicates three of the main hydrogen bonds which hold the packing structure, made by O3, O4 and F3. The methyl-sulphoxide groups of two adjacent molecules (O3···H19A' and O3'···H19A at 2.435 Å) interact to bring the structures into alignment; interactions formed by O4 with H15B' (at a distance of 2.404 Å) and F3' with H10A ensure that the rest of the molecule is held in place in the crystal structure. The intramolecular interaction between O1 and H1A at a distance of 2.323 Å indicates that some intramolecular stabilization of the long, flexible ligand is occurring; this was observed in BDM31343 and BDM41420.

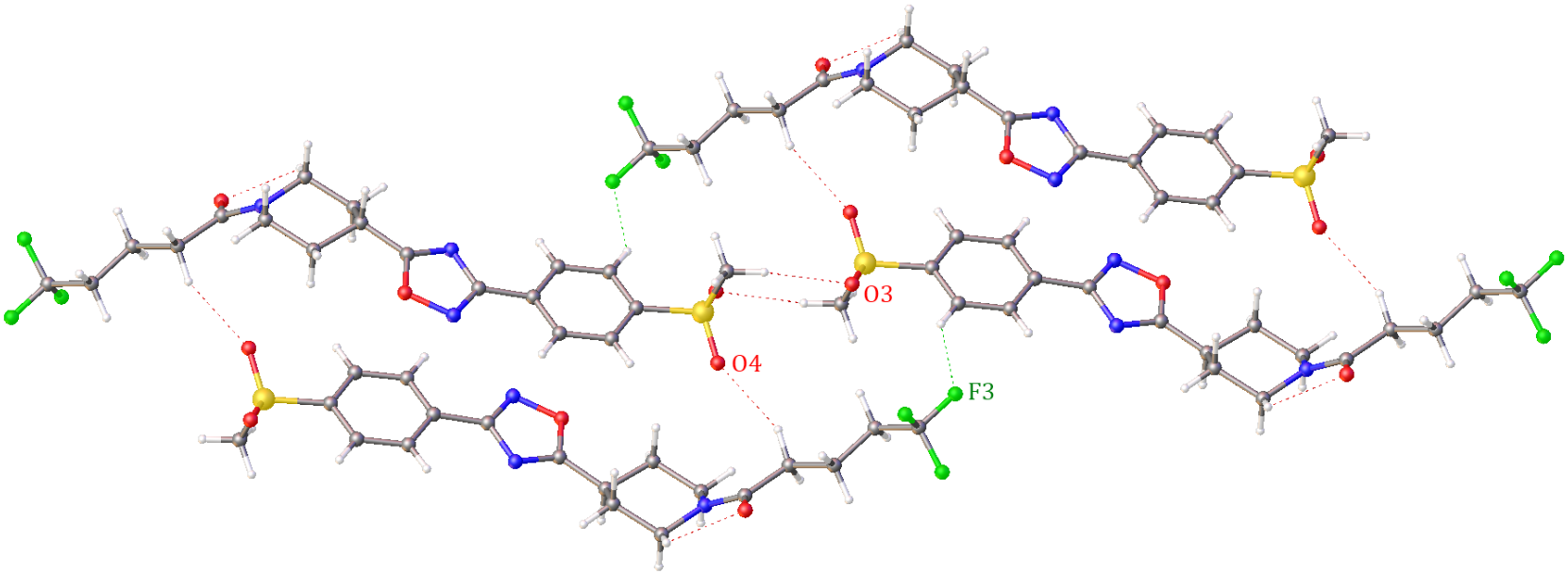


Figure 5.7 | Extension of the intermolecular interactions displays the packing structure.

5.4 | BDM41325

The crystal structure of BDM41325 showed that although it possesses the same core piperidyl-carboxyl link seen in the other three compounds, BDM41325 is structurally distinct. BDM41325 possesses a methoxyl-benzothiazole group connected to the oxadiazole linker; this is the longest ligand of the four crystallised.

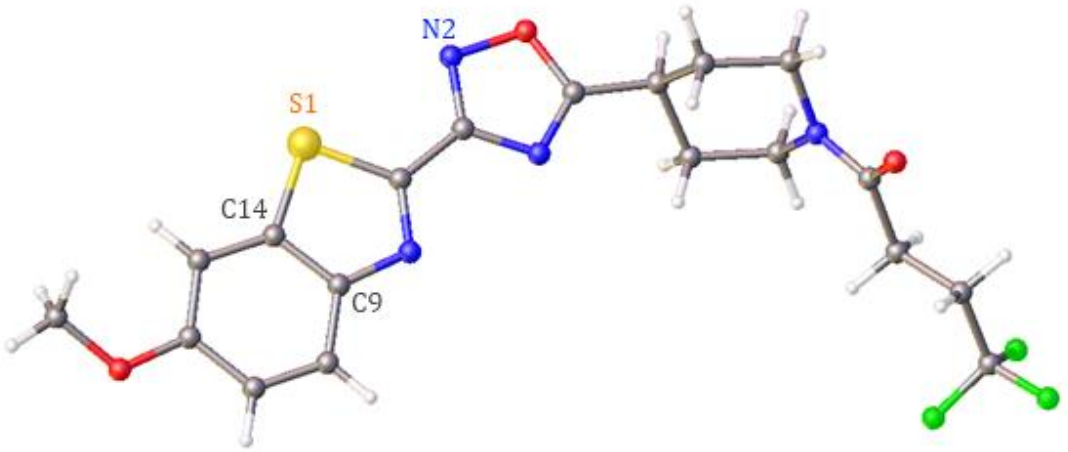


Figure 5.8 | The crystal structure of one molecule of BDM41325.

The structure, from the planar oxadiazole ring (RMSD of 0.003 Å) to the trifluorocarbon group, is similar to previous ligands. The benzothiazole group is practically planar (the phenyl and thiazole rings have a relative angle of 0.2°), while the angle between the thiazole and the oxadiazole is slightly less planar, at 4.9°; the former is more restrained than the latter due to the shared C9-C14 bond. The same potential intramolecular interaction between S1 and N2 is present here as in BDM31343, at a distance of 3.063 Å.

As with the previous ligands, the interaction between the carboxyl oxygen and a hydrogen is present, seen as O1···H5B (2.309 Å), and there is no evidence of π-stacking in the crystal structure. Although BDM41325 the methoxyl group, this atom is not involved in the formation of intermolecular interactions for crystal packing (figure 5.9, overleaf). In fact, despite more opportunities for polar intermolecular bond formation (in S1, N4 and O3), there are only two interactions dictating the packing configuration: F3···H12A' (2.604 Å) and O1···H17A'' (2.335 Å). This is unexpected, and could be a feature of the crystal packing conformation, rather than an inability of these atoms to form such putative interactions.

BDM41325 is the longest of the inhibitors crystallised, and as such is the most likely to cause the opening of the auxiliary pocket in the ligand-binding channel of EthR, if any of the ligands were capable of doing so.

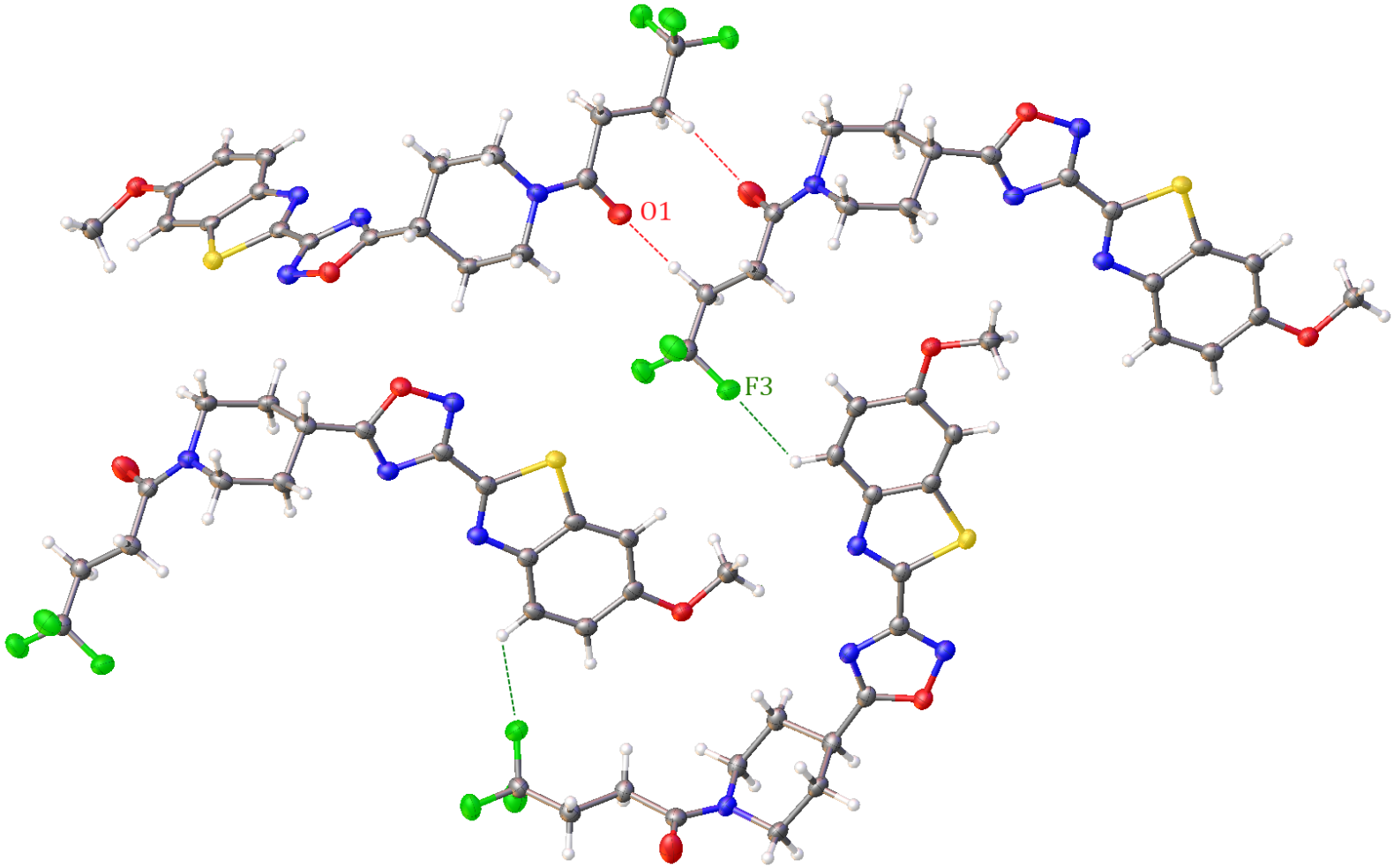


Figure 5.9 | The crystal packing of BDM41325 is distinct from that of the other ligands in that the molecules are less linear. Packing interactions are formed by F3 and O1.

5.5 | Molecular Docking of EthR Inhibitors

Molecular docking was conducted on each ligand with EthR in a ligand-bound conformation with two flexible phenylalanine residues, using *AutoDock Vina.*[10] Details of how this was achieved can be found in chapter three, section 3.3.

5.5.1 | Molecular Docking of BDM31343

The results of the molecular docking algorithm are summarized in table 5.1, below, ranked by calculated binding affinity.

Mode	Affinity (kcal/mol)	Distance RMSD –l.b.	Distance RMSD –u.b.
1	-9.1	0.000	0.000
2	-8.1	1.725	2.229
3	-8.0	2.761	4.953
4	-7.9	1.812	2.760

Table 5.1 The four best binding modes as determined by *AutoDock Vina* for BDM31343. Definitions of lower and upper-bound RMSD can be found in section 3.3.2 in chapter three.

The binding modes suggested are very different, with binding mode 3 showing the highest RMSD values and so being the most distinct from mode 1; this is clearly visible in the graphical representation in figure 5.10.

