

# **Fragment-based interrogation of a bacterial replisome**

**Bas Lamoree**

**Doctor of Philosophy**

**University of York**

**Department of Chemistry**

*August 2020*



## Abstract

Chromosomes are replicated by the replisome, a molecular machine comprised of a set of interacting proteins with interdependent functions. In previous work, a fragment library was screened *in vitro* against a bacterial replisome, reconstituted from thirteen different subunits. In this thesis, the results of this phenotypic assay were validated, analysed and extended with functional and biophysical experiments to explore different mechanisms of inhibition. The work shows that phenotypic fragment screening can be used as a starting point for target identification.

Assays for the individual protein complexes in the replisome revealed that most fragment inhibitors acted through specific proteins. For one fragment inhibitor, the results were consistent with inhibition of primase, an enzyme that synthesises short primers to initiate elongation. The mode of action of this fragment and analogues were studied in detail.

The bacterial primase consists of three domains: a helicase-binding, a zinc-binding, and a polymerase domain. The fragment was not observed to bind to the polymerase domain in X-ray crystallography and nuclear magnetic resonance experiments. Molecular dynamics simulations showed a clear preference for binding to a conserved protein-protein interaction site on the helicase-binding domain, rather than to the zinc-binding domain. This prediction was validated by protein-observed nuclear magnetic resonance experiments and structure-activity and -affinity relationships.

Some of the replisome fragment inhibitors acted only on the full machine and do not show inhibition of the available assays for individual components. Methods were developed to generate and test photoreactive fragment hit analogues and to screen covalent cysteine-targeting fragments. This work did not identify any new site of interaction.



# Contents

<b>Abstract</b>	<b>3</b>
<b>Contents</b>	<b>5</b>
<b>Acknowledgements</b>	<b>9</b>
<b>Author's declaration</b>	<b>11</b>
<b>1. Introduction</b>	<b>13</b>
1.1 Fragment-based lead discovery	13
1.1.1 Core concepts of fragment-based lead discovery	14
1.1.2 Fragment detection techniques	17
1.1.3 FBLD in practice	20
1.2 The replisome	23
1.2.1 Characteristics of the whole replisome	25
1.2.2 The polymerase holoenzyme	26
1.2.3 Helicase complex	32
1.2.4 Single-stranded DNA-binding protein and other proteins	36
1.3 Antibiotic targets	36
1.4 Aims of the project	38
<b>2 Structural and functional characterisation of hits from a fragment screen against a bacterial replisome</b>	<b>41</b>
Preface	41
2.1 Introduction	41
2.2 Results	47
2.2.1 Compound quality controls	47
2.2.2 Confirmation of screening hits	49
2.2.3 How MB25 and MB99 inhibit the replisome	53
2.2.4 Analysis of structural motifs within the replisome fragment hit set	57
2.3 Conclusion	61
<b>3 Finding the binding site of a primase fragment hit using molecular dynamics simulations</b>	<b>63</b>

3.1	Introduction	63
3.2	MD simulation results	65
3.2.1	Model preparations	65
3.2.2	Simulation results	67
3.3	Verification of simulation results by NMR	72
3.4	Conclusion	81
<b>4</b>	<b>Elaboration of primase fragment hits using NMR spectroscopy and X-ray crystallography</b>	<b>83</b>
	Preface	83
4.1	Introduction	83
4.2	Evaluation of fragments by NMR	84
4.2.1	Fragment screening against primase by ligand-observed NMR	84
4.2.2	Confirmation of fragment hits	84
4.3	Evaluation of primase fragment hits using X-ray crystallography	88
4.3.1	Crystal structures of RNA polymerase domain in complex with nucleotide analogues	89
4.3.2	Crystal structures of RNA polymerase domain in complex with fragments	91
4.4	Evaluation of primase fragment hits using NMR	97
4.4.1	Production of the helicase-binding domain	97
4.4.2	Characterisation of helicase-binding domain fragment hits by NMR	98
4.4.3	Binding and activity of HBD fragment hit analogues	104
4.5	Conclusion	116
<b>5</b>	<b>Covalent modification of the replisome for target identification</b>	<b>117</b>
	Preface	117
5.1	Introduction	117
5.2	Photoreactive replisome hits	119
5.2.1	Design of diazirine derivatives of replisome fragment inhibitors	119
5.2.2	Testing of diazirine derivatives	121
5.3	Targeting replisome cysteines	126
5.3.1	Feasibility of replisome inhibition by thiol-targeting fragments	127

5.3.2	Screening of a small electrophilic library	130
5.4	Conclusion	133
<b>6</b>	<b>Conclusions</b>	<b>135</b>
<b>7</b>	<b>Materials and methods</b>	<b>137</b>
7.1	Primase production	137
7.2	DNA replication assay	138
7.2.1	Template preparation	138
7.2.2	Replication reaction	138
7.2.3	Measurement of replication products	139
7.2.4	Calculation of inhibition constants	140
7.3	Primase activity assay	141
7.4	NMR	142
7.4.1	Compound quality controls	142
7.4.2	Ligand-observed experiments	142
7.4.3	Protein-observed experiments	143
7.5	MD simulations	143
7.6	SEC-MALS	144
7.7	Mass spectrometry	144
7.8	Crystallography	145
7.9	Synthesis of diazirine analogues and UV crosslinking experiments	146
7.10	Click reactions	147
7.11	Covalent fragment screening	147
<b>Appendix</b>		<b>149</b>
	Appendix A – Structures of 25 replisome screen fragment hits	149
	Appendix B – Structures of 35 primase NMR screen fragment hits	151
<b>Abbreviations</b>		<b>153</b>
<b>References</b>		<b>155</b>



## **Acknowledgements**

I would like to thank Professor Roderick Hubbard for giving me the opportunity to work on this PhD and for supporting me during these studies. I would also like to thank Dr. Jamieson Howard, Dr. John Darby, Dr. James Firth, Dr. Simon Tanley, and Jim Brannigan for discussing various parts of this research with me and for contributing to some of the results presented here. Finally, I would like to thank all members of the York Structural Biology Laboratory, the McGlynn group at York, the Barril laboratory at the Universitat de Barcelona, and the Keserú laboratory at the Research Centre for Natural Sciences in Budapest for making my studies even more interesting both scientifically and socially.



## **Author's declaration**

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.



# 1. Introduction

## 1.1 Fragment-based lead discovery

Modern Western drugs most often contain single, rationally designed active ingredients. The majority of these active ingredients can be classed into three groups: small organic molecules, large organic molecules, and infectious single particles such as virions and bacterial cells; the first being the largest group. Rational design of novel small organic molecules for therapeutic use depends heavily on scientific methods, so that newly developed methods have often contributed to development of new drugs.

Drug development projects start with a predefined target product profile: *e.g.* a successful drug candidate will cure disease *X* while side effects remain below level *Y*. The complete drug development process is often long (4-20 years) and has a low chance of success. It involves many different steps requiring many different scientific and other types of expertise. The very first step in the process is to hypothesise a mechanism for interfering with the disease, based on available biological information. The hypothesis is almost always based on the central idea that biological effects are caused by the physical interaction of two molecules, most often a protein and a small organic molecule. This first step is not always required, as will be discussed later. The second step in the process is to identify one or more drug candidates, or 'leads'. These early stages are called lead discovery. Many other steps after this will focus on selecting and transforming leads into actual small molecule drugs, the active ingredients in most medicines.

When the target and desired effect are known, the discovery process often starts with the screening of a diverse library of organic compounds. Screening random compounds may seem like a rather blunt method, but currently it remains the most effective approach to find hits. Other approaches are mostly based on mimicking or repurposing known drugs and natural molecules, which do not give much novelty in either medical or business sense. Screening of large collections of compounds (up to several millions) can be done efficiently by using robotics and automation. This is called high-throughput screening (HTS) and has become a standard procedure in the pharmaceutical industry since the 1980s. However, HTS is not accessible to anyone outside the pharmaceutical industry due to the high costs of maintaining not only

screening infrastructure but also compound collections themselves. There are two main alternatives to HTS, both developed more recently and used by many academic groups and small as well as large enterprises. One is virtual screening, which does not require much physical equipment but does require sophisticated software for screening and analysis. The other is fragment screening (Erlanson *et al.*, 2016), the term ‘fragment’ meaning a small part of an average-sized screening or drug molecule. Fragment-based screening and lead development (FBLD) does not require much physical equipment but does require more work to develop a fragment hit into an attractive lead compound. The ideas and methods underlying FBLD have been under development since the 1970s and will be discussed below.

### 1.1.1 Core concepts of fragment-based lead discovery

A core concept of FBLD is that chemical space, the total number of possible small organic molecules, is smaller as the number of atoms in a compound decreases (Hall *et al.*, 2014). Thus, the usefulness of screening libraries containing  $10^5$ - $10^7$  potential lead candidates, historically only available to big pharmaceutical companies, can be rivalled by libraries of just  $10^2$ - $10^4$  fragment molecules (Shi & von Itzstein, 2019). The differences between drug-like HTS compounds and fragments are illustrated by the crude ‘rules’ of five (Lipinski *et al.*, 2001) and three (Congreve *et al.*, 2003; Jhoti *et al.*, 2013), referring to preferred values of selected physicochemical properties of organic screening compounds (*e.g.* molecular weight, hydrophobicity, number of hydrogen bond-forming atoms and rotatable bonds). Another useful way of comparison is to look at the number of non-hydrogen atoms (heavy atoms, HA): while HTS compounds average HA counts of >25, fragment hits often start off with <15 and fragment library members are limited to 20 at most (Keserü *et al.*, 2016).

Not only do fragments offer more efficient chemical space coverage, their hit rates are also higher compared to HTS libraries. This is best thought of in terms of ‘molecular complexity’, a concept developed by Hann *et al.* (Hann *et al.*, 2001; Leach & Hann, 2011). Both a screening compound and a receptor pocket can be viewed as a collection of chemical features (*e.g.* hydrogen bond donors/acceptors) in a certain spatial configuration, that either attract or repel each other. Hits are molecules that complement the receptor features and thus favour binding. Bigger molecules have more complex feature patterns, which decreases the probability of finding the right

complementary pattern. In reality, protein-ligand interactions are more diverse than simply plus or minus, and binding is modulated by more than just complementarity. Nevertheless, the general outcomes from the complexity model hold true: simpler molecules have a higher chance of binding to any target.

The trade-off of decreasing molecular complexity is that the maximum potency (either a biochemical effect or a biophysical binding event) becomes weaker. Magnitudes of biochemical or functional effects are usually expressed in units of compound concentration: the concentration at which 50% of positive control activity is observed is called the 50% inhibitory concentration,  $IC_{50}$ . Binding strengths are usually indicated by stating the equilibrium dissociation constant ( $K_D$ ), which coincidentally also has the formal unit of compound concentration in the most common case of a first-order process (one-step binding). Fragment hits typically have  $K_D$ s in the  $10^{-4}$ - $10^{-2}$  M range (close to the lower detection limit of most instruments), and  $IC_{50}$ s even lower or zero. Attractive leads typically have (sub-)nanomolar  $IC_{50}$ s. This adds the extra step of optimisation of fragment hits to lead-like molecules. Theoretically, fragment hit-to-lead optimisation can be accomplished by linking or merging two fragment hits binding to different parts of a binding pocket, or by growing one fragment hit to fill its binding pocket (Kirsch *et al.*, 2019). Although improving the binding affinity of a weak fragment by six orders of magnitude may seem like an almost impossible task, the underlying physics of protein-ligand binding can often show the way. The use of biophysical techniques and a structure-based approach to lead design are standard practice by now (Renaud *et al.*, 2016). Ligand binding energetics were first explored and discussed within the biochemical community by Jencks and others in the 1970s (Jencks, 1981, 1975). Two important concepts for FBLD are: 1) the relationship between binding affinity and binding energy: the two are related not linearly but exponentially via  $\Delta G = -RT\ln(K_D)$ , meaning that small changes in binding energy can have larger effects on affinity; and 2) entropic effects play a bigger role at lower affinities, meaning that linking or merging two fragment hits can increase affinity by more than just the product of parts. Indeed, several spectacular fragment linking successes are known (Ichihara *et al.*, 2011). However, it is more often the case that fragment linking fails to improve affinity. This is mostly due to the difficulty of designing a linker moiety that can preserve the exact binding orientations of the initial two fragments while not itself subtracting from binding

affinity. Merging of two fragments that share part of the binding pocket does not suffer from these problems, but it does not often happen that two fragments share atoms that exactly overlap in the binding pocket (especially when the originating fragment library was designed to be diverse). In the end, fragment growing following scaffold optimisation appears to be the most popular approach to hit optimisation.

Fragment hit-to-lead optimisation guided solely by physical parameters has several pitfalls. Firstly, a physical binding event does not necessarily correspond to an equivalent biochemical activity. Most FBLD practitioners now insist on measuring both  $K_D$  and  $IC_{50}$  for all optimisation compounds. Secondly, practice has shown that making hits bigger and more lipophilic will often increase binding affinity and biochemical activity regardless of selectivity and specificity (Ferenczy & Keserű, 2016). To make sure that any addition to a hit is worth the cost of increasing molecular weight and lipophilicity, the ligand efficiency (LE) metric was adopted. It is defined as the fraction of binding energy contributed by each heavy atom of the ligand on average:  $LE = \Delta G/HA$ . The number is usually between 0.3 and 0.5 kcal mol<sup>-1</sup> for initial fragment hits as well as later-stage compounds, and ideally it should not decrease during hit-to-lead optimisation. Several variations on LE exist and can be more appropriate during some stages of drug development. For example, LE adjusted specifically for lipophilicity may become more interesting during later stages when drug activity, absorption, distribution, metabolism, and excretion parameters need to be optimised simultaneously (Leeson & Springthorpe, 2007).

Ligands can be modified almost at will (depending on feasibility and tractability of synthesis), but their targets (proteins or polynucleotides) are fixed. In early work analysing proteins using atomic resolution crystal structures, researchers used ‘functional groups’ to computationally probe protein surfaces for energetically favourable binding sites (Goodford, 1985; Miranker & Karplus, 1991). These probes are essentially very small fragments, which would be gaseous or liquid *in vitro*. Experimental work used solvents (which are again very small fragments) to probe protein surfaces *in crystallo* (Allen *et al.*, 1996; Mattos & Ringe, 1996). Both of these approaches (virtual and ‘wet’) have seen more recent iterations as tools in FBLD, for example: FTMap (Kozakov *et al.*, 2015) which probes static protein structures for interaction hot spots with small solvent molecules, MDmix (Alvarez-Garcia & Barril, 2014) which does the same but runs molecular dynamics simulations to incorporate

protein flexibility, and MiniFragments (O'Reilly *et al.*, 2019) that experimentally probe crystallised proteins to find new binding (sub)pockets.

Compared to bluntly screening millions of drug-like molecules, the characterisation of fragment-sized molecules, their targets, and their interactions is more precise and usually requires more experiments. However, the past twenty years have shown that it is worth the efforts (see section 1.1.3), as the field of FBLD has matured from mostly academic curiosity into a standard platform in most pharmaceutical companies. The fragment-based approach is used under various circumstances: as back-up hit-finding method when HTS fails; to overcome hurdles during lead optimisation; or simply as a main approach against certain target classes.

### 1.1.2 Fragment detection techniques

While going smaller increases the chance of binding, it also decreases the chance of experimentally detecting such binding. It follows that there is an optimum size of screening molecules where promiscuity and affinity balance each other (Hann *et al.*, 2001). Molecules of this size turn out to be larger than solvents but smaller than HTS compounds: fragments. Since fragments are often not active enough to be picked up by functional assays, reliable detection of fragment hits depends heavily on the use of biophysical techniques, which have seen much method development and improvements over the past three decades (Renaud *et al.*, 2016; Coyle & Walser, 2020; Erlanson *et al.*, 2019). The three main techniques used for fragment screening are thermal shift assay (TSA), surface plasmon resonance (SPR), and nuclear magnetic resonance (NMR). Recently, innovations in infrastructure have made X-ray crystallography a fourth attractive screening technique (Erlanson, 2019).

TSA relies on a change in protein stability upon ligand binding, assessed by the temperature at which the protein unfolds (Lo *et al.*, 2004). The unfolding temperature is measured as an increase of fluorescence. The effect depends on the presence of a hydrophobic dye that binds to unfolding protein hydrophobic cores, exhibiting an increase in fluorescence efficiency in hydrophobic environments. As it only requires a rtPCR machine and consumes very small amounts of protein, it is cheap, quick, and easy to execute. TSA results are easy to analyse but have low information content that may be over-interpreted. A fundamental drawback of this technique is that the observed ligand binding often takes place at temperatures well above room

temperature, which could give wrong ideas of binding strengths if overlooked. Especially in case of fragments, TSA has a relatively high rate of false negatives and artefacts owing to their weak affinity and high concentrations. A popular alternative to carrying out a full fragment screen by TSA is to quickly screen a small set of well-behaved, diverse compounds to assess protein ligandability.

SPR, NMR, and X-ray crystallography are more expensive and offer lower throughput but more detailed and reliable information. SPR relies on a change in refractive index close to a solid surface, assessed by the resonance angle of a totally internally reflected beam of light at the other side of the surface (Jonsson *et al.*, 1991; Danielson, 2009). Thus, when protein molecules are covalently linked to a surface, fragments that bind will cause a change in refractive index dependent mainly on their concentration and molecular weight. SPR measurements are taken in real time by using a flow cell, allowing determination of kinetic binding parameters in addition to affinity and stoichiometry. SPR is commonly used during fragment screening as well as optimisation stages.

NMR spectroscopy is an extremely versatile technique that relies on changes in the local magnetic field of any atomic nucleus with non-zero spin. The most popular application is in imaging of biological tissues. In drug discovery, two types of NMR approaches are distinguished: ligand-observed (LO) and protein-observed (PO) NMR (Gossert & Jahnke, 2016). The first measures signals from ligand protons (or, less commonly,  $^{19}\text{F}$  nuclei), whereas the second measures signals from coherent protein protons,  $^{13}\text{C}$  nuclei, and  $^{15}\text{N}$  nuclei in any permutation of two or three. LO-NMR is most often used for fragment screening and hit validation. Three often-employed experiments are saturation transfer difference (STD) (Mayer & Meyer, 1999), water ligand-observed via gradient spectroscopy (WaterLOGSY) (Dalvit *et al.*, 2001), and transverse relaxation filtering with the Carr-Purcell-Meiboom-Gills (CPMG) sequence (Hajduk *et al.*, 1997; Meiboom & Gill, 1958). All three experiments will show a change in signal intensity of a ligand when it binds to a protein. In STD and WaterLOGSY, protein protons (preferably at shifts smaller than 1 ppm to avoid overlap with small molecules) and water protons (around 4.8 ppm) respectively are selectively irradiated, after which they transfer some of the energy to the ligand via the nuclear Overhauser effect. In the CPMG experiment, protein and protein-associated protons are filtered out of the spectrum based on their bigger particle size

and thus higher transverse relaxation rates compared to unbound small molecules. LO methods give information about whether a fragment binds and can also determine affinity through titrations and binding site identification through competition experiments. LO-NMR can work with protein concentrations as low as 1  $\mu\text{M}$ , but (like most NMR experiments) requires large sample volumes (usually 500  $\mu\text{L}$ ) which results in high protein consumption during screening. Because standard NMR spectra are always of high enough resolution to allow observation of chemically unique nuclei, a single binding experiment can simultaneously be used for compound identification and quality control (sometimes even indicating solubility issues). LO methods are limited to the fast-exchange regime of compound binding (usually not stronger than 0.1-1.0  $\mu\text{M}$   $K_{\text{DS}}$ ) because of the large excess of ligand over protein. PO methods can yield much more detailed information, including direct binding site and binding mode identification, detection of slow-exchange binding, and conformational changes. Unlike LO-NMR and almost any other biophysical technique, PO experiments rarely yield false positives or false negatives. Recent advances in NMR hardware and experimental methods have made the simple 2-dimensional heteronuclear single quantum coherence experiment viable for screening (Schanda *et al.*, 2005), although this still requires large quantities of high-quality, isotope-labelled protein. However, PO-NMR is mainly used to obtain high-resolution binding information, especially for challenging projects where protein structure and function are impaired by any kind of deviation from its standard solvated state (*e.g.* high temperature, surface immobilisation, chemical labelling, or crystallisation).

X-ray crystallography is the most important technique in structural biology. It relies on well-ordered crystals of pure analyte (most often proteins or stable protein complexes) to determine their 3-dimensional electronic structures at atomic resolution. In drug discovery, it is the preferred approach for validating and guiding early ligand optimisation, provided that high-resolution structures can be obtained. Some research groups use crystallography as their main fragment screening platform. However, this is only possible if high quality crystals (*i.e.* well-diffracting, resistant and permeable to solvated screening compounds) can be grown reliably. The main drawback of crystallography as a fragment screening technique is the high concentrations of ligand needed, which can lead to false positives and does not distinguish between strong and weak binders. Furthermore, crystal packing can sometimes occlude possible binding

sites, which can sometimes but not always be overcome by co-crystallising ligands and proteins instead of soaking ligands into crystals.

Many other biophysical techniques exist that can measure the interactions of proteins and ligands, some of which are sporadically used for screening but are not part of the standard screening platform of most research groups. In particular, isothermal titration calorimetry (ITC) is a technique that gives detailed information on fundamental binding parameters including changes in enthalpy and entropy separately. Like NMR, ITC experiments are carried out label-free and in solution. The high material requirements and low throughput make ITC unamenable to screening, but it is used during optimisation to test hypotheses about protein-ligand binding energetics based on structural information. Lastly, mass spectrometry (MS) is a standard (bio)physical technique that can be adapted to carry out screening. Depending on the required throughput, MS can give detailed information about protein-ligand complexes, including ligand identity, stoichiometry, approximate binding site, and affinity. An example of covalent fragment screening by MS will be discussed in the next section.

### **1.1.3 FBLD in practice**

One of the first traditional examples of fragment-based lead generation used NMR to identify, optimise, and finally link two fragments binding close to each other on the same protein (Shuker *et al.*, 1996). It is a remarkable example of successful fragment linking, of which unfortunately not many others exist. Against the background of the more theoretical works on the interactions between proteins and fragment-sized molecules in the 1980s and 1990s (discussed earlier), this work stood out as the first demonstration of how to develop weak hits (in this case,  $K_{DS}$  of 2  $\mu\text{M}$  and 800  $\mu\text{M}$ ) into a high-affinity (19 nM) lead compound. The two hits bound to two adjacent sites on the FK506-binding protein (FKBP), an enzyme that plays an important role in T-cell activation. After slightly improving the affinity of the weaker fragment hit from 800 to 100  $\mu\text{M}$ , a 3D binding model of the two adjacent fragments was constructed from PO-NMR data. This model suggested which vectors could be used to link the two fragments, and the length of the linker was then quickly optimised experimentally to produce the 19 nM linked lead compound. The work presents a combination of earlier fragment-based ideas put into practice in a way that is now recognised as the standard FBLD approach.

Although FBLD is mainly a way to generate early-stage lead candidates, many FBLD practitioners proudly mention fragment-derived leads that were progressed through the clinic onto the market. As of 2020, four such fragment-based drugs exist: vemurafenib (Bollag *et al.*, 2012), pexidartinib (Zhang *et al.*, 2013), venetoclax (Souers *et al.*, 2013), and erdafitinib (Murray *et al.*, 2019) (Figure 1.1). Venetoclax stands out as an example of unusual chemistry and a difficult target. It inhibits the functional interaction of its family of target proteins, the prosurvival B-cell CLL/lymphoma 2 (BCL-2) family, with other proteins to cause tumour cell death. Inhibition of protein-protein interactions (PPIs) is seen as challenging because PPI surfaces have few pocket-like features or other handles for small molecules. The development of venetoclax depended heavily on structure-based methods including the identification of two weakly binding, non-competitive fragments by PO-NMR (Petros *et al.*, 2006) and the discovery of a new protein crystal form during lead optimisation (Souers *et al.*, 2013).

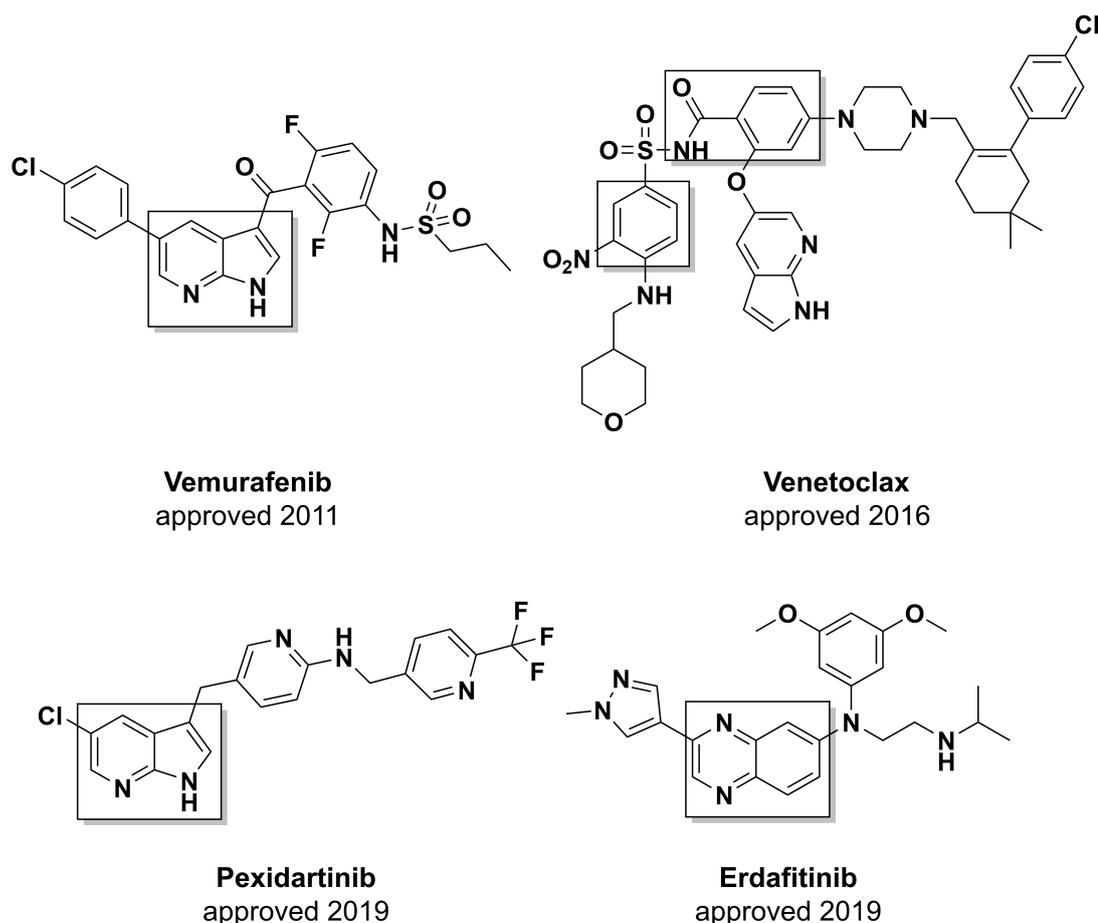


Figure 1.1. All current FDA-approved drugs with a fragment-based origin. The moiety corresponding to the initial fragment hit is indicated.

Nowadays the field of FBLD is a well-established technology with standard approaches and practices as described. Scientists from public as well as private organisations continue to innovate and find new ways of using fragments. Recent examples include screening for enzyme activators (Darby *et al.*, 2017; Tautermann *et al.*, 2019), new ways of fragment library design and synthesis (Keserú *et al.*, 2016; Firth & O'Brien, 2018), and the use of fragments that form specific covalent bonds with their targets (Parker & Pratt, 2020; Keeley *et al.*, 2020). Covalent fragments can be divided into two groups: intrinsically reactive fragments and fragments that contain an inducible covalent linker. Fragments in the first group are mild electrophiles that can react with nucleophilic protein residues (C, K, and activated S, T, Y, H, or E). The reactivities of such fragments are chosen so that covalent bond formation will mostly occur when the fragment is favourably bound and oriented, avoiding false-positive hits from non-specific reactions. Protein nucleophilic reaction parameters are well characterised and can also be used to screen electrophilic fragments *in silico* (Scarpino *et al.*, 2018). The second group of reactive fragments are molecules that contain a non-reactive, variable fragment moiety linked to a reactive functional group that can be 'switched on'. Such inducible reactive groups have been popular tools in chemical biology for decades and include alkyne/azide click chemistry (Patterson *et al.*, 2014; Parker & Pratt, 2020) and photoreactive probes (Murale *et al.*, 2016). Chemical groups such as benzophenone and diazirine are sensitive to ultraviolet light, so that upon exposure they will convert into highly reactive intermediates attacking anything in their immediate vicinity. Thus, if irradiated while bound, a chemical probe containing such a photoreactive group will crosslink to its target protein. The resulting complex can then be detected by MS. This approach can identify and quantify probe-target interactions in complex mixtures or even *in vivo*. Combining this approach with fragments seems counterintuitive since fragments are not very specific probes, but it was recently shown that photoreactive fragments can actually have clearly selective target profiles in human cells (Parker *et al.*, 2017; Wang *et al.*, 2019). Using proteomics methods, the researchers showed that a small set of fragment probes formed *bona fide* interactions with many proteins in human cell culture, each fragment having a different selectivity profile. They could also determine approximate binding sites for some protein-fragment pairs, show functional modulation of two high-occupancy protein-fragment pairs, and apply their methodology to find adipogenesis-promoting target-fragment pairs by phenotypic screening. The work is a remarkable

example of using fragments in a phenotypic setting for target discovery, since FBLD has so far been an overwhelmingly target- and structure-based venture.

This thesis, too, presents a fragment-based approach towards inhibitors of a defined target using biophysical, structure-based methods. However, the target in this case is not a single biomolecule, but a molecular machine consisting of many different parts which help to perform a single function. This introduces a phenotypic aspect to the project that requires novel ways of thinking and problem-solving. The structural and functional characteristics of this molecular machine will be described in the next section.

## 1.2 The replisome

The biomolecular target of an FBLD campaign is usually predefined and is usually a single polypeptide or -nucleotide. Non-fragment-based campaigns, requiring less or no atomic-resolution structural information, may also simply target a whole cell-level or organism-level phenotypic outcome. The target of this project lies somewhere in between the two, as it is a set of proteins that have different structures and functions yet work together to produce a single measurable outcome. The nuances will be described in this section.

The target of this project is the replisome of *E. coli*. ‘Replisome’ is a loose term for the body of proteins that are directly involved in replication of DNA. Unlike other molecular machines, *e.g.* ribosomes or proteasomes, replisomes are not stable complexes, but sets of dynamic and transiently interacting subcomplexes. Genetically, replisomes can be classed as either eukaryotic/archaeal or prokaryotic/viral. Although there is no evolutionary conservation between the two classes, most replisome subunit functions and even some structures are highly similar. The basic DNA replication mechanism consists of three steps: initiation, elongation, and termination.

The initiation step is a separate process, mediated by a different set of proteins called the primosome in bacteria or the pre-replicative complex in eukaryotes, that melts the origin of replication and loads a replisome on each template strand.

The elongation step replicates the whole template, producing one copy, and is mediated by the replisome and various accessory proteins such as gyrases and ligases. *In vivo*, additional essential factors are those that remove blocks on DNA, repair

lesions, and restart replisomes when they stall. Both template strands are replicated simultaneously. DNA duplexes are anti-parallel, and all known replicative polymerases can only extend polynucleotides in the 5'-3' direction. It follows that synthesis of the lagging strand happens discontinuously in small steps as the replication fork moves forward. This mechanism may seem unnecessarily complex compared to simply replicating two leading strands, yet all life replicates dsDNA this way. It requires dedicated priming, distributing, ligating, and stabilising proteins to guide the process, but the extra work is preferable to leaving one template strand exposed as ssDNA for long periods of time. An alternative solution would be to use a reverse polymerase, capable of elongating in the 3'-5' direction. However, while examples of such enzymes exist (Jackman *et al.*, 2012), the absence of a processive reverse polymerase for templated lagging strand elongation implies that there are serious disadvantages that prevent evolution of such mechanisms (Ballanco & Mansfield, 2011). The two possible mechanisms of 3'-5' polymerisation are or were not efficient enough for use as an alternative to discontinuous lagging strand replication: nucleophilic attack of a 5'-triphosphate primer end by an incoming 3'-OH would stall in case of spontaneous triphosphate hydrolysis or of mismatch repair, and nucleophilic attack of an incoming 3'-triphosphate by a 5'-OH primer end would require additional machinery to synthesise and maintain sufficient levels of 3'-NTPs that are much less stable than 5'-NTPs. Furthermore, other (unknown) causes may have contributed to a situation in which 5'-3' polymerisation was the only viable mechanism for very early life forms, leaving only this mechanism available for inheritance (founder effect). Thus, discontinuous lagging strand synthesis is now the only known mechanism of dsDNA replication.

The termination step involves halting and disassembly of the replisome, then filling any small remaining gaps and disentangling the terminal region by mechanisms that are not yet fully elucidated but involve topoisomerases and recombination repair proteins. Thus, the replisome itself is dedicated solely to the simple task of elongating primers on a dsDNA template. Replisomes are the fastest and most accurate polymerase systems known, with typical synthesis rates of 500-1,000 bases per second and error rates of one mismatch per  $10^8$ - $10^{10}$  base pairs. How the *E. coli* replisome achieves this will now be described.

### 1.2.1 Characteristics of the whole replisome

There are many PPIs within the replisome, with the clamp loader complex being the central hub that keeps the polymerases associated with the replication fork (Figure 1.2). Several proteins are stably complexed during replication, but most interactions between replisome subunits are weak, transient, and stochastic. For example, *E. coli* primase has a binding affinity of approximately 1-5  $\mu\text{M}$  for helicase (Oakley *et al.*, 2005) and slightly stronger than 1  $\mu\text{M}$  for ssDNA (Swart & Griep, 1995). Thus, replisome activity is mainly dependent on local subunit concentrations. The enzymatic activities of many subunits are dependent on other replisome components, so that their activities while isolated from the replisome are much lower to the point of becoming inactive. This characteristic is also important during replication as a way of keeping the replisome together. For example, the leading strand polymerase can randomly pause and lag behind the replication fork, at which point the helicase becomes less active, giving the polymerase time to catch up (Graham *et al.*, 2017). As a result, the replisome is a flexible and robust molecular machine with the highest efficiency of any nucleic acid replication system.

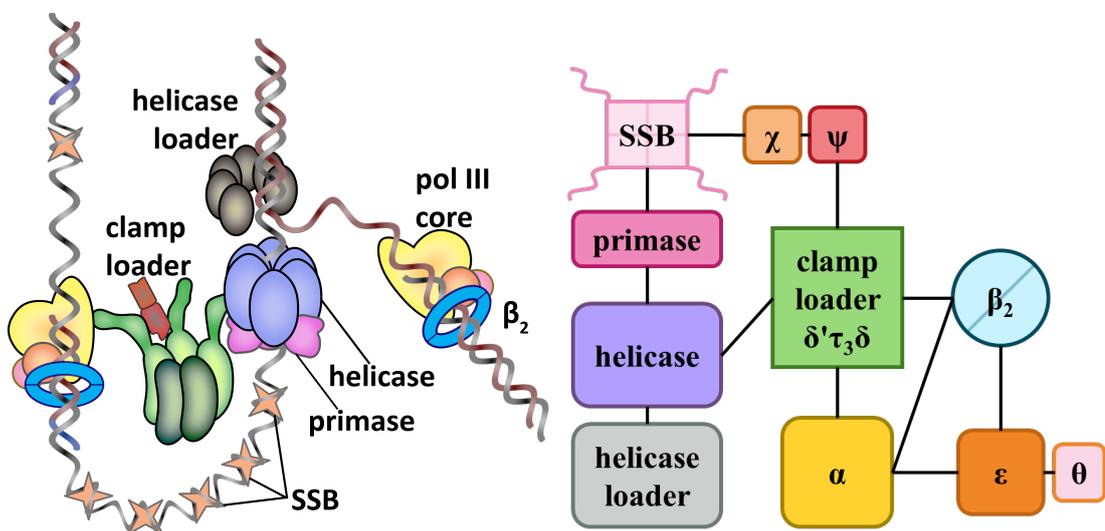


Figure 1.2. Schematic representation (left) of a snapshot of an active replisome at a replication fork (leading strand DNA template: red, lagging strand DNA template: grey, RNA primers: blue). All components are also listed in Table 1-1. Right: PPIs within the replisome. All do not take place at the same time.

### 1.2.2 The polymerase holoenzyme

The *E. coli* replisome consists of a central large protein complex called the polymerase III holoenzyme (pol III HE). There are several additional proteins that are essential parts of the replisome but are not part of this complex. Each single subunit of pol III HE is functional on its own but is extremely inefficient compared to most other enzymes with identical functions. Only when the whole replisome comes together, they become efficient. A functional pol III HE consists of 3 different complexes: the clamp loader containing seven subunits of which five unique, the pol III core containing three unique subunits, and the  $\beta$  sliding clamp (Table 1-1). These three subcomplexes are true structural complexes, while PPIs between them and with other replisome subunits are transient. The pol III HE is called the 'holoenzyme' based on early *in vitro* biochemical experiments using an ssDNA template (Hurwitz & Wickner, 1974; McHenry & Kornberg, 1977), without considerations for its native template or structural arrangements. It is the minimal set of proteins needed to achieve the typical high speed and processivity of the replisome, however without the helicase complex the pol III HE can only synthesise a leading strand. The functions and structures of each subunit are described in detail in the next sections.

Table 1-1. *E. coli* replisome subunits.

Subcomplex	Gene name	Gene product name	Function	Stoichiometry
<b>Pol III core</b>				<b>1-3</b>
	<i>dnaE</i>	$\alpha$ (pol III)	polymerase	1
	<i>dnaQ</i>	$\epsilon$	proofreading exonuclease	1
	<i>holE</i>	$\theta$	unknown	1
<b>Clamp loader</b>				<b>1</b>
	<i>dnaX</i>	$\tau/\gamma$	structural; ATPase	3
	<i>holA</i>	$\delta$	structural	1
	<i>holB</i>	$\delta'$	structural; clamp loader	1
	<i>holC</i>	$\chi$	structural	1
	<i>holD</i>	$\psi$	primer site recognition	1
<b>Sliding clamp</b>	<i>dnaN</i>	$\beta$	Anchoring of pol III core	2
<b>Helicase complex</b>				<b>1</b>
	<i>dnaB</i>	DnaB	helicase	1 hexamer
	<i>dnaG</i>	primase	primase	2-3
<b>Helicase loader</b>	<i>dnaC</i>	DnaC	helicase loader	<b>1 hexamer</b>
<b>SSB</b>	<i>ssb</i>	SSB	ssDNA binding	<b>many tetramers</b>

### 1.2.2.1 Pol III core

Pol III core is the heterotrimer of  $\alpha$ ,  $\epsilon$ ,  $\theta$  (Figure 1.3). The  $\alpha$  subunit is the replicative polymerase. Its catalytic core has a classic right hand-shaped structure (Lamers *et al.*, 2006). While the structure of the ~200 amino acid (aa) C-terminus is not known at atomic resolution, its sequence, its structure of a bacterial homologue (Liu, Lin *et al.*, 2013), and medium-resolution cryo-EM structure (Fernandez-Leiro *et al.*, 2015) indicate that it is composed of a flexible  $\beta$ -binding peptide followed by an oligonucleotide-binding (OB) domain. This C-terminus is essential for  $\alpha$  to gain its boosts in speed (~1 kb/s (Bloom *et al.*, 1997)) and processivity through binding to  $\beta$

and  $\tau$ . In addition, private observations show that even as a monomer, full-length  $\alpha$  is faster than its catalytic core alone, possibly because the OB fold increases the affinity of  $\alpha$  for ssDNA. The  $\alpha$  N-terminal polymerase/histidinol phosphatase (PHP) domain may have been a proofreading exonuclease domain earlier in evolution (Stano *et al.*, 2006; Barros *et al.*, 2013). Its current function is suggested by some to be pyrophosphatase (Fabio Lapenta *et al.*, 2016), though that hypothesis was challenged by others earlier (Lamers *et al.*, 2006).

In many Gram-positive bacteria, the *dnaE* gene product is also an essential polymerase of the replisome, but another polymerase (PolC) is responsible for the greater part of DNA synthesis (Dervyn *et al.*, 2001; Sanders *et al.*, 2010; Rannou *et al.*, 2013).

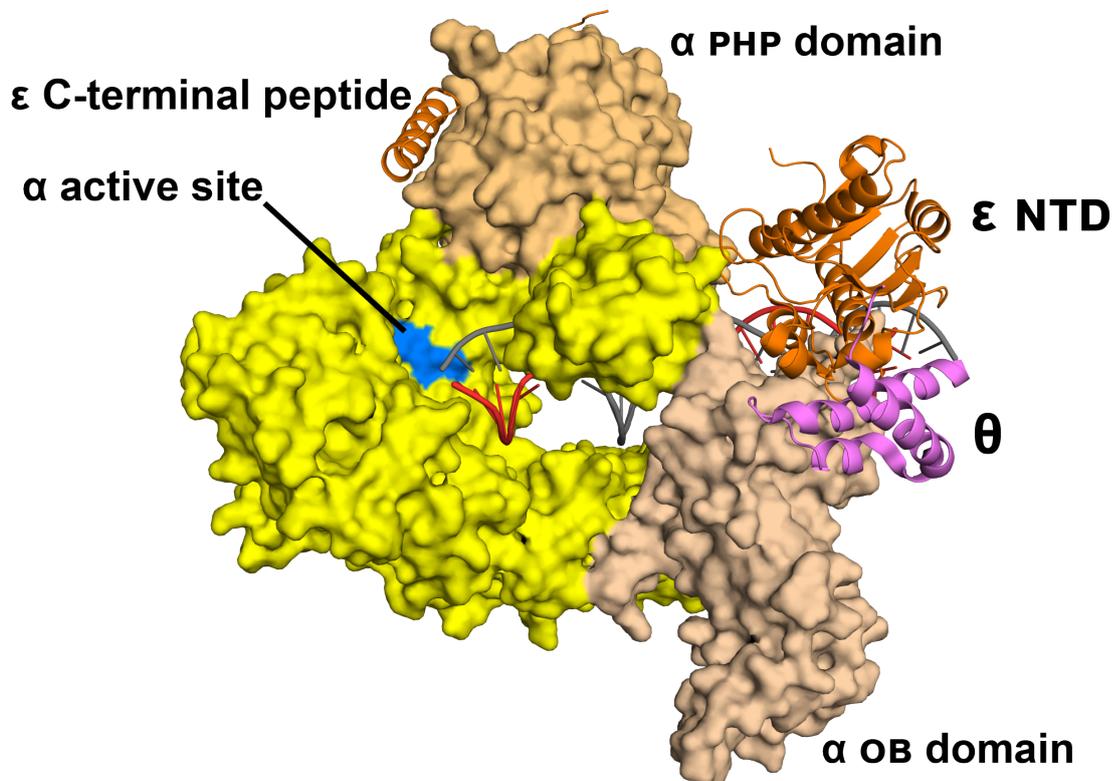


Figure 1.3. Model of the pol III core in complex with DNA. **Subunit  $\alpha$**  (surface representation) consists of an N-terminal PHP domain (beige), a polymerase domain (yellow) that wraps around dsDNA (red template strand, grey nascent strand), and an OB domain (beige). **Subunit  $\epsilon$**  (orange secondary structure representation) consists of an N-terminal exonuclease domain and a long, partially disordered C-terminal peptide that binds to both  $\alpha$  and  $\beta$ . **Subunit  $\theta$**  (pink secondary structure representation) binds to  $\epsilon$ . Model compiled from a cryo-EM structure of an  $\alpha$ - $\epsilon$ - $\beta$ -DNA complex (5fkv (Fernandez-Leiro *et al.*, 2015)) aligned to an NMR model of the  $\epsilon$ - $\theta$  complex (2xy8 (Schmitz & Bonvin, 2011)).

The actual proofreading in the pol III core is carried out by  $\epsilon$ . This 3'-5' exonuclease has an N-terminal catalytic domain dependent on two bivalent cations (Hamdan *et al.*, 2002), and a C-terminal helix that binds the  $\alpha$  PHP domain (Bressanin *et al.*, 2009) (Figure 1.3). Between the N- and C-terminus, a long flexible linker peptide contains a  $\beta$ -binding peptide like  $\alpha$  does. Interestingly, the crystal structure of the  $\epsilon$  exonuclease domain (Hamdan *et al.*, 2002) shows density for two nucleotides at sites where it is hard to imagine a DNA strand could be positioned based on the overall structure of  $\beta$ -pol III core (Fernandez-Leiro *et al.*, 2015). More recently it was shown that, in fact, these densities overlap almost perfectly with the position of a mismatched DNA strand in the pol III core complex (Fernandez-Leiro *et al.*, 2017). In addition to its proofreading function,  $\epsilon$  also has a stabilising effect on  $\alpha$  (Jergic *et al.*, 2013).

The small, non-essential  $\theta$  subunit has a large binding interface to  $\epsilon$  so that  $\epsilon$  sits in between  $\alpha$  and  $\theta$  in the pol III core (Schmitz & Bonvin, 2011) (Figure 1.3). Other than a slight stabilising effect (Taft-Benz & Schaaper, 2004), it may have functions outside the replisome (Dietrich *et al.*, 2014).

There are no known inhibitors of pol III core. Recently, the natural product nargenicin has been reported to bind to and inhibit  $\alpha$  (Painter *et al.*, 2015) based on analyses of mutations in the *S. aureus* homologue. However, we have been unable to show binding of nargenicin to *E. coli*  $\alpha$  by thermal shift analysis, nor consistent inhibition in our polymerisation assay. Additionally, the proposed binding region of nargenicin is based on a single aa mutation (S765L) which has however been reported earlier to have a stabilising effect as a rescue of *dnaQ* deletion (Lancy *et al.*, 1989). Therefore, it is possible that nargenicin has a destabilising effect on pol III HE that is not caused by binding to  $\alpha$ .

Within the pol III core, small molecule-mediated inhibition of either of the two active sites is possible. However, it might be difficult to develop fragments into specific binders as these sites are intrinsically non-specific to nucleotides. Disruption of PPIs within pol III core or between pol III core and the sliding clamp or the clamp loader might be more promising.

### **1.2.2.2 Clamp loader**

Loading and unloading of the sliding clamp on the lagging strand template-primer duplex happens approximately 2,000 times per replication cycle, since the *E. coli*

genomic DNA is ~4.7 Mbp and the average Okazaki fragment is ~1 kb. The clamp loader is a heteropentameric ring consisting of  $\delta'\tau_3\delta$  (in counterclockwise order when N-termini point towards the viewer) (Figure 1.4). The three subunits have very similar folds, despite low sequence similarity. Any number of  $\tau$  subunits can be replaced by  $\gamma$ , the gene product of an alternative *dnaX* transcript that encodes only the first three out of five protein domains (Flower & McHenry, 1990; Tsuchihashi & Kornberg, 1990; Blinkova *et al.*, 1993). Domains I and II are responsible for ATPase activity, and domain III is the oligomerisation domain. In contrast to most related ATPases, domains I of  $\delta'$  and  $\gamma/\tau$  contain zinc-binding C4 motifs of unknown function. Domain I of  $\delta$  contains a  $\beta$ -recognition motif. Domains IV and V of  $\tau$  bind to replisome subunits DnaB and  $\alpha$ , respectively. Therefore, clamp loaders containing only  $\gamma$  can load and unload sliding clamps but cannot form part of the replisome. There is enough space inside the clamp loader ring to bind a primed DNA template, which can enter the ring laterally through a gap between  $\delta'$  and  $\delta$ . The structure of the open clamp loader ring (Simonetta *et al.*, 2009) resembles that of other AAA+ family proteins, with ATP-binding sequence motifs present in the  $\tau$  N-terminal domains so that there are three functional ATP-binding sites on the interfaces between the five subunits. A possible mechanism of action for clamp loading is the following series of events: 1) ATP binds, 2) the clamp loader recognises and opens  $\beta_2$ , 3) ssDNA binding near a primed site induces conformational change reminiscent of screwing from the flat symmetrical apo structure (Jeruzalmi, O'Donnell *et al.*, 2001) to one where the  $\tau$  and  $\delta$  subunits shift out of the plane, their NTDs sliding along the primed DNA backbone (Figure 1.4), and 4) ATP is hydrolysed,  $\beta_2$  closes and the clamp loader is released. Clamp unloading does not require ATP hydrolysis, and can be catalysed by free  $\delta$  (Leu *et al.*, 2000) or by free clamp loaders that are not part of the active pol III HE.

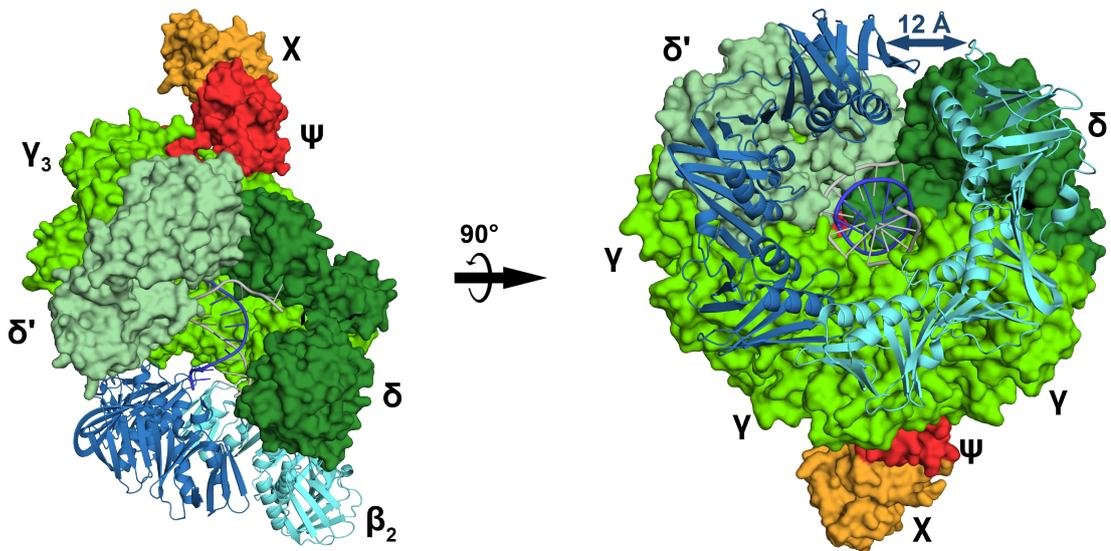


Figure 1.4. Model of the ATP- and DNA-bound state of the *E. coli* clamp loader in complex with an open sliding clamp, representing step 3) described in the text. The  $\gamma$  complex (surface representation, different shades of green for different gene products) wraps around the DNA template strand (grey) primed with a 10-base DNA primer (blue), placing an open  $\beta_2$  dimer (cyan and blue) slightly ahead of the primer. Model compiled from crystal structures of the  $\gamma$  complex with bound DNA,  $\psi$  N-terminal peptide, and ADP-BeF<sub>3</sub> (3gli (Simonetta *et al.*, 2009)),  $\delta$  domain I in complex with  $\beta$  (1jql (Jeruzalmi, Yurieva *et al.*, 2001)), the closed  $\beta_2$  dimer (4k3s (Yin *et al.*, 2013)), and the  $\chi\psi$  complex (3sxu (Marceau *et al.*, 2011)).

A second, more important function of the clamp loader is its role as a central distributive interaction hub. The  $\tau$  subunit C-terminus can bind both helicase and  $\alpha$  (Dallmann, 2000). Although observations of multimeric pol III core have been made long ago (McHenry, 1982), and structures of the  $\tau$  pol III binding CTD are known (Fernandez-Leiro *et al.*, 2015; Liu, Lin *et al.*, 2013), no structures of pol III HE or any combination of two of three major bacterial replisome subcomplexes (pol III core, clamp loader, and helicase) have been observed. This is not surprising, since pol III HE must be highly dynamic, many PPIs are transient depending on which function is being carried out, and most interaction domains are connected through flexible linkers.

An additional distributive function of the clamp loader is mediated by the non-essential subunits  $\chi$  and  $\psi$ . The  $\psi$  subunit serves as a linker between the clamp loader and  $\chi$ . Its N-terminal peptide (aa 2-28) binds the pentameric clamp loader ring with nanomolar affinity by inducing a  $\beta$ -strand alignment with the  $\gamma$  C-terminus (Simonetta *et al.*, 2009), and C-terminal side of  $\psi$  has a large interaction surface with  $\chi$  (Figure 1.4). Opposite this interaction surface,  $\chi$  has a binding site for the highly conserved

SSB C-terminal peptide (Marceau *et al.*, 2011). This SSB C-terminus is a recognition site for many DNA-binding proteins (Shereda *et al.*, 2008), including one other component of the replisome, primase (Naue *et al.*, 2013). This suggests that  $\chi$  and  $\psi$  enable the clamp loader to recognise primed lagging strand template sites.

### **1.2.2.3 Sliding clamp**

Sliding clamps have similar function and shape in all organisms, despite the lack of sequence similarity (Indiani & O'Donnell, 2006). Homodimers (in case of *e.g. E. coli*) or -trimers (in case of *e.g. humans*) form a doughnut shape around dsDNA with a non-specific positively charged inner surface. Sliding clamps serve as anchor points for many different proteins that operate on DNA. *E. coli*  $\beta$  monomers have sub-nanomolar affinity for each other, so that even without DNA,  $\beta$  is mostly dimeric in solution. Its binding half-life on dsDNA has been reported to be between 1-2 h (Yao *et al.*, 1996; Leu *et al.*, 2000).

The  $\beta_2$  dimer of *E. coli* has one binding pocket per monomer, to which replisome proteins  $\epsilon$ ,  $\delta$ ,  $\alpha$ , and various other proteins such as ligase and non-replicative polymerases can bind. It is also the target site of a few reported replication inhibitors, like the auto-inhibitory bacterial gene product SocB (Aakre *et al.*, 2013), the bacterial natural product griselimycin (Kling *et al.*, 2015), and several fragments recently identified by crystallographic screening (Yin *et al.*, 2014) and virtual screening (Pandey *et al.*, 2017). The sliding clamp is clearly a good antibacterial target, though not a novel one.

### **1.2.3 Helicase complex**

The DnaB replicative helicase sets the pace of the replisome and is one of only two replisome components that do not exchange throughout a normal replication cycle (the other being the leading strand sliding clamp). During replication initiation, DnaB forms part of the primosome complex together with the helicase loader DnaC, the primase DnaG, and other proteins. During elongation, DnaB defines the location of the replication fork, forming transient complexes with the clamp loader and primase (Figure 1.5).

### **1.2.3.1 DnaB helicase**

The DnaB helicase is a homohexamer with moderate sequence conservation between bacteria. There are no high-resolution structures of the *E. coli* helicase, but several crystal structures from Gram-positive homologues are known. The inactive structure is a flat ring with six fold symmetry of the CTDs but threefold symmetry of the NTDs (Wang *et al.*, 2008; San Martin *et al.*, 1995). Its CTDs interact with  $\tau$  (Gao & McHenry, 2001) and the helicase loader (Liu, Eliason *et al.*, 2013), while its NTDs interact with primase (Bailey *et al.*, 2007) and the replication initiation factor DnaA (Seitz *et al.*, 2000). Helicase moves along the lagging strand template (5'-3' direction), using ATP hydrolysis to displace the leading strand template with its CTDs. Its active form has a spiral conformation (Itsathitphaisarn *et al.*, 2012; Arias-Palomo *et al.*, 2013).

### **1.2.3.2 DnaC helicase loader**

The helicase loader, DnaC in *E. coli*, belongs to the AAA+ family and has a mechanism of action similar to that of the clamp loader. However, DnaC is a monomer in its inactive solution form (Galletto *et al.*, 2004) and a homohexamer in its active, ATP- and DnaB-bound form. Helicase loading has two different mechanisms depending on phylum: loading of a preassembled helicase hexamer by one protein (*e.g. E. coli* DnaC (Arias-Palomo *et al.*, 2019)) or loading and assembly of six monomers by two proteins (*e.g. Bacillus subtilis* DnaB/DnaI (Velten *et al.*, 2003)). In the first mechanism, the DnaBC complex twists into a spiral conformation, opening a crack through which ssDNA can diffuse into the central channel. DNA-binding then triggers ATP hydrolysis. Some researchers have observed that loader release in *E. coli* requires primase or RNA primer binding (Makowska-Grzyska & Kaguni, 2010). DnaC ATP hydrolysis thus perhaps only locks the complex on ssDNA but does not trigger loader release.

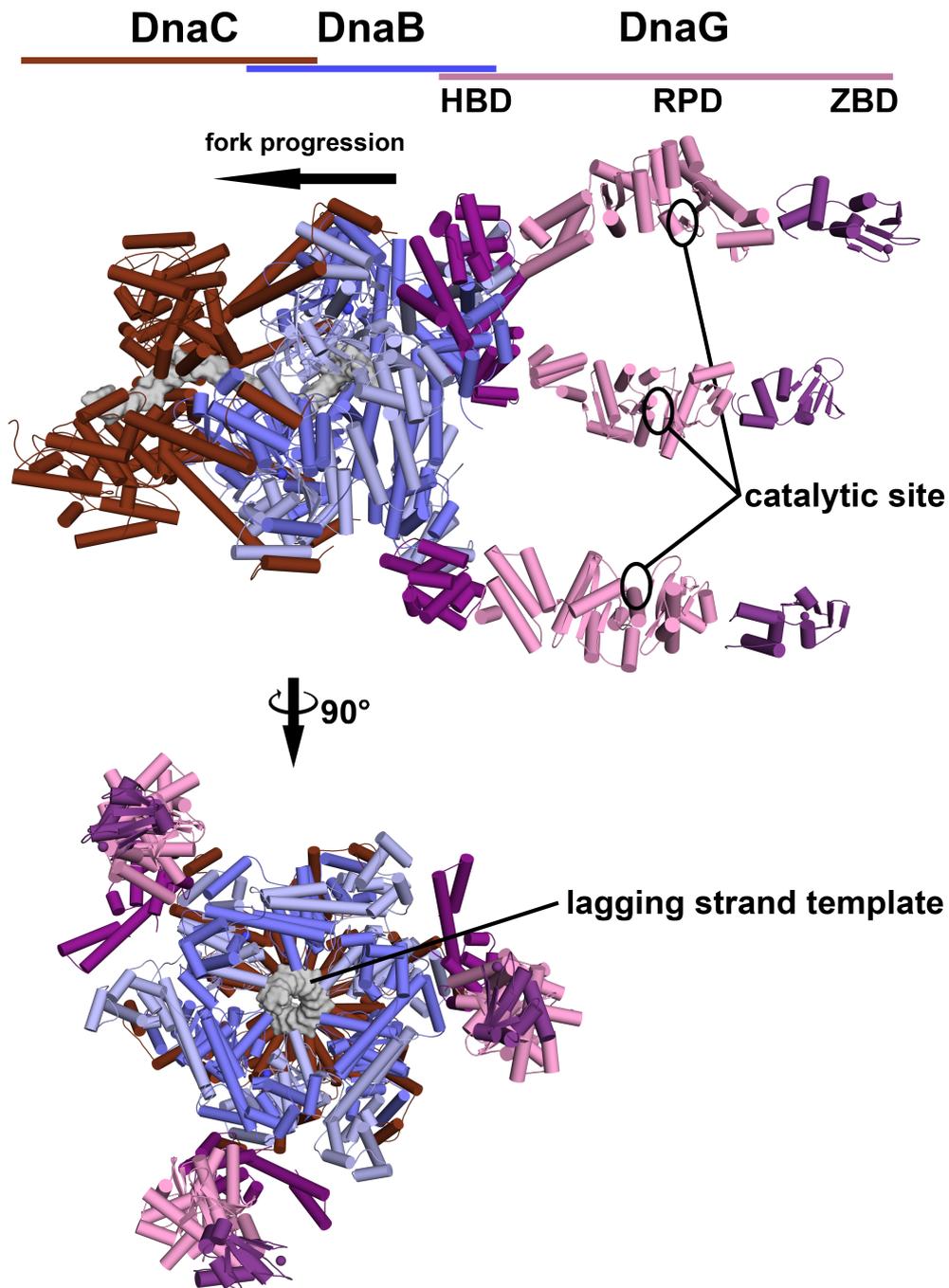


Figure 1.5. Model of the *E. coli* DnaB helicase (blue) in complex with DnaC helicase loader (brown), DnaG primase (purple & pink), and ssDNA (surface representation, grey). The DnaBC-ssDNA model (6qem (Arias-Palomo et al., 2019)) is in the closed state after DnaC ATP hydrolysis. Primase models based on structures of its three domains: helicase-binding (2haj (Oakley et al., 2005)), RNA polymerase (1dde (Keck et al., 2000)), and zinc-binding domain (homology model derived from 1d0q (Pan & Wigley, 2000)). Relative positioning of the helicase NTD and primase HBD based on the homologous complex of *Geobacillus stearothermophilus* DnaB and DnaG HBD (2r6c (Bailey et al., 2007)).

### 1.2.3.3 *DnaG primase*

The helicase NTDs interact with DnaG primase, so that a maximum of three primase molecules can be bound to the helicase at any one time (Figure 1.5). Primase consists of three flexibly linked domains that have different functions to increase the overall activity of the enzyme. Primase binds to helicase through a helical hairpin structure of its C-terminal helicase-binding domain (HBD) (Su *et al.*, 2006). In addition, the HBD has an SSB peptide-binding site on the surface of its small,  $\alpha$ -helical hydrophobic core sub-domain. The helical hairpin of some homologues, including *E. coli* (Oakley *et al.*, 2005), is mostly observed as ‘extended’ with respect to the hydrophobic core, forming a single  $\sim 30$  aa  $\alpha$ -helix. Other homologues show a break in the middle of this helix, adopting a conformation that is likely unable to bind helicase (Catazaro *et al.*, 2017). Possibly, there is a common mechanism by which binding of an effector switches the conformation from active to inactive or *vice versa*. The middle, large catalytic domain, the RNA polymerase domain (RPD), synthesises a short (10-12 nucleotides in *E. coli*) RNA primer. The open, solvent-accessible catalytic site is formed by two subdomains: the conserved topoisomerase-primase (Toprim) domain coordinates two or three  $Mg^{2+}$  ions (Pansegrau & Lanka, 1992; Godson *et al.*, 2000; Aravind, 1998), while the other subdomain presumably discriminates the sugar and provides a basic ridge for template binding. The primase N-terminal zinc-binding domain (ZBD) is required for recognition of priming initiation sites, which are different for each organism (3'-GTC for *E. coli* (Yoda & Okazaki, 1991)). After primer synthesis, primase remains attached to the primed site, presumably to prevent primer degradation by other proteins (Yuzhakov *et al.*, 1999), until the sliding clamp and pol III core are loaded onto the primer. Since both primase and the clamp loader bind the SSB C-terminal peptide, competition for SSB binding is a likely mechanism of securely replacing primase by sliding clamp on the RNA primer. Interestingly, the T7 primase and helicase are part of the same polypeptide, causing the lagging strand template to trail behind the helicase, creating a loop until the primer is released (Hamdan *et al.*, 2009). It is likely that the same ‘trombone’ mechanism (Alberts *et al.*, 1983) also occurs in *E. coli* (Tougu & Marians, 1996).

Many compounds inhibit helicases and primases, but most of these are either non-specific or cytotoxic. Examples are the natural product myricetin which inhibits DnaB helicase (Griep *et al.*, 2007), and the endogenous signalling nucleotide ppGpp which

inhibits primase (Maciąg *et al.*, 2010; Rymer *et al.*, 2012). High-throughput screening against Gram-positive helicases has yielded a specific inhibitor (Aiello *et al.*, 2009; Li *et al.*, 2013), yet without any reported cellular activity. Recently, fragment screens by NMR identified fragments binding to an allosteric site on the T7 primase catalytic domain (Ilic *et al.*, 2016) and to the SSB peptide-binding site of the *E. coli* HBD (Chilingaryan *et al.*, 2018).

#### **1.2.4 Single-stranded DNA-binding protein and other proteins**

Single-stranded DNA-binding protein (SSB) functions as a general protecting agent of ssDNA. As this is also an essential part of chromosome replication, and SSB is required *in vitro*, it is mentioned here as part of the replisome. It is a homotetrameric protein with an oligomerising and DNA-binding NTD and a long (65 aa) disordered C-terminal peptide (Savvides *et al.*, 2004). The conserved six ultimate C-terminal residues (DDDIPF) serve as a tether for many different DNA-binding proteins, including primase and  $\chi$ . In the absence of bound ssDNA, this C-terminal tail is in fast exchange with the DNA binding sites (Shishmarev *et al.*, 2014). Small molecule mimics of this C-terminal peptide have been reported to bind to the *E. coli* SSB interaction partner ExoI (Lu *et al.*, 2010), but none have been developed as therapeutic agents.

Other proteins that are essential to bacterial DNA replication, but do not interact with the described replisome, are: certain ligases, nucleases and polymerases to correct template strand errors, remove RNA primers *etc.*; initiation (DnaA (Shimizu *et al.*, 2016)) and termination (Tus (Mulcair *et al.*, 2006)) proteins; proteins involved in stabilisation or reloading of stalled replisomes (McGlynn & Guy, 2008; Yeeles *et al.*, 2013); and topoisomerases, which relieve torsional strain in the DNA double helix generated by helicase. The latter class of enzymes are the targets of the classic quinolone and aminocoumarin classes of antibiotics, against which resistance is known.

### **1.3 Antibiotic targets**

As described in the previous section, there are few proteins within the replisome that have been proven to be good targets for antibiotics, even though all but one were found to be essential gene products (Baba *et al.*, 2006). The best-characterised antibiotic agent targeting the replisome is griselimycin, a natural product that targets the sliding

clamp. However, it should be noted that the sliding clamp serves as a general anchor point for many different proteins that operate on DNA, and thus also has functions outside the replisome.

Historically, natural products have helped us identify the best antibiotic targets for therapeutic intervention. The most important antibiotic target classes against which effective, natural product-derived antibiotics exist nowadays are ribosomes (targeted by *e.g.* tetracyclines, aminoglycosides), RNA polymerases (rifampicin), gyrases (aminocoumarins), tRNA synthetases, and various enzymes involved in cell wall synthesis ( $\beta$ -lactams) (Silver, 2013). Other, non-protein targets are cell wall peptidoglycans (vancomycin) and the Gram-negative outer membrane (polymyxins). Several of these have been targeted by different antibiotics independently, indicating that the functions of these targets are both essential and susceptible to interference. In some cases, target location (intracellular, periplasmic, or extracellular) also contributes to susceptibility. Intracellular targets are more difficult to reach because of the barely permeable bacterial outer membrane and peptidoglycan layers and because of efficient drug efflux pumps that were poorly characterised until recently. These two barriers are difficult to overcome especially when developing small molecule antibiotics *de novo*, noted by *e.g.* researchers at AstraZeneca (Tommasi *et al.*, 2015). Furthermore, in case of bacteria-derived natural product antibiotics (*e.g.* from *Streptomyces* screening), the ability of the producing organism to resist its own antibiotics is an additional factor affecting targetability. In a similar but more general sense, this may also be part of the reason why the replisome is not a major antibiotic target: DNA replication is too important and too similar in all forms of life to enable efficient and selective inhibition.

Most antibiotic natural products evolved to be effective in their own micro-environments, which is a very different type of use compared to human therapeutic interventions (Davies, 2006). Less than a century after the discovery of the first antibiotic, humans are now facing widespread resistance to classic antibiotics. These resistance mechanisms evolved in response to overuse in humans and livestock. Since this problem arose in human environments, it is less likely that natural products will be able to solve it for a second time. Our list of requirements for ‘good’ antibiotics (broad spectrum, high efficacy, high tolerance, ease of use) is not shared by the micro-organisms from which we derived most first-in-class antibiotics. Especially the

requirement that our antibiotics be effective against a broad spectrum of bacterial species limits the list of possible targets, as was found by researchers at GSK (among others) in the early 2000s (Payne *et al.*, 2007). As a result of these strict product requirements, discovery and development of antibiotics has been very slow over the past several decades. The most recent novel antibacterial compound class that is currently on the market (oxazolidinones) was discovered in the early 1980s (Slee *et al.*, 1987), although its target (the ribosome) is not new.

New drugs are clearly needed to fight bacterial infections and antibiotics resistance. However, new approaches are needed not only in drug discovery but also in the clinic and at the economic and political level. Novel antibiotics (even natural product-derived ones) are within reach, but they are unlikely to be as powerful as the antibiotics of the previous century. Furthermore, it is likely that there are more good antibiotic targets to be discovered rationally, without the help of natural products (for example  $\beta$ -lactamases). Narrow spectrum antibiotics, which only work against a few bacterial species, should be used only against infections with those species, requiring routine diagnostics in and outside the hospital. Furthermore, to prevent the emergence of pan-resistant bacterial strains, effective antibiotics should be used correctly and sparingly. To promote research and development of such products, different economic incentives are needed that do not depend on sales by volume (Roope *et al.*, 2019). Thus, discovery of novel antibiotics is only a small part of the solution to the current antibiotics resistance problem. Recent reports of public organisations such as the World Health Organization and many national government commissions including that of the United Kingdom (O'Neill, 2016) describe these strategic problems and solutions in more detail. However small the part of novel antibiotics research may be, it is clearly essential and still full of opportunities regarding discovery of new antibiotics against known targets as well as of new targets.

#### **1.4 Aims of the project**

*In vivo* replisome activity is influenced by many factors including chaperones, signalling molecules, and DNA damage. In fact, most information about how replisomes work comes from *in vitro* experiments with vastly simplified conditions. The *E. coli* replisome is one of few that have been successfully reconstituted *in vitro* (other examples are the replication machinery of bacteriophage T7 (Lee & Richardson,

2011; Gao *et al.*, 2019), *Bacillus subtilis* (Sanders *et al.*, 2010), and *Saccharomyces cerevisiae* (Georgescu *et al.*, 2014)). Comparisons with measurable *in vivo* parameters show that an *in vitro* replisome is a good, if difficult and complex, model system. Because of its essential function, the replisome is a good antibiotic target in principle (Baba *et al.*, 2006), and the availability of a reconstituted system makes it attractive for target discovery and validation. The single function of the replisome provides a good readout for screening assays. However, the fact that it is made up of no fewer than 28 proteins (of which 13 unique) gives it the characteristics of a phenotypic system requiring target identification.

This thesis will describe efforts to validate and deconvolute fragment screening results against the reconstituted *E. coli* replisome. The fragment-based approach and the replisome are an exciting combination because of the novelty of screening against such a complex system that is nevertheless well-characterised in terms of both function and structure.



## 2 Structural and functional characterisation of hits from a fragment screen against a bacterial replisome

### Preface

The author of this manuscript gratefully acknowledges the contributions of Dr. J.A.L. Howard (McGlynn group) to results presented in this chapter, *i.e.* the screening and initial confirmation of the York fragment library in the replisome assay described in section 2.1. The author also gratefully acknowledges the contributions of Dr. S.W.M. Tanley (Hubbard group) to results presented in this chapter, *i.e.* the production of  $\alpha$  and the testing of hits and analogues for inhibition of  $\alpha$ .

### 2.1 Introduction

The bacterial replisome is a collection of protein complexes that together replicate double-stranded DNA (dsDNA) (described in detail in section 1.2). The McGlynn group at the University of York is able to reconstitute a functional version of the replisome *in vitro*. Dr. J.A.L. Howard of the McGlynn group conducted an initial screen of the York fragment library for fragments that inhibited replisome activity. The work described in this chapter follows up on the results of this initial screen, which will first be summarised here.

The York fragment library was designed and maintained by the Hubbard group and contained 1,004 fragments that were selected from the commercial Maybridge Ro3 collection (see introduction section 1.1.1). Criteria for selection were to maximise chemical diversity based on molecular fingerprints (Hassan *et al.*, 2006) and 3D shape (Sauer & Schwarz, 2003). At the time of use, the library was formatted as 200 mM stocks in dimethylsulfoxide (DMSO) that had been kept in dark, ambient conditions for approximately two years.

The replisome against which the fragment library was screened consisted of thirteen different essential subunits: Pol III core complex ( $\alpha\epsilon\theta$ ), clamp loader complex ( $\tau\delta\delta'\psi\chi$ ), sliding clamp ( $\beta$ ), primase, helicase, helicase loader, and single-stranded DNA-binding protein (SSB). All proteins were produced from native *E. coli* K-12-derived genes cloned into bacterial overexpression vectors, except for the helicase loader. The helicase loader gene contained a mutation in codon 176 (known as

*dnaC810* (Xu & Marians, 2000; Sandler *et al.*, 1996)) that increases the activity of the enzyme, making the presence of other initiation factors (essential *in vivo*) no longer necessary.

This version of the replisome will replicate a dsDNA template but requires the template to have a pre-formed fork and leading strand primer because of the absence of initiation and priming factors (see introduction section 1.2). Here, an artificial template was used that satisfies these requirements: a circular, M13-based dsDNA plasmid containing a nick with single-stranded overhang, similar to M13 replicative form II (Pratt & Erdahl, 1968; Lin & Pratt, 1972). The overhang allows the loading of helicase, and the nick provides a primer for leading strand elongation. This template facilitates ‘rolling circle’ replication, which in the absence of control proteins will generate a linear leading strand of many times the length of the original circle. This generated leading strand also serves as the lagging strand template. The process is drawn schematically in Figure 2.1. Thus, once assembled, the replisome functions *in vitro* like it does during the *in vivo* elongation phase.

The complexity of the reaction makes it inefficient, and only a fraction of available template plasmids will contain stable, active replisomes. To still reliably quantify reaction products, a small amount of  $\alpha$ -<sup>32</sup>P-labelled deoxycytidine triphosphate (dCTP) substrate is added to the reactions. The P-32 nuclide undergoes  $\beta$ -decay into stable S-32 and high-energy electrons, providing a sensitive read-out that does not add any chemically or biologically unnatural parts to the reaction.

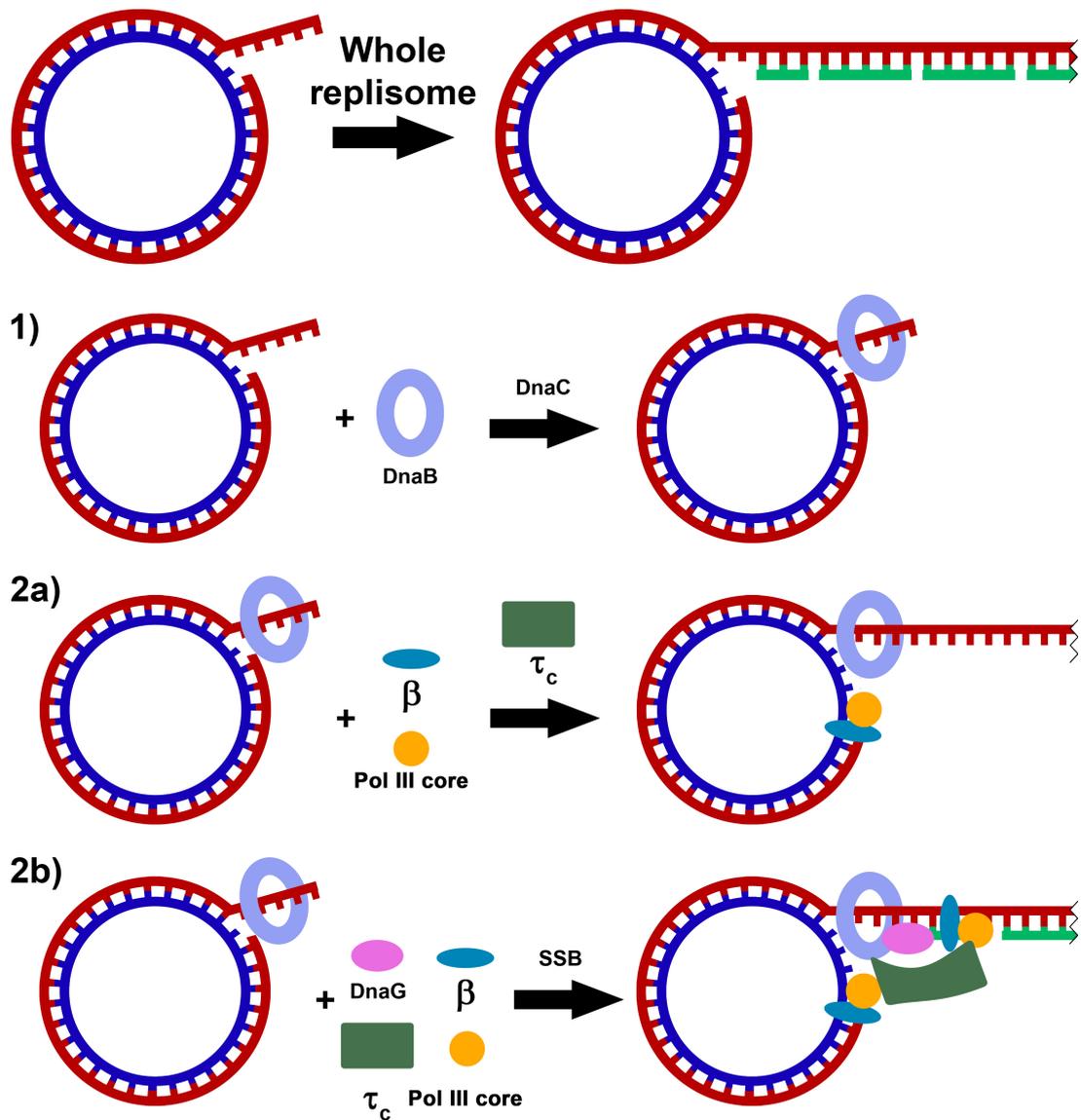


Figure 2.1. The M13 replicative form II template facilitates rolling circle replication, yielding a single leading strand (red) and a distribution of shorter lagging strand Okazaki fragments (green). The reaction used during screening contains all subunits listed in Table 1-1 and elongation proceeds as described in section 1.2. Briefly, three activities contribute to the observed final products: **1)** the DnaB helicase is loaded onto the M13 template by DnaC helicase loader. **2a)** processive leading strand elongation by the pol III core, kept in place by the  $\beta$  sliding clamp that was loaded by the  $\tau$  complex ( $\tau_c$ ) clamp loader. **2b)** discontinuous lagging strand synthesis initiated by de novo RNA primer synthesis by the DnaG primase. The  $\beta$  sliding clamp is loaded onto each new primer-template duplex by the  $\tau_c$  clamp loader which needs SSB to locate these new primers. The clamp loader also acts as an interaction hub, keeping the helicase and the two polymerases together.

Screening was carried out in tube-based format over the course of several months. Details of the screening procedure can be found in section 7.2. Briefly, together with each set of 10 fragments (0.5 mM) as singletons, a positive control (DMSO) and a negative control (omitting Pol III core) were run to determine maximum and background signals within each set. A control fragment inhibitor of the sliding clamp, 4-(*p*-fluorophenyl)benzoic acid, identified from literature (Yin *et al.*, 2014), was also tested each day. A plot of the control fragment response over the course of the screening campaign shows that there was no drift over time (Figure 2.2), indicating that assay conditions, enzymes, template, and reagents were stable over time.

The average replication in the presence of control fragment ( $75\% \pm 9$ ) was also used as a hit cut-off criterium. This yielded an initial hit set of 120 fragments (12%), a hit rate typical for fragment screening against enzymes. One likely cause of false-positive replisome inhibition is fragment binding to the template non-specifically. Therefore, the initial 120 hits were tested for DNA intercalation in a topoisomerase I assay which relaxes negatively supercoiled dsDNA. This identified 21 fragments that partially or completely inhibited supercoil relaxation, leaving 99 fragments as ‘true’ replisome inhibitors (Figure 2.3). Interestingly, many among the top inhibitors were found to be DNA intercalators.

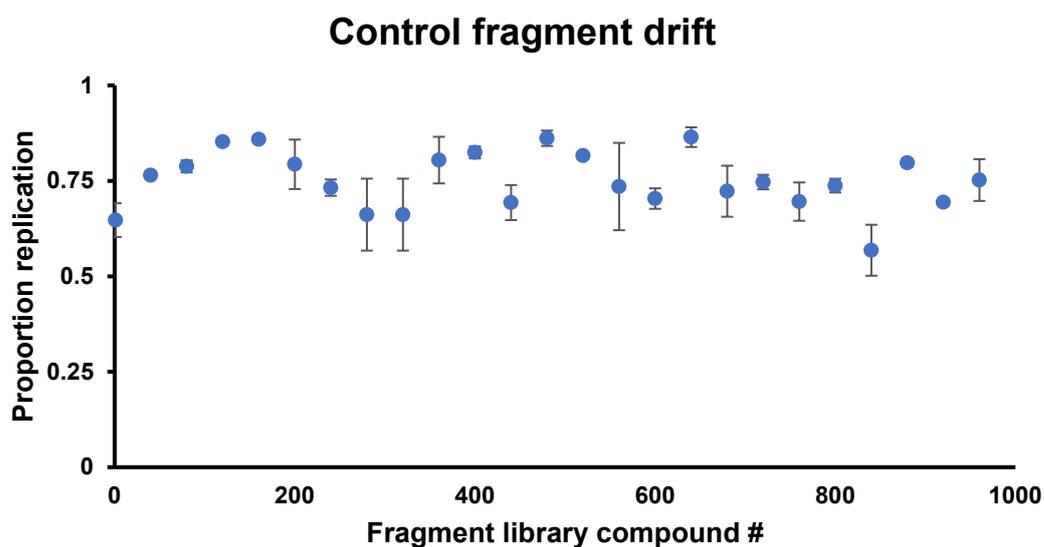


Figure 2.2. Inhibition of replication by the control fragment inhibitor at regular intervals during the screening of the 1,004-fragment library. The average replication in presence of the control fragment inhibitor was  $75\% \pm 9$ . No trends or anomalies in the drift of measured inhibition was observed, indicating no structural problems in screening workflow.