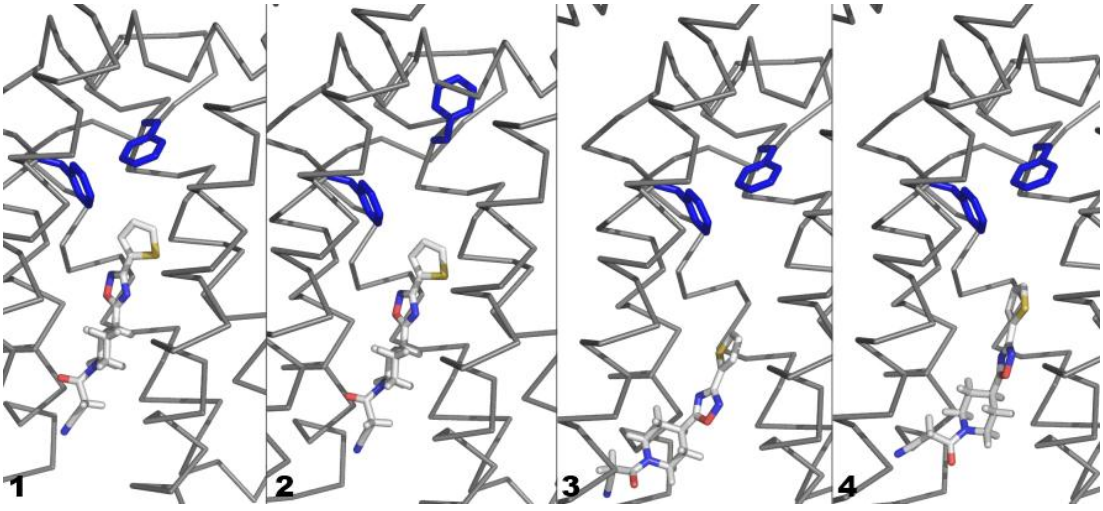


Figure 5.10 | Comparison of the four binding modes as determined by *AutoDock Vina*. Numbers 1-4 in image refer to modes in table 5.1; flexible Phe residues are shown in blue.

Figure 5.11 | Overlay of binding mode one of BDM31343 with the original ligand (BDM31381 and Phe residues shown in black lines) of 3G1M (PDB).

Figure 5.11, right, shows a superimposition of binding mode 1 with the original ligand position in the protein structure. The flexible phenylalanine residues computed, Phe114 and Phe184, adopt very similar positions in both the docked and original crystal structure. The docked ligand occupies a near-identical position as the original ligand, however the orientation has been reversed so the carboxyl tail is towards the aperture of the pocket rather than the top of the channel.

Binding mode 1 is shown below as figure 5.12, with the orientation altered for clarity. The key residues likely to be involved in this binding event are two asparagine side-chains and a phenyl alanine. The ligand is oriented such that the thienyl points to the top of the channel and the nitrile to the aperture. There are few side-chains projecting into the channel at the bottom of the channel; despite the potential for polar contacts between the ligand and protein from the carboxyl, these do not appear to occur. Instead, the thienyl sulphur atom is oriented to interact with the carboxyl of the Asn176 back-bone, while the carboxyl of the Asn176 side-chain is able to interact with N3 of the oxadiazole ring. The oxadiazole ring also forms a potential hydrogen bond with the NH_2 group of the side-chain. There is also some suggestion of π-stacking between the oxadiazole ring and Phe110.

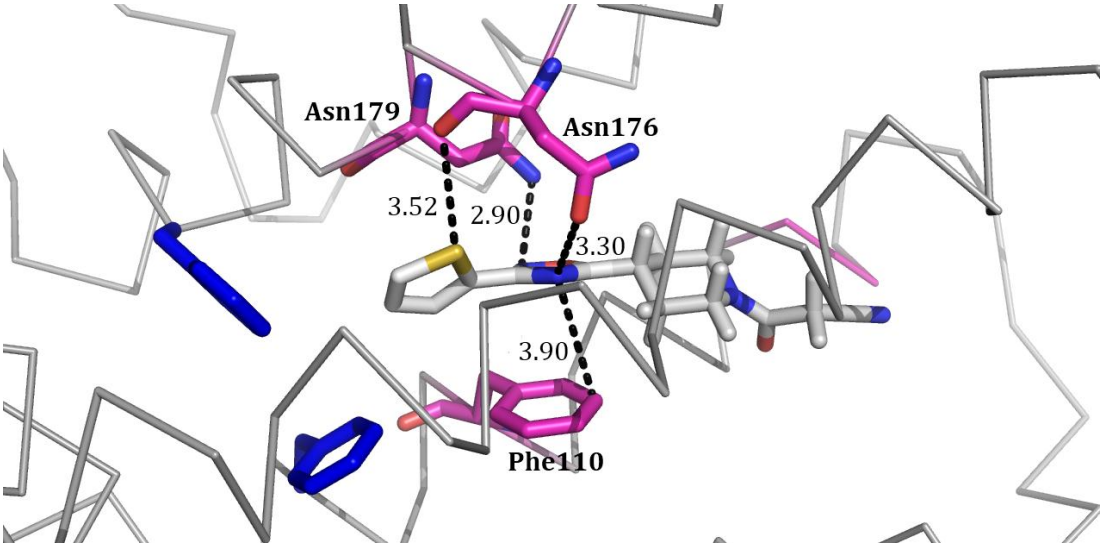


Figure 5.12 | Image of docked BDM31343, with key interacting residues and relative distances displayed.

5.5.2 | Molecular Docking of BDM41420

Though there were two structurally distinct molecules in the crystal structure of BDM41420, only one set of co-ordinates was needed for the docking run as the differences arise from the changes in torsion angles – which are variable in docking.

Mode	Affinity (kcal/mol)	Distance RMSD – l.b.	Distance. RMSD – u.b.
1	-9.4	0.000	0.000
2	-9.0	5.396	8.533
3	-8.8	4.073	7.372
4	-8.4	3.640	7.441

Table 5.2 The four best binding modes as determined by *AutoDock Vina* for BDM41420.

Four binding modes for BDM41420 were calculated with affinities within 1 kcal mol^{-1}. The RMSD values for each binding mode show modes 2, 3 and 4 are statistically, significantly different from mode 1 (table 5.2). The molecular docking results for BDM41420 show very little movement in the phenylalanine residues from the original protein structure, as can be seen in figure 5.13, right; as the shortest ligand, this would indicate there is more opportunity to bind in the channel without expending energy to open up the auxiliary binding pocket.

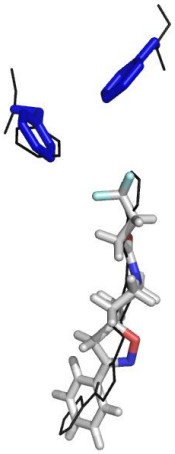


Figure 5.13 | Overlay of the original ligand and Phe residues with binding mode 1.

The differences in binding position shown by the RMSDs in table 5.2 are more clearly represented in figure 5.14, below.

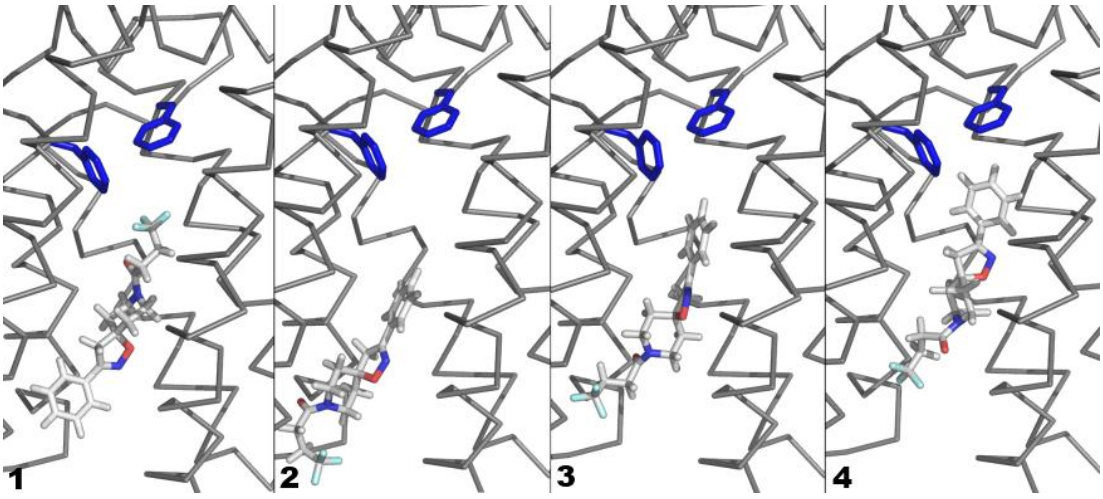


Figure 5.14 | Comparison of the four binding modes as determined by *AutoDock Vina*. Numbers 1-4 in image refer to modes in table 5.2, above.

As figure 5.14 (right) also shows, BDM41420 occupies a position in the best binding mode which corresponds to the position of the original ligand. Despite the overall movement of the ligand throughout the four binding modes, the mid-point of the pocket is always occupied, whether by the phenyl ring in mode 2, the oxazole ring in mode 3 or the piperidyl ring in mode 4; this suggests the conformational change in the overall protein structure is reliant upon occupation of this area by a ligand.

The highest ranked binding mode of BDM41420 according to *Vina*, shown below in figure 5.15, forms both polar and non-polar intermolecular interactions with similar residues to BDM31343 (figure 5.12), but the orientation of the ligand results in these contacts occurring with different ligand atoms. The ligand itself has the phenyl-oxazole end of the ligand at the bottom of the channel, towards the wide, more open area; the trifluorocarbon tail is extended toward the top of the pocket.

The carboxyl oxygen atom of the ligand is in a position to form a hydrogen bond with the NH_2 of the asparagine side chain (Asn179) at a distance of 3.04 Å. The carboxyl of Asn176 is 3.59 Å from N1 in the piperidyl ring, which can be a source of intermolecular interaction. However, the oxazole ring has no discernable side-chain interactions, nor is it able to rotate freely to do so as the oxazole shares C3 with the piperidyl ring. Potential hydrophobic van der Waals interactions can be made between Phe110 and C15 of the carbon tail.

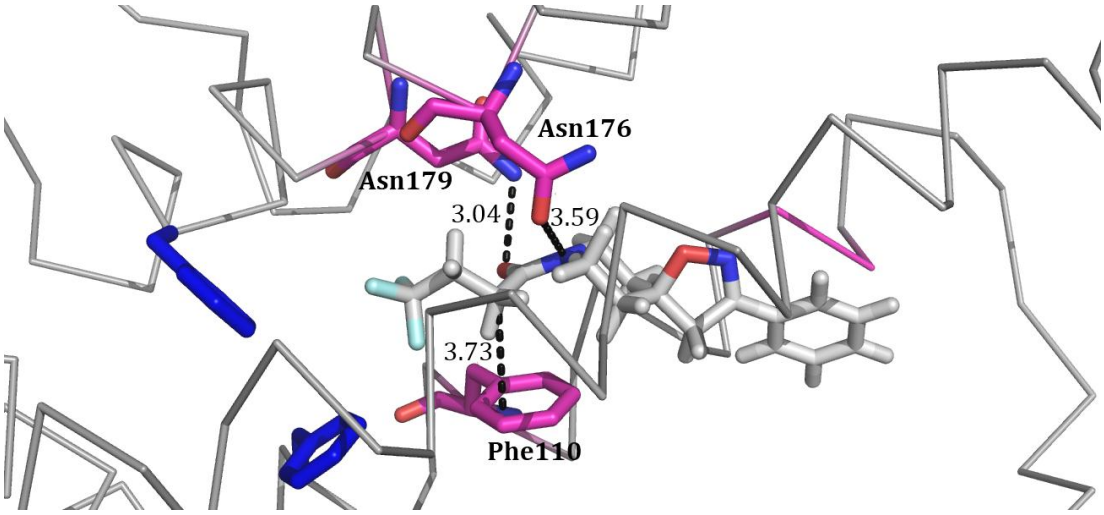


Figure 5.15 | Image of the top docking result for BDM41420, with key interacting residues labelled with relative distances. The orientation has been altered for a clearer image.

5.5.3 | Molecular Docking of BDM41907

The four binding modes suggested share very different orientations within the ligand-binding channel, with the computed flexible Phe114 and Phe184 showing very little movement (figure 5.16).

Mode	Affinity (kcal/mol)	Distance RMSD – l.b.	Distance RMSD – u.b.
1	-10.6	0.000	0.000
2	-10.6	0.647	1.575
3	-9.8	2.592	3.083
4	-9.2	2.001	2.907

Table 5.3 The four best binding modes as determined by *AutoDock Vina* for BDM41907.

The best binding mode (though all four suggested modes share this orientation) shows the ligand with the methyl-sulphoxide group pointed towards the top of the ligand-binding channel. The position of the mode 1 in the ligand-binding channel corresponds directly to the position of the original ligand (figure 5.16).

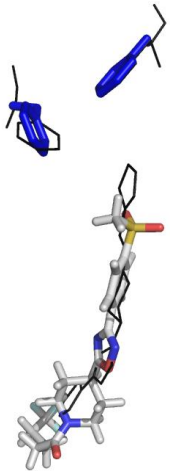


Figure 5.16 | Overlay of the original ligand and mode 1.

Modes 1 and 2 are calculated to have the same binding affinity (-10 kcal mol^{-1}); the sulphoxide groups are in near-identical positions, but the carbon chain is twisted further up in mode 2. Mode 3 shows movement in both the carbon tail and the methyl group, whereas mode 4 shows the ligand slightly higher in the pocket, with alternative torsions to the planar phenyl and oxadiazole rings, as well as the piperidyl ring.