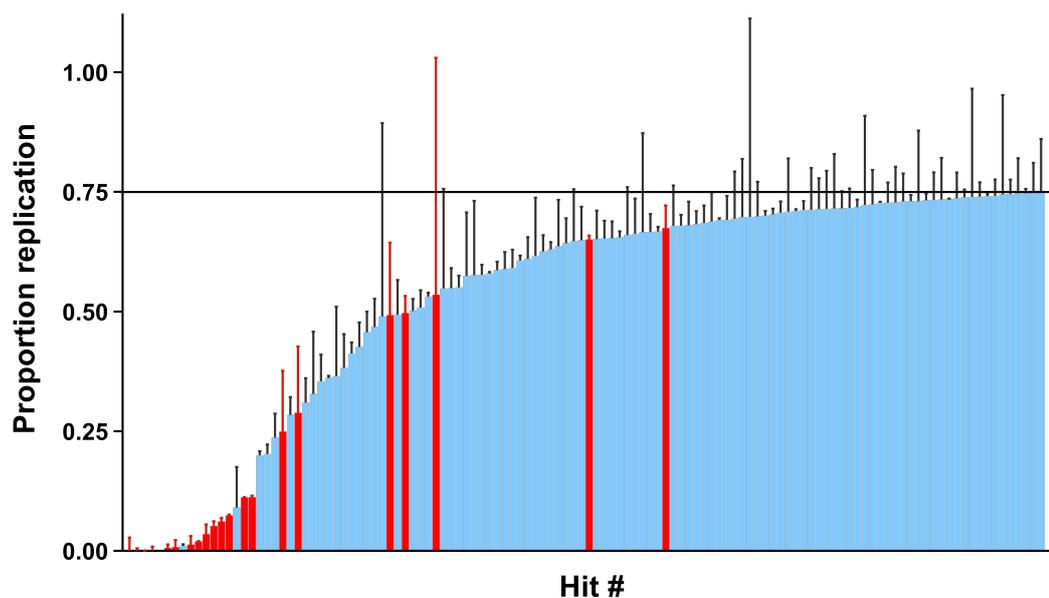


Figure 2.3. Chart of 120 initial fragment hits sorted by activity. The 21 DNA-intercalating hits are coloured red. A hit cut-off of 75% replication was used based on the control fragment inhibitor. Interestingly, most of the top 10-20 inhibitors are intercalators.

Eliminating intercalators leaves 20 hits that inhibit the replisome 50% or more, and 79 that inhibit the replisome between 25-49%. These top 20 hits, and five other, randomly chosen hits, were selected for confirmation and subsequent assays against individual replisome components (structures of these 25 fragments are given in Appendix A). For confirmation, the hits were tested against the whole replisome again at three different concentrations (20, 100, 500  $\mu\text{M}$ ) to check for concentration-dependent activity. Out of 25 tested hits, three did not show concentration-dependent inhibition and indeed less inhibition at the highest concentration than was initially observed during screening. Comparing the 0.5 mM data points from screening and confirmation assays (Figure 2.4), 14 out of 25 hits showed similar activities (within error), 10 showed less activity after confirmation, and one hit (MB63) did not show inhibition anymore. Five out of 25 tested hits fell below the hit threshold of 25% inhibition after retesting.

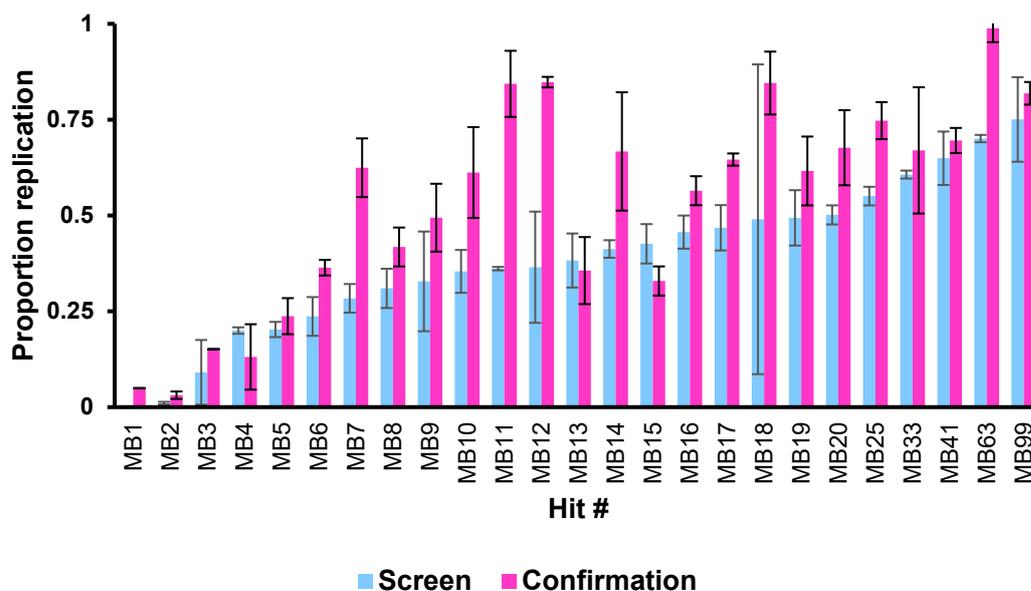


Figure 2.4. Comparison of hit activities against the whole replisome observed during screening and subsequent confirmation. Most hit activities could be reproduced.

Finally, these 25 selected hits were tested against various individual replisome components, namely: pol III subunit  $\alpha$ , pol III holoenzyme ( $\alpha\epsilon\theta\tau\delta\delta'\psi\chi\beta$ ), helicase, helicase + helicase loader, and helicase + primase (it must be noted that some of these assays, especially helicase and primase, were quite insensitive and therefore gave ambiguous results for some compounds). Most hits inhibited multiple of these assays, but four were active against only one replisome component. For example, hit MB19 inhibits the whole replisome and pol III holoenzyme, but not any other component (Figure 2.5). Interestingly, one hit (MB7) inhibited the whole replisome but not any of the tested individual components.

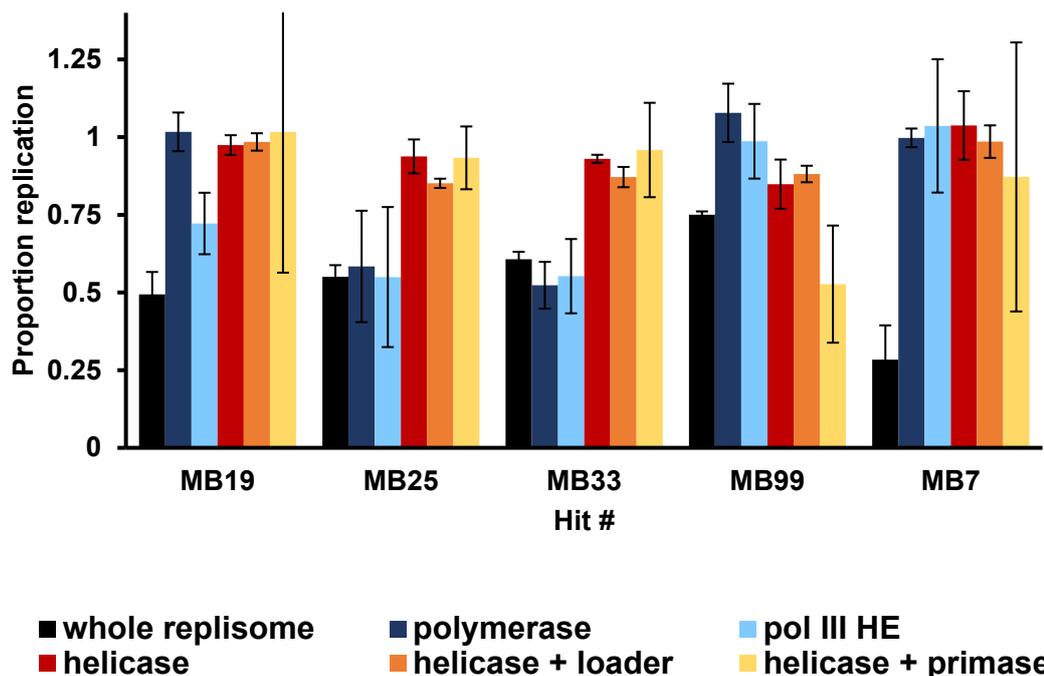


Figure 2.5. Activities of five hits in the whole replisome assay and in assays of different replisome components. Four hits are active against the whole replisome and one component of it, while hit MB7 is only active against the whole replisome.

## 2.2 Results

The screening results presented in the Introduction section were available at the start of this project. The rest of this chapter describes results from follow-up experiments aimed at identifying the most interesting hit compounds for which to investigate their mechanisms of action.

### 2.2.1 Compound quality controls

Every screening technique has different susceptibilities to yielding false-positive and -negative data, and biochemical assays are well-known to be sensitive to false-positive interfering compounds (Baell & Holloway, 2010). Furthermore, compound aggregation or precipitation are common causes of artefacts and false-positives in biochemical screening (McGovern *et al.*, 2003). As noted in section 2.1, the York library stocks used for screening were rather old. Organic compounds dissolved at high concentrations can precipitate or degrade over time, sometimes accelerated by water, oxygen, light, or high temperature (Keserú *et al.*, 2016). Therefore, the 25 selected fragments (see Appendix A) stocks were checked for compound integrity and

solubility. Compound integrity was assessed by proton nuclear magnetic resonance (NMR), using samples containing one fragment each at 0.5 mM in phosphate-buffered saline. Compound solubility was assessed by checking for precipitation during sample preparation, as well as by NMR (insoluble or aggregating compounds will show unexpected peak intensities, chemical shifts, or line widths in their 1D proton spectra). The WaterLOGSY experiment (used in section 2.2.3) can identify aggregators too, since the signal strength is dependent on the rotational correlation time of observed species (Dalvit *et al.*, 2006). Although the NMR buffer (sodium phosphate) was quite different from the replisome assay buffer (HEPES, nucleotides, BSA), compound solubilities were also monitored during the second set of confirmation assays (Figure 2.7) and were found to match the observations made during NMR sample preparations.

One compound contained precipitate in its DMSO stock (MB9). Six compounds (MB1, 2, 4, 14, 16, 20) precipitated upon dilution, four out of which also showed little to no signal in their spectra. Three further compounds (MB8, 17, 18) showed differences in their spectra compared to their references, which were recorded when the library stocks were first prepared. For example, the spectrum of MB18 compared to its reference spectrum showed a decrease in intensities of some peaks with simultaneous increase in total number of peaks, indicating the conversion of part of the original sample into one or multiple new chemical species (Figure 2.6).

It was decided to exclude these potential false positives from all further experiments, as there was a considerable number of hits that passed the control experiments and none were structurally related to those that failed. Interestingly, most problematic compounds in this screen cluster towards the top-ranked inhibitory compounds. Since half-maximum inhibitory concentrations ( $IC_{50}$ s) of fragments are typically in the 0.1-10 mM range, the highest inhibition that could be reasonably expected at a screening concentration of 0.5 mM would be roughly 80%. This trend perhaps serves as a reminder that true fragment hits are supposed to be weak inhibitors requiring elaboration.

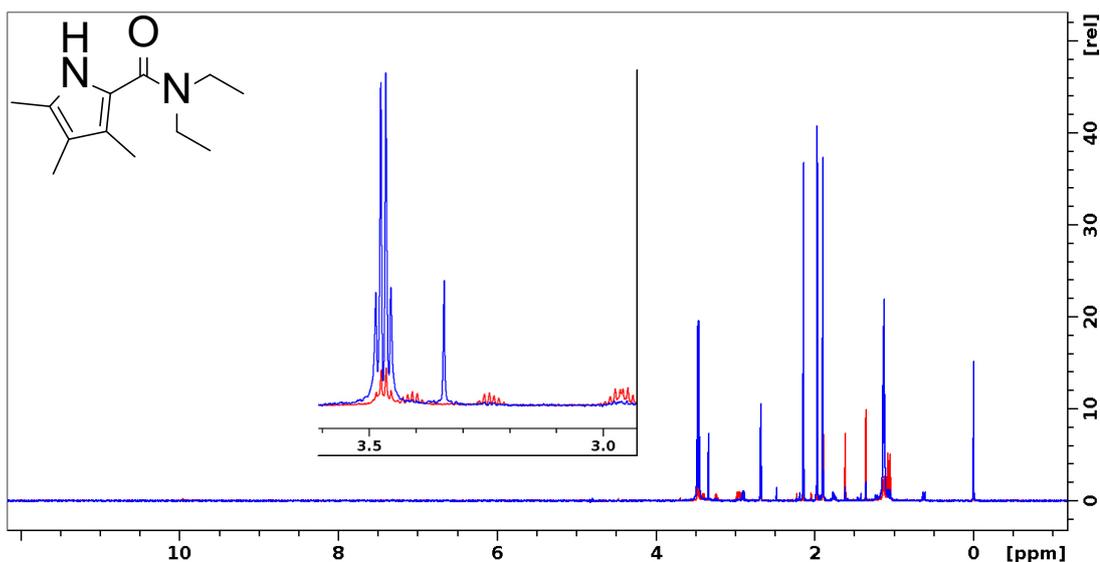


Figure 2.6. Quality control of fragment hit MB18. The initial  $^1\text{H}$  NMR spectrum (blue) contains all expected peaks with only minor impurities. The spectrum of the compound stock after screening (red) contains many additional peaks, suggesting compound degradation. Inset shows a small region of the spectra in more detail. The initial chemical structure is shown.

### 2.2.2 Confirmation of screening hits

Proton NMR analysis of compounds can identify significant changes to the compounds themselves, but not the presence of any minor or non-organic contaminants (*e.g.* metal ions) in DMSO stocks. To further assess whether these hits are true-positives, all initial hits (Figure 2.4) were repurchased, their stocks were made fresh and their quality was checked by NMR again. Reactions were repeated independently with both old and new compound stocks. While results for the control inhibitor fragment could be reproduced, results for many initial hits could not be reproduced (Figure 2.7). Several attempts were made, but no obvious causes were identified. The real causes might be related to subtleties in compound or protein handling, stock preparation, or work-up methods of replication reaction products. On the other hand, repetitions of the same experiment can also be useful as an indicator of hit robustness. For example, like the control fragment, MB25 behaved reproducibly in all experiments. Several other fragments with reproducible inhibition (MB1, 16, 17) had been confirmed as insoluble compounds from both old and new compound stocks. The only fragment hits that could be reconfirmed as potentially promising replisome inhibitors were MB3, 5, 6, 15, 25, 33. Remarkably, only four true positives targeted a

single replisome component (Figure 2.5). A fifth fragment (MB7) inhibited the replisome but was inactive against any of the six tested components. Although this hinted at the exciting possibility of binding to a type of interface site that is only formed when the whole replisome is active, MB7 was deprioritised because of non-reproducibility. Since the activities of MB25 and MB33 in different replisome component assays (Figure 2.5) suggest that each inhibits a specific subunit, these hits were marked as the most interesting hits to follow up.

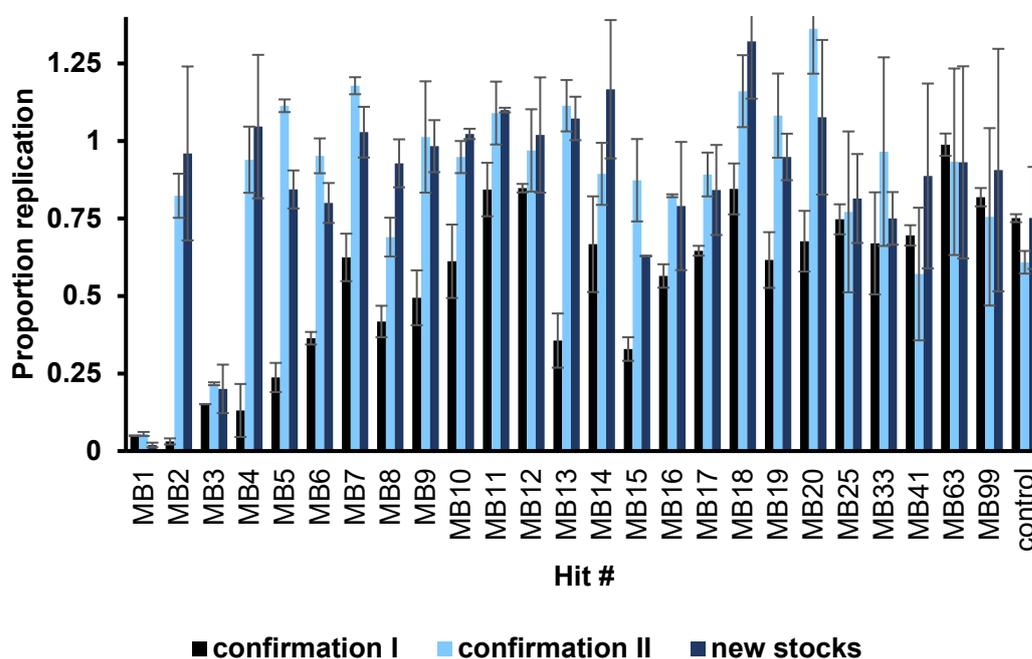


Figure 2.7. Comparison of hit activities between the original screen confirmation data (confirmation I, see Figure 2.4), independent confirmation reactions using the same compound stocks (confirmation II), and independent confirmation reactions using fresh compound stocks. Many hit activities could not be reproduced.

Next, the selected fragment hits shown in Figure 2.5 were assessed for concentration-dependent replisome inhibition. Reaction products (leading and lagging strands) were also separated on gel to check for hints of inhibition mechanisms. As expected, MB7 and MB19 did not appear to inhibit replication even at the highest tested concentration (5 mM). MB99 did show up to 48% inhibition, but still none at the concentration used for screening. MB25 and MB33 showed concentration-dependent inhibition up to 86%. Standard curves could be fitted to these titration data to estimate IC<sub>50</sub> values of

just over 1 mM (Figure 2.8). Interestingly, these two fragments show different effects on reaction product profiles. MB33 only appears to affect the overall amount of reaction products. MB25 additionally causes an increase in average size of lagging strand Okazaki fragment distributions (that is, the peak retention factor decreases; see bottom half Figure 2.8). This suggests that MB25, in addition to inhibiting the polymerase (Figure 2.5), is modulating a distributive component of the replisome (clamp loader, primase, SSB). The effect is most reminiscent of a loss of primase function: indeed, when titrating primase into the whole-replisome reaction, the opposite effect of decreasing Okazaki fragment size can be observed (Tougu *et al.*, 1994).

MB33 seemed to be a reproducible hit targeting the polymerase subunit  $\alpha$ , but unfortunately it was later found to be a likely false positive by others in the Hubbard group (discussed in section 2.2.4). Therefore, MB25 remained as the sole confirmed hit with reproducible activity, with MB99 showing signs of weak inhibitory activity at high concentrations.

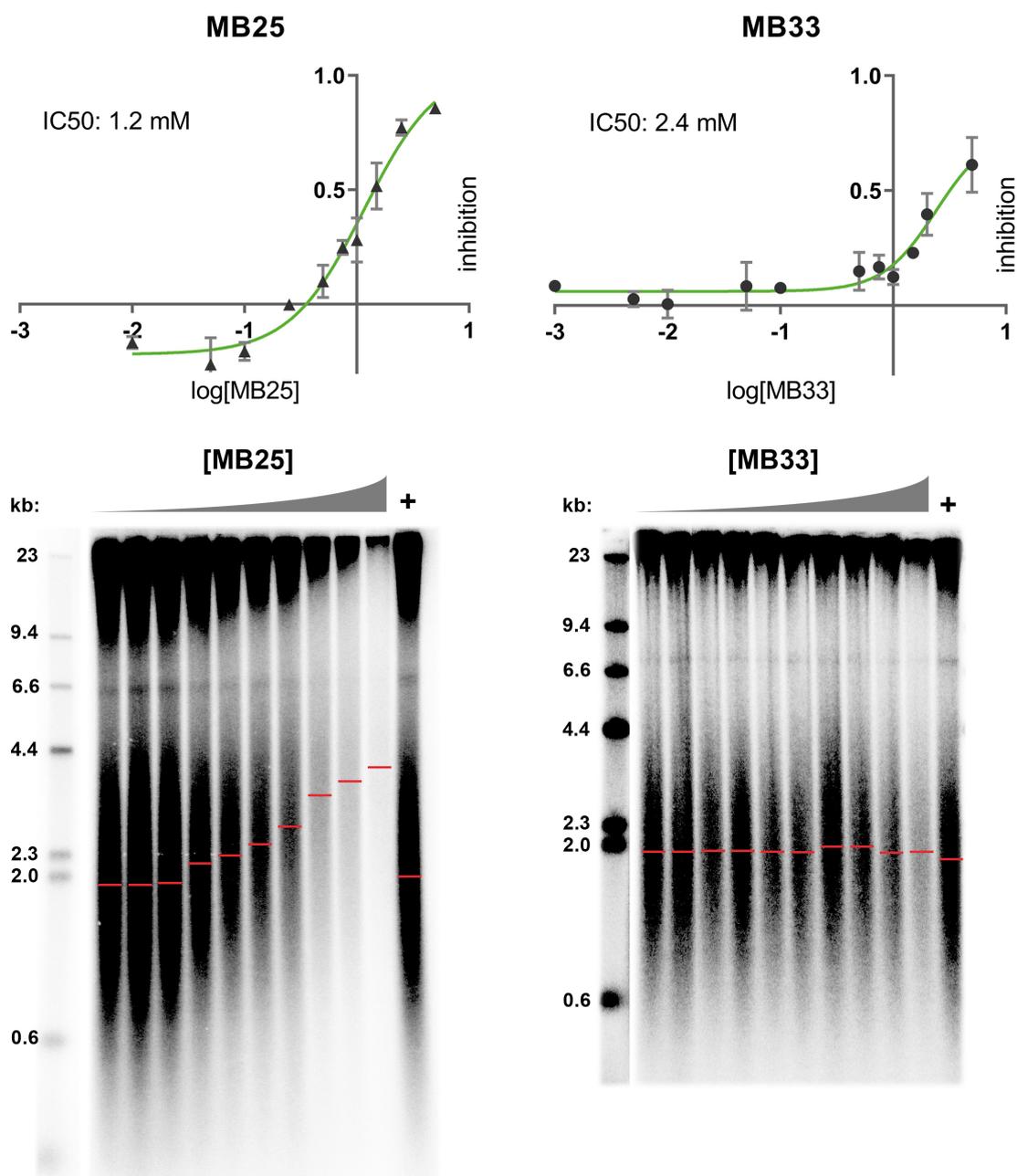


Figure 2.8. Titration data and fitted curves for replisome inhibition by MB25 and MB33. IC<sub>50</sub> values calculated from fit are shown. Inhibition was measured by counting pixel intensities per gel lane (below) as a fraction of the total intensity of the positive control lane. Below, representative phosphor gel images of each titration are shown. Leading and lagging strands are separated based on size (see Figure 2.1). Peak positions based on lane pixel intensities are indicated by red bars. The average lagging strand size increases with addition of MB25 but not with MB33.

### 2.2.3 How MB25 and MB99 inhibit the replisome

The question whether primase inhibition can explain the effect MB25 has on the replisome was further investigated, since MB25 was initially indicated as an  $\alpha$  inhibitor. Literature was searched for small molecule inhibitors of *E. coli* primase, but the only promising candidate was the endogenous signalling molecule guanosine 3',5'-bispyrophosphate (ppGpp). This nucleotide analogue is produced by bacteria including *E. coli* in response to stress signals, and inhibits many cellular processes including transcription and replication (Srivatsan & Wang, 2008). It inhibits primase enzymatic activity (Maciag *et al.*, 2010), most probably by binding to its active site (Rymer *et al.*, 2012). When titrated into the replisome reaction, ppGpp indeed inhibited DNA replication ( $IC_{50}$  1.4 mM), but surprisingly did not affect Okazaki fragment size (Figure 2.9). Possibly, ppGpp inhibits another nucleotide-binding replisome component with greater efficacy, so that the effect on primase becomes negligible and does not appear after gel electrophoresis.

As an alternative more specific probe for primase function, it was hypothesised that a non-hydrolysable ribonucleotide analogue could be used. Guanine is the second base in each *E. coli* RNA primer and as such is needed for primer synthesis initiation (see introduction section 1.2.3.3). Guanosine 5'-( $\alpha,\beta$ -methylene)triphosphate (GpCpp or GMPCPP) should specifically inhibit primer initiation, allowing the initiation complex to form but not to react. The only other replisome component known to use GTP is helicase, however it hydrolyses the  $\beta,\gamma$  phosphoester bond and can use ATP with similar efficiency (Roychowdhury *et al.*, 2009). GpCpp inhibited the replisome with an  $IC_{50}$  of  $>2$  mM (in the presence of GTP at 200  $\mu$ M) (Figure 2.9). Crucially, in addition to inhibiting total reaction products, it also caused an increase in Okazaki fragment size. The gel images further suggest that the total amount of Okazaki fragments decreases faster than the total amount of leading strands. However, separate quantification of each band shows only slightly faster inhibition of Okazaki fragment synthesis at lower GpCpp concentrations (Figure 2.9, right graph). This is expected since leading and lagging strand synthesis are coupled processes. It remains possible that GpCpp also inhibits helicase, since helicase can use both ATP and GTP, but in this assay it likely uses mostly ATP because it is present at high (2 mM) concentrations. Moreover, it is possible, though unknown, that helicase could hydrolyse the  $\beta,\gamma$ -phosphoester bond of GpCpp functionally. Therefore, the most

likely explanation for the overall inhibitory effect on both leading and lagging strand synthesis by MB25 and GpCp is that the two processes are coupled, so that inhibiting one also slows down the other. The similar effects of GpCp and MB25 on the replisome are consistent with inhibition of primase function.

To further confirm the hypothesis that MB25 inhibits the replisome by acting on primase, NMR was used to detect their physical interaction. Recombinantly produced active primase was incubated with excess MB25 and NMR spectra were recorded using the ligand-observed WaterLOGSY method (Dalvit *et al.*, 2000). This method differentiates binding from non-binding ligands based on a change in sign of water-ligand magnetisation transfer efficiency. The magnitude of the ligand signals depends on several experimental parameters, but bound ligands generally show positive peaks while non-binding molecules show negative signals due to differences in their rotational correlation times. Figure 2.10 (bottom, blue spectrum) shows that MB25 binds to the primase full-length (FL) protein. Additionally, a sample of one of the isolated primase domains, the active site-containing RNA polymerase domain (RPD), was also available. Interestingly, when incubated with only the primase RPD, MB25 does not show binding (middle, red spectrum), suggesting that it binds to one of the non-enzymatic domains of primase. Primase has two other domains with accessory functions (see introduction section 1.2.3.3). Inhibition of either of these functions is consistent with overall inhibition of primer synthesis activity. It has been shown previously that inhibition of the C-terminal domain (a protein-protein interaction domain) causes an increase in Okazaki fragment size (Tougu *et al.*, 1994). The next steps for elaboration of MB25 should be to identify its binding site and/or mechanism of action, and then to optimise the scaffold for higher affinity.

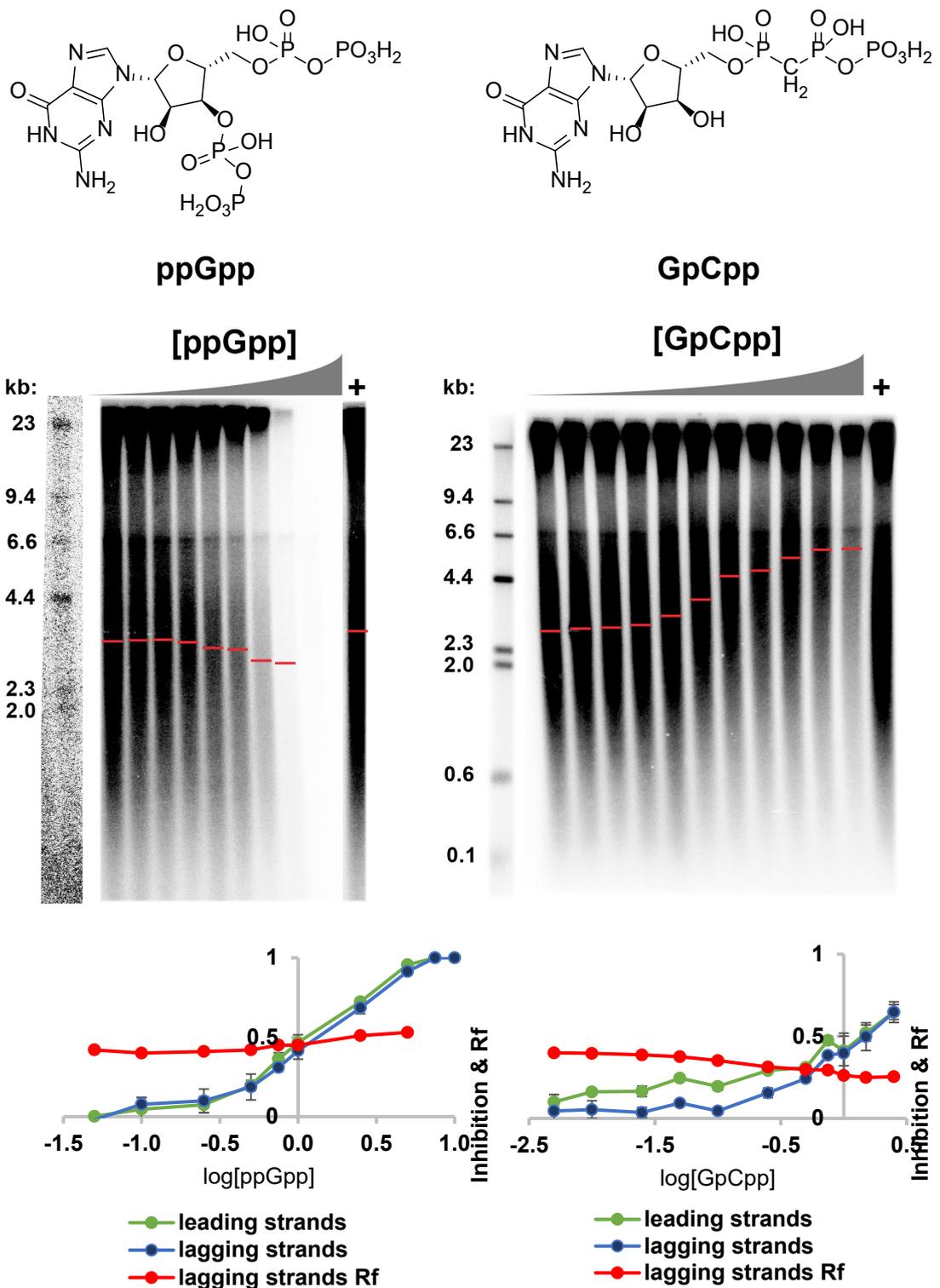


Figure 2.9. Representative gel phosphor image of ppGpp and GpCpp titrations. Leading and lagging strands are separated based on size (see Figure 2.1). Peak positions based on lane pixel intensities are indicated by red bars. The average lagging strand size increases with addition of GpCpp. The average inhibition is shown for each band (leading or lagging strands) separately, as well as the average Rf.

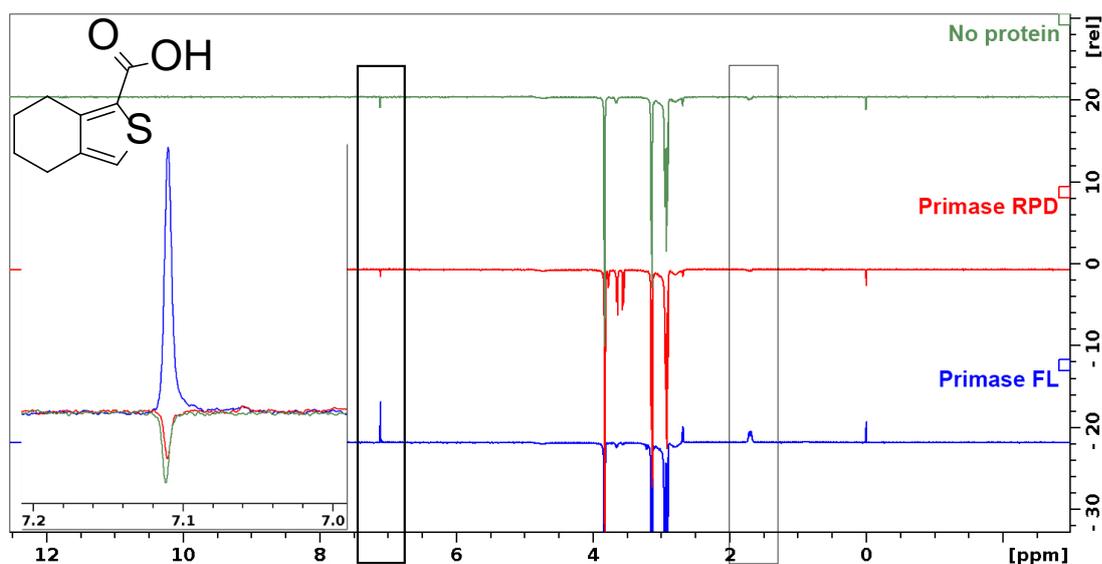


Figure 2.10. WaterLOGSY spectra of MB25 (1 mM) together with: full-length primase (blue, bottom), primase RNA polymerase domain (red, middle), or buffer only (green, top). MB25 seems to bind to full-length primase but not the RPD alone. Ligand peaks are highlighted, and the peak at 7.11 ppm is shown as detailed overlay. Other peaks originate from buffer components: HEPES, DTT, DSS, DMSO, and glycerol. The structure of MB25 is shown.

Finally, since MB99 was identified as another possible primase inhibitor (Figure 2.5), this compound was also tested for primase binding. MB99 indeed appears to bind full-length primase. Furthermore, it also appears to bind to the RPD, although weakly. The measured inhibitory activity and the WaterLOGSY peak intensities thus appear to agree with each other to indicate that MB99 is a weaker binder and inhibitor than MB25. It would be interesting to see whether MB99 also causes an increase in Okazaki fragment size. Unfortunately, the solubility of this fragment was too low to allow titration up to >50% inhibition. MB99 was later used for primase follow-up experiments by crystallography (see section 4.3.2).

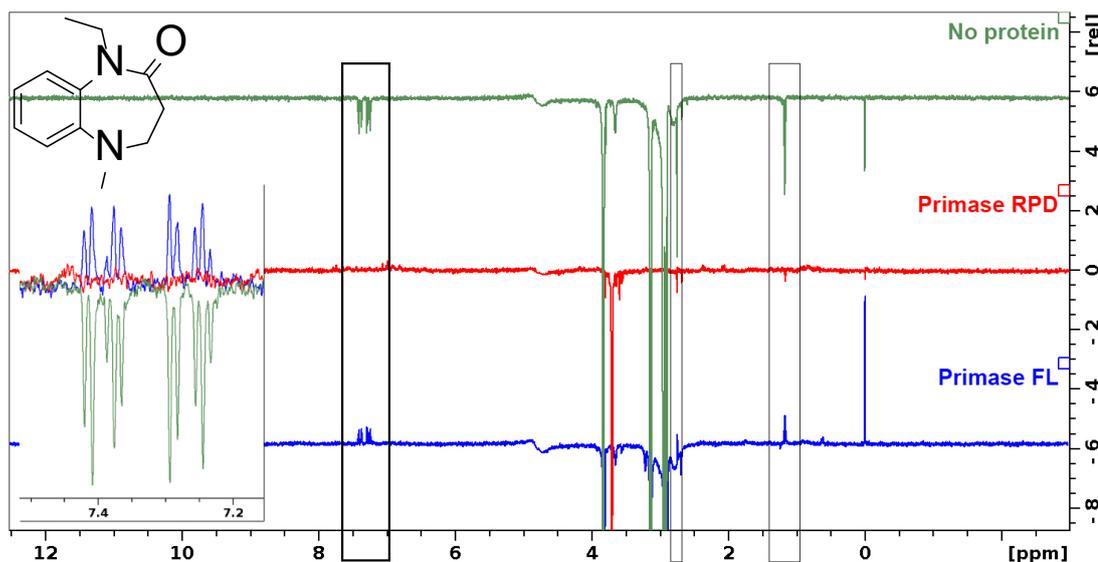


Figure 2.11. WaterLOGSY spectra of MB99 (0.5 mM) together with: full-length primase (blue, bottom), primase RNA polymerase domain (red, middle), or buffer only (green, top). MB99 seems to bind to both full-length primase and to the RPD alone. Some ligand peaks are highlighted, other peaks fall under solvent peaks. The peaks around 7.3 ppm are shown as detailed overlay. Other peaks originate from buffer components: Tris, HEPES, DTT, DSS, DMSO, and glycerol. The structure of MB99 is shown.

#### 2.2.4 Analysis of structural motifs within the replisome fragment hit set

Even though the York fragment library was designed with maximum structural diversity in mind, it contains subsets of compounds that contain identical or very similar substructures. This is helpful when wanting to quickly assess whether a hit is attractive for follow-up. For each of the five selected hits (Figure 2.5), the screening library was searched for substructure matches. Any matching analogue that was also a non-intercalating hit was subjected to quality control by NMR as described earlier.

Table 2-1 lists the analogues that either passed quality control or were not active. Several insights can be gained from these comparisons to aid in prioritising hits for follow-up experiments. MB19 could not be confirmed as a hit earlier, and this appears to agree with the screening data: the same compound was present as a sample made up from its potassium salt form which did not inhibit the replisome. The initial hit MB19 in its free acid form was therefore likely a false-positive.

The other unconfirmed hit, MB7, did not have sufficiently close analogues to draw any conclusions. However, the identification of its methylcinnoline analogue as a strong DNA intercalator (see Figure 2.3) was seen as a warning.

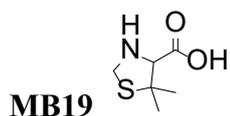
MB99 was found to inhibit the replisome but only at higher concentrations (see section 2.2.2). Two close analogues could be identified in the screening data, but only one was marked as a hit. There are no obvious differences between the three structures that could explain the results. This hit could be interesting to follow up on, but would need careful investigation of its exact binding mode. A co-crystal structure would be preferable in order to identify possible growth vectors for structure-activity relationships (SAR) series. The target of MB99, the primase RPD, has been crystallised before (Keck *et al.*, 2000). Co-crystallisation and soaking experiments with the RPD and MB99 were conducted later (see section 4.3.2).

MB33 has only two analogues within the library despite its simple structure. Interestingly, replacing the carboxylic acid by a primary amine (polar opposites in terms of functional groups) abolishes its activity, suggesting that a negative charge might be important for inhibition of its target (likely polymerase subunit  $\alpha$ , Figure 2.5). Since this scaffold is common, up to ten close analogues preserving the carboxylate group could be purchased and tested against polymerase  $\alpha$  by Dr. S.W.M. Tanley (Hubbard group). Surprisingly, none of the analogues were active. Furthermore, the same compound MB33, purchased from two other suppliers than Maybridge, appeared to have no activity against the replisome nor against  $\alpha$ . The most likely explanation for the initial activity of MB33 is the presence of a contaminant, either at low concentrations or of inorganic nature to escape detection by  $^1\text{H}$  NMR. Polymerases are  $\text{Mg}^{2+}$ -dependent and are known to be sensitive to different cations at varying concentrations (Miyaki *et al.*, 1977; Hermann *et al.*, 2013). Alternatively, it is possible that the contaminant was in fact a chelating agent used during compound purification to remove such ionic impurities. Although a chelator at low concentrations is unlikely to inhibit polymerase in the presence of 10 mM of  $\text{Mg}^{2+}$ , it could interfere with other replisome components coordinating  $\text{Zn}^{2+}$  (primase, clamp loader) (Mendes *et al.*, 2011). Therefore, this hit was deprioritised.

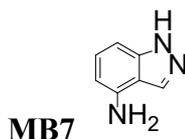
MB25 is an uncommon scaffold with few close analogues within the library. Still, three out of seven distant analogues (either preserving the thiophene 2-carboxylate

scaffold or the benzo[c]thiophene scaffold) were active against the replisome. Since there is nothing to suggest that MB25 was a false-positive hit, and it is the most potent hit for which a target could be suggested, it is an attractive hit to move forward.

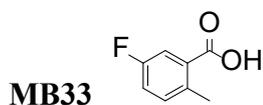
Table 2-1. Screening data for selected hit analogues. Rank indicates amount of inhibition, rank 1 being the most inhibitory fragment. The first 100 fragments are hits.



Structure	Rank	Proportion replication ( $\pm$ SE)	Comment
	292	0.87 (0.03)	Salt form
	597	0.97 (0.03)	



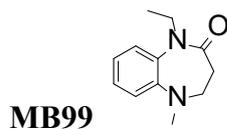
Structure	Rank	Proportion replication ( $\pm$ SE)	Comment
	616	0.98 (0.08)	
	2	0.00 (0.01)	DNA intercalator



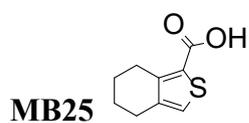
Structure	Rank	Proportion replication ( $\pm$ SE)	Comment
	542	0.95 (0.13)	
	464	0.93 (0.06)	

(continued next page)

Table 2-1 (continued)



Structure	Rank	Proportion replication ( $\pm$ SE)	Comment
	45	0.57 (0.13)	
	667	1.00 (0.11)	



Structure	Rank	Proportion replication ( $\pm$ SE)	Comment
	50	0.59 (0.04)	
	656	0.99 (0.07)	
	94	0.72 (0.04)	
	735	1.03 (0.12)	
	276	0.85 (0.07)	
	83	0.70 (0.07)	No QC
	545	0.95 (0.03)	

### 2.3 Conclusion

All experiments described in this chapter were conducted with the purpose of hit confirmation and prioritisation. Out of the top 25 hits from the original screen, 10 compounds did not pass simple quality control tests due to compound degradation or insolubility. Out of the 15 remaining hits, only 6 reproducibly inhibited the replisome. Out of these 6 confirmed hits, 2 (MB25 and MB33) had previously shown targeted inhibition of a single replisome component. These two were then titrated into the replisome assay to determine their  $IC_{50}$ s, where it was discovered that MB25 has a remarkable effect on replication products. MB25 caused a concentration-dependent increase in lagging strand size distributions, consistent with inhibition of the primase subunit of the replisome. While MB25 was not marked as a primase inhibitor during the replisome component assays, it was shown by NMR experiments that MB25 (as well as MB99 which did show early signs of primase component inhibition) physically binds to primase. Furthermore, the effects of MB25 on the whole replisome reaction were similar to those of GpC<sub>pp</sub>, a tool compound designed to specifically inhibit primer synthesis initiation. Therefore, primase was identified as an interesting target with the fragment inhibitors MB25 and possibly MB99 as starting points for more detailed structure-based elaboration of these hits.



### 3 Finding the binding site of a primase fragment hit using molecular dynamics simulations

#### 3.1 Introduction

Chapter 2 described the identification of an inhibitor (MB25) from fragment screen against the replisome that was shown to bind to primase (see section 2.2.3). The effect of MB25 on replisome activity was consistent with inhibition of the primase subunit, but crystallographic efforts failed to find a binding site for MB25 on the primase RPD. While a mechanism of primase inhibition by binding to the active site domain was the most straightforward hypothesis to test, it is not the only possible mechanism that would explain the observed effects of this fragment. The RPD of bacterial primases is preceded by a zinc-binding domain (ZBD) and followed by a helicase-binding domain (HBD) (Frick & Richardson, 2001) (Figure 3.1), which modulate the RNA polymerase activity of primase. As discussed in section 1.2.3.3, the ZBD controls initiation of primer synthesis by facilitating the formation of the initiation complex (Corn *et al.*, 2005). It contains a CHC2 motif capable of coordinating a zinc ion that most likely stabilises a  $\beta$  sheet fold that can recognise specific template trinucleotide sequences (Larson *et al.*, 2010) (Figure 3.2). The HBD controls the activity of primase by targeting it to the replication fork through interaction with the replicative helicase (Tougu *et al.*, 1994). It is entirely  $\alpha$ -helical, containing a small globular domain with hydrophobic core and a long C-terminal helix hairpin that is involved in helicase binding (Figure 3.2). Inhibition of ZBD-DNA interactions or of HBD-helicase interactions is therefore also expected to decrease the RNA polymerase activity of primase in the context of the replisome.

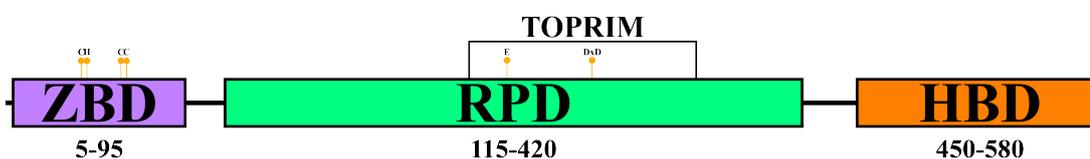


Figure 3.1. Schematic representation of primase domains. The conserved zinc-binding CHC2 and magnesium-binding E-DxD motifs are indicated. Domain boundaries are given in residue indices.

High-resolution structural information for full-length primase is currently not available and would be difficult to obtain, because the flexibility of the protein makes

it unamenable to crystallisation, and the size of the protein is above the practical limit for assignment of NMR resonances. To find out whether MB25 binds to either of these non-catalytic domains, each would have to be produced separately. Structural information for the isolated domains is available: a crystal structure of the ZBD of *Geobacillus stearothermophilus* primase (50% sequence identity to the *E. coli* DnaG ZBD) (Pan & Wigley, 2000), and a crystal structure (Oakley *et al.*, 2005), solution NMR structure (Su *et al.*, 2006), and solution NMR backbone assignments (Naue *et al.*, 2013) for the *E. coli* primase HBD (Figure 3.2). As these structures were already available, it was hypothesised that *in silico* predictions could help prioritise production of and biophysical experiments with one of the two domains and so facilitate an overall more efficient workflow.

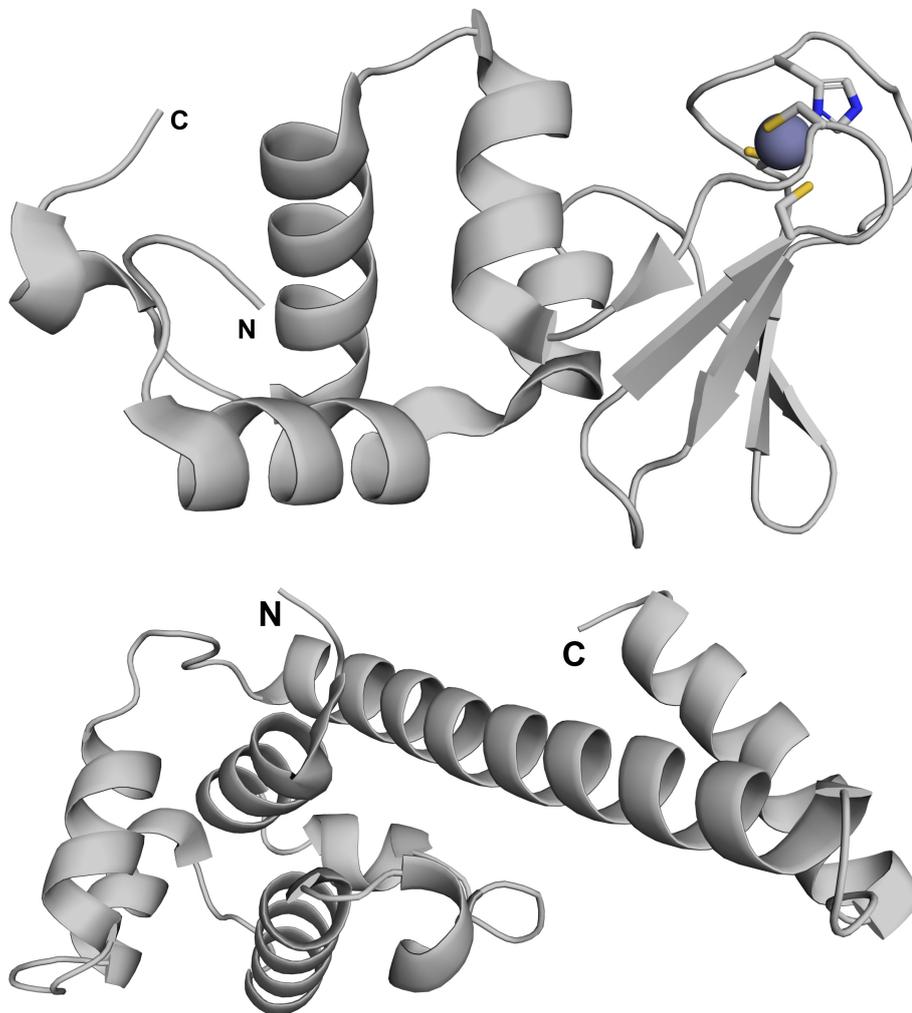


Figure 3.2. Structures of the *G. stearothermophilus* primase ZBD (Pan & Wigley, 2000) (top) and the *E. coli* HBD (Su *et al.*, 2006) (bottom). The ZBD zinc ion and coordinating CHC2 residues are shown.

Computational tools are widely used in structure-based drug design. The most popular methods are docking of small molecules into protein binding pockets as a way of virtual (pre-)screening and assessing structure-activity relationships (SAR), and molecular dynamics (MD) simulations of protein-ligand systems as a way of assessing protein dynamic behaviour and ligand binding kinetics and thermodynamics (Śledź & Caffisch, 2018). Docking can be done very efficiently by relying on general assumptions about protein rigidity and solvent effects, whereas MD simulations are low-throughput but relatively unbiased and can give much more information. Continuous hardware and software developments over the last decade or two have made MD simulations for large systems more accessible and more reliable (Harvey & De Fabritiis, 2012; Bottaro & Lindorff-Larsen, 2018). Forcefields prescribing the behaviour of atoms in standard biomolecular systems such as proteins and polynucleotides have become highly accurate. However, one should remember that these parameter sets are based on historical experimental observations, and can sometimes still be wrong (Auffinger *et al.*, 2007) or incomplete (Pinheiro *et al.*, 2017). Simulation results can be used to formulate a hypothesis which should then be tested experimentally.

Here, MD simulations were used to predict which primase domain is most likely to contain the binding pocket for MB25, taking advantage of the fact that simulations can be run without requiring any prior information about possible binding pockets. The expected result would be a clear preference of MB25 for either the ZBD or the HBD of primase, and identification of the protein residues involved in forming the binding pocket. These predictions were then tested by looking for protein chemical shift perturbations (CSPs) by NMR.

## **3.2 MD simulation results**

### **3.2.1 Model preparations**

Two systems were prepared for simulations using Amber (Case *et al.*, 2014). Each system contained water, NaCl, one molecule of MB25 and one molecule of either ZBD or HBD.

The published NMR structure of the entire HBD (547-581) was used (Su *et al.*, 2006). No model of the entire ZBD (2-101) is available, so a homology model was prepared. There are two known structures of homologous ZBDs: a crystal structure of the *G.*

*stearothermophilus* primase ZBD (Pan & Wigley, 2000), and a crystal structure of the *Aquifex aeolicus* primase ZBD linked to its RPD (Corn *et al.*, 2005). Sequence alignment of these two ZBD templates to the *E. coli* target (Figure 3.3) shows that the *G. stearothermophilus* ZBD has the highest sequence identity with the target (50%) and contains no alignment gaps.

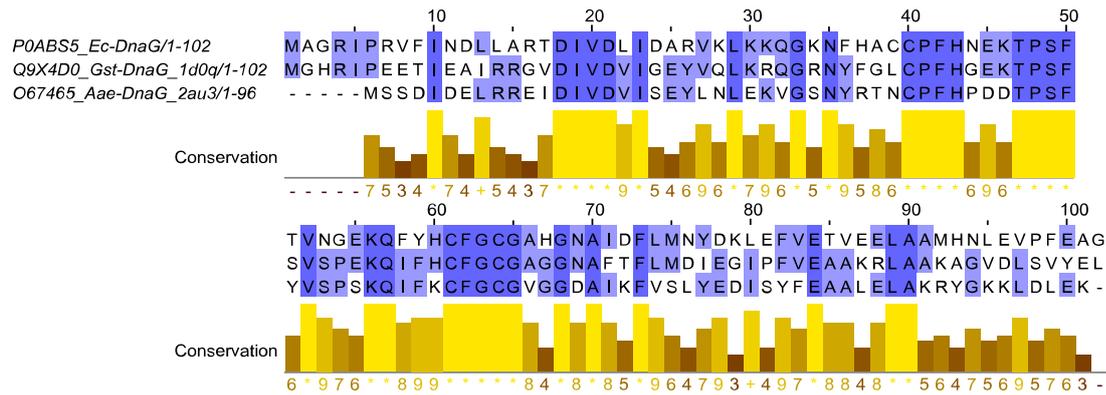


Figure 3.3. Alignment of *E. coli* primase ZBD with homologues of which structures are known. The *G. stearothermophilus* model 1d0q was chosen as a template for homology modelling of the *E. coli* ZBD based on higher sequence identity compared to *A. aeolicus*.

An *E. coli* homology model based on this template (chain A from PDB 1d0q (Pan & Wigley, 2000)) was generated using the homology modelling tool in the MOE software package. The model has an average C $\alpha$  root-mean-square deviation (rmsd) from the template of 0.47 Å after protonation and energy minimisation. Its QMEAN Z-score (Benkert *et al.*, 2011) is -1.16, suggesting that the model is reasonably similar to experimental protein structures in the same size range in terms of torsion, solvation, and atom and residue interaction potentials.

While forcefield parameters for biological macromolecules are generally accurate, those for other types including small organic molecules are less well-described. A thorough and accurate description of MB25 electrostatic potentials was deemed necessary here in order to best predict the weak interactions between the fragment and its protein binding site. Therefore, partial atomic charges for MB25 with deprotonated carboxylate group were calculated using *ab initio* quantum mechanical methods, then normalised and converted into Amber-compatible atom types.

Eight different systems were prepared for parallel simulation runs: four for each protein domain. Each system contained one protein molecule and one ligand molecule in different orientations at least 5 Å away from each other.

### 3.2.2 Simulation results

MD simulations were performed for 500 ns per system, *i.e.* a total of 2 μs per protein domain. Proteins were quite dynamic, with an average backbone rmsd of 2.1 Å for the ZBD and 3.7 Å for the HBD after realignment. The N-terminal globular part of the HBD (residues 450-529), which contains a hydrophobic core, is less dynamic than the long C-terminal helix hairpin (residues 530-579). This is revealed by analysis of the backbone rmsd for each part separately, after aligning all residues or only the globular domain. Alignment of the globular domain only results in a decrease of the rmsd for that domain from 3.5 Å to 2.6 Å, while the helix hairpin backbone rmsd for residues 547-579 increases from 4.3 Å to 10.0 Å. This flexibility is most likely real, and important for helicase binding or unbinding, since some primase HBD homologues adopt helicase-unbound conformations in which the helix is broken in the region correlating to residue 547 (Syson *et al.*, 2005; Bailey *et al.*, 2007; Catazaro *et al.*, 2017).

Analysis of simulations focussed on the movements of MB25 relative to the protein in each 500 ns trajectory separately. During most trajectories MB25 prefers to stay near a limited set of residues, which is different for each ZBD run but the same for each HBD run (Table 3-1). Each set forms a shallow binding site. The binding sites of the ZBD cover four different sides of the domain (Figure 3.4A). The binding site of the HBD is located at the N-terminal globular part (Figure 3.4B).

Table 3-1. Proximity of MB25 to protein residues as a fraction of total trajectory time (top 10 most occurring for each simulation run). Residues in bold appear in the top 10 of more than one run.

### ZBD

	#1	#2	#3	#4			
M92	1.00	<b>R4</b>	0.74	L14	0.35	T51	0.64
L22	1.00	I5	0.71	N11	0.35	N35	0.63
T17	0.98	V97	0.66	R7	0.34	H60	0.55
H93	0.96	V86	0.58	N69	0.27	F58	0.54
D18	0.96	P6	0.56	A15	0.23	N53	0.51
D21	0.96	E87	0.53	Q57	0.23	<b>K56</b>	0.50
L89	0.88	P98	0.51	F82	0.18	Y59	0.47
R16	0.76	V83	0.48	I10	0.18	V52	0.46
R26	0.41	F99	0.48	<b>R4</b>	0.13	F62	0.37
L74	0.04	F9	0.48	<b>K56</b>	0.12	K34	0.19

### HBD

	#1	#2	#3	#4			
<b>R452</b>	0.98	<b>R452</b>	0.76	<b>R452</b>	0.64	<b>R452</b>	0.61
<b>I455</b>	0.98	<b>T515</b>	0.72	<b>L519</b>	0.61	<b>L519</b>	0.52
<b>T515</b>	0.97	<b>I455</b>	0.69	<b>T515</b>	0.41	<b>T515</b>	0.49
<b>G481</b>	0.94	<b>L484</b>	0.68	<b>I455</b>	0.37	<b>I455</b>	0.41
<b>L519</b>	0.93	<b>L519</b>	0.66	<b>G481</b>	0.35	<b>L484</b>	0.40
<b>L484</b>	0.92	<b>G481</b>	0.63	<b>L484</b>	0.35	<b>G481</b>	0.37
<b>M451</b>	0.87	<b>M451</b>	0.60	<b>M451</b>	0.32	<b>M451</b>	0.36
<b>P480</b>	0.60	<b>P480</b>	0.44	W522	0.30	<b>N512</b>	0.35
<b>L516</b>	0.38	<b>N512</b>	0.43	I453	0.28	<b>L516</b>	0.34
<b>N512</b>	0.31	<b>L516</b>	0.36	G456	0.25	Y507	0.25

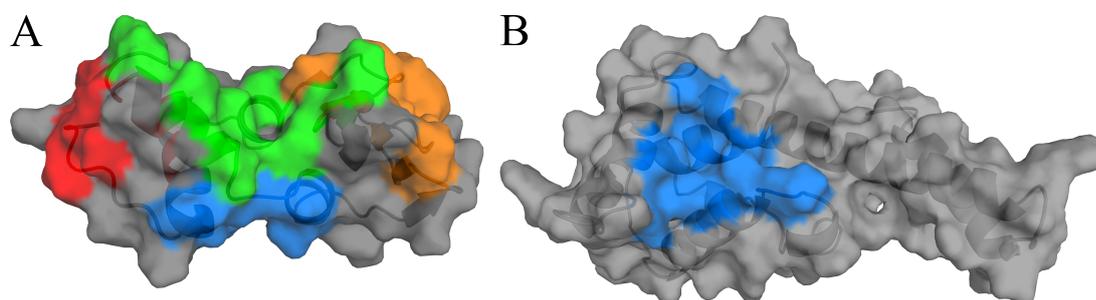


Figure 3.4. MB25 interacts with different surfaces of the ZBD during the four different simulation runs (run 1: blue, run 2: red, run 3: green, run 4: orange) (A), but only one site on the HBD (B).

Protein-ligand proximity analysis indicates whether MB25 is engaging the protein but does not say anything about the quality of interactions. Hydrogen bonds often contribute significantly to the binding energy of ligands, especially in case of fragments. Therefore, H bond analysis was performed. MB25 has two H bond acceptors, which are part of the anionic carboxylate. Table 3-2 shows that MB25 accepts H bonds mainly from cationic arginine side chains, which is also a favourable ionic interaction. During each run, the most frequent H bond donors are residues that were also identified in the proximity analysis (Table 3-1). However, the fraction of simulation frames during which protein-ligand H bonds are present is smaller than during which protein and ligand are near each other. This could be in part due to the stricter distance restraints for H bonds ( $<3.5 \text{ \AA}$ ) compared to those used for the proximity analysis ( $<4 \text{ \AA}$ ), and to the fact that fewer ligand atoms can participate in H bonds. It is also possible that different types of interactions contribute more to binding. MB25 has hydrophobic aliphatic and aromatic features that can interact favourably with hydrophobic protein sidechains, and an aromatic sulfur atom that can interact favourably with carbonyl oxygens (Zhang *et al.*, 2015). The interaction sites identified in simulations do contain clusters of hydrophobic/aromatic side chains likely to contribute to MB25 binding, especially ZBD residues I5, V97, and F9 in simulation #2 (*e.g.* Figure 3.5A), and HBD residues I455, T515, L519, and L484 in all HBD simulations (*e.g.* Figure 3.5B). A search for intermolecular sulfur-oxygen bonds did not reveal any such interactions that were present for more than one percent of any simulation.

Table 3-2. Protein hydrogen bond donors to MB25, by fraction of simulation time during which at least one protein-ligand H bond is present.

### ZBD

#1		#2		#3		#4	
R16 Nη2	0.09	R4 Nη2	0.47	N69 Nδ2	0.11	N35 Nδ2	0.23
R16 Nη1	0.07	R Nε	0.34	R7 Nη1	0.04	K34 Nζ	0.08
R16 Nε	0.06	R4 N	0.14	R7 Nε	0.04	K56 Nζ	0.03
R26 Nη1	0.02	R7 Nη1	0.03	R4 Nη2	0.02	K30 Nζ	0.03
R26 Nη2	0.02	G3 N	0.03	R7 Nη2	0.01	N53 Nδ2	0.01
		R7 Nε	0.02				
		R4 Nη1	0.02				
		A2 N	0.02				
		I5 N	0.01				

### HBD

#1		#2		#3		#4	
R452 Nη1	0.12	R452 Nε	0.10	T450 Oγ1	0.16	R452 Nη2	0.23
R452 Nε	0.12	R452 Nη2	0.10	M451 N	0.15	R452 Nε	0.21
R452 Nη2	0.11	R452 Nη1	0.07	R452 Nε	0.08	R452 Nη1	0.06
N512 Nδ2	0.04	N512 Nδ2	0.05	R452 N	0.06	R448 Nε	0.03
R452 N	0.03	R568 Nη1	0.02	R452 Nη1	0.06	K528 Nζ	0.01
K518 Nζ	0.03	T450 Oγ1	0.01	R452 Nη2	0.06		
		R452 N	0.01				

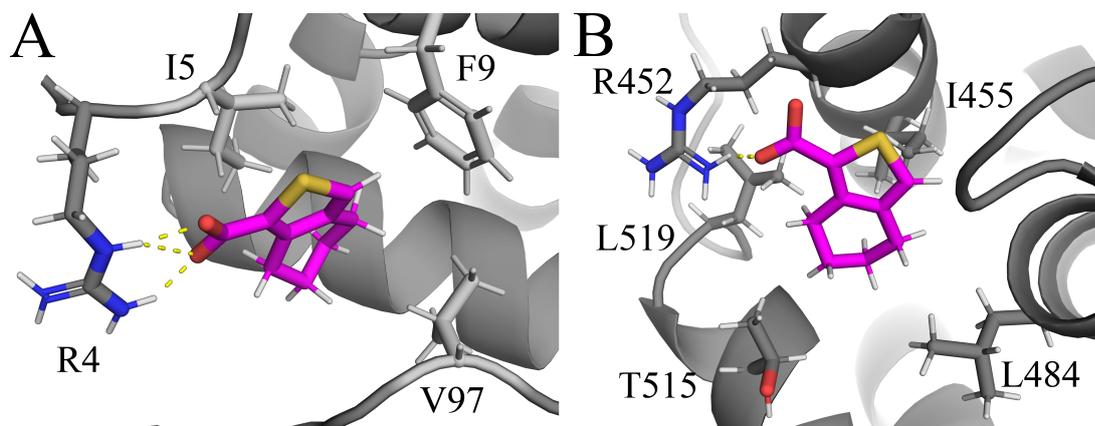


Figure 3.5. Example snapshots of MB25 binding modes with the ZBD (A) or HBD (B), showing hydrogen bonds (yellow dashed lines) and hydrophobic residues.

While MB25 interacted with the ZBD for large fractions of time in each simulation run, and the interactions seem arguably favourable from a physicochemical perspective, the four runs did not converge to any common binding site or set of residues. Furthermore, it is difficult to imagine any of the interaction sites being functionally relevant, except the site identified in run #4 which is the DNA interaction surface containing F58 (important for initiation site recognition (Larson *et al.*, 2010)) and conserved residues N35, K56, and F62. Interestingly, ionic and H bond interactions by the nearby K34 and K56 with the MB25 carboxylate are not frequently observed, with N35 being the major H bond donor. The pharmacophore of this site seems to be suboptimal for interaction with MB25. The presence of a significant physical interaction between MB25 and the ZBD should be tested by NMR (see section 3.3).

In contrast, each run with the HBD showed MB25 interacting with the same set of residues, of which R452 is the most important one. The geometry of the site allows MB25 to sit comfortably at the bottom formed by I455, L484, L516, and L519 while still being able to interact with the R452 guanidinyll group (2-dimensional representation of a snapshot from run #4 shown in Figure 3.6). However, it should be stressed that these simulation data do not converge upon any single binding mode. Indeed, the exact example shown in Figure 3.6 is not maintained for more than several tens of ns. The absence of a single major binding mode in the simulation data may simply be a consequence of the system: the combination of a small, flexible protein, a weakly binding fragment, and a shallow binding site makes it inherently very dynamic. Furthermore, even for more rigid, well-defined binding pockets, fragment binding mode prediction is thought to require much longer simulation times of up to 40  $\mu$ s (Linker *et al.*, 2019). The current results merely suggest the binding site of MB25, not its binding mode.

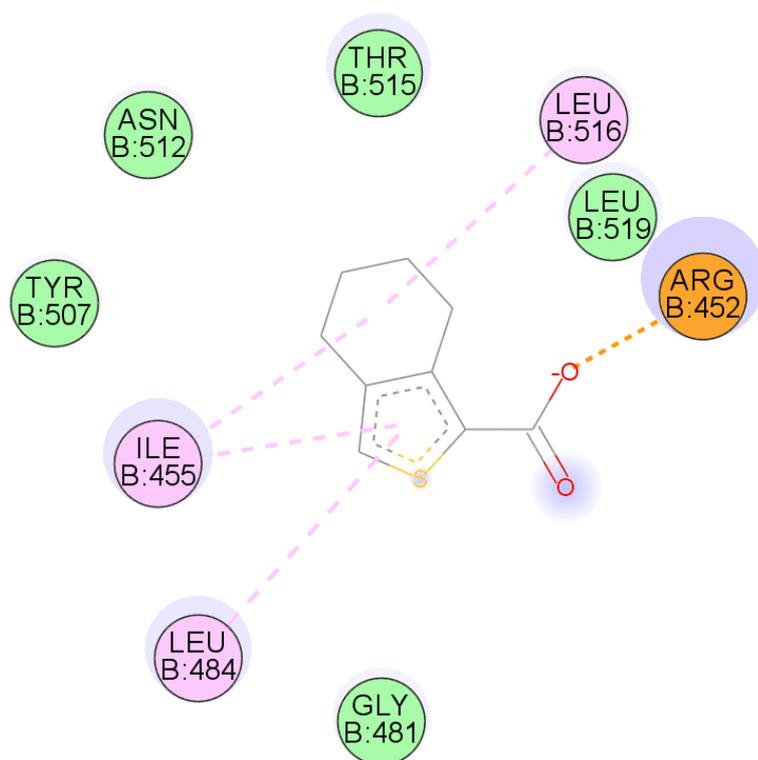


Figure 3.6. Example of a binding mode of MB25 to the HBD, observed in simulation run #4. Orange dashed line: salt bridge; pink dashed lines: hydrophobic interactions; blue background spheres: solvent-accessible surface (scaled). Image prepared using Biovia Discovery Studio.

### 3.3 Verification of simulation results by NMR

The simulation results described in the previous section suggest that MB25 prefers to bind to the HBD, because all four runs show MB25 associating with the same binding pocket and making favourable interactions. To verify this prediction biophysically, the two domains were produced with uniform  $^{15}\text{N}$  labelling for CSP analysis by NMR.

For protein characterisation purposes, the ZBD was first produced in rich media without added  $^{15}\text{N}$ . The protein was well-behaved throughout the purification. Analysis of this material (concentration:  $6\text{ mgmL}^{-1}$ ) by size-exclusion chromatography followed by multi-angle light scattering (SEC-MALS) indicated that it contained about 96% monomeric ZBD, with the remaining fraction being a dimer (Figure 3.7). Mass spectrometry analysis of the intact, unfolded protein revealed a molecular weight

of 11,762.8 Da (expected: 11,762.8 Da). Although the  $^{15}\text{N}$ -labelled protein was expressed in different growth media, it is assumed that it behaves similarly.

The 2D  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear single quantum coherence (HSQC) spectrum of the ZBD shows good dispersion of NH resonances (Figure 3.8A), indicating that it is properly folded under the experimental conditions. Out of 116 expected NH cross-peaks, 112 were observed. In the presence of MB25 at ten-fold molar excess, no CSPs are observed (Figure 3.8B), indicating that MB25 does not bind to this domain.

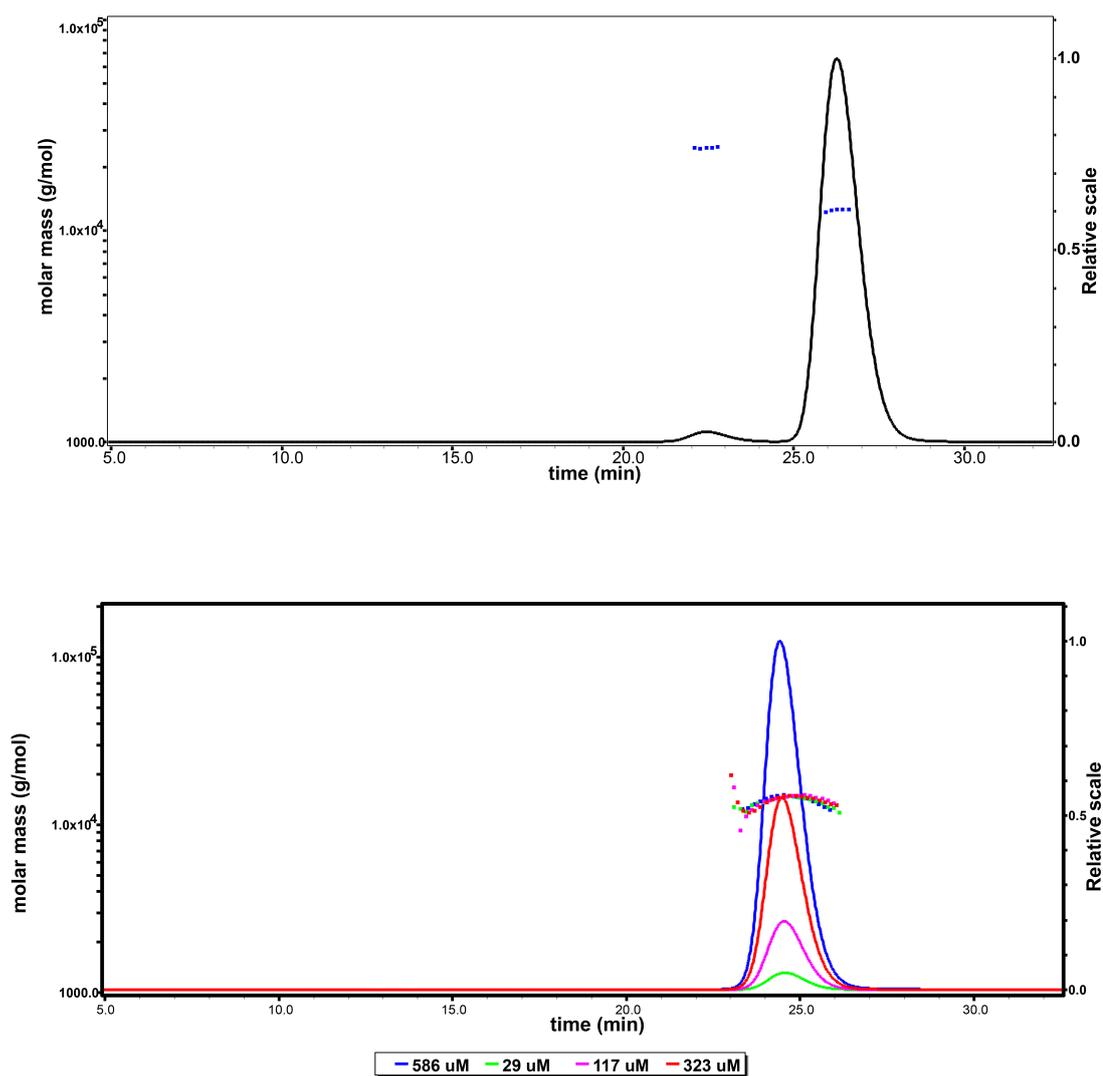


Figure 3.7. SEC-MALS spectra of the ZBD ( $6 \text{ mgmL}^{-1}$ ) (top) and the HBD (various concentrations) (bottom), showing  $>95\%$  monomeric proteins. Flow rate was  $0.5 \text{ mLmin}^{-1}$ .

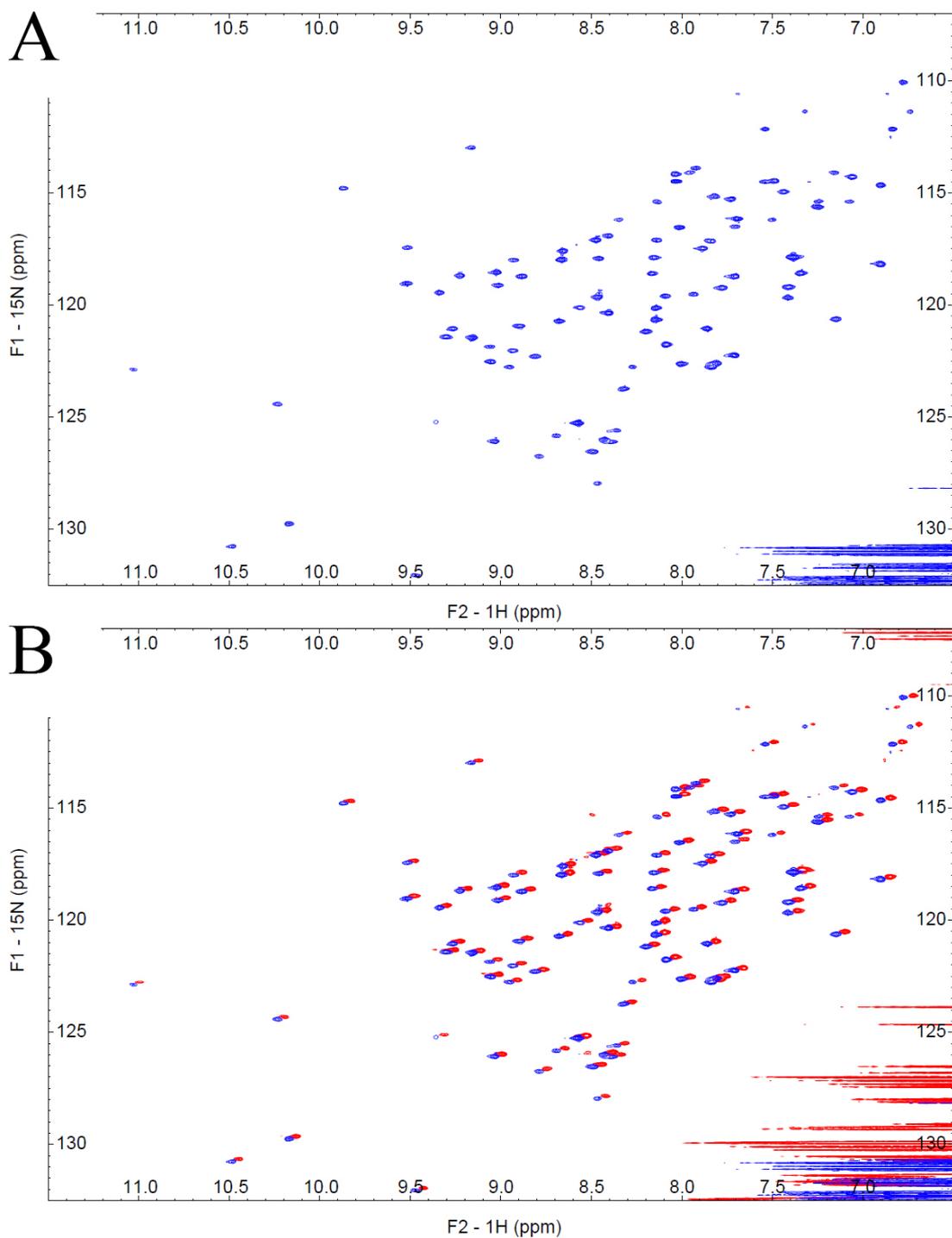


Figure 3.8. (A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of uniformly  $^{15}\text{N}$ -labelled ZBD at  $50\ \mu\text{M}$ . (B) Overlay of the spectrum from A (blue) with the spectrum recorded in the presence of MB25 ( $500\ \mu\text{M}$ ) (red). The red spectrum was artificially shifted slightly upfield in both dimensions in order to compare peaks between the two spectra. No compound-induced shifts were observed.

Similarly, the HBD was first produced without labelling. SEC-MALS analysis of this protein revealed a single population with an estimated molecular weight consistent with a monomer. Because the SEC peak retention time was slightly earlier than expected for proteins in this size range, and there have been reports of possible dimerisation of this protein (Oakley *et al.*, 2005), further analysis was performed over a range of protein concentrations. Between 30-600  $\mu\text{M}$ , only the single monomeric peak was observed (Figure 3.7). The unexpected retention time, as well as the estimated hydrodynamic radius of 2.2 nm, is likely a result of the non-globular, elongated shape of the protein. The molecular weight of the protein was 17,053.7 Da (expected: 17,054.3 Da) as determined by mass spectrometry. The measured molecular weight of the  $^{15}\text{N}$ -labelled HBD was 17,258.1 Da, consistent with 97% labelling efficiency assuming the protein was not chemically modified.

The 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the HBD shows good dispersion of NH resonances (Figure 3.9A), indicating that it is properly folded under the experimental conditions. Out of 177 expected NH cross-peaks, 155 were observed. In the presence of MB25 at ten-fold molar excess, significant CSPs of a small number of cross-peaks are observed (Figure 3.9B). The observed CSPs did not occur upon addition of either dimethyl sulfoxide (DMSO) or acetic acid instead of MB25. Upon addition of acetic acid to hundred-fold molar excess, a small number of cross-peaks shifted, only two of which also shift with MB25. Together, these results indicate that MB25 is binding to the HBD specifically.

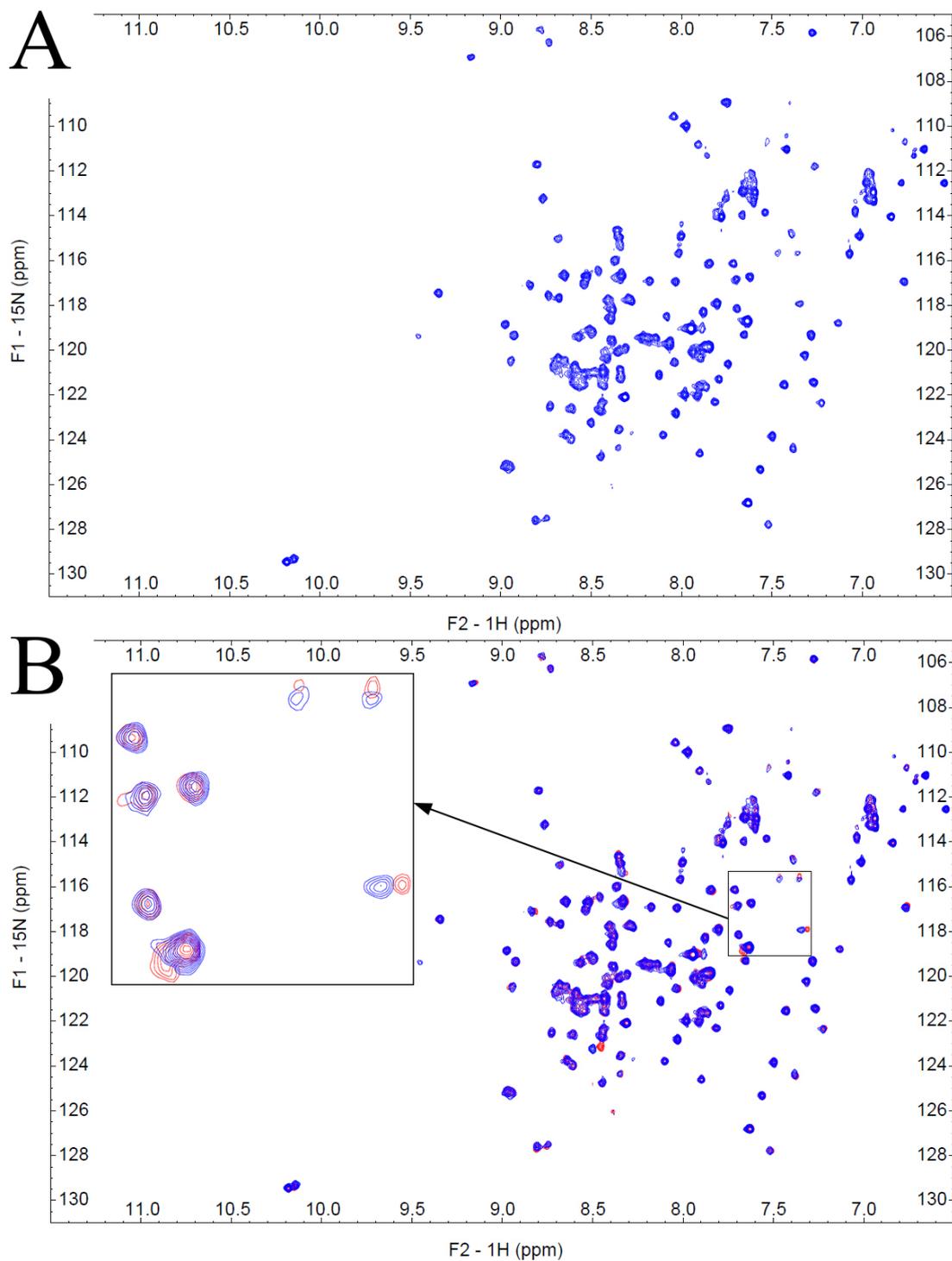


Figure 3.9. (A)  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of uniformly  $^{15}\text{N}$ -labelled HBD at 50  $\mu\text{M}$ . (B) Overlay of the spectrum from A (blue) with the spectrum recorded in the presence of MB25 (500  $\mu\text{M}$ ) (red). The inset shows an enlarged example of a region containing cross-peaks with compound-induced CSPs.