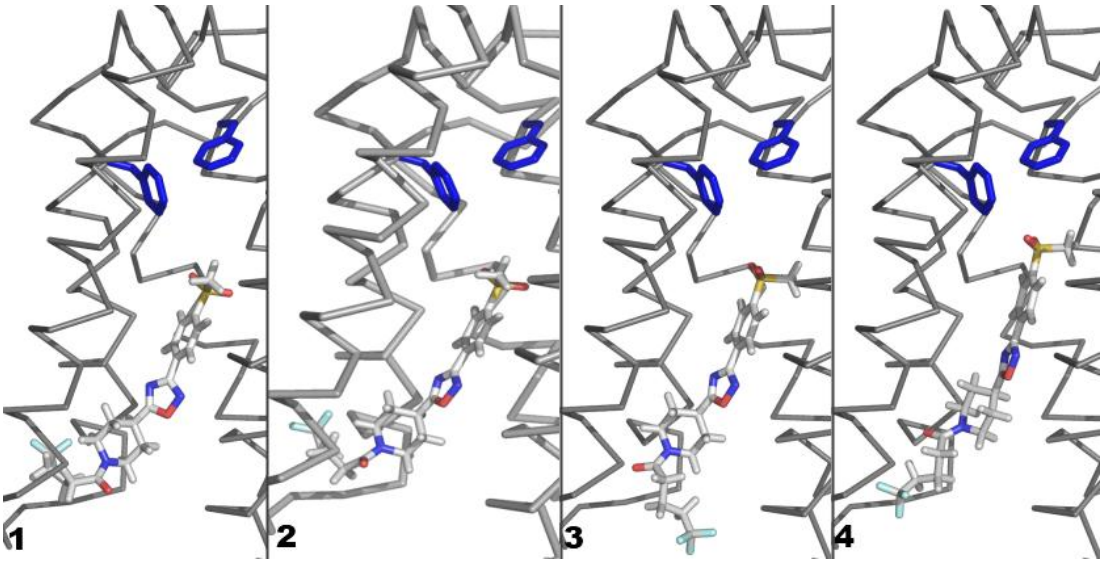


Figure 5.17 | Comparison of the four binding modes as determined by *AutoDock Vina*. Numbers 1-4 in image refer to modes in table 5.3, above.

In the ligand-binding channel, BDM41907 has key interactions with Asn176 and Asn179, which were indicated to have a similar role in the previous two ligands (figure 5.18); here they both potentially form polar interactions with O3 – Asn179 at a distance of 3.15 Å and Asn176 at the slightly shorter distance of 2.98 Å. Phe110, which was shown to have potential interaction with the previous ligands is not indicated to do so here, instead, Tyr148 is implicated. The carboxyl oxygen atom O1 in BDM41907 is in a position to form a hydrogen bond with the OH functional group of the tyrosine residue, an interaction so far unique to this ligand. Despite the size and availability of the trifluorocarbon group, no residues around the bottom of the pocket are within 4-5Å for interaction.

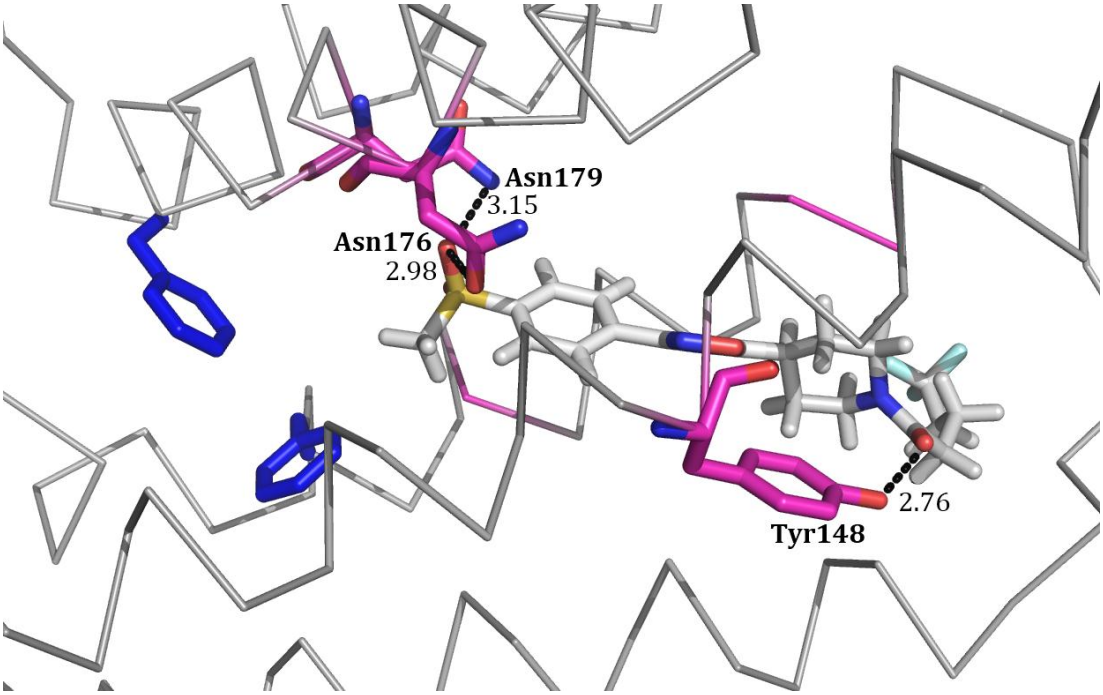


Figure 5.18 | Close-up of the best binding mode in the ligand binding site with key residue interactions highlighted. The orientation has been altered for a clearer view of the residues and distances.

5.5.4 | Molecular Docking of BDM41325

Binding mode 1 is the most "energetically favourable" as it has been ranked the highest by *Vina* with the greatest binding affinity (table 5.4).

Mode	Affinity (kcal/mol)	Distance RMSD – l.b.	Distance RMSD – u.b.
1	-10.4	0.000	0.000
2	-10.2	1.582	2.721
3	-10.1	5.843	10.619
4	-10.0	1.290	1.846

Table 5.4 The four best binding modes as determined by *AutoDock Vina* for BDM41325.

The RMSDs of modes 2, 3 and 4 show they adopt distinctly different positions to mode 1, but it is mode 3 which differs the most. Figures 5.19 (right) and 5.20 (below) show mode 3 is the only mode which involves the movement of Phe184 to allow the ligand to occupy a position higher in the ligand binding channel. Figure 5.19 shows the mode 1 ligand bound at a similar position as the original ligand, with little movement in the phenylalanine residues chosen as flexible for this analysis. This is consistent with the best binding modes seen for the previous three ligands. The overlay of binding mode 3 shows a clear shift in the ligand position, such that it binds deeper in the pocket – for an energetic demand. Though mode 1 is most likely to occur naturally, both are worth further analysis as no ligand so far as been able to produce this effect *in silico*.

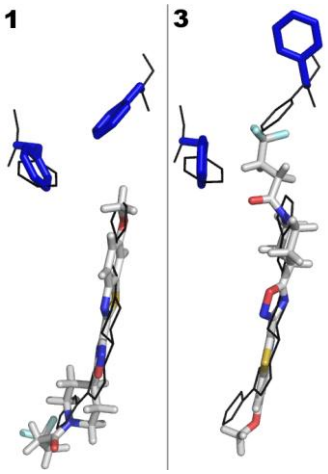


Figure 5.19 | Overlay of the original ligand with binding modes 1 and 3.

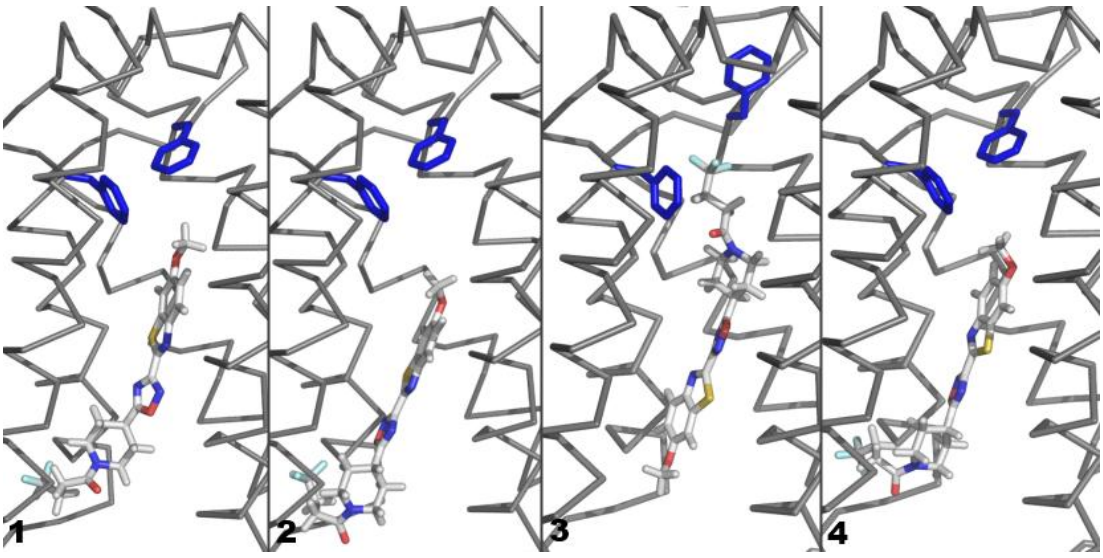


Figure 5.20 | Comparison of the four binding modes as determined by *AutoDock* Vina.

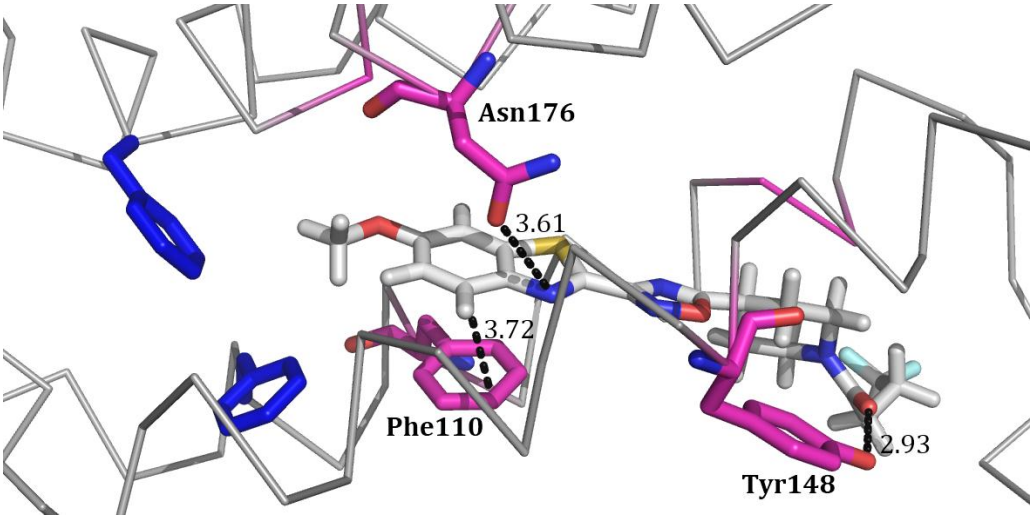


Figure 5.21 | Key residue interactions of binding mode 1.

As BDM41325 binding mode 1 holds a similar position in the ligand-binding channel as previous ligands (specifically BDM41907), similar residue interactions can be deduced (figure 5.21). The carboxyl group of the ligand appears likely to form the same hydrogen bond with Tyr140, and some stacking interaction is seen between Phe110 and the benzyl-thiazole ring system of BDM41325. Though no interactions were suggested between Asn179, unlike in previous ligands, there is potential polar interaction between the carboxyl of the Asn176 side chain and N4 in the thiazole ring. When docked, this ligand adopts a conformation which prevents intramolecular interaction between S1 and N2 (section 5.4).

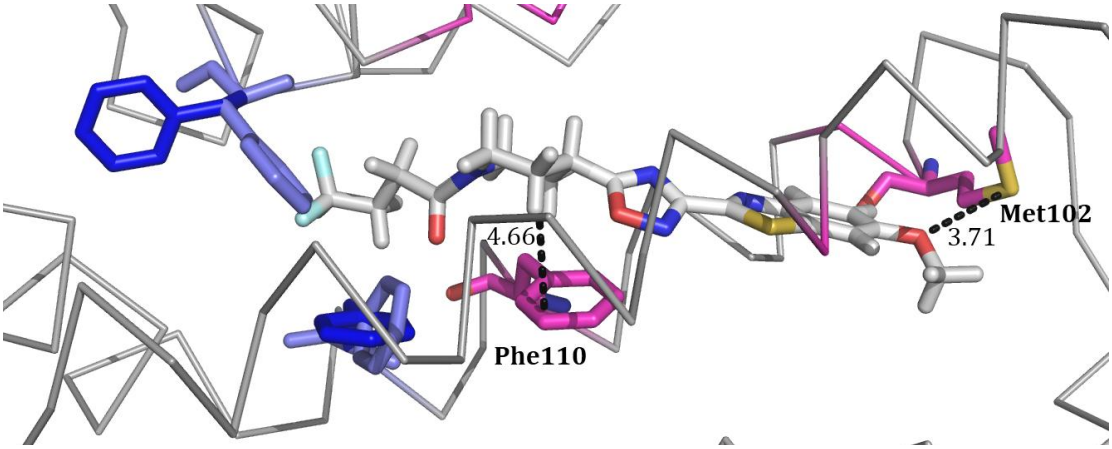


Figure 5.22 | Key residue interactions of binding mode; original Phe orientations shown in pale blue.

Binding mode 3 results in some potential π-stacking between Phe110 and the piperidyl ring, and O3 may form a polar interaction with the sulphur atom of Met102. The inclusion of the original phenylalanine residues shows that this ligand orientation is impossible without the movement; the CF_3 group on the carbon tail would not be able to take up that position. The movement of Phe 184 in particular opens out the channel and results in no discernable polar interactions for this CF3 group within 4 Å.

CHAPTER 6

CONCLUSIONS & MOVING FORWARD

6.1 | Peptoid Precursors

The three-dimensional structure given by crystallographic analysis of the precursor allows inference of the potential structure and intramolecular interactions of the resulting peptoid.

The anionic precursor form exemplifies alternate packing interactions to the protonated precursor form [82]; each structure is distinct but the information taken together gives insight into what interactions the final peptoid could make. For example, the N1 and N2 atoms are available for ionic and hydrogen bond co-ordination. However, the NO_2 group would not be present in the final peptoid, despite using localised charge to form ionic and hydrogen bonds in the anionic structure. Though there is no evidence of π-π stacking interactions within or between molecules, there is still the potential for stacking in a peptoid oligomer, depending upon the size and nature of the groups in the putative oligomer.

The crystallisation and structural characterization of these first two intermediates is a beginning towards creating a range of well-defined peptoids with designed, desired features for bonding and non-bonding interactions. This is a step forward in developing a toolbox of known peptoid conformations; a potential database of monomers which can be linked *in silico* to give theoretical ligands for virtual screening, which would be a powerful tool for future drug discovery.

6.2 | EthR Inhibitors

6.2.1 | Docking Results

Based on the work done on EthR, the crystal packing structure of the inhibitor molecules allows inference of the potential interactions which could be made upon protein-ligand binding;[7,9] therefore, the work detailed here would suggest that the protein-ligand binding is dominated by polar interactions and intermolecular hydrogen bonding.

The S1···N2 interaction seen in the single crystal structures of BDM31343 and BDM41325 were not upheld in the docking conformations derived from *AutoDock Vina*;[10] this could suggest that while this is favourable in the packing structure, there

are overriding intermolecular interactions in the ligand-binding conformations which negate this specific bond. However, this cannot be deduced for certain because although empirical data is used to calculate the force fields, there is no term within *Vina* to define this kind of interaction. Therefore, this interaction being absent from the docking structures could be a result of a more favourable interaction between the ligand and protein, or simply a clear limitation of the docking algorithm.

Though the CF_3 group has a demonstrated role in the packing structure of all three molecules which possess it, they play no role in the docking structures suggested – possibly because this group occupies the wider, open regions of the ligand-binding channel, either at the top or the aperture at the bottom, where no polar residues are within 4 Å. Therefore, while these groups do not show a role in ligand-binding, they could potentially contribute to penetrance into the mycobacterium, a proven role of fluorination.[57]

The only common source of intermolecular interaction displayed in the packing structures of all four structures is the formation of hydrogen bonds using the carboxyl atom O1. In the docking analysis, this atom forms interactions with Asn179 (BDM41420) or Tyr148 (BDM41907 and BDM41325); BDM31343 shows this interaction in the packing, though not in the docking analysis, which suggests that while influential, this interaction is not solely responsible for protein-ligand binding. Alternatively, Asn176 plays an apparently critical role in all four docked structures, interacting with different parts of the molecule in each case. Similarly, in three out of the four structures, Phe110 shows potential non-polar interactions to contribute to the ligand binding. Overall, this indicates that no single interaction can be indicated as being responsible for inducing the inhibitory conformation of the protein; therefore, this is no potential gain in curtailing the ligand structure to specific atoms for interaction.

It is important to note that these residues are not conserved within the TetR-like repressor family. There is a high sequence and three-dimensional structural similarity in the helix-turn-helix domain which binds DNA, which is not unexpected; however, there is no conservation of sequence in the non-DNA-binding domains.[49] The range of functions covered by members of this repressor family require different control mechanisms, which accounts for this difference in ligand binding regions.

Though the opening of the auxiliary pocket (by the movement of Phe114 and Phe184) was demonstrated in the docking of BDM41325, there was a corresponding compromise in binding affinity. The inhibitor with the highest binding affinity (-10 kcal mol^{-1}) was BDM41907, neither the shortest nor the longest ligand of those

crystallised, suggesting that while the auxiliary pocket can be utilized, there is no benefit to a longer ligand with more atoms included for polar interactions.

6.2.2 | Evaluation of *Autodock Vina*

Each of the *Vina* runs is informative, even if only to give context to the highest ranked binding mode suggested; this is clearly demonstrated by the results for BDM41325, which showed there is an energetic compromise for the opening of the auxiliary binding pocket. For the other docking runs, where the differences were shown to be more subtle but no less significant, the results demonstrate that statistically minor changes in binding position can have a drastic effect on the binding affinity. Confidence in the highest ranked result can be drawn from the observation that the second-, third- and fourth-ranked results are feasible and still presented in the ligand-binding channel.

In my experience *Vina* is a powerful tool for molecular docking with a variety of user-definable parameters. The program can be adjusted to be comprehensive or fast depending upon the needs of the user, and repeated runs with the same data and parameters gives identical results; this allows the user to be confident in the derived values.

However, there are limitations to the program; the interface is not user-friendly and the preparation of the ligand and protein files is not trivial – tutorials are available but cover only the basic set-up and do not include instruction on creating a run with flexible side-chains. Also, it is insufficient to simply prepare the files, define a search grid and trust the data presented. The investigations into the program conducted in the course of this project indicated that the default exhaustiveness may be increased and give vastly different results; similarly the definition of the search grid was crucial. Though an increased exhaustiveness and a search grid encompassing the entire protein caused the run time to increase, the results were far more accurate and were reproducible. Iterative runs were required to find a compromise on how many flexible residues could be used in the docking in a reasonable length of time; a target of thirty minutes was used to ensure that the results could be achieved quickly, accurately and would be reproducible.

The most pressing limitation to *Vina* is that, although based upon empirical data, some parameters in the model do not have a corresponding term within the algorithm to include in the docking run. This was shown with the S1···N2 interaction in BDM31343 and BDM41325. This docking work bears repeating in a program with that additional coding to have absolute confidence in the results from *Vina*.

In docking studies, small molecule crystal structures are essential. Three-dimensional co-ordinates are an absolute requirement for the process, and while these can be achieved through *ab initio* computational approaches such as AMBER [89], a theoretical construct is no substitute from empirical data – particularly when the docking algorithm calculates force fields based upon empirical data. This approach is especially warranted when the target protein is difficult to crystallise or access to the protein is not possible; all that is needed is the PDB file and a small molecule structure for any scientist to evaluate the binding of a putative ligand. This has the added advantage of opening up research, particularly in the drug design field where one breakthrough can make a vast difference to the development process.

6.2.3 | Moving Forward

From here, the way forward is clear: biophysical characterization of the highest scoring inhibitors. This would require continued crystallisation of inhibitors in isolation, including those provided for this project, and the same analysis by docking; co-crystallisation of these inhibitor-protein complexes would confirm the accuracy of the docking results to give proven confidence to the software-derived results.

This approach is not limited to EthR, or even the TetR family of repressors; steric hindrance as a mode of control can be adapted to a variety of regulatory proteins. Databases of small molecule structures such as the Cambridge Crystallographic Database, maintained by the CCDC (www.ccdc.cam.ac.uk), could be used to virtually screen ligands for target proteins, either as activators or inhibitors. DNA-binding proteins are a profound target for combating pathogens, as they have a direct effect on the life cycle or regulatory processes of the organism itself. A well-chosen and defined target would benefit from the approach detailed in this thesis.

Structure-based drug design, be it for a known target (EthR inhibitors) or for the purpose of creating lead fragments or adaptable monomers (the peptoid precursors) is a systematic and iterative method for the development of new drugs. X-ray crystallography, while time and resource consuming, is a powerful method which feeds into this design process in ways no other structure-determination method is capable. Together, they allow us to find novel ways to combat long-standing known diseases, and can give us an arsenal against future and developing health crises.

REFERENCES

1. Simon, R.J. *et al.* Peptoids: a modular approach to drug discovery. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 9367-71 (1992).

2. Udugamasooriya, D.G., Dineen, S.P., Brekken, R.A. & Kodadek, T. A peptoid "antibody surrogate" that antagonizes VEGF receptor 2 activity. *Journal of the American Chemical Society* **130**, 5744-52 (2008).

3. Wender, P.A. *et al.* The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13003-8 (2000).

4. Fowler, S.A. & Blackwell, H.E. Structure-function relationships in peptoids: recent advances toward deciphering the structural requirements for biological function. *Organic & Biomolecular Chemistry* **7**, 1508-24 (2009).

5. Dye, C. & Williams, B.G. The population dynamics and control of tuberculosis. *Science* **328**, 856-61 (2010).

6. Koul, A., Arnoult, E., Lounis, N., Guillemont, J. & Andries, K. The challenge of new drug discovery for tuberculosis. *Nature* **469**, 483-490 (2011).

7. Frénois, F., Engohang-Ndong, J., Locht, C., Baulard, A.R. & Villeret, V. Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. *Molecular Cell* **16**, 301-7 (2004).

8. Frénois, F., Baulard, A.R. & Villeret, V. Insights into mechanisms of induction and ligands recognition in the transcriptional repressor EthR from Mycobacterium tuberculosis. *Tuberculosis* **86**, 110-4 (2006).

9. Willand, N. *et al.* Synthetic EthR inhibitors boost antituberculous activity of ethionamide. *Nature Medicine* **15**, 537–544 (2009).

10. Trott, O. & Olson, A.J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multi-threading. *Journal of Computational Chemistry* **31**, 455-461 (2010).

11. Zuckermann, R. & Kodadek, T. Peptoids as potential therapeutics. *Current Opinion in Molecular Therapeutics* **11**, 299-307 (2009).

12. Kwon, Y.-U. & Kodadek, T. Quantitative evaluation of the relative cell permeability of peptoids and peptides. *Journal of the American Chemical Society* **129**, 1508-9 (2007).

13. Romani, B., Engelbrecht, S. & Glashoff, R.H. Functions of Tat: the versatile protein of human immunodeficiency virus type 1. *Journal of General Virology* **91**, 1-12 (2010).

14. Chen, G. *et al.* HIV-Tat-mediated delivery of an LPTS functional fragment inhibits telomerase activity and tumorigenicity of hepatoma cells. *Gastroenterology* **140**, 332-43 (2011).

15. Gobbo, M. *et al.* Substitution of the arginine/leucine residues in apidaecin Ib with peptoid residues: effect on antimicrobial activity, cellular uptake, and proteolytic degradation. *Journal of Medicinal Chemistry* **52**, 5197-206 (2009).

16. Xiao, X., Yu, P., Lim, H.-S., Sikder, D. & Kodadek, T. Design and synthesis of a cell-permeable synthetic transcription factor mimic. *Journal of Combinatorial Chemistry* **9**, 592-600 (2007).

17. Shibuya, M. Differential roles of vascular endothelial growth factor receptor-1 and receptor-2 in angiogenesis. *Journal of Biochemistry and Molecular Biology* **39**, 469-478 (2006).

18. Hara, T., Durell, S.R., Myers, M.C. & Appella, D.H. Probing the structural requirements of peptoids that inhibit HDM2-p53 interactions. *Journal of the American Chemical Society* **128**, 1995-2004 (2006).

19. Horwell, D.C. The "peptoid" approach to the design of non-peptide, small molecule agonists and antagonists of neuropeptides. *Trends in Biotechnology* **13**, 132-4 (1995).

20. Reddy, M.M. *et al.* Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening. *Cell* **144**, 132-42 (2011).

21. Wu, C.W. *et al.* Structural and spectroscopic studies of peptoid oligomers with alpha-chiral aliphatic side chains. *Journal of the American Chemical Society* **125**, 13525-30 (2003).