To determine whether the shifting peaks belong to residues that could form a binding site together, peak assignments are necessary. The  $^1\text{H}$ - $^{15}\text{N}$  peak assignments belonging to the published structure of the HBD (Su *et al.*, 2006) do not match the peaks observed here, likely due to differences in buffer pH (pH 6.1 vs. 7.5) or composition. Assignments for 134 backbone amides at pH 7.4 were previously determined (Naue *et al.*, 2013) and generously provided by Dr. P. Schmieder. These assignments closely matched, although two cross-peaks could not be observed. Furthermore, based on the shared data, asparagine side chain amide NH peak pairs could also be assigned, and these assignments also closely matched the peaks observed here. Several peaks could not be unambiguously assigned because of partial or complete overlap with cross-peaks of other residues. Some cross-peak pairs that are overlapped in the apo spectrum show perturbation of one of the peaks after addition of MB25. To reveal the identity of the perturbed residue, the two overlapping peaks must be distinguished by some other feature in which they differ. The shared data indicated that most of these residues have non-overlapping resonances of the carbonyl carbon atoms belonging to the backbone amide (formally the carbonyl carbon of the preceding residue,  $i-1$ ). Since these spin systems are chemically connected, a 3-dimensional experiment would be able to link the backbone amide H, N, and C atom resonances. Thus, even if the H and N frequencies of two different residues are identical, they can be distinguished by differing  $i-1$  CO resonances. Therefore, the CSP experiment with MB25 was repeated and HNCO spectra (that correlate the resonances of these three nuclei) were recorded to identify the perturbed residue out of each pair displaying overlapping  $^1\text{H}$ - $^{15}\text{N}$  cross-peaks (Figure 3.10).

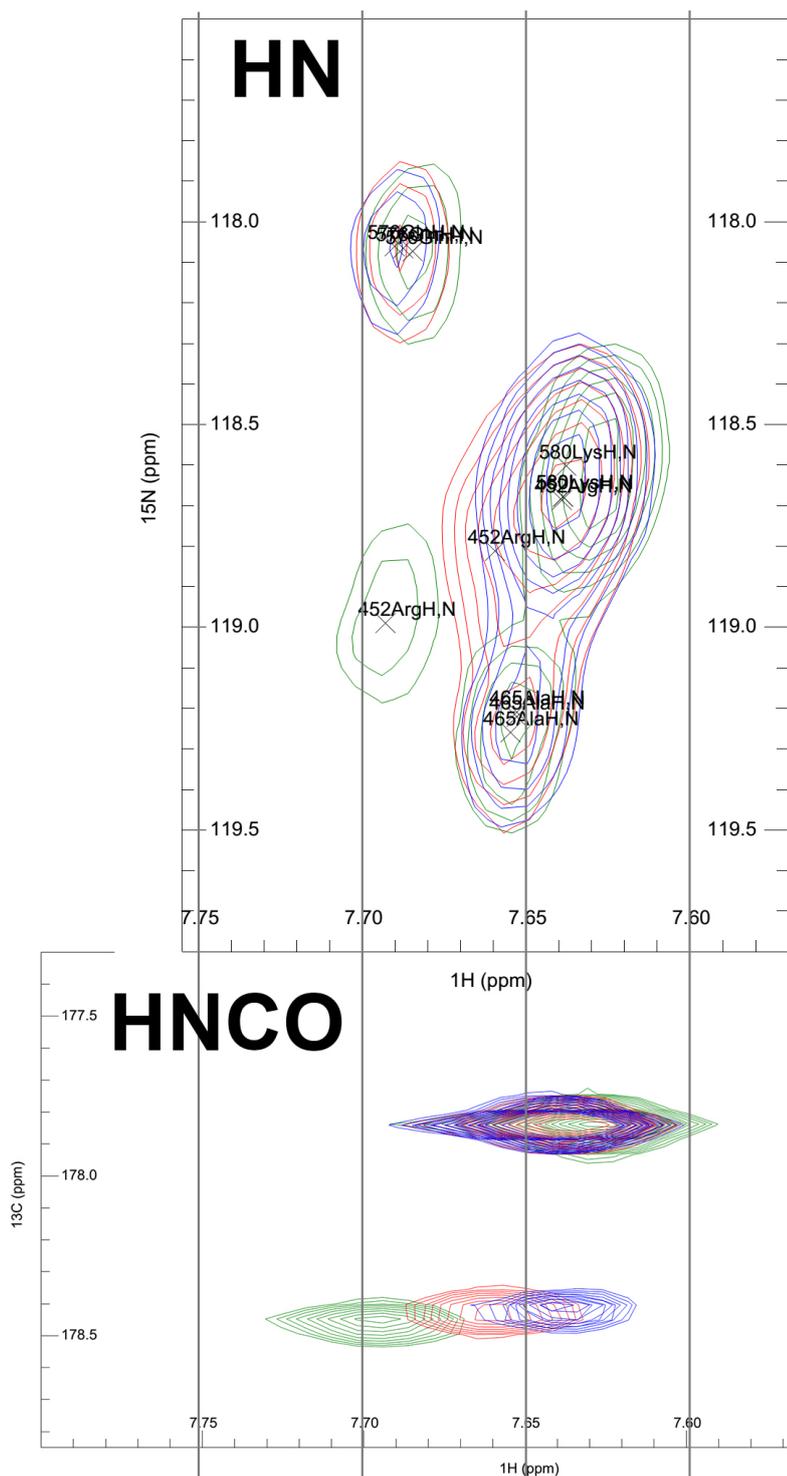


Figure 3.10. Comparison of  $^1\text{H}$ - $^{15}\text{N}$  HMQC (top) and  $^1\text{H}$ - $^{15}\text{N}$ - $^{13}\text{C}$  HNCO (bottom) spectral regions of the R452 backbone peak. Overlay of three spectra of each experiment, containing MB25 at 0 (blue), 0.5 (red), or 2 mM (green). The HNCO spectrum is shown at 118.5-119.0 ppm in the  $^{15}\text{N}$  dimension, therefore only R452 and K580 peaks are visible. While resonances of these two residues completely overlap in the H and N dimensions at 0 mM, they are separated in the C dimension, allowing assignment of the perturbed peak by comparison with the previously obtained resonance assignments.

Combining the matching assignments with the CSP analysis clearly shows that MB25 is binding to the site predicted by MD simulations (Table 3-3, Figure 3.11). Out of the top ten furthest shifting peaks, seven belong to residues that were present in MD simulations proximity analysis top ten (Table 3-1). Two of these residues, R452 and N512, were also identified as possible H bond donors to MB25 (Table 3-2). The two residues that are also perturbed by addition of acetic acid, D524 and R559, are the only two residues that are not part of the predicted binding site.

Table 3-3. Top 10 compound-induced CSPs of the HBD.

<b>MB25 (0.5 mM)</b>			<b>AcOH (5.0 mM)</b>		
NH	CSP (ppm)	$\sigma$	NH	CSP (ppm)	$\sigma$
G481	0.089	9.0	H541	0.030	3.1
L482	0.046	4.6	N540	0.021	2.1
I455	0.034	3.4	D524	0.016	1.6
D524	0.031	3.2	T560	0.014	1.4
R452	0.030	3.0	R559	0.014	1.4
M451	0.023	2.3	E505	0.014	1.4
N512 NH $\delta$ 21	0.022	2.2	Q460	0.014	1.4
N512 NH $\delta$ 22	0.022	2.2	N540 NH $\delta$ 21	0.012	1.2
R559	0.020	2.0	N540 NH $\delta$ 22	0.012	1.2
L454	0.020	2.0	N512 NH $\delta$ 21	0.012	1.2

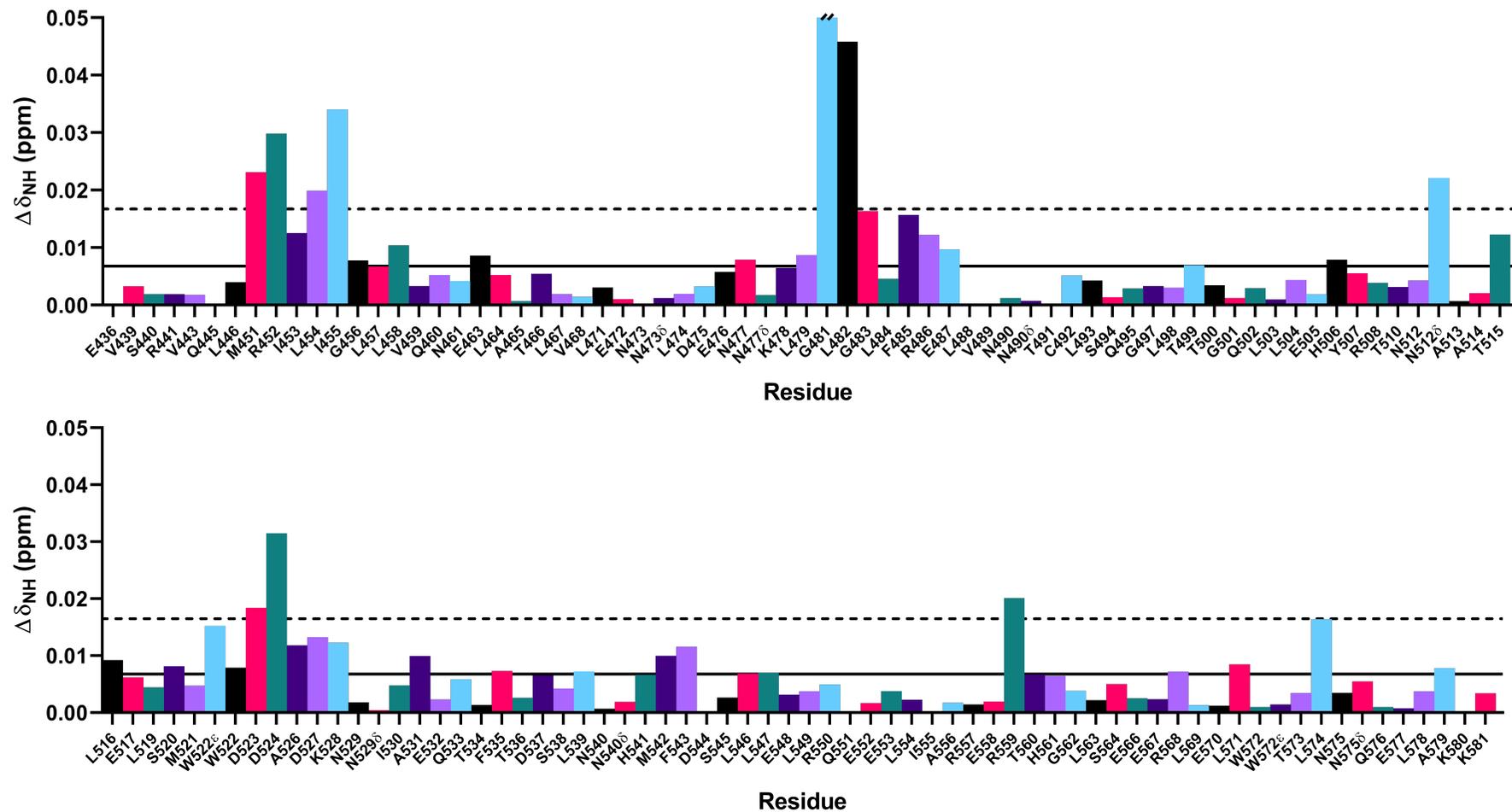


Figure 3.11. MB25-induced CSPs of assigned HBD NH peaks. Solid line indicates the mean  $\Delta\delta$ , dashed line indicates  $1\sigma$  above the mean. The G481 NH CSP extends to 0.089 ppm outside the axis upper limit. All significantly shifted peaks except R559 are part of the same pocket.

### 3.4 Conclusion

The data presented in this chapter provide a rare example of the correct prediction of a fragment hit binding site by unbiased MD simulations. Fragment MB25 was previously identified as a hit in a screen for inhibitors of the replisome, and primase was indicated as the most likely target based on functional and biophysical data. MD simulations were used to quickly assess which primase domain would be most likely to bind MB25 and thus should be the focus of future experiments. Few meaningful interactions were observed during simulations of MB25 with the ZBD, and indeed no interaction whatsoever is observed by NMR. In contrast, NMR experiments showed that MB25 perturbs the backbone amide cross-peaks of several HBD residues. The ensemble of perturbed residues forms a binding site on the N-terminal globular part of the HBD. The NMR data are consistent with a binding mode such as is seen in a snapshot of simulation #4 (Figure 3.6). The main features of MB25, its partially aliphatic character and the negative charge of its carboxylate, are likely to contribute to its affinity for the HBD.

The identified binding site is also a binding site for the SSB disordered C-terminal peptide (sequence: DDDIPF) (Naue *et al.*, 2013). Both this peptide and its HBD binding site are conserved between bacterial species. Strikingly, MB25 has both its hydrophobic and charged characteristics in common with this peptide. Indeed, a crystal structure of the *E. coli* HBD shows a crystallographically related C-terminal DDIPF sequence interacting with the R452 side chain and the hydrophobic pocket (unpublished; PDB: 6cbr). It is likely that the identified binding site has functional significance, because SSB is essential for replisome activity. However, the exact roles of the SSB C-terminal peptide within the replisome are not known. One of its functions is to serve as an anchor point for the clamp loader complex at newly synthesised lagging strand primers (Marceau *et al.*, 2011). Here it cannot be ruled out that the inhibitory effect of MB25 on the replisome could be because of blocking of the SSB-clamp loader interaction (since the clamp loader binding site must have similar features as the primase HBD), and binding to the HBD would have no functional consequence. Primase would then stay bound to the primer-DNA duplex, decreasing the effective primase concentration at the replication fork (Yuzhakov *et al.*, 1999). This would then lead to the decreased rate of primer synthesis required to explain the observed MB25-induced increase in lagging strand size (chapter 2). However, a direct

inhibitory effect of binding to the conserved HBD binding site is more likely. It is possible that the SSB C-terminal peptide plays a role in primase-helicase dissociation (this has not previously been investigated by others but is tested in section 4.4.3). In this case, there are two possible mechanisms of replisome inhibition. The first is that MB25 also induces primase-helicase dissociation, but in an uncontrolled manner (*e.g.* before a primer is fully synthesised). The second is that MB25 inhibits this SSB-induced primase-helicase dissociation, causing primase to stay bound to the replication fork with a non-functional, primer-DNA duplex-bound RPD. Since SSB has multiple functions within the replisome and mutating SSB will affect all of them, MB25 might offer an easier way of probing specific aspects of the complex interplay between primase, helicase, SSB, and clamp loader within the replisome.

## 4 Elaboration of primase fragment hits using NMR spectroscopy and X-ray crystallography

### Preface

The author of this manuscript gratefully acknowledges the contributions of James A. Brannigan to the work presented in chapter 4, *i.e.* the expression and purification of full-length primase for NMR screening and confirmation experiments, and of primase RPD for NMR confirmation experiments and for crystallography; and the initial crystallisation and structure refinement of apo RPD. The author also gratefully acknowledges the contributions of Dr. John F. Darby to the work presented in chapter 4, *i.e.* the preparation and data collection of the full-length primase NMR fragment screen.

### 4.1 Introduction

The previous chapters described the identification and biochemical characterisation of fragment inhibitors of the replisome, and the identification of the primase helicase-binding domain (HBD) as the binding site of one of these inhibitors. The half-maximum inhibitory concentration (IC<sub>50</sub>) of this hit (MB25) for inhibiting the whole replisome is approximately 1.5 mM, which is quite weak. This chapter will describe attempts to evolve this and other fragment hits into more potent probes of primase function within the replisome.

A functional primary screen such as the one introduced in chapter 2 would not have identified fragments that bind to replisome proteins without inhibiting their functions. As discussed in chapter 1, fragments that bind to a target do not necessarily modulate its function. Knowledge of such fragments is nevertheless valuable during structure-based elaboration of initial hits, because they may reveal new pockets or binding modes that can be combined with others to increase potency. To detect such fragments, measurement of a biophysical binding event is necessary. Here, NMR spectroscopy is used to find fragments from our screening library that bind to primase, and to guide a small campaign to understand structure-activity relationships (SAR) that can be used to develop high-affinity probes.

## 4.2 Evaluation of fragments by NMR

### 4.2.1 Fragment screening against primase by ligand-observed NMR

Out of the initial set of twenty-five replisome fragment hits, described in chapter 2, two compounds were already tested against primase by ligand-observed NMR (LO-NMR) to confirm that biochemical and biophysical data agreed with each other (Figure 2.10 and Figure 2.11). One hit seemed to bind to the HBD and one to the RNA polymerase domain (RPD). However, it would be desirable to have a larger hit set for each binding site to guide SAR investigations. Therefore, our fragment library was screened against full-length primase by LO-NMR. Three experiments were run with each sample for higher confidence: saturation transfer difference (STD), water ligand observed via gradient spectroscopy (WaterLOGSY), and Carr-Purcell-Meiboom-Gills sequence (CPMG) (see introduction section 1.1.2). Since the initial fragment library was already old when it was screened against the whole replisome in 2016 (see section 2.2.1), a subset of refreshed compound stocks was used for this screen, a total of 558 fragments.

Compounds were screened at 0.5 mM final concentrations in cocktails of six, against full-length wild-type primase at 15  $\mu$ M. A compound was marked as a hit when its spectra showed clear signs of binding in all three experiments. The CPMG experiment was found to be not as sensitive as the other two, often showing spectra with only small decreases in compound peak areas when STD and WaterLOGSY showed clear signs of binding for the same compounds. Therefore, individual spectra were marked with scores of 0, 0.5, or 1, where 0.5 indicates a ‘dubious’ or hard-to-interpret signal. A hit was defined as any compound scoring 2.5 or 3 out of three experiments, allowing one such spectrum with an unclear result. With these criteria, 59 initial hits were found, a 10% hit rate. For fragment screens a 10% hit rate is relatively high but not uncommon (Chen & Hubbard, 2009), especially against enzymes with well-defined pockets. Full-length (FL) primase consists of two protein-macromolecule interaction domains and one enzymatic domain that contains a shallow active site, which make a 10% hit rate slightly higher than expected.

### 4.2.2 Confirmation of fragment hits

Hits were confirmed by testing them in the presence of FL primase as single compounds instead of in their cocktails. Furthermore, they were tested against primase

RPD to get a first indication of which domain contained their binding site. At the time of screening, the primase zinc-binding domain (ZBD) and HBD were not available in sufficient amounts to enable confirmation experiments against them. A selection of hits was later tested against the HBD (see section 4.4.2.2). Out of 59 initial hits, 35 (6%) could be reconfirmed to bind to full-length primase with slightly relaxed criteria (score of 2 or higher). Out of these 35 confirmed hits, 24 showed binding to full-length primase as well as its RPD. With strict criteria (score of 2.5 or higher), 19 out of 59 were reconfirmed to bind to full-length primase, 14 of which also showed good binding to primase RPD. Figure 4.1 shows example confirmation spectra of two hits in the presence of primase RPD: MB106 binds to primase RPD and MB129 does not. The STD experiments (top, purple) show clear peaks above the noise for all but one resonance of MB106, while only the methyl group of MB129 shows an STD difference peak. The CPMG experiments (bottom; blue: reference, red: filtered) show at least 50% signal reduction for all MB106 peaks, but not for all MB129 peaks. The WaterLOGSY experiments (middle, green) show negative peaks for both compounds, indicating no binding. However, the relative intensities of MB106 peaks (compared to the WaterLOGSY dimethylsulfoxide peak and the 1D compound peaks) are much lower than expected, indicating that they are the result of a balance between positive and negative signals. This means that, during signal acquisition, the sample contained a population of compound molecules that did not bind to the protein. This can be explained by 1) slow binding kinetics, 2) weak binding, or 3) too short a mixing time delay before acquisition. In the case of this fragment screen, explanation 2) is the most likely, because fragments rarely display slow binding or high affinity, and because the mixing time was set to the same default value (1.5 s) for all screening samples. Since the primary goal of this screen was not to identify high-affinity fragments, but to find new diverse hits to support development of SAR, such weak binders were evaluated as *bona fide* hits.

For some of these confirmed hits there were concerns about their aqueous solubility based on their predicted partition coefficients (calculated using the AlogP method (Ghose *et al.*, 1998)) or on observations during sample preparation. Compounds near their maximum solubility sometimes form micelles or soluble aggregates, which could appear as false-positives in LO-NMR because they behave as large particles. To test for such false-positives, the same screening experiments are run in the absence of

protein, and any signs of ‘binding’ must be artefacts. Additionally,  $^1\text{H}$  spectra of compound at single and double concentrations can be recorded to compare peak areas. Out of 19 confirmed hits, 4 compounds failed at least one of these tests and were therefore marked as potentially false positives.

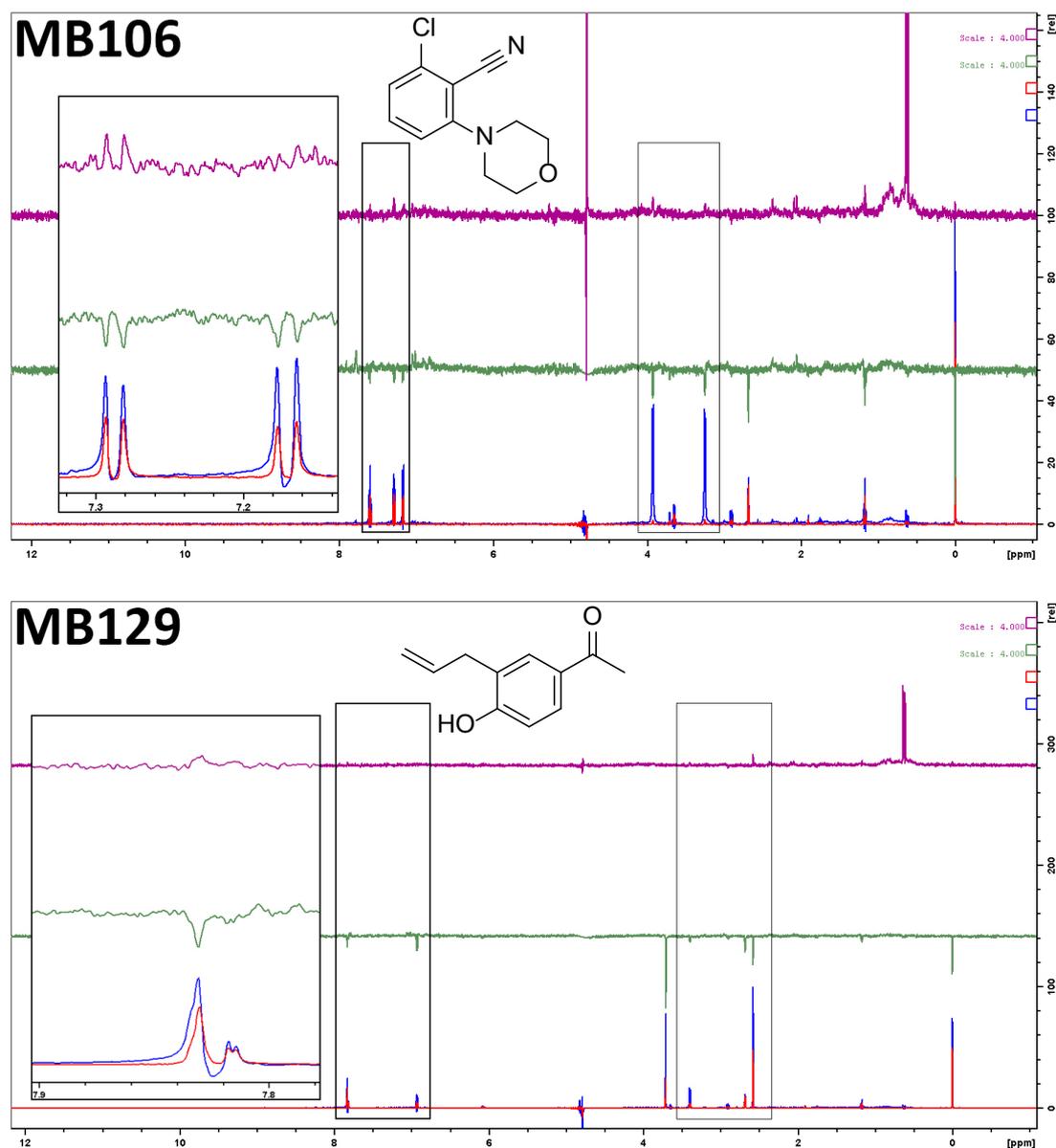


Figure 4.1.  $^1\text{H}$  NMR spectra of two fragment hits (0.5 mM) in the presence of primase RPD (15  $\mu\text{M}$ ). For each sample three experiments are shown: STD difference (top, purple), WaterLOGSY (middle, green), and CPMG (bottom; reference: blue, filtered: red). MB106 shows binding to primase RPD in STD and CPMG experiments, MB129 does not.

Table 4-1 summarises the final screening results after confirmation. Structures of all NMR screening hits against primase can be found in Appendix B. Interestingly, neither MB25 nor MB99 (see section 2.2.3) were identified as primase binders. Nevertheless, MB25 (in its screening cocktail) received a score of 2 and showed clear signs of binding in STD and WaterLOGSY experiments, and MB99 (in its screening cocktail) received a score of 1.5 and showed signs of binding in the WaterLOGSY experiment and reduction of some but not all of its peaks in the CPMG experiment. Furthermore, they were previously confirmed to bind to primase as single compounds. When also recalling that these compounds showed inhibition of replisome activity, a low but non-zero score in a binding screen does not disqualify these compounds from being good starting points for optimisation of primase inhibitors.

The 31 hit compounds that were confirmed as true-positive hits show a large diversity of chemical scaffolds. Two scaffolds are enriched: 3-phenylpyrazole (MB555, MB636, MB911; 3 hits out of 3) and quinoline (MB163, MB500, MB543; 3 hits out of 5). However, it is unlikely that hits in either of these sets target the same binding sites, because each set contains hits that bind to primase RPD as well as hits that do not. Finally, although MB25 was not picked up as a hit in this screen, the other two fragments containing the tetrahydrobenzo[c]thiophene core (MB50 and MB655) were. However, again, these two hits were found to bind to primase RPD in contrast to MB25, casting doubts on their usefulness in SAR. Possibly, binding to the HBD site predicted in the previous chapter was prevented by the absence of a negative charge to pair with the conserved R452. Notably, this NMR screen found no carboxylic acid-containing hits except MB307 which bound to the RPD.

In summary, this NMR screen found 31 primase binders, 22 of which target the RPD. Five hits were also identified as replisome inhibitors previously. There is no indication of binding strength, but since few of these NMR hits inhibit the replisome by more than 25% (the cut-off criterium during replisome screening), many could be weak binders or target non-functional sites. No obvious SAR appears from the current results alone, therefore higher resolution information about the binding domain and binding site of these hits is needed.

Table 4-1. Confirmed fragment hits from the NMR screen against primase. Highlighted hits were judged as binding to full-length primase but not RPD.

Index	Fragment	Score (FL)	Score (RPD)	Replisome screen inhibition	Potential false-positive
1	MB15	2	2	57%	
2	MB31	2	2	41%	
3	MB50	3	1.5	33%	
4	MB51	3	3	32%	CPMG
5	MB85	2	2	27%	
6	MB108	2.5	0	24%	
7	MB142	2	0	20%	
8	MB194	2.5	2	17%	
9	MB228	3	2	16%	Solubility
10	MB286	2	1	13%	
11	MB305	2.5	1.5	12%	
12	MB313	2	0	12%	
13	MB343	2	1	11%	
14	MB345	2.5	1.5	11%	
15	MB361	2.5	0.5	10%	
16	MB371	3	1.5	10%	
17	MB387	2.5	1.5	9%	
18	MB389	3	0.5	9%	WaterLOGSY
19	MB418	2.5	0.5	8%	CPMG
20	MB458	3	2	7%	
21	MB470	1.5	1.5	7%	
22	MB479	3	2	6%	
23	MB497	2	0	6%	
24	MB522	2	2	5%	
25	MB534	3	2	5%	
26	MB615	2	1	2%	
27	MB634	2	1.5	1%	
28	MB660	2	1	0%	
29	MB662	1.5	1	0%	
30	MB666	2.5	1.5	0%	
31	MB715	2	0.5	0%	
32	MB850	3	1	0%	
33	MB870	3	2	0%	
34	MB890	2	0.5	0%	
35	MB980	3	1.5	0%	

### 4.3 Evaluation of primase fragment hits using X-ray crystallography

FL primase consists of three flexibly linked domains (see sections 1.2.3.3 and 3.1) and is not amenable to crystallisation (Zhou *et al.*, 2017). Fortunately, the target domains

of many of our fragment hits have been determined, allowing crystallography efforts to focus on those isolated primase domains rather than full-length primase. To develop these hits into more potent probes, high resolution information about their binding modes and protein interactions is needed. Therefore, attempts were made to determine their crystal structures in complex with their target domain.

#### 4.3.1 Crystal structures of RNA polymerase domain in complex with nucleotide analogues

Initial protein crystallisation conditions were identified from a commercial screen, and these were optimised towards the published crystallisation conditions (PDB: 1dde (Keck *et al.*, 2000)) around pH 5. The first crystals of the RPD construct comprising amino acids (aa) 111-433 (Figure 4.2A) diffracted to 1.5 Å and were resistant to at least 10% dimethylsulfoxide (DMSO). The crystals are tightly packed in the P212121 space group. Another screening condition, at pH 9, yielded a single crystal in the I222 space group after seeding. In this crystal the solvent content is higher, making this form more attractive for soaking experiments. However, subsequent attempts to reproduce the I222 crystal form yielded only the former P212121 crystal form again, albeit with a different morphology (Figure 4.2B). A slightly different RPD construct comprising aa 111-429 also crystallised under the same conditions, but only in the form of needles (Figure 4.2C) with weak diffraction.

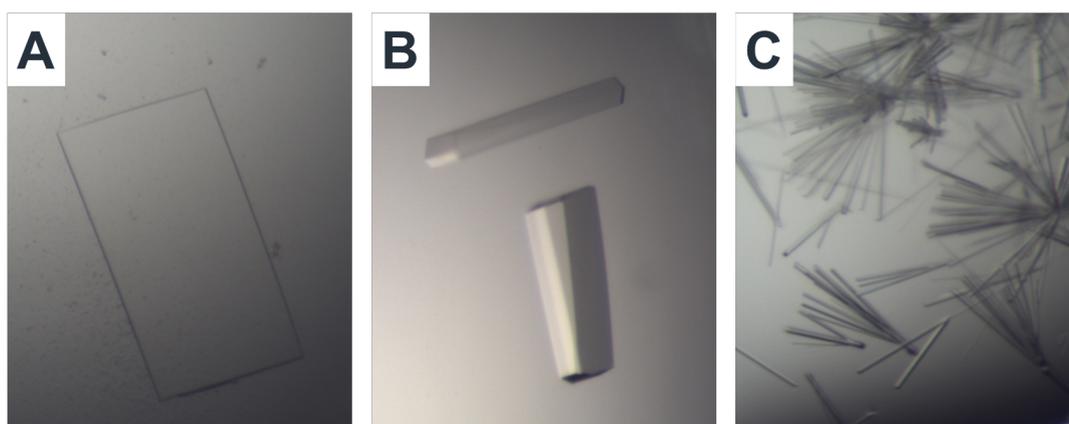


Figure 4.2. Different crystals of the RPD construct in the P212121 space group. **A)** Thin plates grew initially at pH 5 and pH 8.5-9. **B)** Optimisation of conditions and seeding led to crystals in the same space group but different morphology. **C)** RPD construct 111-429 grew into needles under the same conditions as in B.

The structures of the primase RPD (aa 111-433) grown at pH 8.5 in the absence or presence of DMSO (10%) were solved by molecular replacement and refined at 1.5 Å. There were no major differences between our structure and published structures (C $\alpha$  root-mean-square deviation: 0.55 Å). Partial density for a DMSO molecule was present in a small pocket above the active site. All subsequent, soaked crystals were isomorphous, enabling an efficient workflow directly producing difference density maps using these initial apo or DMSO-soaked coordinates. A small number of compounds caused crystals to crack or dissolve upon soaking. In some cases, after data collection unit cell dimensions appeared to have changed significantly, and structures had to be solved by molecular replacement.

Repeated soaking experiments with ATP, adenosine  $\alpha,\beta$ -methyleno-triphosphate (ApCpp or AMPCPP), guanosine  $\beta,\gamma$ -imido-triphosphate (GppNHp or GMPPNP), and guanosine 3',5'-bispyrophosphate (ppGpp), did not yield interpretable difference density for these compounds. However, several datasets showed strong density in the template-binding groove adjacent to the active site, consistent with a di- or triphosphate moiety (Figure 4.3). Possibly, some material contained small amounts of hydrolysed phosphates, or these nucleotides were hydrolysed during soaking. Alternatively, since the RPD has no sequence specificity and binds oligonucleotides weakly, the bases do not interact with the protein and do not adopt specific conformations without a complementary template present. Others have noted the same with *E. coli* primase RPD (Corn *et al.*, 2008) as well as other bacterial homologues (Rymer *et al.*, 2012; Hou *et al.*, 2018). Although these crystals could be packed in a way that occludes the incoming nucleotide subsite, the observation of phosphate density in the template binding site shows that most of the active site space is accessible to small molecules.

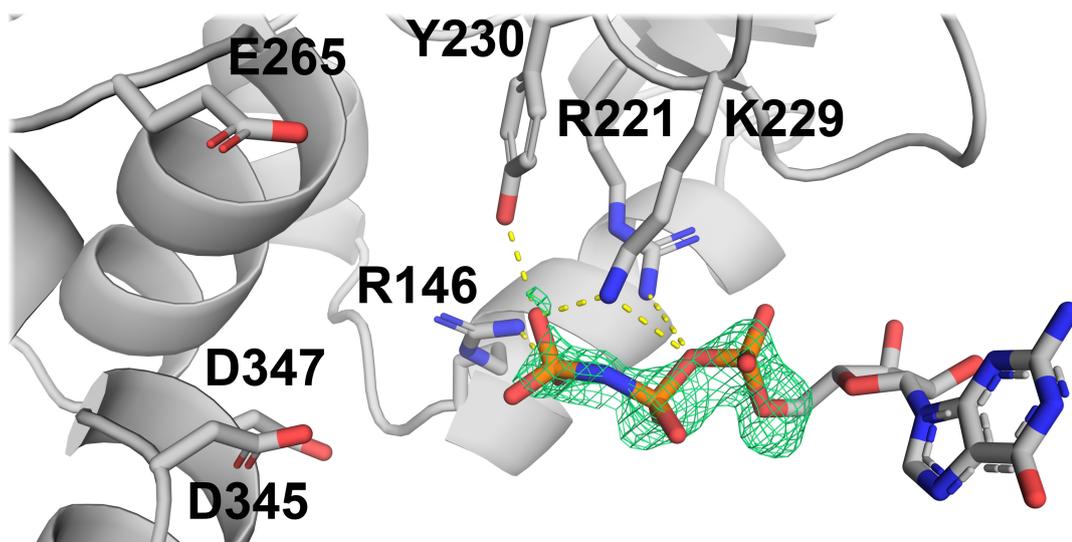
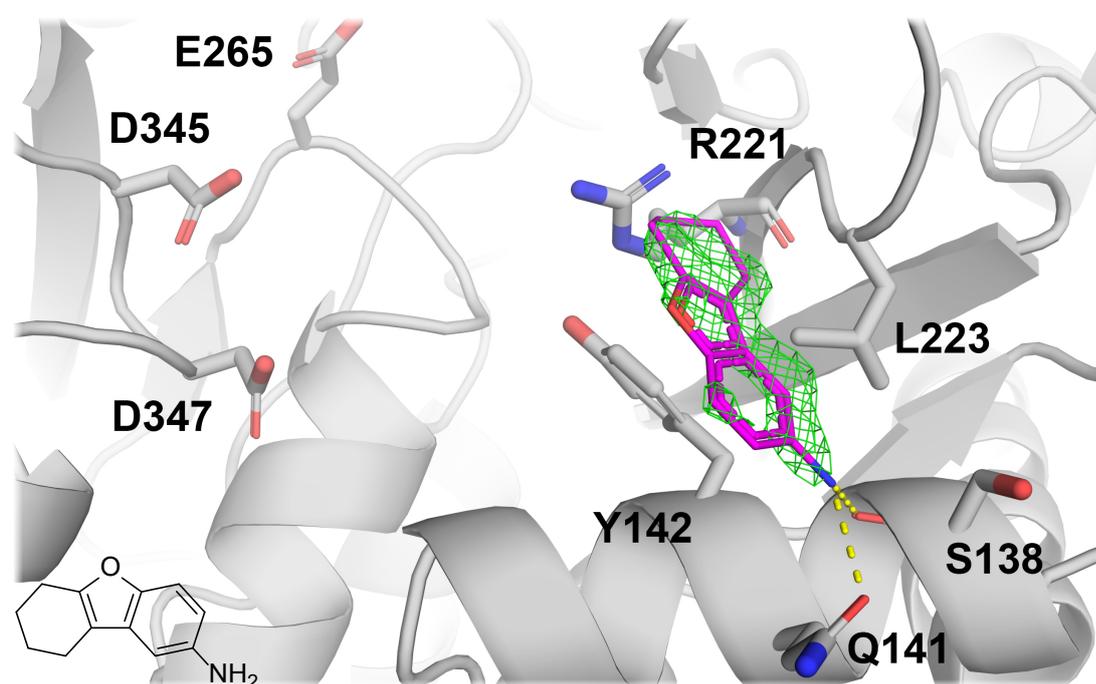


Figure 4.3. Positive difference density (green mesh, contoured at  $\sigma=3$ ) in the active site groove of primase RPD crystals soaked with GppNHp, consistent with the presence of a triphosphate moiety. The sugar and base are not observed and are here modelled to extend into the solvent. E265, D345, and D347 are catalytic residues. Dashed yellow lines represent intermolecular polar contacts.

#### 4.3.2 Crystal structures of RNA polymerase domain in complex with fragments

Since soaking with nucleotides showed that the crystallised RPD active site is accessible to small molecules, several NMR fragment hits were chosen for soaking experiments. Of fragment hits MB15 through MB715 (as presented in Table 4-1), most compounds binding the RPD (16 in total) were used for at least two soaking experiments. Datasets were obtained for at least one crystal soaked with each fragment hit, although dataset quality varied because some compounds caused crystals to crack or dissolve. High resolution limits used for refinements varied between 1.4-2.5 Å. Unfortunately, no clear difference densities large enough to accommodate ligands were found, except in one case. For a selection of compounds prioritised based on their inhibitory activity, solubility, and availability of analogues (compounds MB15, MB31, MB85, MB615), different soaking concentrations and durations were tested, as well as co-crystallisation instead of soaking, but to no avail. Crystals soaked with MB666 did show density consistent with partial occupancy of one MB666 molecule. MB666 binds to a shallow pocket beneath the template-binding groove, its hydrophobic character possibly accommodated by Y142 and L223, and its aniline interacting with Q141 and the backbone carbonyl of S138 (Figure 4.4). It is difficult to imagine this site being functionally important to the RPD, although it is possible

that this surface patch is occupied by DNA template during primer synthesis initiation or elongation. Interestingly, although MB666 does not inhibit the replisome, its tetrahydrodibenzofuran ring structure is analogous to that of replisome hit MB16, a tetrahydrocarbazole. Unfortunately, soaking experiments with MB16 did not result in interpretable density for this compound, suggesting that the similarity between MB666 and MB16 is mere coincidence.