22. Chongsiriwatana, N.P. *et al.* Peptoids that mimic the structure, function, and mechanism of helical antimicrobial peptides. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2794-9 (2008).

23. WHO *World Heath Statistics 2011*. *World* (2010).

24. Takayama, K., Wang, C. & Besra, G.S. Pathway to Synthesis and Processing of Mycolic Acids in Mycobacterium tuberculosis. *Clinical Microbiology Reviews* **18**, 81-101 (2005).

25. Russell, D.G., Barry, C.E. & Flynn, J.L. Tuberculosis: what we don't know can, and does, hurt us. *Science* **328**, 852-6 (2010).

26. Flynn, J.L. & Chan, J. What's good for the host is good for the bug. *Trends in Microbiology* **13**, 98-102 (2005).

27. Russell, D.G., Cardona, P.-J., Kim, M.-J., Allain, S. & Altare, F. Foamy macrophages and the progression of the human tuberculosis granuloma. *Nature Immunology* **10**, 943-8 (2009).

28. Parrish, N., Dick, P. & Bishai, W. Mechanisms of Latency in Mycobacterium tuberculosis. *Trends in Microbiology* **6**, 107-112 (1998).

29. Riska, P.F. & Carleton, S. Latent tuberculosis: models, mechanisms, and novel prospects for eradication. *Seminars in Pediatric Infectious Diseases* **13**, 263-72 (2002).

30. WHO *The Global Plan to Stop TB 2011-2015*. 26-33 (2010).

31. TB Alliance Ethionamide. *Tuberculosis* **88**, 106-8 (2008).

32. Winder, F.G. & Collins, P.B. Inhibition by isoniazid of synthesis of mycolic acids in Mycobacterium tuberculosis. *Journal of General Microbiology* **63**, 41-8 (1970).

33. Takayama, K., Wang, L. & David, H.L. Effect of isoniazid on the in vivo mycolic acid synthesis, cell growth, and viability of Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy* **2**, 29-35 (1972).

34. Marrakchi, H., Lanéelle, G. & Quémard, A. InhA, a target of the antituberculous drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS-II. *Microbiology* **146**, 289-96 (2000).

35. Vilchèze, C. *et al.* Inactivation of the inhA-encoded fatty acid synthase II (FASII) enoyl-acyl carrier protein reductase induces accumulation of the FASI end products and cell lysis of Mycobacterium smegmatis. *Journal of Bacteriology* **182**, 4059-67 (2000).

36. Banerjee, A. *et al.* inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. *Science* **263**, 227-230 (1994).

37. Quémard, A. *et al.* Enzymatic characterization of the target for isoniazid in Mycobacterium tuberculosis. *Biochemistry* **34**, 8235-41 (1995).

38. Dessen, A., Quémard, A., Blanchard, J.S., Jacobs, W.R. & Sacchettini, J.C. Crystal structure and function of the isoniazid target of Mycobacterium tuberculosis. *Science* **267**, 1638-1641 (1995).

39. Baulard, A.R. *et al.* Activation of the pro-drug ethionamide is regulated in mycobacteria. *Journal of Biological Chemistry* **275**, 28326-31 (2000).

40. Vannelli, T.A., Dykman, A. & Ortiz de Montellano, P.R. The antituberculosis drug ethionamide is activated by a flavoprotein monooxygenase. *Journal of Biological Chemistry* **277**, 12824-9 (2002).

41. Fraaije, M.W., Kamerbeek, N.M., Heidekamp, A.J., Fortin, R. & Janssen, D.B. The prodrug activator EtaA from Mycobacterium tuberculosis is a Baeyer-Villiger monooxygenase. *Journal of Biological Chemistry* **279**, 3354-60 (2004).

42. Hanoulle, X. *et al.* Selective intracellular accumulation of the major metabolite issued from the activation of the prodrug ethionamide in mycobacteria. *Journal of Antimicrobial Chemotherapy* **58**, 768-72 (2006).

43. DeBarber, A.E., Mdluli, K., Bosman, M., Bekker, L.G. & Barry, C.E. Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9677-82 (2000).

44. Wang, F. *et al.* Mechanism of thioamide drug action against tuberculosis and leprosy. *Journal of Experimental Medicine* **204**, 73-8 (2007).

45. Engohang-Ndong, J. *et al.* EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. *Molecular Microbiology* **51**, 175-188 (2003).

46. Camus, J.-C., Pryor, M.J., Médigue, C. & Cole, S.T. Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. *Microbiology* **148**, 2967-73 (2002).

47. Brennan, R.G. The winged-helix DNA-binding motif: Another helix-turn-helix takeoff. *Cell* **74**, 773-776 (1993).

48. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Structural Biology* **7**, 215-9 (2000).

49. Ramos, J.L. *et al.* The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews* **69**, 326-356 (2005).

50. Dover, L.G. *et al.* Crystal structure of the TetR/CamR family repressor Mycobacterium tuberculosis EthR implicated in ethionamide resistance. *Journal of Molecular Biology* **340**, 1095-105 (2004).

51. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica: Section D. Biological Crystallography* **66**, 486-501 (2010).

52. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Structural Biology* **7**, 215-9 (2000).

53. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **46**, 3-26 (2001).

54. Kolb, H.C., Finn, M.G. & Sharpless, K.B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie (International ed. in English)* **40**, 2004-2021 (2001).

55. Willand, N. *et al.* Exploring drug target flexibility using in situ click chemistry: application to a mycobacterial transcriptional regulator. *ACS Chemical Biology* **5**, 1007-13 (2010).

56. Flipo, M. *et al.* Ethionamide Boosters: Synthesis, Biological Activity, and Structure-Activity Relationships of a Series of 1,2,4-Oxadiazole EthR Inhibitors. *Journal of Medicinal Chemistry* **54**, 2994-3010 (2011).

57. Rentmeister, A., Arnold, F.H. & Fasan, R. Chemo-enzymatic fluorination of unactivated organic compounds. *Nature Chemical Biology* **5**, 26-8 (2009).

58. Marsh, E.N.G., Buer, B.C. & Ramamoorthy, A. Fluorine--a new element in the design of membrane-active peptides. *Molecular Biosystems* **5**, 1143-7 (2009).

59. Sköld, M. & Behar, S.M. The role of group 1 and group 2 CD1-restricted T cells in microbial immunity. *Microbes and Infection* **7**, 544-51 (2005).

60. IUCr Online Dictionary of Crystallography. at <http://reference.iucr.org/dictionary/Main_Page>

61. Woolfson, M.M. *An Introduction to X-Ray Crystallography*. (Cambridge University Press: Cambridge, 1997).

62. IUCr IUCr Definitions: Miller Indices. (2009).at <http://reference.iucr.org/dictionary/Miller_indices>

63. Woolfson, M.M. Direct methods in crystallography. *Reports on Progress in Physics* **34**, 369-434 (1971).

64. Patterson, A. A Fourier series method for the determination of the components of interatomic distances in crystals. *Physical Review* **46**, 372-376 (1934).

65. Hauptman, H. Direct Methods and Anomalous Dispersion (Nobel Lecture). *Angewandte Chemie (International ed. in English)* **25**, 603-613 (1986).

66. Sheldrick, G.M. Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallographica: Section A. Foundations of Crystallography* **46**, 467-473 (1990).

67. Glusker, J.P., Lewis, M. & Rossi, M. *Crystal Structure Analysis for Chemists and Biologists*. 297-301 (VCH Publishers, Inc.: New York, 1994).

68. Glusker, J.P., Lewis, M. & Rossi, M. *Crystal Structure Analysis for Chemists and Biologists*. 397-401 (VCH Publishers, Inc.: New York, 1994).

69. Sheldrick, G.M. A short history of SHELX. *Acta Crystallographica: Section A. Foundations of Crystallography* **64**, 112-22 (2008).

70. Jones, P.G. Crystal structure determination: a critical view. *Chemical Society Reviews* **13**, 157 (1984).

71. Watkin, D. Structure refinement: some background theory and practical strategies. *Journal of Applied Crystallography* **41**, 491-522 (2008).

72. Laaksonen, A., Talanquer, V. & Oxtoby, D.W. Nucleation: Measurements, Theory, and Atmospheric Applications. *Annual Review of Physical Chemistry* **46**, 489-524 (1995).

73. Glusker, J.P., Lewis, M. & Rossi, M. *Crystal Structure Analysis for Chemists and Biologists*. 40-43 (VCH Publishers, Inc.: New York, 1994).

74. Rowland, R.S. & Taylor, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *Journal of Physical Chemistry* **100**, 7384-7391 (1996).

75. Chetina, O.V. How to Grow Single Crystals for X-ray Analysis by Solution Crystallisation. *Unpublished Research Guide* (2009).

76. Jones, P.G. Crystal Growing. *Chemistry in Britain* **17**, 222-225 (1981).

77. Hope, H. Crystallography of Biological Macromolecules at Ultra-Low Temperature. *Annual Review of Biophysics and Biophysical Chemistry* **19**, 107-26 (1990).

78. Bruker SMART, SAINT and XPREP. (2000).

79. Dolomanov, O.V., Bourhis, L.J., Gildea, R.J., Howard, J.A.K. & Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography* **42**, 339-341 (2009).

80. Gohlke, H. & Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie (International ed. in English)* **41**, 2644-76 (2002).

81. van Gunsteren, W.F. *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angewandte Chemie (International ed. in English)* **45**, 4064-92 (2006).

82. Colgin, N., Tatum, N.J., Pohl, E., Cobb, S.L. & Sandford, G. Synthesis and molecular structure of a perfluorinated pyridyl carbanion. *Journal of Fluorine Chemistry* **133**, 33-37 (2011).

83. Serrano-Becerra, J.M., Hernández-Ortega, S., Morales-Morales, D. & Valdés-Martínez, J. Bottom-up design and construction of a non-centrosymmetric network through π–π stacking interactions. *CrystEngComm* **11**, 226 (2009).

84. Rybalova, T. & Bagryanskaya, I.Y. CF... pi, F... H, and F... F intermolecular interactions and F-aggregation: Role in crystal engineering of fluoroorganic compounds. *Journal of Structural Chemistry* **50**, 741–753 (2009).

85. Thalladi, V.R. *et al.* C–H···F Interactions in the Crystal Structures of Some Fluorobenzenes. *Journal of the American Chemical Society* **120**, 8702-8710 (1998).

86. Anderson, K., Afarinkia, K., Yu, H., Goeta, A. & Steed, J.W. When Z'= 2 Is Better than Z'= 1 Supramolecular Centrosymmetric Hydrogen-Bonded Dimers in Chiral Systems. *Crystal Growth & Design* **6**, 2109-2113 (2006).

87. Collins, A. Distribution of molecular pairs in Z'=2 structures. *Acta Crystallographica: Section B. Structural Science* **62**, 897-911 (2006).

88. Pidcock, E. Spatial arrangement of molecules in homomolecular Z'=2 structures. *Acta Crystallographica: Section B. Structural Science* **62**, 268-79 (2006).