*Figure 4.4. Crystal structure of the RPD in complex with MB666. Difference density (green mesh, contoured at  $\sigma=2.5$ ) is consistent with the partial presence of a molecule the size of MB666. In the binding mode modelled here, the aniline nitrogen makes polar contacts with Q141 and the backbone carbonyl of S138. For reference are shown R221 which is part of the basic template-binding groove, and catalytic residues E265, D345, and D347. Dashed yellow lines represent intermolecular polar contacts.*

The soaking experiments also included MB25, although this compound was not an NMR hit and was shown not to bind the RPD. Nevertheless, crystals soaked with MB25 (30-40 mM) overnight showed strong positive difference density consistent with the presence of a molecule of MB25 in the same spot where MB666 binds (Figure 4.5A). MB25 makes an ionic interaction with R221, and its rings lie in almost the same position and orientation as MB666. A second molecule of MB25 binds in a small

hydrophobic pocket on the other side of the  $\alpha$ -helix containing Q141 (Figure 4.5B). This pocket, too, is likely of no functional relevance.

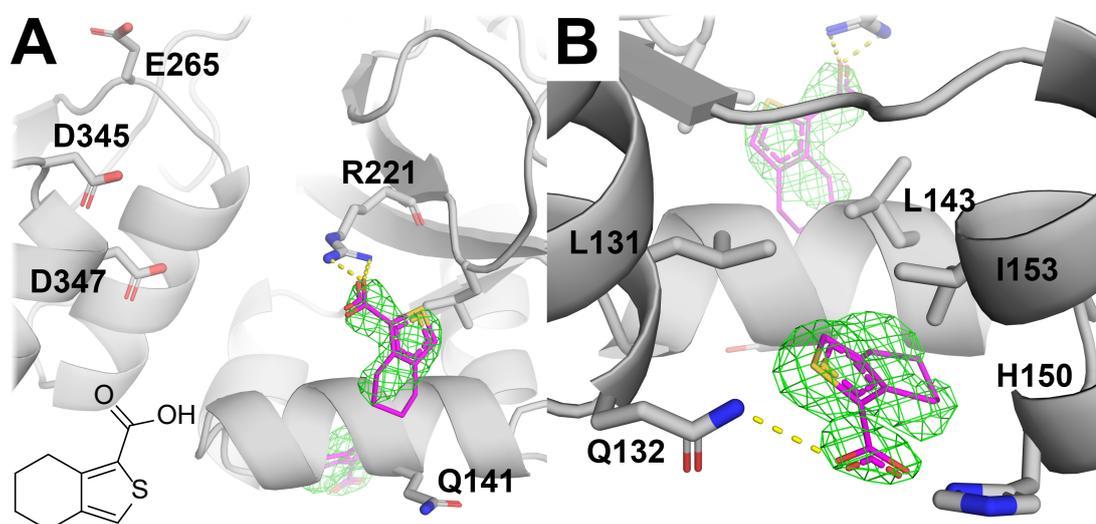
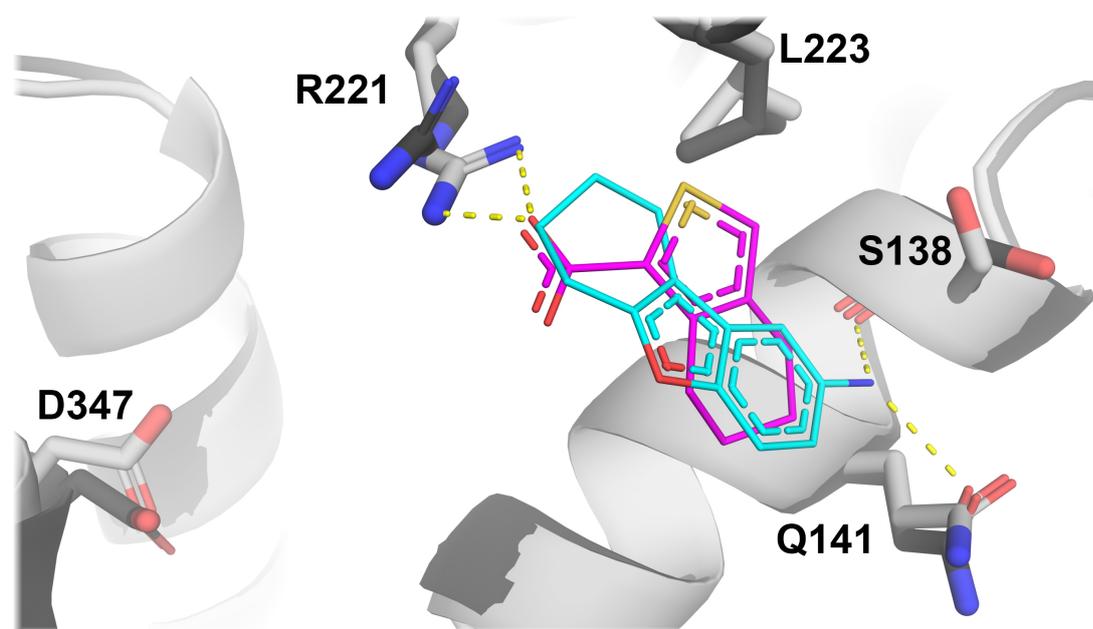


Figure 4.5. Crystal structure of the RPD in complex with MB25. Difference density (green mesh, contoured at  $\sigma=3$ ) is consistent with the presence of two molecules of MB25. **A)** One molecule binds in the same spot as MB666, making polar contacts with R221. **B)** Another copy of MB25 binds on the opposite side of the  $\alpha$ -helix, making polar interactions with Q132 and possibly H150. For reference are shown catalytic residues E265, D345, and D347. Dashed yellow lines represent intermolecular polar contacts.

Since MB25 did not show any signs of binding to the primase RPD during NMR experiments, the observation of the RPD-MB25 complex from crystallography is likely a result of the soaking conditions. The crystallography samples, compared to NMR samples, contained 30 mM of MB25 instead of 0.5 mM, 15% DMSO instead of 0.3%, and 34% ethylene glycol species instead of 0%. Further noting that protein structure is constrained by packing in a crystal, it appears that NMR experimental conditions are much closer to those of the functional assays. Therefore, even though MB25 seems to bind to the RPD *in crystallo*, this must be much weaker binding than to the HBD, and probably does not modulate primase function under normal assay conditions.

The FBLD approach does not require functionally active starting compounds for optimisation. Even if MB666 and MB25 do not inhibit primase in their observed binding modes, compounds evolved from them might show functional activity with

increased affinity. The two partially overlapping fragment binding modes present a good opportunity for fragment merging (Figure 4.6), where the ideally merged compound would interact with both R221 and Q141, resulting in increased affinity.



*Figure 4.6. Overlay of the binding modes of MB25 (magenta, light grey) and MB666 (cyan, dark grey). MB25 interacts with R221 while its rings lie in the same plane as those of MB666 of which the primary amine interacts with Q141. The R221 guanidinium group is pushed slightly away from the cyclohexyl group in the RPD-MB666 crystal structure.*

This idea of merging was tested by purchasing various analogues (as medicinal chemistry support was not available) and testing them by crystallography. Purchased analogues fulfilled the following criteria: 1) a negatively charged (at pH 7.5) H bond acceptor, approximately opposite 2) a H bond donor, separated by 4 or 5 bonds through an aromatic ring. Five such compounds were found available for purchasing and were soaked into RPD crystals (Figure 4.7). At initial soaking concentrations (20-30 mM), RPD crystals were quickly dissolved by the two tetrazole analogues BL3 and BL5. This problem could be overcome for BL3 by lowering its concentration to 5-10 mM, but for BL5 soaking turned out to be impossible. Furthermore, crystals soaked with BL4 did not diffract. After data processing, only maps from crystals soaked with BL3 showed positive difference density in the spot where MB25 and MB666 bind. The density is weak, but matches the two rings of BL3, suggesting partial occupancy (Figure 4.8). It is not possible to deduce the orientation of the BL3 amine and whether

it interacts with Q141. However, even when modelled in the most favourable orientation towards the Q141 sidechain, a H bond is not possible.

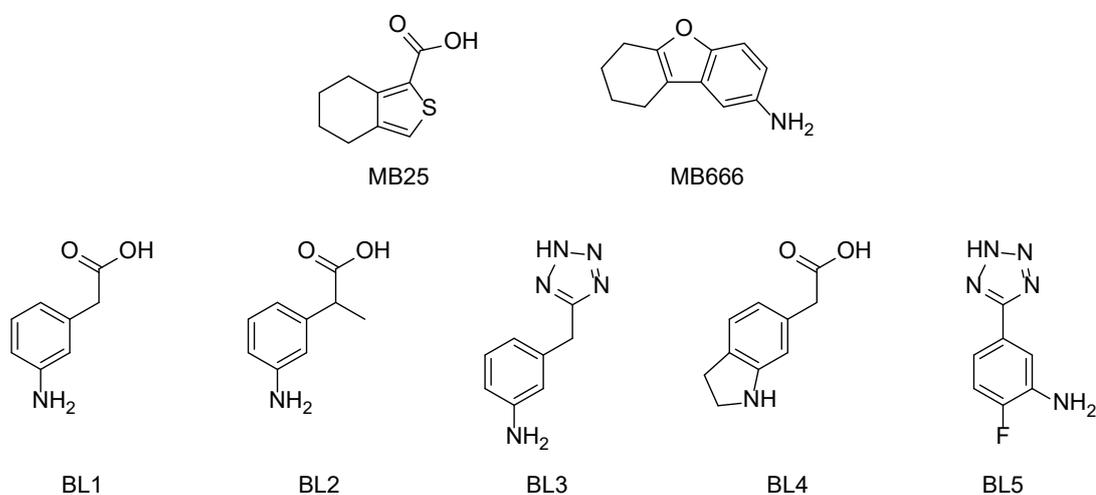


Figure 4.7. Structures of initial fragment hits MB25 and MB666, and of four analogues attempting to merge their functional groups based on their crystal structures. Only BL3 was observed after soaking.

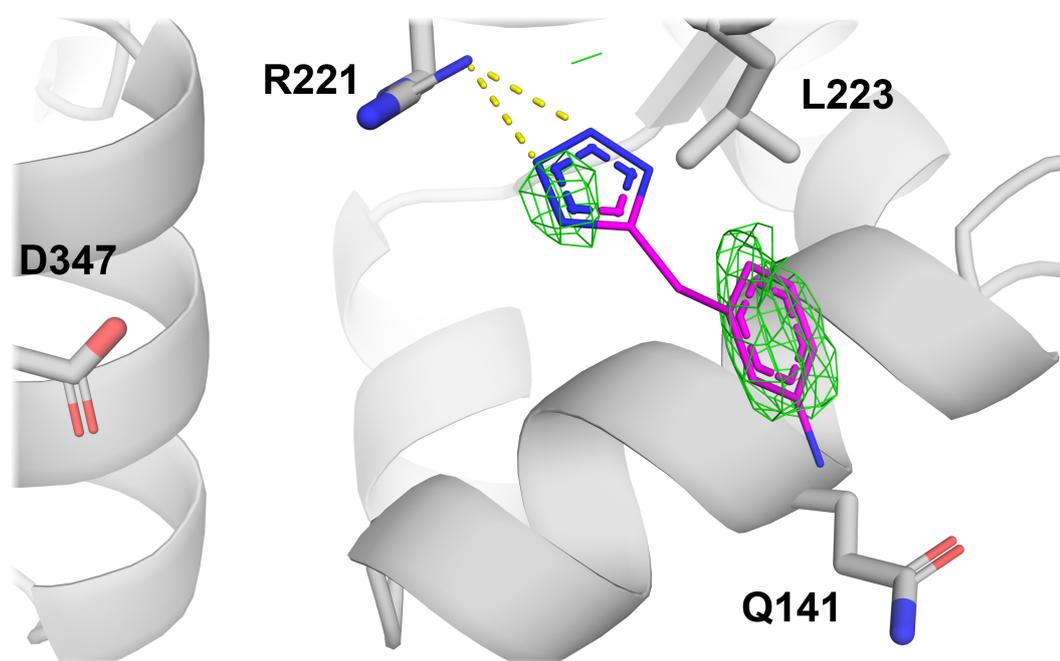


Figure 4.8. Crystal structure of the RPD in complex with BL3. Difference density (green mesh, contoured at  $\sigma=2.3$ ) is consistent with the partial presence of a molecule the size of BL3. In the binding mode modelled here, the tetrazole interacts with R221 but the aniline nitrogen does not make polar contacts with Q141. For reference is shown catalytic residue D347. Dashed yellow lines represent intermolecular polar contacts.

At this point, no options to progress any of the tested hits or analogues by crystallography were available. It was suspected that none of these compounds bound to the RPD active site. Moreover, the site to which MB25, MB666, and BL3 seemed to bind was suspected to be artefactual. Indeed, *in crystallo* this site is part of an artificial small molecule pocket, formed with another protein copy, and forms the end of a solvent channel. Modelled fragments were almost as close to one protein molecule as to the other, although they did not seem to interact directly with the other (*e.g.* Figure 4.9). Shape complementarity could contribute to fragments binding in this spot. The fact that the RPD active site and substrate-binding groove are open and accessible directly next to this spot (as evidenced by the observation of density consistent with phosphates, Figure 4.3) suggests that none of the tested fragments preferred the active site over this crystallographic pocket. Finally, the fact that MB25 was found in this pocket while it did not show binding to the RPD in NMR experiments suggests that this pocket may simply attract any small molecule that fits, because of the presence of the crystallographically related RPD molecule. These doubts, together with the absence of high-resolution structural information on the binding modes of other fragment hits, were enough to halt SAR investigations with the RPD by crystallography. Instead, it was decided to investigate the binding of MB25 to the HBD.

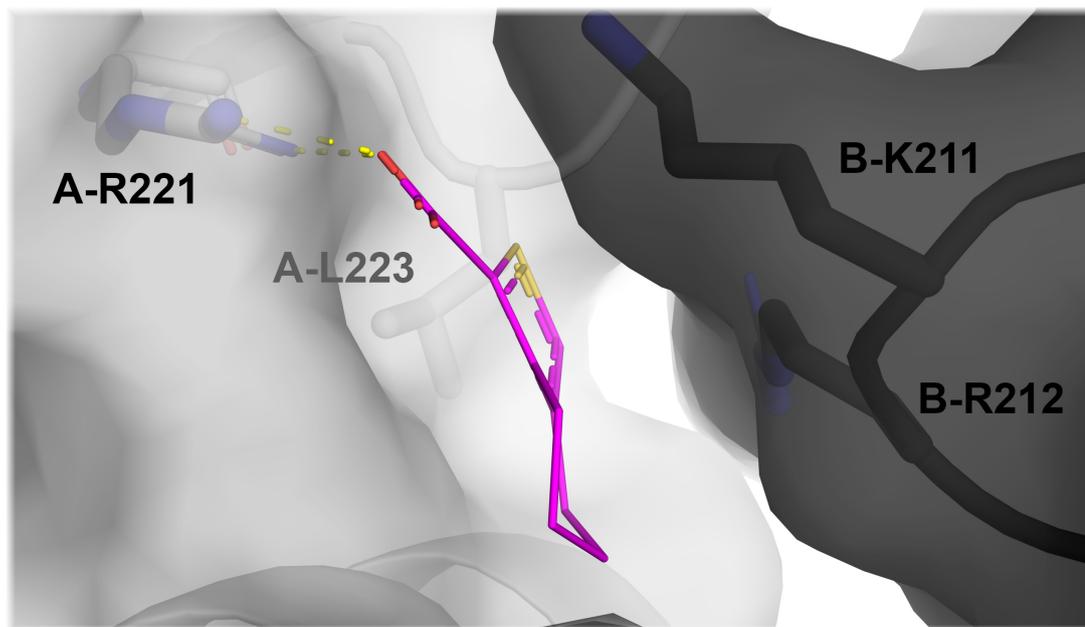


Figure 4.9. Electrostatic surface view of the RPD crystal soaked with MB25 (see Figure 4.5). The next crystallographically related RPD molecule (labelled chain B, dark grey) is very close to MB25. Dashed yellow lines represent intermolecular polar contacts.

#### 4.4 Evaluation of primase fragment hits using NMR

Since the RPD fragments data were not tractable, the HBD was investigated next. Chapter 3 already described that the HBD most likely contains the binding site for MB25, and the putative mechanism of inhibition at this site is consistent with observations from functional assays. Furthermore, several hits from the NMR fragment screen were shown to bind to full-length primase but not the RPD, meaning they could inform SAR investigations for MB25 if they bind to the HBD.

##### 4.4.1 Production of the helicase-binding domain

The HBD by itself would be toxic to cells, since it would compete with full-length primase in the replisome. Therefore, production of *E. coli* HBD constructs by standard over-expression in *E. coli* cells would have very low yields. This concern originally contributed to the focus on the RPD as a target (see sections 4.2 and 4.3) and on molecular dynamics (MD) simulations instead of physical experiments (chapter 3). However, since the simulations showed that the HBD is the target of MB25, and since the RPD fragment hits are not tractable, the HBD is an interesting target to pursue. We found that growth of BL21(DE3) cells transformed with a pET-based vector carrying the HBD gene was indeed very slow (doubling times of approx. 2 hours), suggesting

that low levels of ‘leaky’ expression of HBD was inhibiting replication. Growth improved when steps were taken to prevent such leaky expression: using cells containing the LysS plasmid to prevent T7 polymerase activity, and the addition of glucose to the growth medium to repress the *lac* operon. To improve HBD yields, expression was induced at higher cell density than usual ( $OD_{600} > 1.5$ ). With these steps, average yields (after purification) improved from approx. 1 mg per litre of cell culture to 5 mg, which was high enough to enable an efficient workflow.

#### **4.4.1.1 Helicase-binding domain crystallisation**

The published crystal structure of the *E. coli* primase HBD (PDB: 1t3w (Oakley *et al.*, 2005)) shows a domain-swapped dimer where the C-terminal helix hairpins switch positions and adopt two different conformations. These conformations were shown to be different from the solution structure determined by NMR, and likely an artefact of crystallisation. Therefore, a new crystal form was sought by starting from broad crystallisation screens. Crystals were found in an initial hit condition of sodium formate (4 M) but were not reproducible. One of the crystals was tested at the synchrotron and diffracted to  $\sim 8$  Å showing hints of multiple lattices. To optimise these crystals, a seed stock was prepared and used to screen other conditions. Needle clusters were found in a condition containing sodium acetate (pH 4.6) and ammonium chloride (3.5 M). Unfortunately, optimisation screens for these conditions with or without seeding did not yield any single crystals. Finally, it was hypothesised that addition of the ssDNA-binding protein (SSB) C-terminal peptide (aa sequence [MDFDD]DIPF) as a ligand could stabilise the HBD and improve crystallisation, but this did not yield any results in the original screens. Because HBD crystals were not forthcoming, focus was shifted to protein-observed NMR (PO-NMR) experiments to obtain structural information about fragment binding.

#### **4.4.2 Characterisation of helicase-binding domain fragment hits by NMR**

##### **4.4.2.1 Binding of MB25 to the helicase-binding domain**

The molecular dynamics (MD) simulations and experimental NMR results presented in chapter 3 suggested binding modes of MB25 to the HBD. Most of the experimentally determined protein chemical shift perturbations (CSPs) (Table 3-3) match with the residues predicted by simulations to be in proximity or contact with MB25 (Table 3-2). To further confirm these results, HBD CSPs were measured in the

presence of MB25 at 5 mM instead of 0.5 mM, a concentration at which the protein should be >50% occupied and which is closer to the simulated MB25 concentration at the chosen box volumes (approx. 6.2 mM). Most of the top ten CSPs (Table 3-3) do not change between 0.5 mM and 5 mM, except that R559 is replaced by I530, and I455 becomes ranked below R452, D524, and M451. To visualise the observed CSPs on the HBD structure, relative CSP sizes (*i.e.* the residue showing the largest shift is assigned a value of 100% and all other shifts are set to fractional values relative to 100%) are shown as a colour spectrum (Figure 4.10A). Compared to the top ten residues from MD simulations by proximity or polar contacts (Figure 4.10B), the same binding pocket (site 1) is highlighted. However, additional residues that showed large CSPs, located on the opposite side of the first  $\alpha$ -helix of the HBD, were not predicted by simulations.

The residues between M521-A531 are part of a flexible loop and the beginning of the long helix hairpin, and likely form an interaction surface with helicase (indicated by structural and sequence homology to known bacterial primase-helicase structures). The observation that MB25 perturbs the backbone amides of these residues suggests either that a second molecule of MB25 binds to this separate site (site 2), or that MB25 binding to site 1 causes conformational or H bond network changes that extend to this loop. If the CSP profile is caused by two separate binding events (the former scenario), it is likely that the two binding sites have different affinities for MB25. In that case, the two binding events could be distinguished by performing a titration by PO-NMR to determine saturation rate constants for each NH cross-peak. To this end, sequential  $^1\text{H}$ - $^{15}\text{N}$  heteronuclear multiple quantum coherence (HMQC) spectra of  $^{15}\text{N}$ -labelled HBD were recorded in the presence of MB25 at concentrations between 0.1-5.0 mM. CSPs for all peaks were measured and fitted to a one-to-one binding model to estimate their binding affinities ( $K_{\text{D}}$ s). An overlay of spectra for each titration point shows that a number of peaks is selectively perturbed by MB25 (Figure 4.11). While calculated  $K_{\text{D}}$  values varied between individual peaks (Table 4-2), all were within the same approximate range with an average value of 2.6 mM.

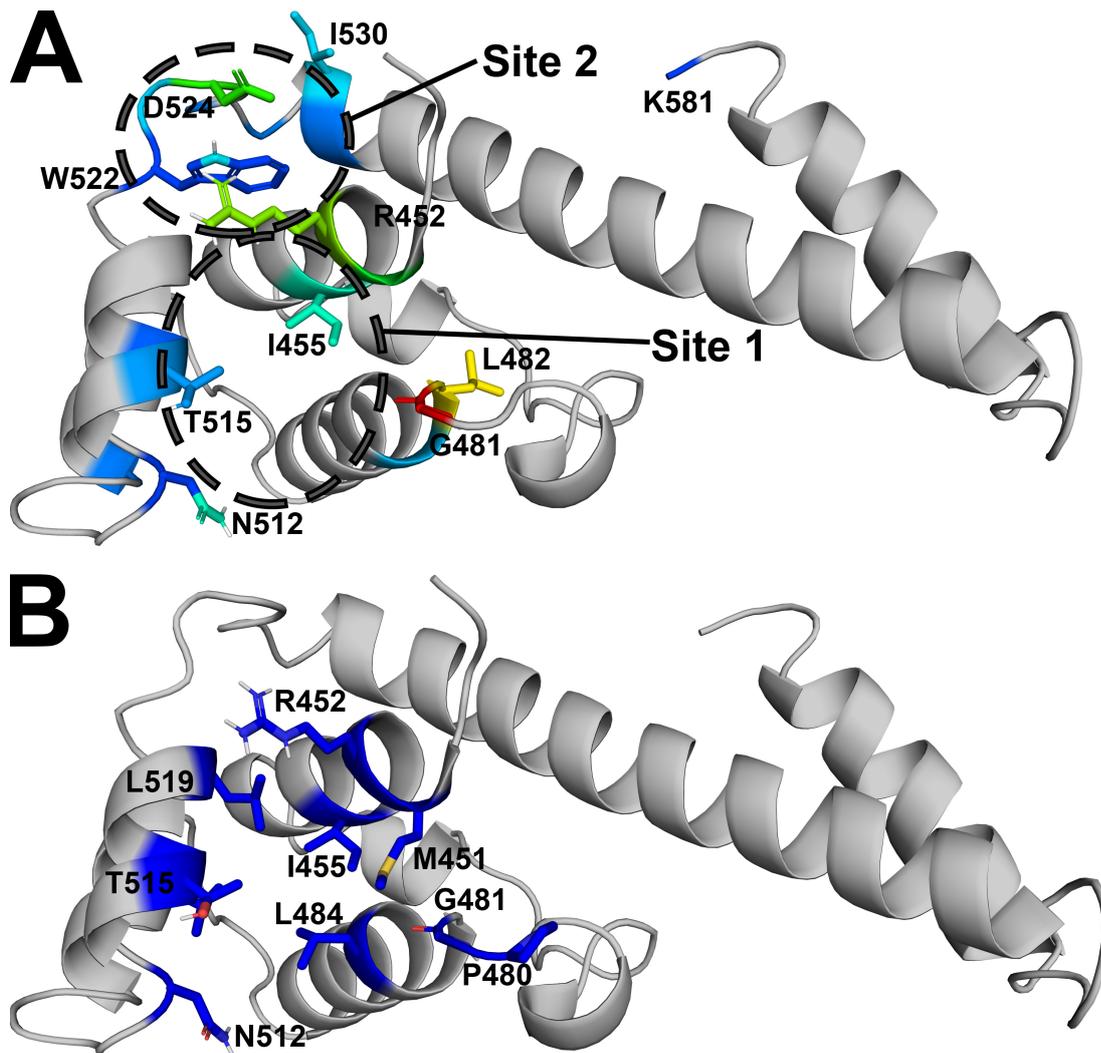


Figure 4.10. Comparison of HBD residues that interact with MB25 (coloured) observed by NMR (A) or MD simulations (B). Residues in A are coloured according to CSP magnitude: grey for negligible shifts, dark blue for small shifts, through longer wavelengths for larger shifts. The largest shift observed was for G481 (red). Note that N512 and W522 sidechain NH cross-peaks showed larger shifts than their backbone amides. Shifts in site 2 were not predicted by simulations.

Most peak trajectories were of the same length as the distance between their first and last point, suggesting that these shifts are caused by a single event. Notably, there is no distinction between fitted values for residues belonging to one or the other site. However, shift distances for residues belonging to site 2 were all below one standard deviation, except for that of D524. Taken together, these results do not suggest the occurrence of two separate binding events of MB25 to site 1 and 2. It is possible that the shifts of residues in site 2 are caused allosterically. Since site 2 is most likely involved in helicase binding, this would support a hypothesis that MB25 inhibits the

replisome by binding to site 1 of the primase HBD. This hypothesis is further supported by the fact that binding of the SSB C-terminal peptide (aa sequence MDFDDDIPF), its most important interactions being between its carboxylates and K447/R452/K518 (Naue *et al.*, 2013), shows a much larger number of shifts than could be caused by direct interactions, including in site 1. Furthermore, the MB25 HBD dissociation constant (2.6 mM) and its replisome IC<sub>50</sub> (1.5 mM) are in reasonable agreement. Thus, MB25 seems to have the same putative mechanism of action as the SSB C-terminal peptide. Finally, this hypothesis is also supported by primer synthesis experiments with helicase and primase (see section 4.4.3.5).

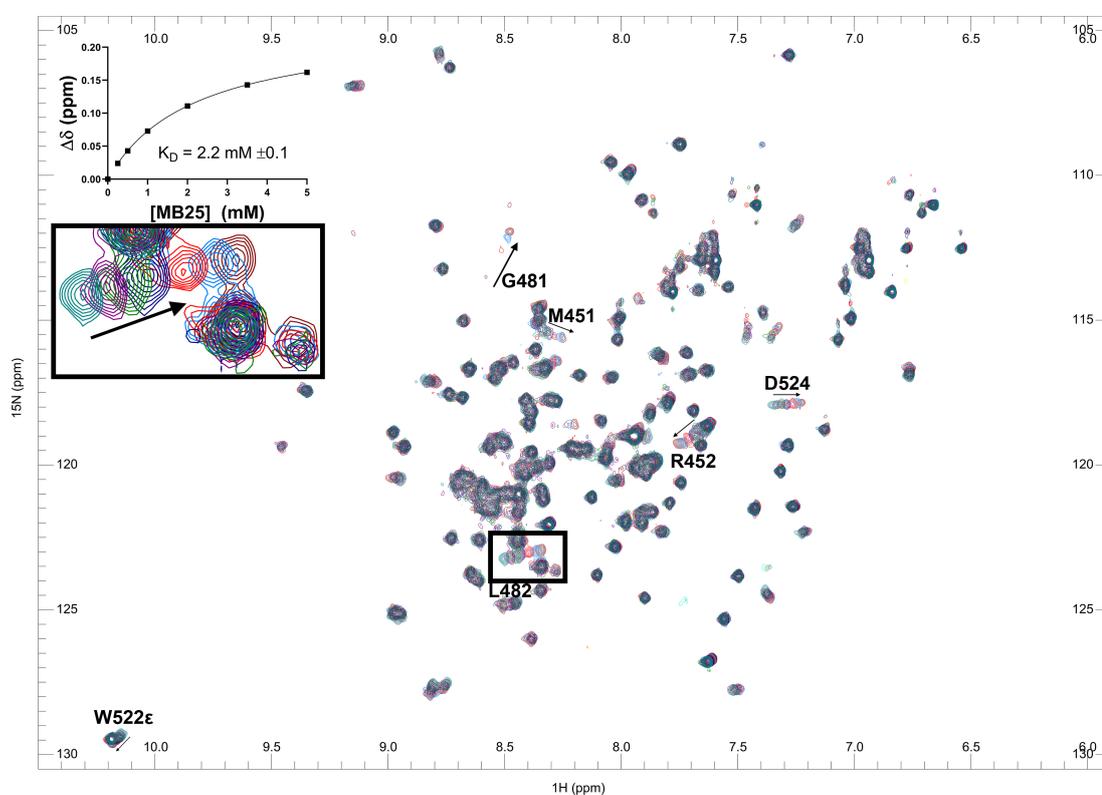


Figure 4.11. Titration of MB25 with <sup>15</sup>N-labelled HBD by HMQC (overlay of seven spectra). Several peaks are highlighted. One of the shifting peaks (L482) is shown in detail and its shifts plotted against ligand concentration to derive a  $K_D$  value of 2.2 mM for this peak.

Table 4-2. CSP measurements and fit parameters for the titration of MB25. The average shift distance of all visible peaks was 0.023 ppm with a standard deviation of 0.032 ppm. Only CSPs of  $+1\sigma$  or larger were used for calculation of average  $K_D$ . Only fitted  $K_{DS}$  with fit errors of less than 30% were used for calculation of average  $K_D$ .

NH	Trajectory Distance (ppm)	Shift Distance (ppm)	$K_D$ (mM)	$K_D$ error
G481	0.269	0.262	3.4	0.9
L482	0.164	0.162	2.1	0.1
R452	0.147	0.134	1.6	0.2
D524	0.117	0.116	2.3	0.2
M451	0.100	0.096	4.0	0.6
I455	0.074	0.073	7.4	4.0
N512δ2a	0.084	0.071	2.5	0.4
N512δ2b	0.082	0.069	2.1	0.4
I530	0.054	0.053	7.3	2.4
W522ε	0.049	0.049	1.9	0.4
T515	0.044	0.041	1.4	0.3
K581	0.027	0.027	7.4	1.5
W522	0.031	0.026	1.7	0.3

#### 4.4.2.2 Binding of NMR screening hits to the helicase-binding domain

All original NMR screening hits (Table 4-1) were also tested by LO-NMR against unlabelled HBD for direct confirmation of binding. As expected, most hits did not show any signs of binding to the HBD. Although the size of the HBD (17 kDa) is close to the lower limit of LO methods, potentially causing some false-negative results, several hits showed binding to the HBD in at least two out of three experiments (Table 4-3). Six fragments showed signs of binding to both RPD and HBD and three fragments showed preferential binding to the HBD. There appears to be no common structural motif among any of these HBD binders.

Table 4-3. NMR screening hits that were found to bind to the HBD. Highlighted hits were previously judged as binding to full-length primase but not RPD.

Index	Fragment	Score (FL)	Score (RPD)	Score (HBD)	Replisome screen inhibition
1	MB15	2	2	2.5	57%
2	MB50	3	1.5	1.5	33%
3	MB108	2.5	0	1.5	24%
4	MB194	2.5	2	1.5	17%
5	MB305	2.5	1.5	2	12%
6	MB343	2	1	2	11%
7	MB345	2.5	1.5	1.5	11%
8	MB497	2	0	1.5	6%
9	MB666	2.5	1.5	2	0%

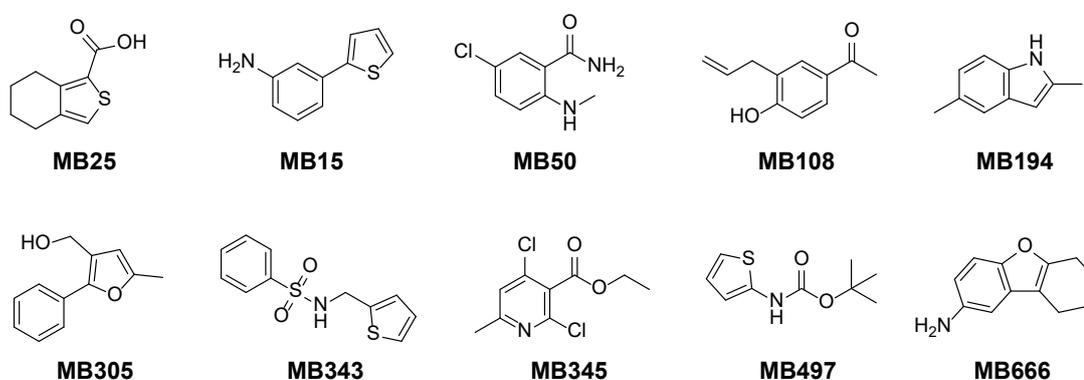


Figure 4.12. Structures of NMR fragment hits that showed signs of binding to the HBD in LO-NMR experiments. MB343 was further investigated because it showed preferential binding to the HBD.

Of these nine fragment hits, MB343 appeared to be the most interesting compound to take forward, because it showed the clearest preferential binding. Unfortunately, MB343 was not soluble at concentrations higher than the screening concentration (0.5 mM), making  $K_D$  determination by NMR impossible. The HMQC spectra of MB343 with the HBD showed that MB343 perturbs many of the same amides as MB25 does (see Table 4-5). Thus, MB343 possibly interacts with R452 by accepting a H bond with its sulfonamide group while one or both aromatic rings point towards the HBD hydrophobic core residues L482, L519, and others.

### 4.4.3 Binding and activity of HBD fragment hit analogues

Full medicinal chemistry support was not available for this project, but fortunately several MB25 and MB343 analogues were commercially available. To track SAR, PO-NMR was used to gain low-resolution information about which residues are involved in compound binding, and functional assays were used to assess whether changes in binding were coupled to changes in compound activity. The standard whole-replisome replication assay, also used for screening and hit confirmation (chapter 2), was used to check activity of compounds against the whole replisome, a prerequisite since the primary goal of this project is to find probes for the whole replisome. A primer synthesis assay was used to assess primase activity specifically. In this assay (Figure 4.13), helicase is present to boost primase activity, which improves signal accuracy and assay throughput.

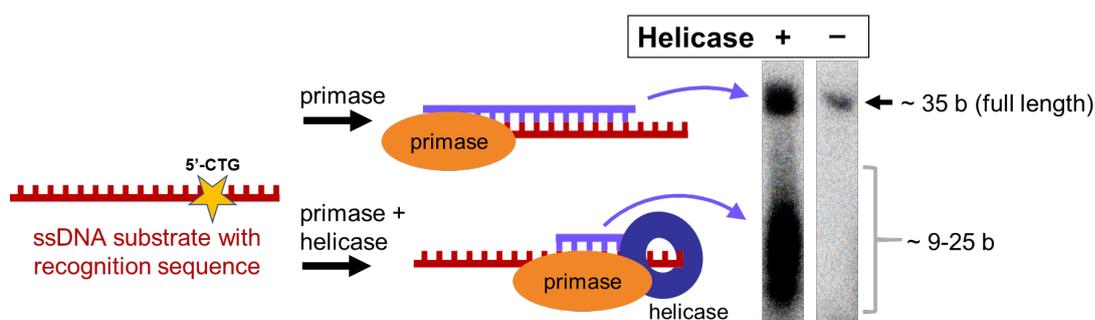


Figure 4.13. Primer synthesis assay used to assess activity of analogues against primase. Reaction products were separated by polyacrylamide gel electrophoresis. Different reaction profiles are observed in the presence or absence of DnaB helicase (the exact cause of the difference is unknown but assumed to be dependent on primase-helicase interaction). Detection of reaction products by quantifying pixel intensities is based on incorporation of radioactive nucleotides.

As the DNA template in this assay is short (40 b) and single-stranded, helicase enzymatic activity does not play a role in primer synthesis. The increase in primase activity with addition of helicase likely results from the increased affinity of the helicase-primase complex for the template. The appearance of shorter reaction products is not easily explained, but similar profiles have been reported by others (Johnson *et al.*, 2000; Maciąg *et al.*, 2010). If the synthesis of these short primers depends on an interaction between primase and helicase (a reasonable assumption), they could be used to deduce the mechanism of action of an inhibitor: if an inhibitor

directly inhibits primase activity at the active site, then the amounts of all reaction products should decrease equally, whereas if an inhibitor disrupts the primase-helicase interaction, then the full-length products should remain less affected. Interestingly, addition of the SSB C-terminal peptide (sequence MDFDDDIPF), which binds to the HBD globular domain, to this reaction results in effects similar to removal of helicase (Figure 4.14). This result suggests that the SSB C-terminal peptide blocks the primase-helicase interaction.

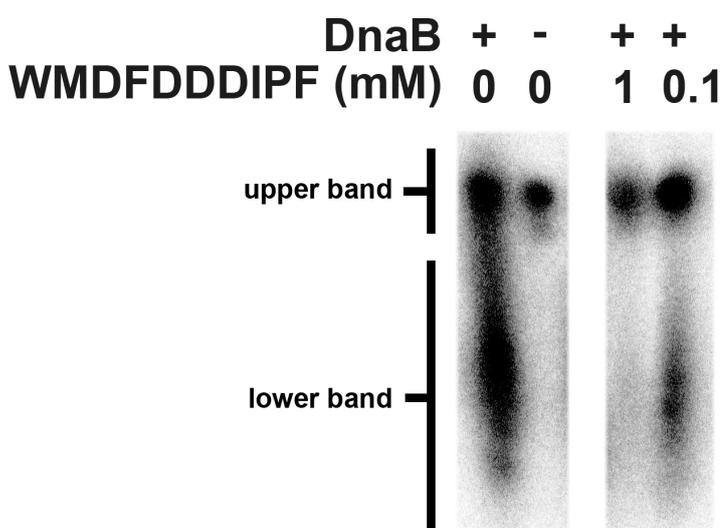


Figure 4.14. Effect of the SSB C-terminal peptide (sequence MDFDDDIPF) on the primase-helicases primer synthesis reaction. The peptide binds to the primase HBD and could interrupt the primase-helicase interaction. DnaB was present at 160 nM. DnaG was present at 200 nM. An N-terminal tryptophan was added to the peptide for quantification purposes.

#### 4.4.3.1 MB25 analogues

First, the MB25 scaffold was investigated in detail to gain an understanding of the importance of its pharmacophore features. Of a more practical concern, it was noted that benzo[c]thiophene compounds were less common among commercially available fragments than benzo[b]thiophenes, and that benzothiophenes with partially hydrogenated benzene rings were less common than fully aromatic ones. Therefore, fragments with small variations in each of these properties were tested by NMR and the two assays. Most analogues were initially tested for HBD binding by LO-NMR to save on  $^{15}\text{N}$ -labelled protein, after which the positives were also tested by HMQC at

single concentration. The first set of analogues (Figure 4.15) investigated the sulfur position, benzene hydrogenation, and presence of the benzene ring. Binding and activity data (Table 4-4) seem to agree that an aromatic benzene and either position of the sulfur in the ring are tolerated (BL6, BL7). However, replacing the six-membered fused ring by smaller single substituents (BL9, BL10) causes loss of binding and activity. Keeping the six-membered ring while making it more hydrophilic (BL11) has the same detrimental effect, indicating that a hydrophobic group is necessary in that position. Interestingly, the replacement of a carbon by a nitrogen next to the sulfur (BL8) causes decreased binding and activity compared to the original hit. The affinity of BL8 for the HBD was determined by HMQC titration up to 5 mM, and an average  $K_D$  of 4.8 mM was extrapolated. This value is based on only four peaks that showed significant shifts (R452, G481, L482, D524). The addition of a nitrogen makes the ring slightly less aromatic and slightly more polar, but it is hard to imagine how this could affect binding. The nitrogen replaces not only a carbon atom but also its hydrogen. It is known that aromatic rings containing sulfur can form intermolecular interactions with carbonyl oxygens, with bond strengths approaching that of H bonds under ideal circumstances (Zhang *et al.*, 2015). Such interactions require a CH next to the sulfur and are formed between an oxygen lone pair and a  $\sigma$  hole located approximately along the S-C bond. Such an interaction would explain why changing the sulfur position has lower impact on HBD binding than the replacement of CH by N: the  $\sigma$  hole moves less than one bond length. A candidate for the interacting carbonyl could be that of G481, although this is entirely speculative based on MD simulation binding poses.

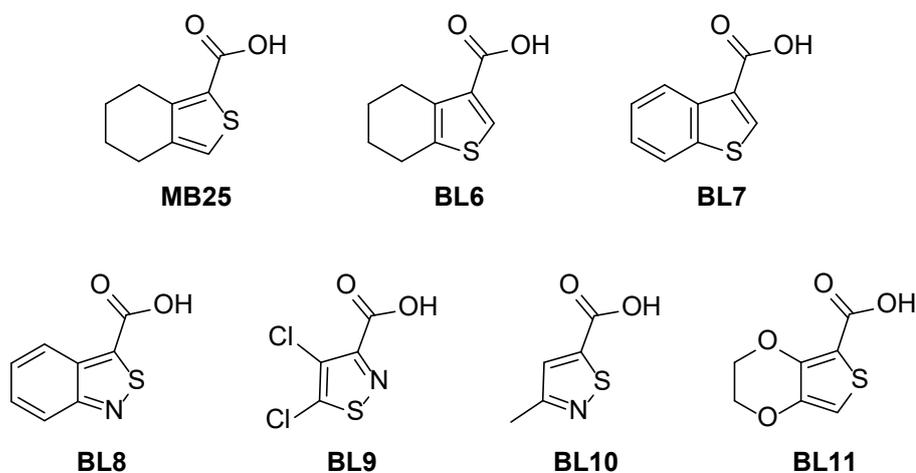


Figure 4.15. First set of MB25 analogues tested for activity and HBD binding (see Table 4-4).

Table 4-4. Binding and activity results for fragment analogues. Average binding affinities are given for analogues that were titrated by HMQC. Amounts of inhibition determined at 1 mM compound concentration. N.D.: not determined.

Index	Fragment	LO-NMR binding	PO-NMR binding	Replisome inhibition	DnaG-DnaB inhibition
1	WMDFDDDIPF	yes	yes	89%	72%
2	MB25	yes	2.6 mM $\pm$ 0.8	28%	35%
3	MB343	yes	yes	16%	N.D.
4	BL6	yes	yes	21%	43%
5	BL7	yes	yes	24%	32%
6	BL8	N.D.	4.8 mM $\pm$ 2.6	16%	N.D.
7	BL9	N.D.	no	4%	N.D.
8	BL10	N.D.	no	0%	N.D.
9	BL11	N.D.	no	0%	N.D.
10	BL12	no	no	0%	N.D.
11	BL13	no	2.4 mM $\pm$ 0.6	9%	N.D.
12	BL14	no	yes	19%	39%
13	BL15	yes	yes (weak)	26%	N.D.
14	BL16	yes	2.2 mM $\pm$ 0.8	27%	N.D.
15	BL17	no	N.D.	N.D.	0%
16	BL18	no	N.D.	6%	N.D.
17	BL19	no	N.D.	13%	N.D.
18	BL20	no	N.D.	7%	N.D.
19	BL21	yes	N.D.	12%	27%
20	BL22	yes	2.2 mM $\pm$ 0.3	22%	45%
21	BL23	N.D.	yes	29%	N.D.
22	BL24	N.D.	no	0%	0%
23	BL25	no	N.D.	29%	39%
24	BL26	yes	yes	58%	N.D.
25	BL27	yes	1.4 mM $\pm$ 0.3	92%	29%

Next, several analogues were tested that represented slightly larger variations on MB25 (Figure 4.16, Table 4-4) to find ways to improve binding affinity. As expected, the presence of a negative charge is necessary for binding as replacing it with a neutral H bond acceptor of similar size (BL12) abolishes binding and activity. Repositioning the carboxyl group (on a fully aromatic scaffold) (BL13, BL14) is tolerated but does not improve binding or activity. Adding a H bond donor next to the sulfur as a replacement for the hypothesised S-O bond (BL15) is also tolerated. On the benzene side of the scaffold, unfortunately very few analogues were commercially available. Addition of a single heavy atom (BL16) does not appear to improve potency. However, it is reconfirmed that the polarity of this side of the ligand is important, as

addition of an amine seems to abolish binding. The data for this second set of analogues suggests that larger additions to the scaffold are needed to improve potency, provided they preserve the hydrophobic-negative character of the ligand. Unfortunately, no such larger molecules were commercially available that would also make a comprehensive SAR series.

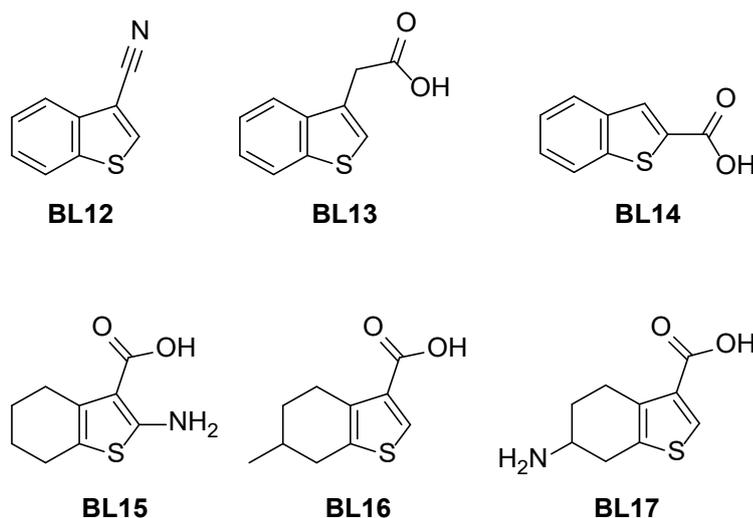


Figure 4.16. Second set of MB25 analogues tested for activity and HBD binding (see Table 4-4).

#### 4.4.3.2 MB343 analogues

Because a way forward with the MB25 scaffold was not clear, SAR for hit MB343 was investigated first. Availability of MB343 analogues with modifications on the phenyl group was low, but several compounds with substituents on the sulfonamide were commercially available (Figure 4.17). Since MB343 is not soluble above 1 mM in HBD NMR buffer, analogues with lower predicted hydrophobicity were prioritised. Unfortunately, analogues BL18, BL19, and BL20 containing charged groups did not show binding or inhibition (Table 4-4). BL21, however, showed slight activity in the whole-replisome assay. BL22, closely resembling MB343, was the most active analogue in this set. From these data, it appears that a hydrophobic character is still needed, similar to the MB25 analogues series. It is unclear whether the carboxylate on BL22 strengthens binding or simply increases solubility. The increased solubility

allowed titrations of BL22 and an average  $K_D$  of 2.2 mM was calculated. BL22 perturbed many of the same peaks as MB25 did, including G481, D524, and N512 $\delta$  (Table 4-5), which is an improvement over its parent fragment MB343.

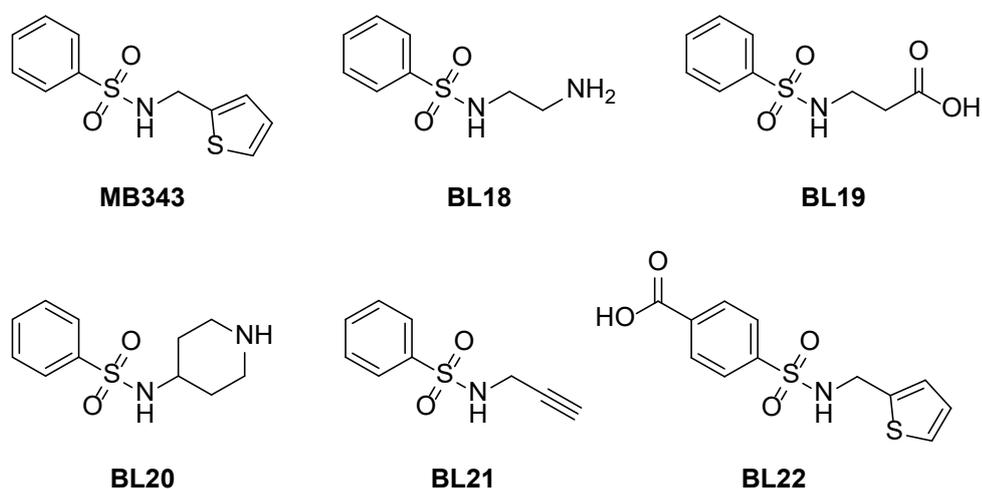


Figure 4.17. First set of MB343 analogues tested for activity and HBD binding (see Table 4-4).

R452 was also shifted by BL22 by more than one standard deviation, but its curve showed a weaker  $K_D$  (6.5 mM) than those of the other peaks. Determining the R452 CSPs is somewhat difficult at low concentrations because its <sup>1</sup>H-<sup>15</sup>N cross-peak is normally overlapped with that of K580, but in the case of BL22 it was clear that neither of these peaks were perturbed at concentrations up to 0.5 mM. The simplest explanation for this differing  $K_D$  would be a secondary binding event, but that would not explain why all other perturbed residues around R452 show similar, lower  $K_D$ s. Alternatively, the observed shift of R452 at high BL22 concentrations could be due to non-specific binding or to a different, weaker, competing binding mode. Indeed, although parent fragment MB343 could not be titrated to high concentrations, its spectrum at 1 mM also clearly shows that R452 is unperturbed. This suggests that binding of the MB343 scaffold does not induce changes to the local environment of the R452 backbone amide. In contrast, all MB25 analogues involve R452 shifts that saturate at the same rate as the other shifting amides. These data suggest that MB25 and analogues contain a moiety that can shift the R452 amide peak, most likely the carboxylate. Furthermore, binding of BL22 depends on a different set of interactions (of similar strength to that of MB25) that does not include an R452-carboxylate

interaction. The SAR data for these two series (MB25 and MB343) raised hopes that it would be possible to combine the two scaffolds in an energetically favourable way.

#### 4.4.3.3 *Further analogues*

Although further elaboration of fragments became increasingly specific, a small number of compounds that fit with the observed SAR so far were commercially available. Firstly, it was determined that the BL22 scaffold was already optimal, as analogues that contained an extra methylene (BL23), a saturated ring (BL24), or a differently oriented carboxylate (BL25) were not more active (Figure 4.18) (Table 4-4). The absence of any signs of binding or activity for BL24 suggests that the aromatic analogues possibly benefit from intramolecular  $\pi$ - $\pi$  stacking to form a single hydrophobic part from which the sulfone protrudes, not dissimilar to the MB25 scaffold. If so, the *p*-carboxylate group would point in opposite direction and would not be able to interact with the R452 side chain, as already hypothesised in the previous section. Why BL25 with a *m*-carboxylate would not show binding to the HBD is less clear. Possibly, the LO-NMR experiments were false-negatives (as for BL14) that unfortunately led to omission of this compound from further PO-NMR experiments. However, judging from its inhibitory activities in the two assays, BL25 would not have a significantly higher affinity for the HBD than BL22.

To combine elements of the two scaffolds MB25 and MB343, analogues BL26 and BL27 (Figure 4.18) were selected. BL26 also builds on the BL16/BL17 analogues (Figure 4.16), adding a bulkier hydrophobic group where a methyl did not improve affinity. Unfortunately, though not unexpectedly, BL26 was insoluble in NMR buffer above 0.5 mM (as evidenced by diminishing compound proton peak integrals). Several lower concentrations below 0.5 mM were tested, but there was no sign of saturation of CSPs, indicating that its  $K_D$  is higher than 0.5 mM.

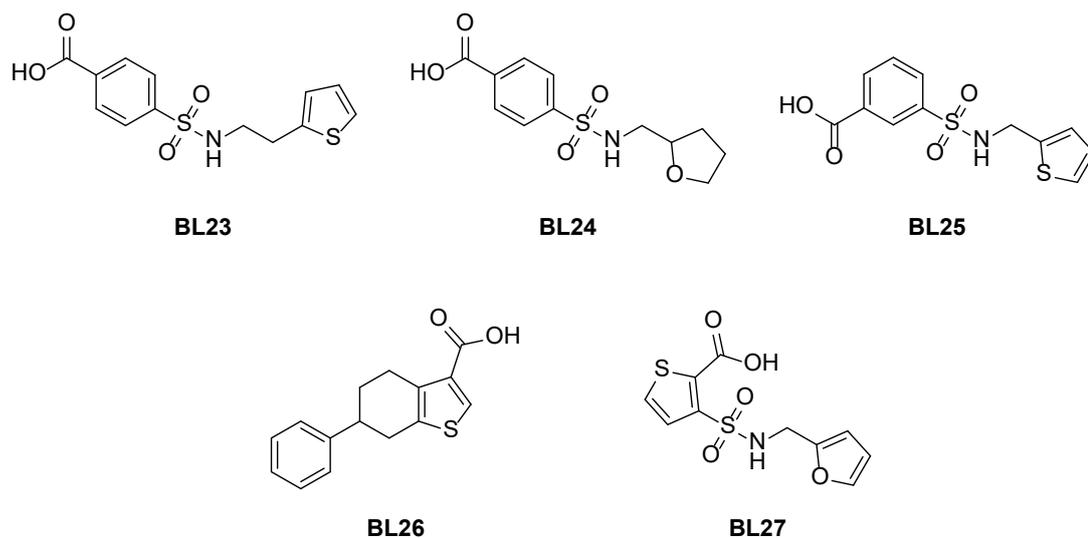


Figure 4.18. Further MB25/MB343 analogues tested for activity and HBD binding (see Table 4-4).

BL27 replaces the phenyl ring with a thiophene, making it the analogue with the most characteristics from both fragment hits MB25 and MB343, only lacking the aliphatic character of MB25. The ‘*ortho*’ carboxylate is a placement that had not been tested yet. BL27 was found to be a superior compound to MB25 and MB343, in terms of both binding ( $K_D$  1.4 mM) and inhibitory activity (see section 4.4.3.5).

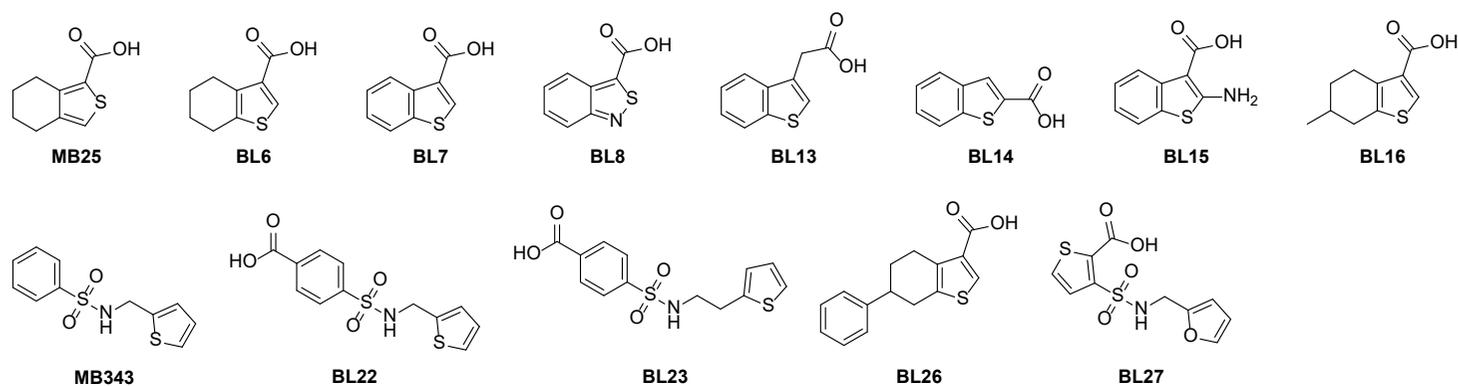
#### 4.4.3.4 Analysis of residue interaction patterns

The 2D NMR spectral data collected for these analogues, given the amide peak assignments, contain detailed information about which analogues interact with which residues. Comparison of these interaction patterns gave several insights into the SAR of these analogue series (Table 4-5). Roughly taken, residues between 450-520 belong to ‘site 1’ (Figure 4.10) and residues between 521-531 belong to ‘site 2’. The SSB C-terminal peptide perturbs almost all backbone amides close to these sites, but not the N512 side chain. The two initial primase hits, MB25 from the functional screen (chapter 2) and MB343 from the NMR screen (see section 4.2.1), perturb most of the non-hydrophobic, solvent-accessible residues of these two sites but not the hydrophobic,  $\alpha$ -helical residues I455 and L519. As mentioned earlier, a notable difference between the two hits is that MB343 does not perturb R452, suggesting that it does not interact with the guanidinium group. Overall, it appears that interaction with residues G481, L482, and D524 are prerequisites for binding. These residues also consistently rank among the highest shift distances, suggesting that these distances

could be used to estimate binding strength (although this suggestion should be treated with caution as it is not possible to rationally predict shift sizes and directions between different compounds). Among the MB25 analogues, BL6 and BL16 perturb the highest number of residues, indicating that they benefit from their aliphatic character. This observation could be useful in further optimisation of BL27. Furthermore, the N512 side chain, amide opposite R452 in site 1, only shows shifts in the presence of BL6, BL16, BL13, and the three MB343 analogues, suggesting that molecular weight or physical size could also play a role. Finally, BL26 and BL27 were the only analogues that induced CSPs of the hydrophobic core residues I455 and L519. Given that both these backbones are part of  $\alpha$ -helices, it should follow that when their chemical environment is changed, that of other residues in the helix could also be changed. Further inspection of neighbouring residues of I455 and L519 revealed that their backbone amides are indeed perturbed more than with other analogues, but not far enough to become significantly above average. BL27 is the analogue that causes the most CSPs and is also the only analogue with noticeably improved affinity and activity. This analogue together with the partially aliphatic MB25 and analogues is also able to interact with the N512 sidechain, the only amide that is not perturbed by the SSB C-terminal peptide. Therefore, N512 provides a possible way of differentiation or selectivity.

Table 4-5. Comparison of HBD-binding compounds and their induced CSPs.  $^1\text{H}$ - $^{15}\text{N}$  HMQC spectra of the HBD in the presence of compound (0.5-1.0 mM) were compared with the presence of only DMSO, and peaks that shifted by more than the average distance are marked as perturbed. N.D.: not determined.  $K_D$  value of SSB Ct was taken from (Naue et al., 2013).

Fragment	Parent hit	$K_D$ (mM)	M451	R452	I455	G481	L482	N512 $\delta$	T515	L519	W522 $\epsilon$	D524	I530
SSB Ct		0.01	X	X	X	X	X		X	X	X	X	X
MB25		2.6 $\pm$ 0.8	X	X		X	X	X	X		X	X	X
MB343		N.D.				X	X					X	X
BL6	MB25	N.D.	X	X		X	X	X	X		X	X	X
BL7	MB25	N.D.		X		X	X		X		X	X	
BL8	MB25	4.8 $\pm$ 2.6	X			X	X					X	
BL13	MB25	2.4 $\pm$ 0.6		X		X	X	X				X	
BL14	MB25	N.D.		X		X	X					X	
BL15	MB25	N.D.	X			X	X					X	
BL16	MB25	2.2 $\pm$ 0.8	X	X		X	X	X	X		X	X	X
BL22	MB343	2.2 $\pm$ 0.3				X	X	X	X		X	X	X
BL23	MB343	N.D.				X	X	X					
BL26	MB25	N.D.	X	X	X	X	X	X	X	X		X	
BL27	MB343	1.4 mM $\pm$ 0.3	X	X		X	X	X	X	X	X	X	X



#### 4.4.3.5 Functional characterisation of BL27

BL27 was titrated against the replisome to determine an  $IC_{50}$  of 0.2 mM (Figure 4.19). This value is significantly different from its HBD binding constant (1.4 mM), suggesting that BL27 may additionally inhibit another part of the replisome. Since the HBD binding site binds the SSB C-terminal peptide, and replisome subunit  $\chi$  also binds this peptide, it is possible that BL27 also binds to the  $\chi$  peptide-binding site, which would inhibit sliding clamp loading. However, BL27 also inhibits the primase-helicase interaction within the replisome, because it has the same concentration-dependent effect on lagging strand sizes as MB25 (Figure 2.8).

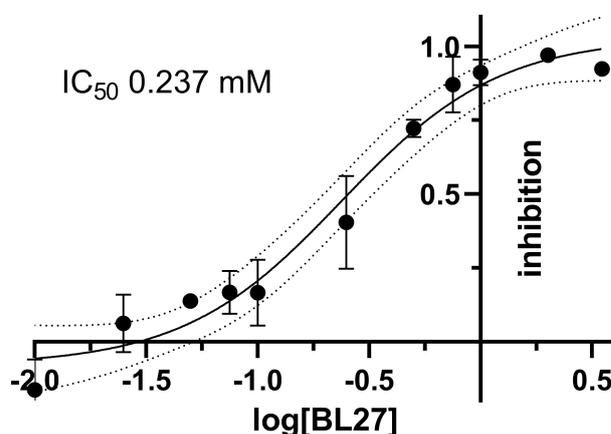


Figure 4.19. Titration of BL27 against the whole replisome. Interpolated sigmoidal curve (solid line) and its 95% confidence intervals (dotted lines) are shown.

Furthermore, while MB25, MB343, and BL22 were too weak to allow accurate  $IC_{50}$  calculation when titrated against the HBD in the primase-helicase primer synthesis assay (Figure 4.13), BL27 showed higher inhibitory activity with an average  $IC_{50}$  of 1.2 mM (Figure 4.20). Most interestingly, BL27 seems to inhibit synthesis of shorter primers (lower band on gel) more effectively, and while these products are reduced to almost zero at the highest BL27 concentration, the amount of full-length products (upper band) never falls below that observed in the absence of helicase. This indicates that primer synthesis is not directly inhibited at the active site, but rather that BL27 counteracts the presence of helicase. The  $IC_{50}$  of BL27 for the lower band products is around 1.0 mM, in between the BL27 HBD affinity (1.4 mM) and whole-replisome



## 4.5 Conclusion

The work presented in this chapter focused on primase as the target of a standard biophysical FBLD project, because primase was previously identified as the target of a whole-replisome fragment hit. An NMR fragment screen against primase alone yielded a reasonable 6% hit rate, with most hits binding to the enzymatic domain of primase. Although the primase active site was the obvious target site, crystallographic efforts did not reveal any active site binders, possibly due to a crystal packing unfavourable for compound soaking. Focus was shifted towards the primase HBD, as it had previously been indicated as the domain binding MB25. The NMR screening hit MB343 showed clear preference for the HBD in LO-NMR binding experiments. SAR was explored for these two hits, MB25 and MB343. In the absence of reproducible crystals, information on fragment binding modes was obtained at the whole-residue level by PO-NMR using residue assignments that were previously obtained (see section 3.3) (Naue *et al.*, 2013). Two important determinants of binding were identified from the first set of MB25 and MB343 analogues: presence of a bulky hydrophobic part (with slight preference for aliphatic) and of an adjacent negative charge. The best analogue, combining elements of the two original hits, showed improved binding and inhibitory activity against the whole replisome and the primase HBD. These SAR investigations mark BL27 as a promising HBD ligand for further optimisation, possibly by incorporating aliphatic elements.

## 5 Covalent modification of the replisome for target identification

### Preface

The author of this manuscript gratefully acknowledges the contributions of Dr. James D. Firth to results presented in this chapter, *i.e.* the synthesis of diazirine and other derivatives of various replisome fragment inhibitors.

### 5.1 Introduction

Fragment-based lead discovery (FBLD) projects often start with an identified target, and there is considerable knowledge and experience with using fragments in biophysical and purified-enzymes assays. The previous chapters described how, after screening the whole replisome for inhibitors, the binding site of hit MB25 was identified through many different experiments involving functional and binding assays and computational simulations. These types of experiments give detailed information about the system investigated and are most often used for single protein targets. However, as described in chapter 1, the replisome is not a single target even though it acts as a single system and consisted here of purified components. Some of the identified fragment inhibitors did not appear to inhibit any of the tested sub-components of the replisome. This chapter describes preliminary research to use chemically reactive fragments to identify the target protein or protein complex to which fragment hits bind. This area of research, where complex biological systems are probed with molecular precision, often using chemical compounds, is called chemical biology.

Chemical probes are molecules that have been designed to bind to a protein, serving as a tool to answer questions about the identity and function of the protein in its native environment (Arrowsmith *et al.*, 2015; Gerry & Schreiber, 2018). Detection of probe-protein complexes is most often done by modifying the probe with groups that enable detection or further work-up procedures, *e.g.* fluorescence microscopy (if the probe can be modified with a fluorescent group), mass spectrometry (MS), or affinity purification (Schenone *et al.*, 2013). For a long time, the chemical complexity of the cell has been seen as requiring high affinity and specificity of effective chemical probes (Bunnage *et al.*, 2013). Fragments were expected to meet neither of these criteria, and

while it may be possible to optimise fragment affinity, this would be pointless when the target list contains hundreds of equally ranked proteins. There have been no published examples of the use of fragments for whole-cell screening until recent years (Parker *et al.*, 2017; Ayotte *et al.*, 2018). In fact, while fragment-protein affinity limits are easily derived from the maximum number of favourable interactions that a fragment-sized molecule can make, the rules for fragment specificity are less clear. These recent studies showed that fragments can indeed be specific towards one phenotype or one or few protein targets, exerting biological effects by modifying the activity of their targets. The approach by Parker *et al.* (Parker *et al.*, 2017; Galmozzi *et al.*, 2018; Wang *et al.*, 2019) uses small sets of fragments that have been modified with photoreactive groups to enable direct target identification by MS-based proteomics. Target identification by MS is based on the observation of unique signals of protein chains that have been covalently modified by reactive compounds. In case of the replisome, such an approach could also be of interest, even in a retrospective way, since the current data on most replisome hits did not reveal their targets (see section 2.1). This chapter will present the results of two approaches for MS-based detection of covalent replisome inhibitors (Figure 5.1). First, several previous replisome hits were modified with a UV-sensitive diazirine group in the hope that they would covalently attach to their binding site during the replication reaction upon irradiation. Second, a small set of thiol-targeting electrophiles was screened against the replisome to find out whether there are native sites within the replisome that can be targeted with covalent inhibitors.

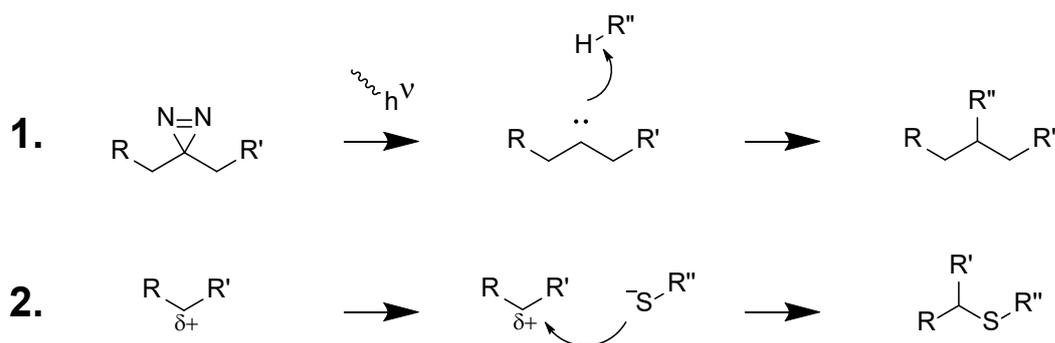


Figure 5.1. Two approaches for covalently modifying proteins with small molecules. The first uses the photoreactive diazirine group to generate carbene species which will insert into any nearby R''H bond (most often CH or OH). The second approach uses electrophiles of varying strength to react with surface-exposed cysteines. In both approaches the R group confers binding site specificity.

## 5.2 Photoreactive replisome hits

Reactive groups can lead to non-specific labelling of proteins if their reactivity cannot be controlled. Diazirine is a chemical group with inducible photoreactivity which can provide a clean way of crosslinking with few drawbacks. In case of fragments, the two main concerns are the relative size of the diazirine-containing group which could significantly alter the affinity or specificity of the whole molecule, and the low original affinity of the fragment which leads to low levels of crosslinking. With sufficiently high-resolution MS methods, crosslinking levels down to a few per cents can still result in good data (Grant *et al.*, 2019). The influence of the diazirine-containing group on fragment binding is not known beforehand. In case of screening this is not of serious concern because it is still probable to find at least one hit that is not influenced. In the present case of known hits to be modified, however, the influence of the diazirine modification on fragment activity should be investigated.

### 5.2.1 Design of diazirine derivatives of replisome fragment inhibitors

Two well-characterised fragment inhibitors, hit MB25 and the literature control fragment used during screening (see section 2.1) (Yin *et al.*, 2014) (hereafter F5), were chosen to assess the influence of diazirine modification on fragment binding and activity. The target of each fragment was known (primase for MB25, sliding clamp for the control fragment), as well as a binding mode allowing the prediction of a position where modification with the diazirine group would result in minimal interference (Figure 5.2). In case of F5, a published crystal structure of F5 in complex with its target (the  $\beta_2$  sliding clamp) (Yin *et al.*, 2014) showed that the two rings are already twisted with respect to each other, therefore any of the four *meta*-positions were good candidates for modification. The *ortho*-position of the benzoate is most directly pointing towards the solvent, but it was feared that modification there could influence H bonding by twisting the carboxylate. Therefore, the 2'-position on the *p*-fluorophenyl ring (*meta* relative to the fluoride) was chosen for modification (Figure 5.2, compound D1). In case of MB25, the optimal attachment position was difficult to predict because previous structure-activity relationships experiments (see section 4.4.3.1) suggested that the thiophene does not have any positions that can be modified without impeding binding. Furthermore, multiple binding modes were sampled during simulations, which did not reveal any of the aliphatic carbons as being more solvent-exposed than the others. To increase chances of at least one productive binding mode,

the conformationally flexible 6-position was chosen for modification, and a synthetic route that would result in a racemic mixture (Figure 5.2, compound D2). The position of the sulfur was changed for greater synthetic accessibility, a change that was previously determined to have little impact on compound activity (compare MB25 and BL6, Table 4-4).

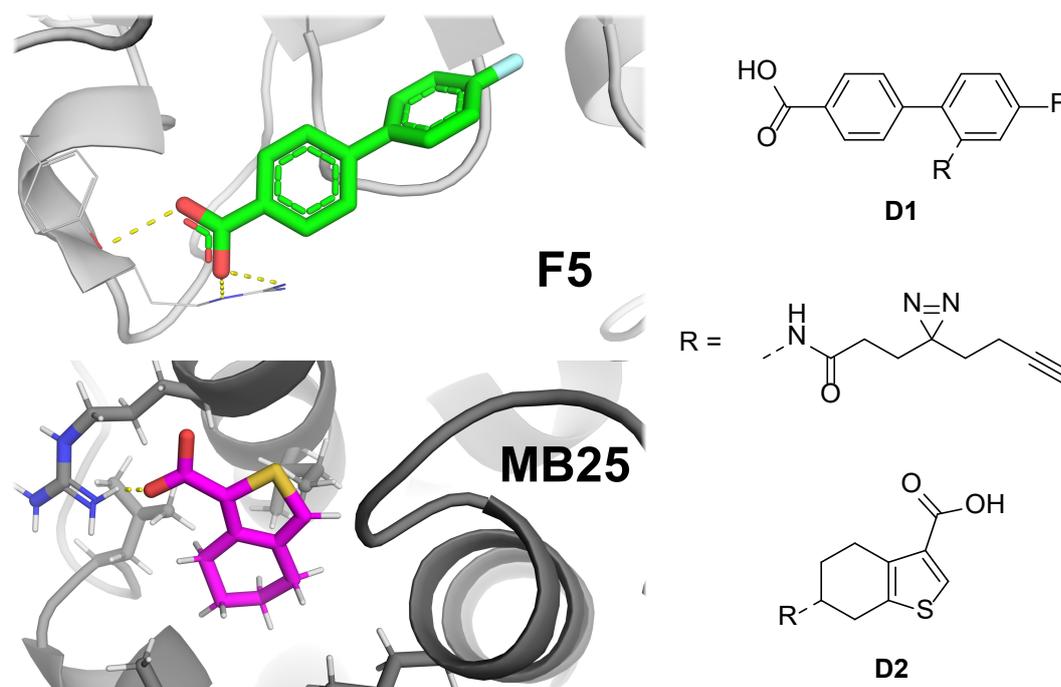


Figure 5.2. Proposed modification positions based on binding modes from co-crystal structure (F5) (Yin *et al.*, 2014) or simulations (MB25) (see section 3.2). For each original fragment F5 and MB25, an example binding mode structure is shown. The binding modes suggest possible vectors for modification with the shown diazirine-containing R group which would least impede fragment binding. Further discussed in the accompanying text.

The modification consisted of a diazirine group flanked by two ethylenes, a terminal alkyne that could serve as a handle for labelling for fluorescence or affinity purification experiments via click chemistry, and a carboxylic acid for nucleophilic acyl substitution by the aminated ligand. The resulting R-group is identical to the one used by Parker *et al.* (Parker *et al.*, 2017). Two control compounds were also synthesised to test the influence of various parts of the diazirine-modified fragments: F5 modified with *n*-hexanamide without diazirine and alkyne groups (D3), and an R group where the fragment part is a single methyl group (D4).

### 5.2.2 Testing of diazirine derivatives

First, the modified fragments and various control compounds were tested for replisome inhibition under standard non-crosslinking conditions (absence of sunlight and other sources of UV light) in the original whole-replisome assay used throughout this work (see section 7.2). It was confirmed that the diazirine-alkyne moiety alone does not inhibit replication. Modification of F5 at the 2'-position did not significantly change its activity, showing that the choice of this position based on the crystallographically determined binding mode was correct. Modification of BL6 also resulted in similar inhibition. Although the 6-position of BL6 is not expected to point directly towards the solvent, and analogue D2 is a racemate, the observed activity suggests that any interactions between primase and the diazirine moiety are neither repulsive nor strong enough to significantly change the activity of the whole molecule. An NMR saturation transfer difference experiment also showed signs of binding of D2 to the primase helicase-binding domain.

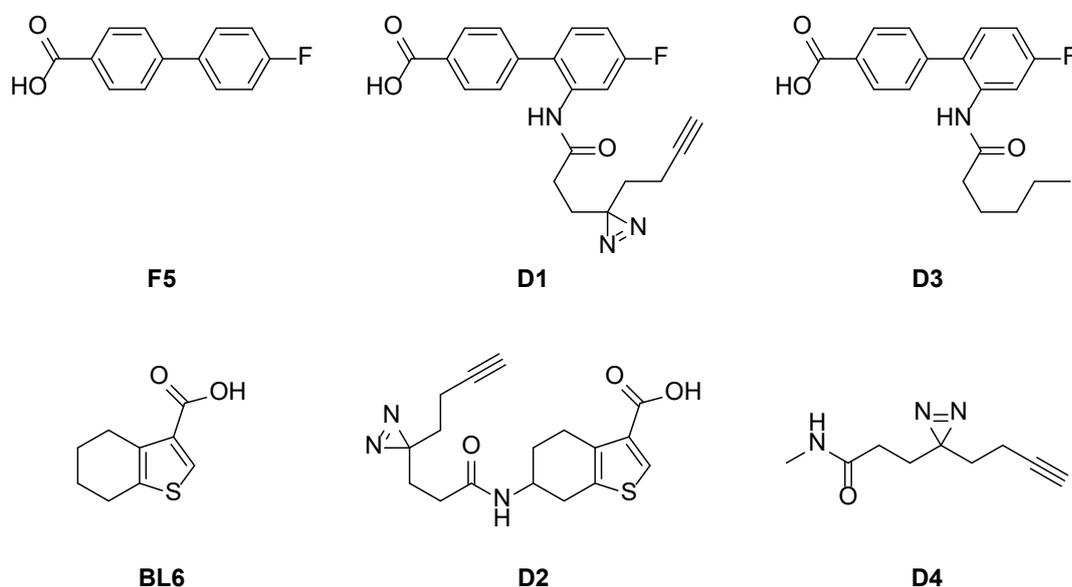
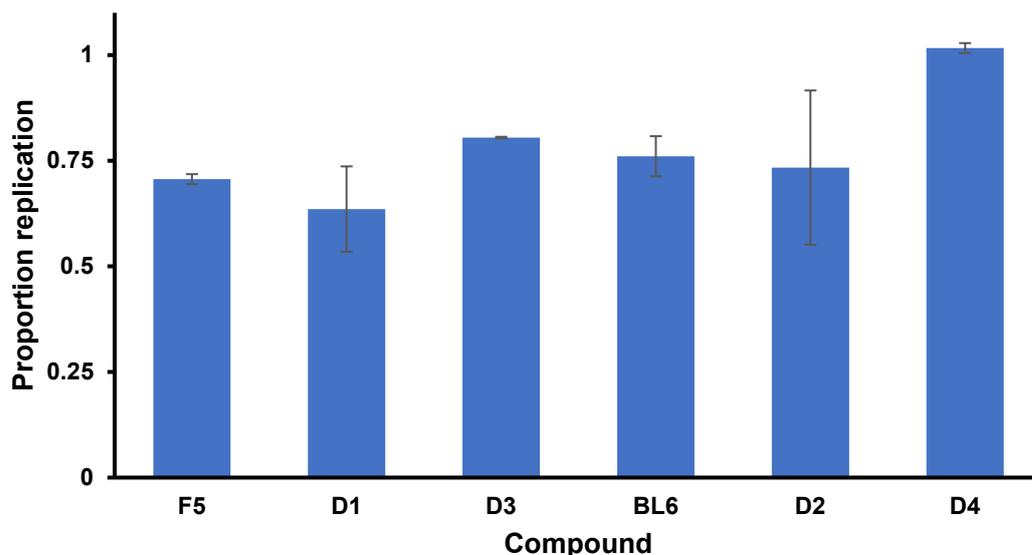


Figure 5.3. Fragment inhibitors and their modified analogues tested for replisome inhibition under standard non-crosslinking conditions. Compounds were tested at 0.5 mM except for BL6 and D2 which were tested at 1 mM. Diazirine derivatives showed similar inhibition to their original fragment parents, and the diazirine-alkyne group alone (D4) did not show any non-covalent inhibition of the replisome, suggesting that the modification does not change the binding and activity of the original fragment parts.

Next, the influence of UV irradiation on compounds and replisome components separately was investigated. At 0.5 mM in phosphate-buffered saline, the three diazirine analogues showed different conversion rates under UV irradiation, as measured by NMR peak areas of their ethylene protons (Figure 5.4). While D4 alone

needs approximately 30 minutes of irradiation before complete conversion, diazine groups attached to fragments (D1 and D2) seem to be more sensitive. The presence of a non-covalent fragment is sufficient to increase sensitivity (see D4 + F5). Possibly, F5 re-emits UV light at wavelengths that are closer to the diazine absorption maximum. The effect seems limited to F5, as D2 is slower to degrade than D1, and the presence of other aromatic molecules such as nucleotides that are present in assay buffer does not affect degradation rates.

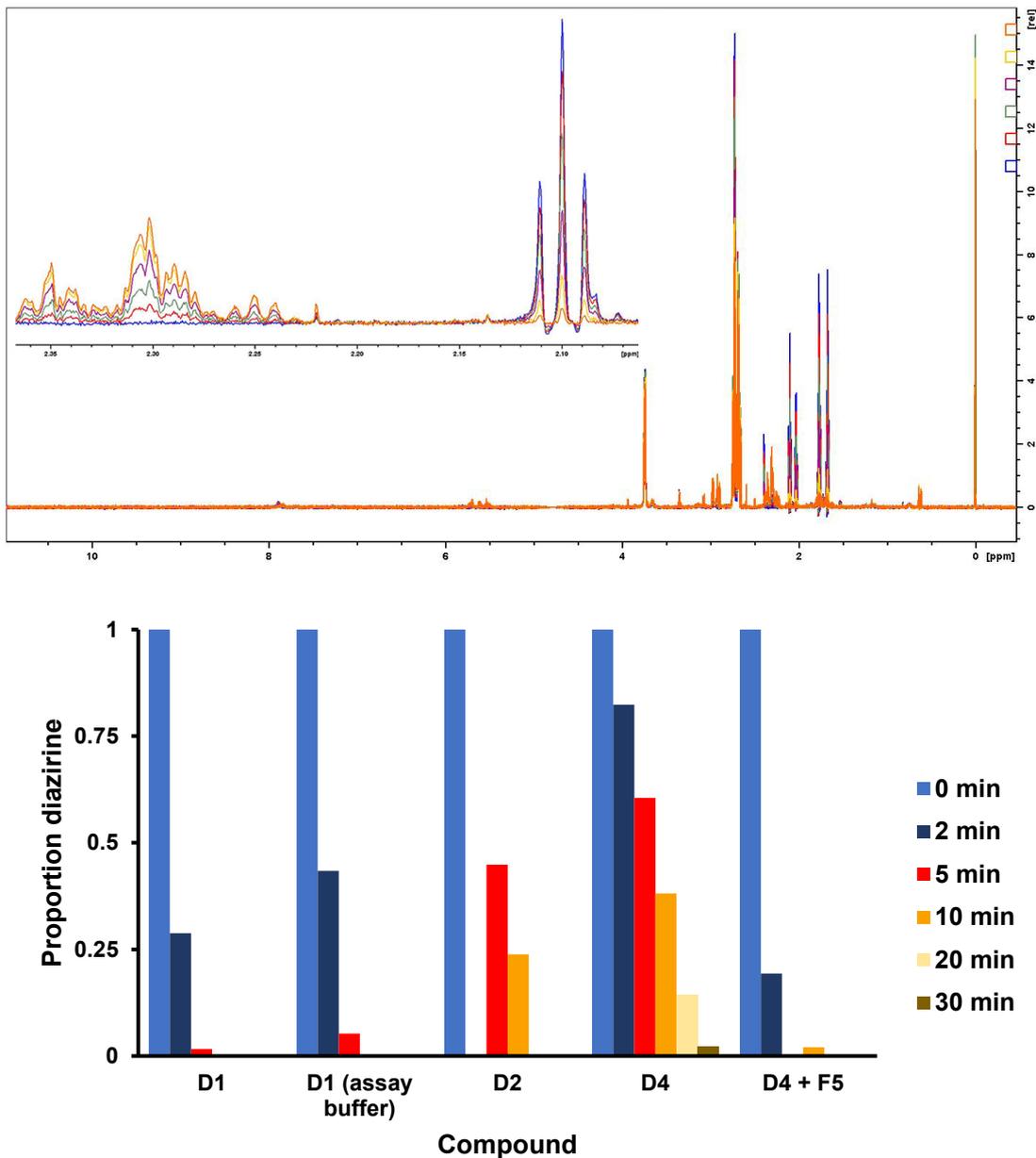


Figure 5.4. Conversion of diazirines under UV irradiation measured by NMR. An overlay of D4 spectra at time points between 0-30 minutes is shown, with an expanded region where ethylene peaks (2.10 ppm) disappear and new peaks (2.25-2.35 ppm) appear with increasing UV irradiation time (blue spectrum = 0 min, orange spectrum = 30 min). The chart shows fractions of remaining diazirine molecules after each time point relative to initial values at time point zero (backgrounds subtracted).