89. Pearlman, D.A. *et al.* AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **91**, 1–41 (1995).

APPENDIX

A-I | Anionic Peptoid Precursor

Table A1 | Crystallographic Data for the Anionic Peptoid Precursor

	Anionic Peptoid Precursor
Empirical Formula	$C_{11}H_4N_3O_4F_8Na$
Formula weight	382.14
Temperature / K	99.69
Crystal System	monoclinic
Space Group	$P2_1/n$
a / Å	5.635(2)
b / Å	16.467(4)
c / Å	16.432(4)
α / °	90.00
β / °	99.29(2)
γ / °	90.00
Volume / Å^3	1504.8(7)
Z	4
ρ_{calc} / mg mm^3	1.687
μ /mm^{-1}	1.925
2θ Range for Data Collection /°	7.66 – 132.86
Index Ranges	$-6 \leq h \leq 5, -13 \leq k \leq 19, -19 \leq l \leq 14$
Reflections Collected	6216
Independent Reflections	2304 [R_{int} = 0.0300]
Data/Restraints/Parameters	2304/4/256
Goodness-of-fit on F^2	1.061
Final R1 Index [I >2σ(I)]	0.0366
Final wR2 Index [I >2σ(I)]	0.0974
Final R1 Index [all data]	0.0445
Final wR2 Index [all data]	0.1023
Largest diff. peak/hole / e Å^{-3}	0.250/-0.224

Figure A1 | Labelled Anionic Peptoid Precursor

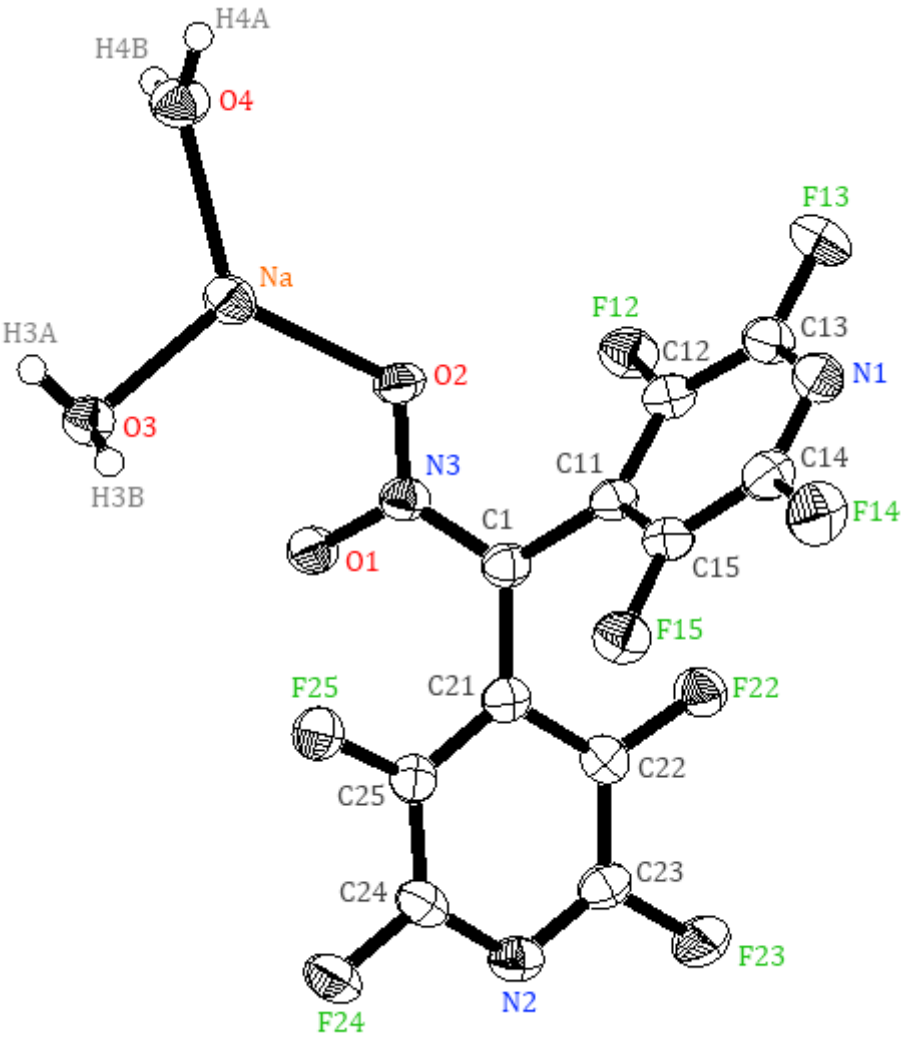


Table A2 | Bond lengths for the Anionic Peptoid Precursor

Atom	Atom	Length / Å	Atom	Atom	Length / Å
O2	Na	2.356(8)	C14	C15	1.373(16)
Na	O3	2.289(8)	C15	F15	1.342(12)
Na	O4	2.241(8)	C15	C11	1.390(14)
N3	O2	1.288(11)	C1	C21	1.443(15)
N3	O1	1.282(11)	C21	C22	1.386(13)
C1	N3	1.338(13)	C22	F22	1.358(11)
C1	C11	1.461(14)	C22	C23	1.386(16)
C11	C12	1.384(15)	C23	F23	1.349(11)
C12	C13	1.391(15)	C23	N2	1.288(14)
C12	F12	1.350(12)	N2	C24	1.323(14)
C13	F13	1.331(14)	C24	F24	1.346(12)
C13	N1	1.344(15)	C24	C25	1.354(16)
N1	C14	1.290(15)	C25	F25	1.339(11)
C14	F14	1.339(12)	C25	C21	1.412(14)

Table A3 | Selected bond angles for the Anionic Peptoid Precursor

Atom	**Atom**	**Atom**	**Angle/ °**	**Atom**	**Atom**	**Atom**	**Angle /°**
O4	Na	O3	118.9(9)	F14	C14	C15	119.2(10)
O4	Na	O2	116.9(4)	C14	C15	F15	119.5(9)
O3	Na	O2	122.4(3)	C14	C15	C11	120.4(10)
Na	O2	N3	121.7(6)	F15	C15	C11	120.1(10)
O2	N3	O1	120.7(8)	C15	C11	C12	114.9(10)
O2	N3	C1	120.3(9)	C1	C21	C22	122.0(9)
O1	N3	C1	119.8(9)	C21	C22	C23	120.9(9)
N3	C1	C11	118.0(10)	F22	C22	C23	119.8(9)
N3	C1	C21	120.3(9)	C22	C23	F23	117.6(9)
C11	C1	C21	121.6(8)	F23	C23	N2	118.1(9)
C1	C11	C12	124.0(9)	C23	N2	C24	115.9(9)
C11	C12	F12	121.0(10)	N2	C24	F24	115.0(9)
C11	C12	C13	120.5(10)	N2	C24	C25	125.5(10)
F13	C13	N1	117.6(9)	C24	C25	F25	120.2(9)
C12	C13	N1	122.7(10)	C24	C25	C21	119.6(9)
C13	N1	C14	116.4(10)	C25	C21	C1	124.0(9)
N1	C14	F14	115.6(10)	C25	C21	C22	113.9(9)

A-II | Protonated Peptoid Precursor

Table A4 | Crystallographic Data for the Protonated Peptoid Precursor

	Protonated Peptoid Precursor
Empirical Formula	$C_{11}HN_3O_2F_8$
Formula weight	359.15
Temperature / K	99.81
Crystal System	monoclinic
Space Group	$P2_1/n$
a / Å	8.552(2)
b / Å	12.540(2)
c / Å	11.831(2)
α / °	90.00
β / °	107.965(5)
γ / °	90.00
Volume / $Å^3$	1206.9(4)
Z	4
ρ_{calc} / mg mm^3	1.977
μ /mm^{-1}	2.022
2θ Range for Data Collection /°	10.56 – 133.28
Index Ranges	$-10 \leq h \leq 10, -14 \leq k \leq 14, -14 \leq l \leq 13$
Reflections Collected	15399
Independent Reflections	2068 [R_{int} = 0.0560]
Data/Restraints/Parameters	2068/0/217
Goodness-of-fit on F^2	1.082
Final R1 Index [I >2σ(I)]	0.0532
Final wR2 Index [I >2σ(I)]	0.1540
Final R1 Index [all data]	0.0561
Final wR2 Index [all data]	0.1570
Largest diff. peak/hole / e $Å^{-3}$	0.531/-0.447

Figure A2 | Labelled Protonated Peptoid Precursor

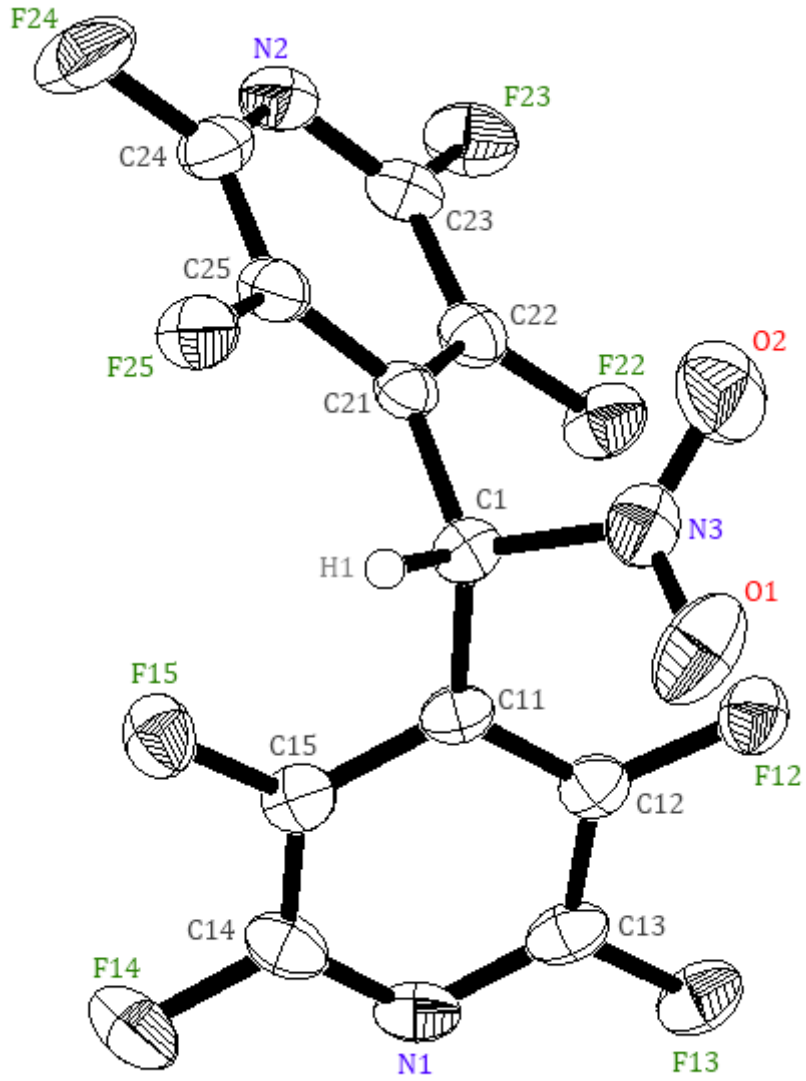


Table A5 | Bond lengths for the Protonated Peptoid Precursor

Atom	Atom	Length / Å	Atom	Atom	Length / Å
O2	N3	1.246(4)	C15	C11	1.389(4)
O1	N3	1.200(4)	C1	C21	1.514(4)
N3	C1	1.487(4)	C21	C22	1.379(4)
C1	C11	1.503(4)	C22	F22	1.336(3)
C11	C12	1.372(4)	C22	C23	1.380(4)
C12	F12	1.331(3)	C23	F23	1.325(3)
C12	C13	1.377(5)	C23	N2	1.307(4)
C13	F13	1.330(4)	N2	C24	1.311(4)
C13	N1	1.297(4)	C24	F24	1.334(4)
N1	C14	1.295(4)	C24	C25	1.376(4)
C14	F14	1.333(3)	C25	F25	1.332(3)
C14	C15	1.371(4)	C25	C21	1.380(4)
C15	F15	1.337(3)			

Table A6 | Bond angles for the Protonated Peptoid Precursor

Atom	Atom	Atom	Angle /°	Atom	Atom	Atom	Angle /°
O1	N3	O2	123.9(3)	F15	C15	C11	120.5(3)
O1	N3	C1	120.4(3)	C15	C11	C12	115.7(3)
O2	N3	C1	115.6(3)	C15	C11	C1	123.8(2)
N3	C1	C11	111.4(2)	C1	C21	C22	125.3(3)
N3	C1	C21	111.9(2)	C21	C22	F22	121.0(2)
C11	C1	C21	113.3(2)	F22	C22	C23	119.3(2)
C1	C11	C12	120.5(3)	C21	C22	C23	119.7(3)
C11	C12	F12	120.8(3)	C22	C23	F23	118.9(3)
C11	C12	C13	119.2(3)	C22	C23	N2	123.8(3)
F12	C12	C13	119.9(3)	F23	C23	N2	117.3(3)
C12	C13	F13	118.2(3)	C23	N2	C24	116.8(3)
F13	C13	N1	117.4(3)	N2	C24	F24	116.4(3)
C12	C13	N1	124.5(3)	N2	C24	C25	124.0(3)
C13	N1	C14	117.1(3)	F24	C24	C25	119.6(3)
N1	C14	F14	117.2(3)	C24	C25	F25	120.5(3)
N1	C14	C15	123.7(3)	C24	C25	C21	119.6(3)
F14	C14	C15	119.1(3)	F25	C25	C21	119.9(3)
C14	C15	F15	119.7(3)	C25	C21	C1	118.6(2)
C14	C15	C11	119.8(3)	C25	C21	C22	116.1(3)

A-III | BDM31343

Table A7 | Crystallographic Data for BDM31343

	BDM31343
Empirical Formula	$C_{14}H_{14}N_4O_2S$
Formula weight	302.35
Temperature / K	99.80
Crystal System	monoclinic
Space Group	$P2_1/c$
a / Å	6.834(2)
b / Å	16.923(3)
c / Å	12.531(2)
α / °	90.00
β / °	104.882(5)
γ / °	90.00
Volume / $Å^3$	1400.6(5)
Z	4
ρ_{calc} / mg mm^3	1.434
μ /mm^{-1}	2.153
2θ Range for Data Collection /°	8.98 – 133.28
Index Ranges	$-8 \leq h \leq 8, -19 \leq k\ 20, -14 \leq l \leq 14$
Reflections Collected	23371
Independent Reflections	2429 [R_{int} = 0.0320]
Data/Restraints/Parameters	2429/0/190
Goodness-of-fit on F^2	1.175
Final R1 Index [I >2σ(I)]	0.0316
Final wR2 Index [I >2σ(I)]	0.0916
Final R1 Index [all data]	0.0326
Final wR2 Index [all data]	0.1003
Largest diff. peak/hole / e $Å^{-3}$	0.286/-0.393

Figure A3 | Labelled BDM31343

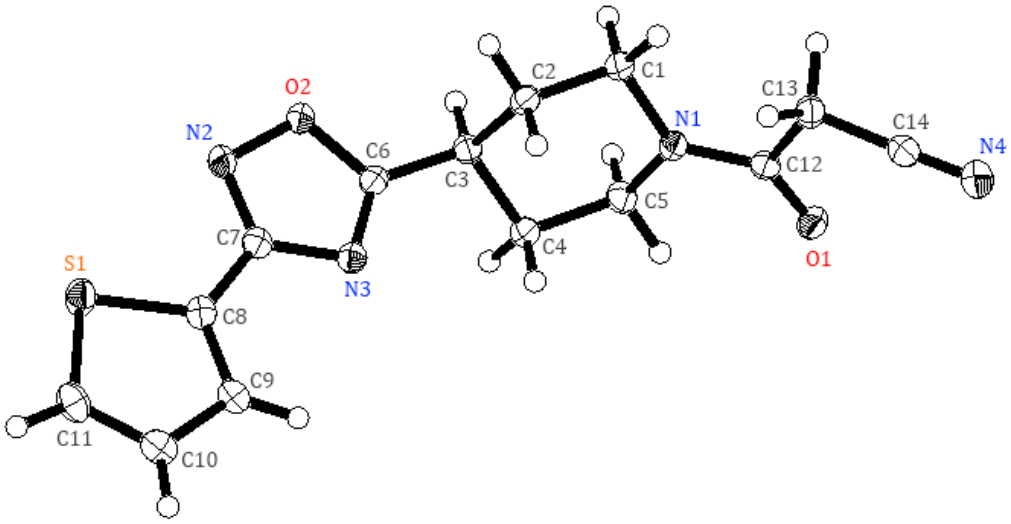


Table A8 | Bond lengths for BDM31343

Atom	Atom	Length /Å	Atom	Atom	Length /Å
S1	C8	1.727(3)	C3	C2	1.532(2)
C8	C9	1.379(2)	C2	C1	1.525(2)
C9	C10	1.425(2)	C1	N1	1.472(3)
C10	C11	1.353(3)	N1	C5	1.465(2)
C11	S1	1.714(9)	C5	C4	1.517(2)
C8	C7	1.450(2)	C4	C3	1.537(2)
C7	N2	1.310(2)	N1	C12	1.343(2)
N2	O2	1.417(4)	C12	O1	1.229(2)
O2	C6	1.351(2)	C12	C13	1.532(2)
C6	N3	1.299(2)	C13	C14	1.463(2)
N3	C7	1.383(2)	C14	N4	1.145(2)
C6	C3	1.485(2)			

Table A9 | Bond angles for BDM31343

Atom	Atom	Atom	Angle /°	Atom	Atom	Atom	Angle /°
S1	C8	C9	111.7(13)	C6	C3	C2	112.5(13)
C8	C9	C10	111.4(16)	C6	C3	C4	109.5(13)
C9	C10	C11	113.3(16)	C4	C3	C2	109.6(13)
C10	C11	S1	112.1(14)	C3	C4	C5	109.6(14)
C11	S1	C8	91.39(9)	C4	C5	N1	110.6(13)
S1	C8	C7	121.7(13)	C5	N1	C1	114.0(13)
C9	C8	C7	126.5(16)	N1	C1	C2	110.9(13)
C8	C7	N2	155.5(16)	C1	C2	C3	110.6(13)
C8	C7	N3	122.5(15)	C5	N1	C12	125.2(14)
N2	C7	N3	114.9(15)	C1	N1	C12	120.2(14)
C7	N3	C6	102.5(14)	N1	C12	C13	120.2(15)
N3	C6	O2	113.3(14)	N1	C12	O1	123.9(16)
C6	O2	N2	106.4(12)	O1	C12	C13	120.2(15)
C7	N2	O2	102.9(13)	C12	C13	C14	111.3(14)
O2	C6	C3	117.2(14)	C13	C14	N4	177.1(19)
N3	C6	C3	129.5(15)				

A-IV | BDM41420

Table A11 | Crystallographic Data for BDM41420

	BDM41420
Empirical Formula	$C_{17}H_{19}N_2O_2F_3$
Formula weight	300.35
Temperature / K	100.0
Crystal System	monoclinic
Space Group	$P2_1/c$
a / Å	10.599(3)
b / Å	29.145(2)
c / Å	10.661(2)
α / °	90.00
β / °	102.669(2)
γ / °	90.00
Volume / $Å^3$	3213.3(8)
Z	8
ρ_{calc} / mg mm^3	1.242
μ /mm^{-1}	0.621
2θ Range for Data Collection /°	6.06 – 136.96
Index Ranges	$-12 \leq h\ 12, -34 \leq k \leq 33, -12 \leq l \leq 12$
Reflections Collected	20222
Independent Reflections	5615 [R_{int} = 0.0213]
Data/Restraints/Parameters	5615/0/449
Goodness-of-fit on F^2	1.294
Final R1 Index [I >2σ(I)]	0.0351
Final wR2 Index [I >2σ(I)]	0.1366
Final R1 Index [all data]	0.0395
Final wR2 Index [all data]	0.1533
Largest diff. peak/hole / e $Å^{-3}$	0.263/-0.229

Figure A4 | Labelled BDM41420 (1)

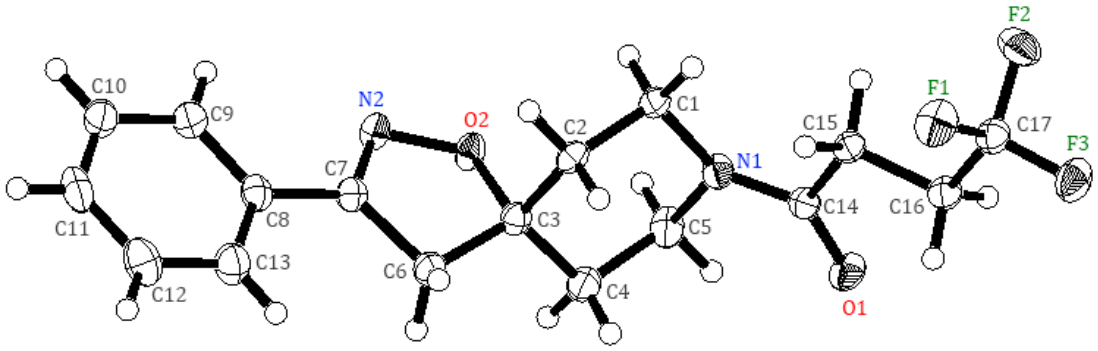


Table A12 | Bond lengths for BDM41420 (1)

Atom	Atom	Length /Å	Atom	Atom	Length /Å
C8	C9	1.385(17)	C2	C1	1.526(17)
C9	C10	1.391(17)	C1	N1	1.464(14)
C10	C11	1.374(18)	N1	C5	1.465(14)
C11	C12	1.383(18)	C5	C4	1.527(15)
C12	C13	1.389(17)	C4	C3	1.526(15)
C13	C8	1.389(17)	N1	C14	1.354(15)
C8	C7	1.472(15)	C14	O1	1.232(14)
C7	N2	1.283(15)	C14	C15	1.522(16)
N2	O2	1.424(12)	C15	C16	1.523(16)
O2	C3	1.475(14)	C16	C17	1.493(16)
C3	C6	1.521(16)	C17	F1	1.345(14)
C6	C7	1.505(16)	C17	F2	1.345(14)
C3	C2	1.521(16)	C17	F3	1.340(15)

Table A13 | Bond angles for BDM41420 (1)

Atom	Atom	Atom	Angle /°	Atom	Atom	Atom	Angle /°
C8	C9	C10	120.6(12)	C2	C3	C4	110.8(9)
C9	C10	C11	120.3(12)	C3	C4	C5	111.2(9)
C10	C11	C12	119.5(12)	N1	C5	C4	109.5(9)
C11	C12	C13	120.4(12)	C1	N1	C5	112.9(9)
C8	C13	C12	120.4(12)	N1	C1	C2	110.4(9)
C9	C8	C13	118.8(11)	C1	C2	C3	112.3(10)
C7	C8	C9	120.1(11)	C14	N1	C5	124.5(9)
C7	C8	C13	121.2(11)	C14	C15	C1	120.0(10)
N2	C7	C6	121.3(10)	O1	C14	N1	122.5(11)
C6	C7	C8	125.3(10)	O1	C14	C15	120.5(10)
N2	C7	C6	113.5(10)	N1	C14	C15	116.9(10)
C7	N2	O2	108.6(9)	C14	C15	C16	110.8(9)
N2	O2	C3	107.9(8)	C15	C16	C17	111.5(9)
O2	C3	C6	102.5(8)	C16	C17	F1	112.1(10)
C3	C6	C7	100.4(9)	C16	C17	F2	113.0(10)
O2	C3	C2	107.7(9)	C16	C17	F3	112.9(10)
O2	C3	C4	107.2(9)	F1	C17	F2	106.5(10)
C6	C3	C2	115.2(10)	F2	C17	F3	105.8(10)
C4	C3	C6	112.8(10)	F1	C17	F3	105.9(10)

Figure A5 | Labelled BDM41420 (2)

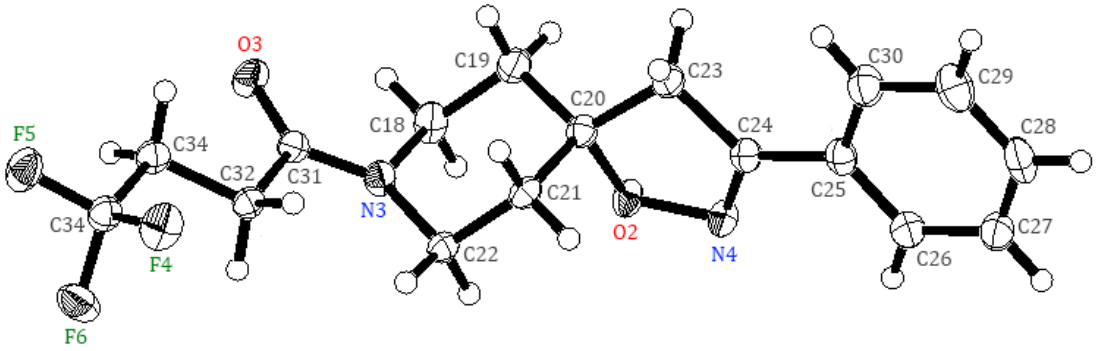


Table A14 | Bond lengths for BDM41420 (2)

Atom	Atom	Length /Å	Atom	Atom	Length /Å
C25	C26	1.395(17)	C19	C18	1.526(16)
C26	C27	1.390(17)	C18	N3	1.471(14)
C27	C28	1.387(2)	N3	C22	1.463(15)
C28	C29	1.379(2)	C22	C21	1.530(18)
C29	C30	1.391(18)	C21	C20	1.523(16)
C30	C25	1.389(18)	N3	C31	1.353(15)
C25	C24	1.471(16)	C31	O3	1.230(14)
C24	N4	1.285(15)	C31	C32	1.522(16)
N4	O4	1.421(12)	C32	C33	1.527(16)
O4	C20	1.475(14)	C33	C34	1.492(16)
C20	C23	1.521(16)	C34	F4	1.347(14)
C23	C24	1.503(16)	C34	F5	1.346(15)
C20	C19	1.528(15)	C34	F6	1.347(14)

Table A15 | Bond angles for BDM41420 (2)

Atom	Atom	Atom	Angle /°	Atom	Atom	Atom	Angle /°
C27	C26	C25	120.3(12)	C21	C20	C19	110.6(9)
C28	C27	C26	120.2(12)	C18	C19	C20	111.3(9)
C29	C28	C27	119.8(12)	C20	C21	C22	111.6(10)
C28	C29	C30	120.3(13)	N3	C22	C21	110.1(9)
C25	C30	C29	120.4(12)	N3	C18	C19	109.7(9)
C30	C25	C26	119.1(11)	C31	N3	C18	124.9(10)
C30	C25	C24	120.2(11)	C22	N3	C18	113.0(9)
C26	C25	C24	120.7(11)	C31	N3	C22	119.6(10)
C25	C24	C23	125.2(10)	O3	C31	N3	122.3(11)
N4	C24	C25	121.3(10)	N3	C31	C32	116.9(10)
N4	C24	C23	113.4(10)	O3	C31	C32	120.8(10)
C24	C23	C20	108.8(9)	C31	C32	C33	110.6(9)
N4	O4	C20	108.0(13)	C34	C33	C32	111.6(9)
C24	C23	C20	100.5(9)	C33	C34	F4	113.1(10)
O4	C20	C23	102.8(8)	C33	C34	F5	112.0(10)
C23	C20	C19	112.2(10)	C33	C34	F6	113.0(10)
O4	C20	C21	107.7(9)	F4	C34	F5	105.8(10)
O4	C20	C19	107.3(9)	F4	C34	F6	106.0(10)
C23	C20	C21	115.5(10)	F5	C34	F6	106.5(10)

A-V | BDM41907

Table A17 | Crystallographic Data for BDM41907

	BDM41907
Empirical Formula	$C_{19}H_{22}N_3O_4SF_3$
Formula weight	371.44
Temperature / K	99.85
Crystal System	monoclinic
Space Group	$P2_1/c$
a / Å	10.702(6)
b / Å	19.853(2)
c / Å	9.873(2)
α / °	90.00
β / °	104.568(2)
γ / °	90.00
Volume / Å^3	2030.4(5)
Z	4
ρ_{calc} / mg mm^3	1.215
μ /mm^{-1}	1.622
2θ Range for Data Collection /°	8.54 – 132.38
Index Ranges	-12 ≤ h ≤ 11, -23 ≤ k ≤ 23, -10 ≤ l ≤ 11
Reflections Collected	14675
Independent Reflections	3493 [R_{int} = 0.0408]
Data/Restraints/Parameters	3493/0/272
Goodness-of-fit on F^2	0.992
Final R1 Index [I >2σ(I)]	0.0386
Final wR2 Index [I >2σ(I)]	0.1163
Final R1 Index [all data]	0.0498
Final wR2 Index [all data]	0.1271
Largest diff. peak/hole / e Å^{-3}	0.395/-0.332

Figure A6 | Labelled BDM41907

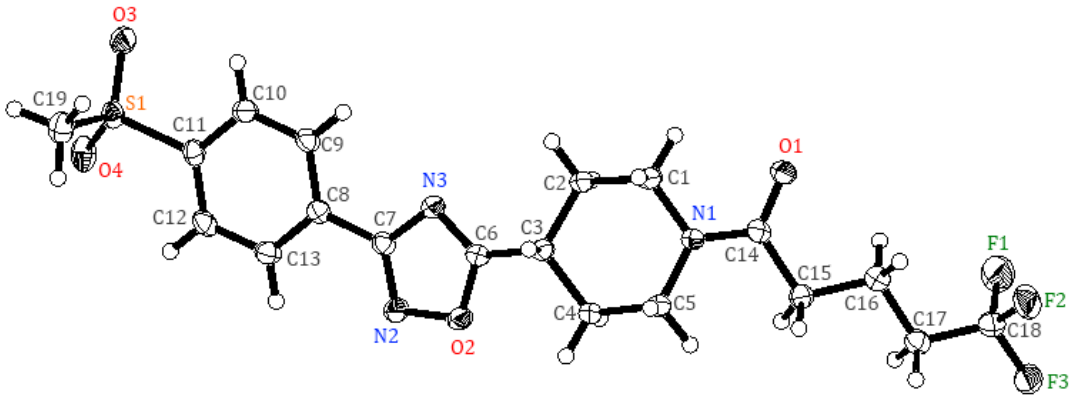


Table A18 | Bond lengths for BDM41907

Atom	Atom	Length /Å	Atom	Atom	Length /Å
C19	S1	1.761(2)	C6	C3	1.492(3)
S1	O3	1.442(15)	C3	C2	1.534(2)
S1	O4	1.432(15)	C2	C1	1.525(3)
S1	C11	1.771(2)	C1	N1	1.462(2)
C11	C10	1.389(3)	N1	C5	1.468(2)
C10	C9	1.381(3)	C5	C4	1.524(3)
C9	C8	1.393(3)	C4	C3	1.536(2)
C8	C13	1.389(3)	N1	C14	1.353(2)
C13	C12	1.387(3)	C14	O1	1.239(2)
C12	C11	1.386(3)	C14	C15	1.525(3)
C8	C7	1.474(3)	C15	C16	1.525(3)
C7	N3	1.381(2)	C16	C17	1.524(3)
N3	C6	1.298(2)	C17	C18	1.492(3)
C6	O2	1.346(2)	C18	F1	1.340(3)
O2	N2	1.409(2)	C18	F2	1.342(2)
N2	C7	1.307(2)	C18	F3	1.351(2)

Table A19 | Bond angles for BDM41907

Atom	Atom	Atom	Angle /°	Atom	Atom	Atom	Angle /°
O4	S1	O3	119.0(9)	C7	N2	O2	103.1(15)
O4	S1	C11	108.1(9)	C2	C3	C4	110.5(15)
O4	S1	C19	108.6(10)	C6	C3	C2	111.0(15)
O3	S1	C11	108.1(9)	C6	C3	C4	111.2(15)
O3	S1	C19	108.0(10)	C1	C2	C3	110.5(15)
C19	S1	C11	104.1(10)	N1	C1	C2	110.0(15)
C12	C11	S1	120.0(14)	C14	N1	C5	125.9(16)
C12	C11	C10	121.0(18)	C14	N1	C1	120.8(15)
C10	C11	S1	119.0(15)	C1	N1	C5	113.2(15)
C9	C10	C11	119.7(18)	N1	C5	C4	110.8(15)
C10	C9	C8	119.8(18)	C5	C4	C3	111.6(15)
C9	C8	C7	119.2(17)	O1	C14	N1	121.3(17)
C13	C8	C9	120.3(18)	O1	C14	C15	120.3(17)
C13	C8	C7	120.4(17)	N1	C14	C15	118.4(16)
C12	C13	C8	120.0(18)	C16	C15	C14	111.5(16)
N2	C7	C8	123.5(16)	C17	C16	C15	111.5(16)
N3	C7	C8	114.7(17)	C18	C17	C16	111.7(18)
N2	C7	N3	121.7(17)	F3	C18	C17	113.2(17)
C6	N3	C7	102.5(15)	F2	C18	F3	105.5(17)
O2	C6	C3	117.8(16)	F2	C18	C17	113.4(17)
N3	C6	O2	113.1(16)	F1	C18	F3	105.7(18)
N3	C6	C3	129.1(16)	F1	C18	F2	105.8(17)
C6	O2	N2	106.6(14)	F1	C18	C17	112.6(18)

A-VI | BDM41325

Table A21 | Crystallographic Data for BDM41325

	BDM41325
Empirical Formula	$C_{19}H_{19}N_4O_3F_3S$
Formula weight	440.44
Temperature / K	99.93
Crystal System	orthorhombic
Space Group	$P2_12_12_1$
a / Å	4.828(1)
b / Å	16.186(5)
c / Å	24.332(7)
α / °	90.00
β / °	90.00
γ / °	90.00
Volume / $Å^3$	1901.56(9)
Z	4
ρ_{calc} / mg mm^3	1.538
μ /mm^{-1}	2.059
2θ Range for Data Collection /°	6.56 – 132.92
Index Ranges	$-5 \leq h \leq 5, -19 \leq k \leq 19, -28 \leq l \leq 28$
Reflections Collected	19327
Independent Reflections	3233 [R_{int} = 0.0436]
Data/Restraints/Parameters	3233/0/272
Goodness-of-fit on F^2	0.979
Final R1 Index [I >2σ(I)]	0.0298
Final wR2 Index [I >2σ(I)]	0.0761
Final R1 Index [all data]	0.0333
Final wR2 Index [all data]	0.0788
Largest diff. peak/hole / e $Å^{-3}$	0.256/-0.165

Figure A7 | Labelled BDM41325

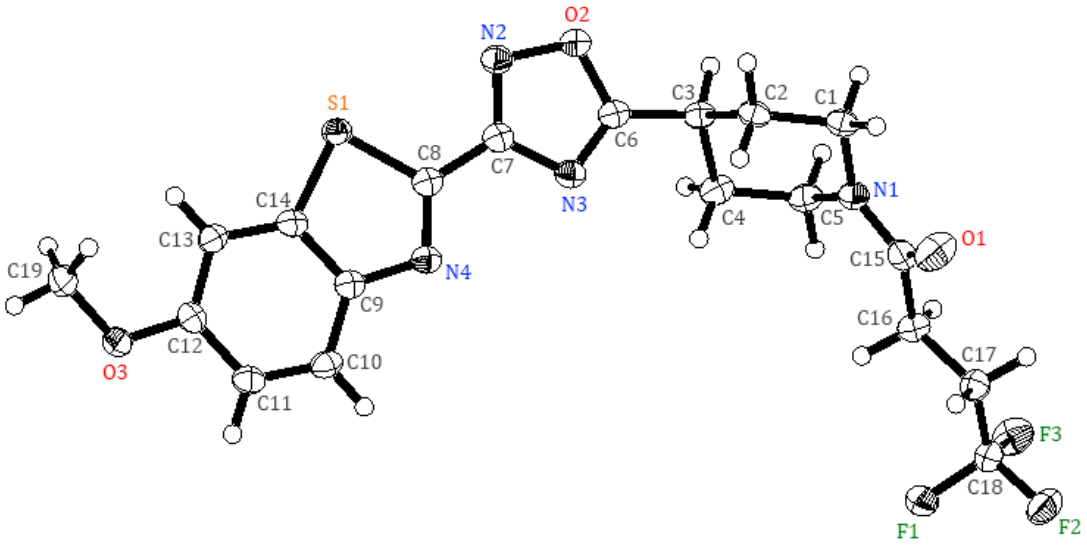


Table A22 | Bond lengths for BDM41325

Atom	Atom	Length /Å	Atom	Atom	Length /Å
C19	O3	1.443(3)	N3	C7	1.376(3)
O3	C12	1.368(3)	C6	C3	1.487(3)
C12	C13	1.368(3)	C3	C2	1.539(3)
C13	C14	1.391(3)	C2	C1	1.522(3)
C14	C9	1.402(3)	C1	N1	1.456(3)
C9	C10	1.406(3)	N1	C5	1.466(3)
C10	C11	1.364(3)	C5	C4	1.522(3)
C11	C12	1.412(3)	C4	C3	1.535(3)
C14	S1	1.735(2)	N1	C15	1.350(3)
S1	C8	1.754(2)	C15	O1	1.224(3)
C8	N4	1.303(3)	C15	C16	1.525(3)
N4	C9	1.389(3)	C16	C17	1.525(3)
C8	C7	1.455(3)	C17	C18	1.488(3)
C7	N2	1.313(3)	C18	F1	1.351(3)
N2	O2	1.417(2)	C18	F2	1.339(2)
O2	C6	1.349(2)	C18	F3	1.348(2)
C6	N3	1.298(3)			

Table A23 | Bond angles for BDM41325

Atom	Atom	Atom	Angle /°	Atom	Atom	Atom	Angle /°
C12	O3	C19	117.0(17)	N3	C6	O2	113.6(19)
O3	C12	C13	124.2(18)	N3	C6	C3	128.1(19)
O3	C12	C11	115.0(19)	C6	N3	C7	102.5(17)
C13	C12	C11	120.8(2)	C6	C3	C4	109.9(18)
C12	C13	C14	117.3(18)	C6	C3	C2	110.3(17)
C9	C14	S1	109.2(16)	C2	C3	C4	110.3(16)
C13	C14	S1	128.1(16)	C1	C2	C3	111.9(17)
C13	C14	C9	122.7(19)	N1	C1	C2	110.4(18)
N4	C9	C14	115.8(18)	C15	N1	C5	119.9(19)
N4	C9	C10	125.4(18)	C15	N1	C1	126.0(17)
C14	C9	C10	118.8(19)	C1	N1	C5	114.1(17)
C11	C10	C9	119.1(19)	N1	C5	C4	110.7(17)
C10	C11	C12	121.3(2)	C5	C4	C3	110.9(19)
C14	S1	C8	88.8(10)	O1	C15	N1	122.1(19)
C7	C8	S1	120.4(15)	O1	C15	C16	120.7(19)
N4	C8	S1	116.1(16)	N1	C15	C16	117.2(19)
N4	C8	C7	123.5(18)	C17	C16	C15	110.7(18)
C8	N4	C9	110.0(17)	C18	C17	C16	111.8(18)
N2	C7	N3	114.9(2)	F3	C18	F1	105.3(18)
N2	C7	C8	122.0(18)	F3	C18	C17	112.6(18)
N3	C7	C8	123.1(18)	F2	C18	F3	106.2(18)
C7	N2	O2	103.0(15)	F2	C18	F1	106.2(18)
C6	O2	N2	106.0(15)	F2	C18	C17	112.7(18)
O2	C6	C3	118.3(18)	F1	C18	C17	113.3(19)