Irradiation of different assay components causes large decreases in subsequent replication activity (Figure 5.5). Not unexpectedly, irradiation of the circular dsDNA template almost completely inhibits subsequent replication reactions. Nucleotides and proteins are slightly more resistant to UV light, but still lose more than half of their

activities. The most important consequence of these observations is that it would be impossible to crosslink diazirines to active replisomes using the current approach, because replisomes would stall and disassemble before diazirines can be completely converted. Unfortunately, this meant that the diazirine crosslinking approach in this form was not suitable for identifying binding sites of hits that did not seem to inhibit any replisome subcomponent (*e.g.* MB7, Figure 2.5), as the expectation for such hits was that they might bind to sites that are only formed when the replisome is active. One possible solution to this problem would be to use high-power UV beams to decrease diazirine-to-carbene conversion times to just several seconds (Horne *et al.*, 2018). Such equipment was not available during the current project. Alternatively, the use of light sources with narrower wavelength profiles could limit damage to the replisome or to dsDNA template, although some forms of DNA are known to take damage from wavelengths up to 365 nm (Tyrell, 1973), and the exact absorption maximum for diazirine groups could be different for each derivative.

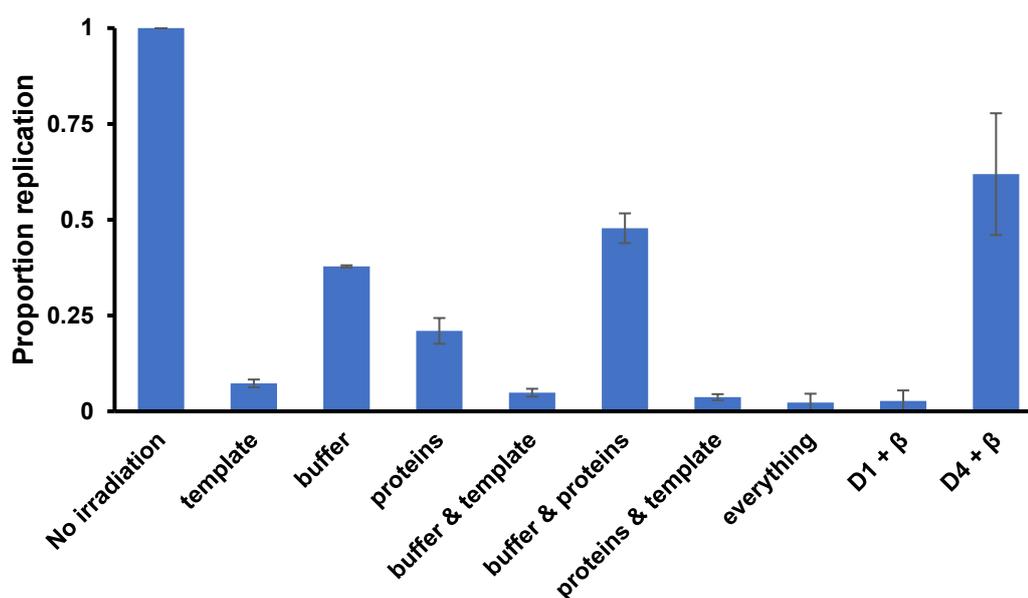


Figure 5.5. Effect of UV irradiation on different replication reaction components. Each component was irradiated for 5 minutes, then added to reaction mixtures as normal. UV pre-irradiation of each component severely impairs assay activity.

Since the targets of the control compounds were known, pre-irradiation of protein and diazirine together could indicate whether crosslinking at the inhibitory site occurs. Pre-incubation and -irradiation of D1 with the  $\beta$  sliding clamp indeed renders the whole replisome inactive (Figure 5.5), as would be expected when the sliding clamp is

inactivated. The effect cannot be explained by simple inactivation of the sliding clamp by UV light directly, because irradiation of D4 and  $\beta$  still permits subsequent replication.

Several attempts were made to directly observe a crosslinked D1 +  $\beta$  product. Unfortunately, the available protein quantities were insufficient for MS experiments. Attaching an azide derivative of a fluorescent dye to the crosslinked product via click reaction (Presolski *et al.*, 2011) showed low levels of fluorescence on polyacrylamide gel at the position expected for  $\beta$  (not shown), but also showed similar levels after crosslinking between D4 and  $\beta$  or between D1 and primase, suggesting that these signals are mainly due to non-specific crosslinking.

In conclusion, diazirine derivatives of two known fragment inhibitors were used to show that this type of crosslinking chemistry can be used without much impact on the functionality of the original fragments. Activity assays showed that crosslinking D1 to its target ( $\beta$ ) enhances inhibitory activity, suggesting that D1 was covalently attached to the protein-protein interaction pocket of  $\beta$ . However, attempts at direct observation of the crosslinked molecule failed. Furthermore, UV irradiation of the replisome using standard equipment was found to be unsuitable for *in situ* crosslinking experiments as neither DNA template nor replisome subunits could tolerate irradiation for the duration that was found necessary for complete conversion of diazirines. For *in situ* crosslinking of inhibitor diazirine derivatives to active replisomes followed by direct observation and identification of crosslinked products by MS, more bespoke equipment would be needed as well as higher protein quantities than were used previously.

### **5.3 Targeting replisome cysteines**

Cysteine thiol groups are often the target of protein covalent modification, as they are the most reactive (when solvent-accessible) and are rare enough that most have unique molecular environments. Thiol-reactive compounds can be developed in two different ways: by introducing an electrophilic group to a known non-covalently binding ligand, or by screening for reactive electrophiles and subsequently adding or modifying non-covalent selectivity groups (Hallenbeck *et al.*, 2017; Lonsdale & Ward, 2018; Erlanson *et al.*, 2004; Keeley *et al.*, 2020). Instead of exchange rate constants, reaction rate constants and specificity determine the largest part of the effectivity of a covalent

ligand. In a fragment-based approach to covalent ligands, the specificity element can be built up in much the same way as for non-covalent ligands, allowing the separate optimisation of covalent bond formation rate and target specificity (Miller *et al.*, 2013; Kathman *et al.*, 2014; Kathman & Statsyuk, 2016). Here, the replisome was probed for native cysteines that could support covalent inhibition of replication.

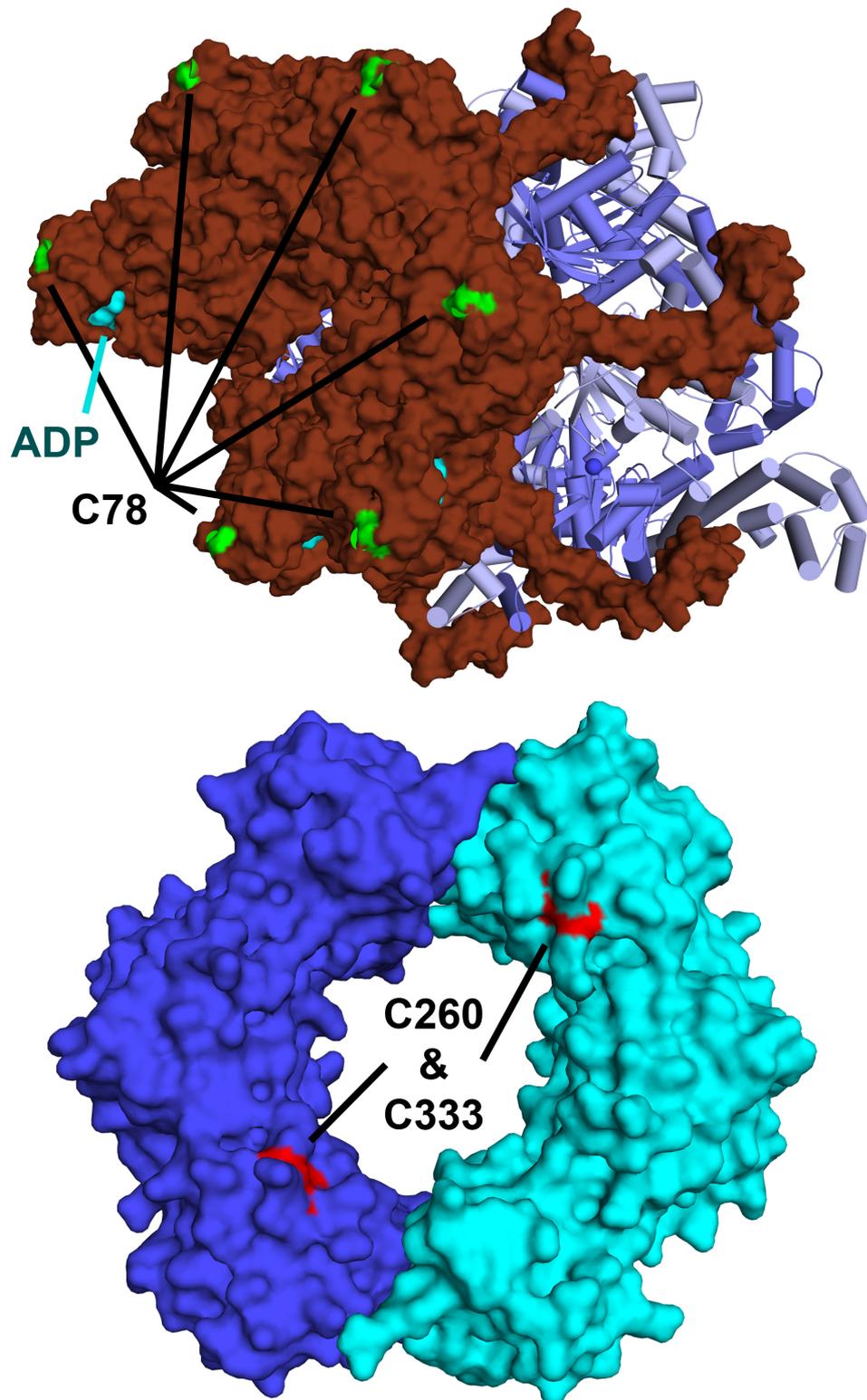
### **5.3.1 Feasibility of replisome inhibition by thiol-targeting fragments**

The replisome contains several solvent-accessible cysteines that could be targeted with electrophilic fragments. With crystal structures of most replisome subunits publicly available, a detailed analysis of replisome cysteines was made using the *Cpipe* web server (Soylu & Marino, 2017). This pipeline calculates solvent-accessible surface areas of cysteines and attempts to make predictions about the functions of reactive cysteines. Five cysteines belonging to four replisome proteins were predicted to be at least somewhat reactive (Table 5-1). Eight other cysteines were highlighted as well, but as these belong to zinc-binding folds they were ignored. A few minor parts of the replisome were not present among the analysed structures: the  $\alpha$  oligonucleotide-binding domain containing one cysteine, the primase zinc-binding domain containing four cysteines (three of which coordinate zinc), the  $\tau$  helicase-binding and polymerase-binding domains containing one cysteine, and the SSB carboxyterminal disordered peptide containing no cysteines.

Table 5-1. Cpipe analysis of published replisome subunit structures. SASA: solvent-accessible surface area; RPD: RNA polymerase domain; HBD: helicase-binding domain.

Protein	PDB ID	Predicted reactive Cys	SASA S $\gamma$ (A <sup>2</sup> )	SASA Cys (A <sup>2</sup> )	Note
helicase	6qem	-			No cysteines
helicase loader	6qem	C78	7.5	36.7	
$\alpha$	2hnh	C154	6.6	12.9	
Primase RPD	1dde	None			
Primase HBD	6cbr	None			
$\beta$	1mmi	C260	6.7	14.8	
		C333	4.6	25.6	
$\epsilon$	1j53	None			
$\tau$	3glf	C64	6.0	14.7	Zinc finger
		C73	8.5	43.8	
		C76	4.4	17.0	
		C79	4.3	10.4	
$\delta$	3glf	None			
$\delta'$	3glf	C50	3.9	17.1	Zinc finger
		C59	6.6	37.0	
		C62	4.0	16.3	
		C65	5.48	11.46	
$\chi$	3sxu	None			
$\psi$	3sxu	C85	5.27	13.9	
$\theta$	2ae9	-			No cysteines
SSB	4mz9	-			No cysteines

Among the five cysteines predicted to be potentially reactive, C78 of the helicase loader is the most interesting one because it is located close (approx. 9 Å) to the ATP-binding site of this enzyme (Arias-Palomo *et al.*, 2019) (Figure 5.6). Two cysteines of the sliding clamp are also of interest because both are located on the inside of the dsDNA-binding ring, possibly giving the opportunity to sterically hinder clamp loading. The two remaining cysteines predicted to be reactive, one belonging to  $\alpha$  and one to  $\psi$ , are of less interest because they do not appear to be of structural or catalytic importance and are less solvent-accessible than the first three cysteines mentioned. These predictions were sufficiently encouraging to conduct an initial screen of electrophilic fragments against the replisome.



*Figure 5.6. Structures of the E. coli helicase loader (brown surface) in complex with the helicase (blue cartoon) (Arias-Palomo et al., 2019) and of the E. coli sliding clamp homodimer (Oakley et al., 2003) with putative reactive cysteines highlighted.*

To enable screening of thiol-reactive electrophilic fragments, the replication reaction conditions needed to be changed to remove any excess reactive thiols. The replisome activity assay that was employed in previous chapters contains large amounts of dithiothreitol (DTT) (10 mM) to maintain the replisome in its native reduced state, and bovine serum albumin (BSA) (1.5  $\mu$ M) to promote general stability of dilute proteins *in vitro*. Omission of BSA did not significantly affect replisome activity (not shown). Omission of DTT reduced the activity of the replisome by approximately 40% (Figure 5.7). Replisome subunits were stocked under reducing conditions, which could explain the remaining 60% activity. The combined concentration of DTT from replisome subunit stocks in the final reaction mixture is 240  $\mu$ M, at unknown stages of oxidation. As the concentration of fragments during screening was set to 500  $\mu$ M, these conditions were an acceptable compromise between replisome activity and free thiol interference.

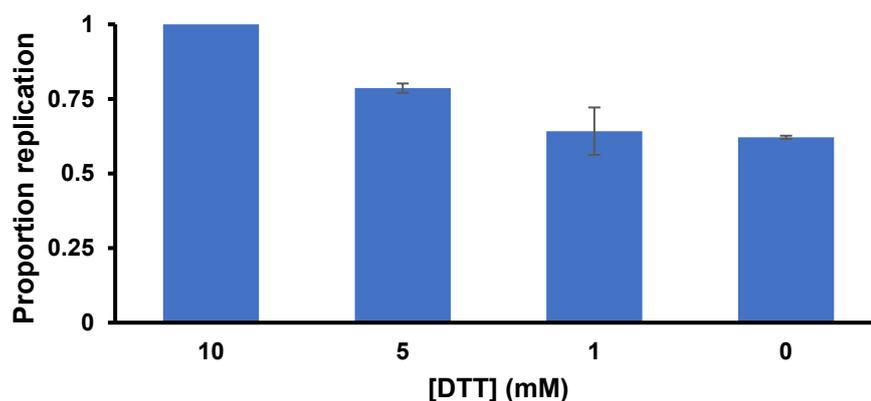


Figure 5.7. Effect of varying concentrations of DTT on replisome activity. A strong reducing environment appears to be important to replisome activity.

### 5.3.2 Screening of a small electrophilic library

A set of 28 electrophilic fragments covering a broad range of electrophilic reactivities (Ábrányi-Balogh *et al.*, 2018) was used to find suitable warheads for replication inhibition. The electrophilic groups were attached to the 3,5-bis(trifluoromethyl)-phenyl scaffold, covered different types of mechanisms (*e.g.* Michael addition, nucleophilic substitution, addition-elimination), and ranged from very reactive to quite

stable as measured by incubation with glutathione. The set also contained two non-reactive control compounds. Structures of all screened electrophiles can be found in Appendix B. To increase chances of covalent bond formation, each compound was incubated with the inactive replisome at room temperature for 30 minutes before starting replication at 37 °C for 13 minutes. The two non-reactive control compounds (ML17B and ML28) were found to weakly inhibit replication, and three electrophiles (ML21, ML25, ML30) showed higher inhibition than these control compounds (Figure 5.8).

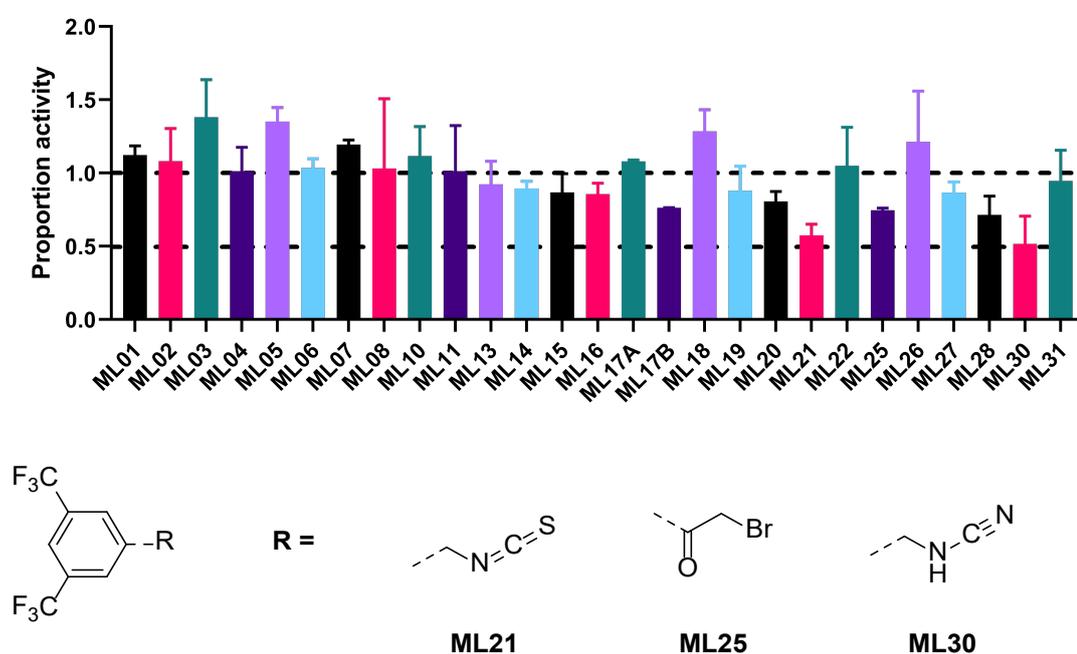
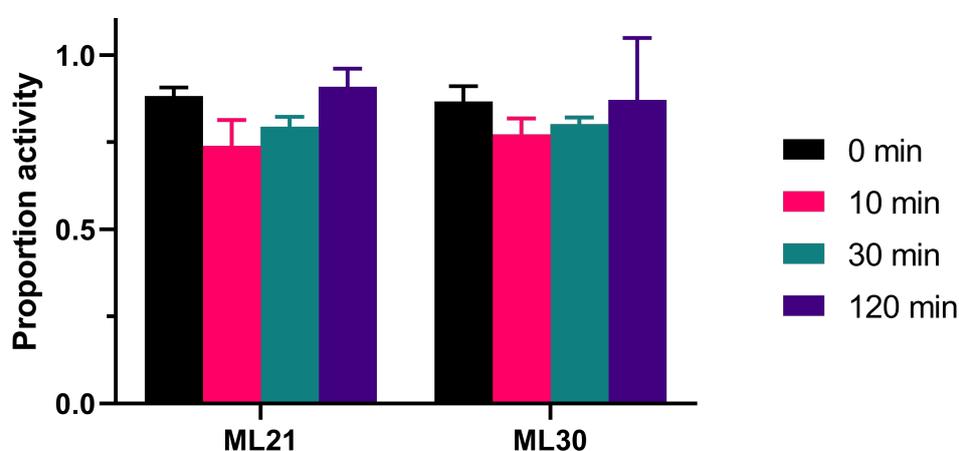


Figure 5.8. Screening results for the electrophilic fragments set against the replisome and structures of the top three hits.

ML25 is one of the most reactive electrophiles, even showing partial degradation in aqueous phosphate buffer during NMR analysis, and thus would be expected to show up as a hit in most screens. However, another known thiol-reactive fragment, the maleimide ML20, did not show increased inhibition over the control compounds. ML21 and ML30 have similar structures and mild reactivities. Isothiocyanates in natural products are associated with various health effects through covalent protein modification mechanisms, but cyanamides are uncommon in bioactive compounds.

The finding that both fragments were hits possibly suggests a specific action on the replisome.

Since ML21 and ML30 were expected to have mild reactivities and slow reaction rates, their inhibitory activities were measured after various incubation times prior to starting the replication reaction. If these hits acted via a covalent mechanism, their activities would be expected to correlate with incubation time. However, a clear correlation was not observed (Figure 5.9), except for a slight increase in inhibition between 0 and 10 minutes of incubation.



*Figure 5.9. Replisome inhibition by electrophilic fragment hits after various incubation times. Inhibition does not appear to be time-dependent, suggesting either a fast reaction or a non-covalent mechanism of inhibition.*

To directly confirm whether ML21 or ML30 covalently bound to one of the cysteines of interest (the helicase loader or the sliding clamp), various attempts were made to observe the expected covalent reaction products by MS. No mass additions were observed for either protein after incubation under conditions similar to those used during activity assays. Several parameters were then changed to make reaction conditions more favourable and observation of covalent modification more likely, although the range of conditions was limited by moderate protein stability and solubility. Increased electrophile concentrations (0.5-5.0 mM), increased reaction times (1-21 h) and temperatures (22-37 °C), and various buffers (HEPES, Tris, phosphate) were tested, but no covalent modifications of the helicase loader or the sliding clamp were observed. These results suggest that ML21 and ML30 may inhibit

replication by a non-covalent mechanism of action, a hypothesis that is supported by the fact that these hits are structurally almost identical.

## 5.4 Conclusion

The experiments described in this chapter failed to produce any covalent protein-fragment complexes that were observable by MS. In case of the cysteine-targeting approach, the most likely explanation is that the replisome does not contain reactive cysteines that can transmit activity changes when covalently modified. Alternatively, it is possible that one or more cysteines were modified but the phenyl scaffold was not able to inhibit replication due to suboptimal structure or size. It would be interesting to screen covalent libraries that contain more variety of scaffolds, or to select a known binder (such as ATP in case of the helicase loader) and grow it towards the closest cysteine. However, these alternative approaches would fall outside the purpose of this project to identify replisome-binding fragments and their targets *ab initio*.

In case of the photoaffinity approach, several results indicated that this approach could be viable in principle. Modification of known fragment binders with a diazirine-containing group did not interfere with their activity, and the putative crosslinking of one fragment to its protein target resulted in increased inhibitory activity. However, the effect of UV light on active replisomes prohibited *in situ* crosslinking. To directly crosslink a fragment inhibitor to an unknown target during replication would require considerable optimisation of reaction conditions to avoid replisome breakdown. This could be achieved by decreasing diazirine-to-carbene conversion times from several minutes to several seconds. Additionally, it would be interesting to test whether there are specific wavelengths that can activate diazirines but do not damage dsDNA and folded proteins.



## 6 Conclusions

This project was based around the first results of a fragment screen for inhibitors of the *E. coli* replisome. The idea of screening against an ensemble of targets was novel, so initially the range of expected possible outcomes included complete inhibition by all fragments due to the high concentration of compound used, as well as no effects by any fragment due to the size and complexity of the replisome. The screening results were exciting because they appeared to be similar to those of most fragment-based campaigns which target single proteins: an initial hit rate of 10% and chemical diversity among hits with some structural motifs showing signs of enrichment.

The number of hits decreased slightly after compound quality control, notably among the hits that were most active. In hindsight, it was surprising to find so many fragments that more than halved the activity of the replisome (39 fragments, of which 18 later appeared to be DNA intercalators). This emphasises that more attention should have been paid to compound behaviour following the initial screen. This would have shifted attention to the eighty fragments that showed moderate inhibition and additional interesting compounds may have been confirmed for further study. However, the fragments hits that were confirmed did show a range of activities against different components of the replisome (these results could be included in a small publication on the *in vitro* screening of fragments against a large molecular machine) and the work in this thesis focused on these.

For one confirmed hit, MB25, target identification was achieved through a combination of functional and biophysical experiments with molecular dynamics simulations. The first hints that MB25 might inhibit primase function came from the whole-replisome activity assay itself, as this fragment increased lagging strand sizes. NMR binding experiments with primase suggested that MB25 did not bind to the active site of primase, which would have been the simplest explanation. Since primase is a flexible protein with multiple separate domains, binding mode identification by crystallography (standard practice in FBLD) was not straightforward. Instead, an attempt to identify the site of action was made by conducting molecular dynamics simulations with the two non-enzymatic domains of primase. These simulations suggested specific binding to a conserved peptide-binding site on the helicase-binding domain, and these results were subsequently confirmed by protein NMR.

Simulation data also proved useful during hit optimisation because crystallisation of the MB25 target domain failed. Protein peak assignments, kindly provided by others (Naue *et al.*, 2013), enabled preliminary structure-based hit optimisation at the residue level. Structure-activity relationships were monitored by both NMR binding data and enzymatic activity data for both DNA replication and RNA primer synthesis. Because the accessible chemical space was limited by commercial availability of hit analogues, a more ‘traditional’ single-target screen with primase was conducted. This yielded a novel scaffold (MB343) that was shown to interact with the MB25 binding site. MB343 itself did not inhibit the replisome, but combination of elements from MB25 and MB343 yielded a compound with improved binding as well as inhibitory activity. This result shows the benefits of collecting both binding and activity data. Fragments may bind at or close to a functionally interesting site on a protein but will not show functional effects until the fragment is optimised. Finally, during the hit optimisation stage a minor but interesting new piece of information from a biochemistry perspective was discovered: that binding to the SSB-binding site of primase also influences primase-helicase binding. The prediction of the MB25 binding site and subsequent mechanistic and fragment optimisation experiments could be published as an example of successful fragment binding site identification and characterisation in the absence of atomic-resolution (crystallographic) data.

Attempts were made to develop a chemical strategy that could identify the exact target of a replisome inhibitor based on recent new ideas in chemical biology. Compounds were synthesised that combine a fragment with a chemical warhead that can be activated to covalently attach to a protein if the fragment binds, using mass spectrometry to identify and characterise the resulting protein-compound conjugate. Pilot work demonstrated proof of principle with an isolated protein from the replisome and activated compounds derived from fragments that are known to bind to it. However, further work is needed to adapt the methods (and conditions) as the activity of the whole replisome was lost in exploratory and control experiments.

In conclusion, this thesis reports successful first steps of using fragments to functionally interrogate a complex molecular machine, contributing a new perspective to recent developments in phenotypic fragment screening.

## 7 Materials and methods

### 7.1 Primase production

Wild-type full-length *E. coli* primase-encoding pET3c based plasmid was received from the McGlynn group (University of York). The gene was subcloned into the pET-based pYSBL3C vector containing a hexahistidine tag followed by the 3C protease recognition site upstream from the insertion site. Linear *dnaG* and pYSBL3C were prepared by PCR amplification using primers 5'-gttctgttcaggaccagcaATGGCTGGA CGAATCC and 5'-catatgtgaggagaaggcgcgtaCTTTTTCGCCAGCTC for *dnaG*, and 5'-GGAAGTTCTGTTCCAGGGACCAGCAA and 5'-CGCGCCTTCTCCTCACAT ATGGCTAGC for pYSBL3C. The two linear fragments were then combined using the HiFi DNA Assembly Cloning kit (NEB) per the manufacturer protocol, using a plasmid to insert ratio of 1:2 and ~40 fmol of plasmid. BL21 competent cells were transformed with these plasmids and grown on LB agar containing kanamycin (35  $\mu\text{g mL}^{-1}$ ) to select for successfully transformed cells. Presence of the insert was checked by colony PCR, and correctness of the inserts was confirmed by Sanger sequencing. Plasmid for primase RNA polymerase domain (RPD) (111-433), zinc-binding domain (ZBD) (1-102), and helicase-binding domain (HBD) (433-581) were prepared in the same way by modifying the primers to anneal to the wanted regions of *dnaG*.

For unlabelled protein, constructs were expressed in *E. coli* BL21(DE3) (Agilent) or BL21(DE3) pLysS (HBD), grown in Luria broth (1 L in 2.5 L baffled flasks) at 37 °C to an optical density of 0.6, by addition of 1-isopropylthiogalactopyranoside to 0.2 mM. After 2-4 h at 37 °C or overnight at 18 °C, cells were harvested and stored at -80 °C until further use.

For isotope-labelled protein, constructs were expressed in minimal media (M9 salts containing  $^{15}\text{NH}_4\text{Cl}$ , supplemented with  $\text{MgSO}_4$ ,  $^{12}\text{C}$ - or  $^{13}\text{C}$ -glucose, essential trace elements, and Gibco vitamins mix) to obtain uniformly  $^{15}\text{N}$ -labelled or  $^{15}\text{N}$ - $^{13}\text{C}$ -labelled protein.

Purification of primase was performed in three steps. First, resuspended cells with added protease inhibitors (Roche) were lysed by sonication or by passage through a French press, the cleared lysate was applied to a HisTrap column (GE) pre-equilibrated

with buffer A (Tris pH 8 50 mM, NaCl 300 mM, 2-mercaptoethanol 5 mM, imidazole 10 mM), and the bound primase eluted by a gradient up to 500 mM of imidazole in buffer A. Second, the affinity tag was removed by incubation with HRV 3C protease (50:1 molar ratio) at room temperature for 3 h and subsequent reapplication to a HisTrap column. Third, the flow-through from the previous step was concentrated to 1% of the column volume of a Superdex 75 (for ZBD and HBD) or 200 (for RPD and full-length) size exclusion column, then applied to it and eluted with buffer B (HEPES pH 7.5 20 mM, NaCl 50 mM, DTT 1 mM, EDTA 1 mM). Peak fractions of the final product were combined, concentrated, and stored at 4 °C or -20 °C. Full-length primase used for NMR screening and confirmation experiments still contained its N-terminal tags. The primase construct with attached tags had activity comparable to wild-type primase in the whole-replisome assay.

## **7.2 DNA replication assay**

### **7.2.1 Template preparation**

The template for this replication reaction was bacteriophage M13 in replicative form (RF) II, which facilitates a rolling circle mechanism of replication. It was prepared by annealing a partially overlapping oligonucleotide (5'-(dT)<sub>36</sub>(f1-complement)<sub>30</sub>) to single-stranded M13mp18, then generating M13 RFII by complete extension of the primer by T4 polymerase lacking its exonuclease domain. This produced a double-stranded plasmid with a single nicked site at the primer. T4 polymerase was permanently inactivated by heating to 70 °C for 15 minutes.

### **7.2.2 Replication reaction**

All replisome subunits were previously produced and purified by the McGlynn group and stored at >15x final assay concentrations in storage buffer (Tris pH 8 50 mM, NaCl 100 mM, EDTA 1 mM, DTT 1 mM, glycerol 50%) at -80 °C. The replication reaction mix was assembled on ice by mixing all replisome components, template, buffer, DTT, BSA, Mg<sup>2+</sup>, and all nucleotides except dATP and dTTP (Table 7-1). The experiment was started by addition of the reaction mix to the fragment (0.5 mM) or to DMSO (the final concentration of DMSO was 2.5% in all reactions). The reaction mix was incubated at 37 °C for 3 minutes for equilibration. Then, the reaction was started by the addition of dATP and dTTP to 40 μM. The reaction was allowed to proceed for 3 minutes in order to prevent the initial burst of unstable replisome complexes from

contributing to the signal. Then,  $\alpha$ - $^{32}\text{P}$ -dCTP (33 nM) was added to the reaction mix for subsequent detection of reaction products. After another 10 minutes, the reaction was terminated by addition of either EDTA to 45 mM or  $\text{NH}_4\text{OAc}$  to 2.3 M, depending on the method of detection (Figure 7.1).

Table 7-1. Components of the whole-replisome DNA replication assay.

Component	Final concentration (monomers)	
Pol III core ( $\alpha\epsilon\theta$ )	50	nM
Clamp loader ( $\tau_3\delta\delta'\psi\chi$ )	25	nM
Helicase DnaB	160	nM
Helicase loader DnaC810	160	nM
SSB	1	$\mu\text{M}$
Primase DnaG	200	nM
Sliding clamp ( $\beta$ )	80	nM
BSA	0.1	$\text{mgmL}^{-1}$
M13 RFII	2	nM
HEPES pH 8 (4 °C)	40	mM
DTT	10	mM
$\text{Mg}(\text{OAc})_2$	10	mM
ATP	2	mM
GCUTP	0.2	mM
dCdGTP	40	$\mu\text{M}$
dAdTTP	40	$\mu\text{M}$
$\alpha$ - $^{32}\text{P}$ -dCTP	0.03	$\mu\text{M}$

### 7.2.3 Measurement of replication products

To measure  $^{32}\text{P}$  incorporation into all reaction products, reaction mix terminated with EDTA was precipitated by addition of ice-cold trichloroacetic acid (TCA) (5%). TCA will precipitate all icosamer and larger oligonucleotides. Precipitated reaction product suspensions were filtered through glass microfibre filters, which were then washed with ice-cold TCA (1%) and ethanol, and finally air-dried. Reaction products were quantified by measuring Cherenkov radiation using a liquid scintillation spectrometer (Packard). The workflow is represented schematically in Figure 7.1.

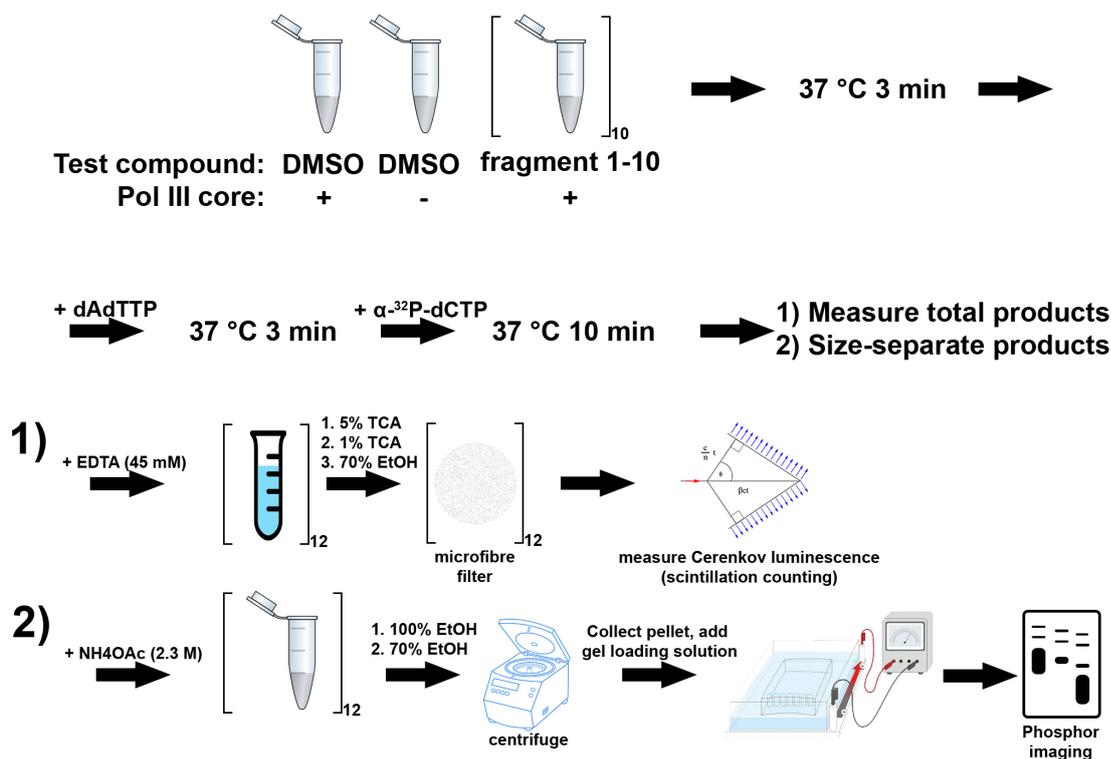


Figure 7.1. Schematic representation of the workflow used during screening of the fragment library in the whole-replisome assay and subsequent confirmation experiments. The DNA elongation reaction is explained in section 1.2 and Figure 2.1. From the start, all reaction tubes contained all components listed in Table 7-1 except those mentioned explicitly in this figure. For screening and confirmation, total products were quantified as in 1). For quantification and visualisation of separate leading and lagging strands, method 2) was used.

To separate leading strand products from lagging strand products, reaction mix terminated with NH<sub>4</sub>OAc was precipitated by adding ice-cold ethanol to 70%, followed by a wash (70% ethanol) of the pellet to remove most of the remaining free  $\alpha$ -<sup>32</sup>P-dCTP. The precipitated reaction products were run on alkaline (50 mM NaOH) agarose gel (0.7%) overnight, which was then dried and imaged onto a photostimulable phosphor plate for detection (Personal Molecular Imager, Bio-Rad). Data were quantified using ImageQuant TL software version 8 (GE) by counting total pixel intensities for each row of pixels down each manually demarcated lane or band, yielding 2D graphs representing amount of product per Rf.

#### 7.2.4 Calculation of inhibition constants

Total amounts of detected reaction products at each concentration of fragment were transformed into fraction of the control reaction (DMSO), which were then averaged over 2-4 independent repeats. IC<sub>50</sub> values were calculated by fitting the averaged data

points to the following sigmoidal dose-response model using Graphpad Prism version 8:

$$Inhibition = Y_0 + \frac{(Y_{max} - Y_0)}{1 + 10^{(\log IC_{50} - \log[fragment]) - p}}$$

where  $Y_0$  and  $Y_{max}$  are the bottom and top of the curve, respectively,  $IC_{50}$  is the X value halfway between  $Y_0$  and  $Y_{max}$ , and  $p$  is the Hill slope. The Hill slope was fixed to 1.

### 7.3 Primase activity assay

Primase alone has insufficient primer synthesis activity to yield reliable product signal even after washing and gel electrophoresis, therefore DnaB helicase was included in this assay. The primer template was ssDNA of sequence (CT)<sub>17</sub>GCAAAGC, resulting in a full-length complementary RNA primer product of 35 b. This template or template-primer duplex is not expected to be a substrate to helicase, therefore the only role of helicase in this assay is to bind primase.

Assay conditions were kept as close as possible to those of the whole replisome reaction. The primer synthesis reaction mix was assembled on ice by mixing template (0.2  $\mu$ M), buffer, DTT, BSA,  $Mg^{2+}$ , all (deoxy-)nucleotides (ATP final concentration was lowered from 2 mM to 0.2 mM), and  $\alpha$ -<sup>32</sup>P-GTP (33 nM/35  $\mu$ Ci). The experiment was started by addition of the reaction mix to the fragment or to DMSO (the final concentration of DMSO was 2.5% in all reactions). The reaction mix was incubated at 37 °C for 3 minutes for equilibration. Then, the reaction was started by the addition of primase to 200 nM and helicase to 160 nM. The reaction was allowed to proceed for 1 h, after which the reaction was terminated by addition of  $NH_4OAc$  to 2.3 M. reaction product was precipitated by ethanol precipitation in a test tube, followed by a wash of the pellet to remove most of the remaining free  $\alpha$ -<sup>32</sup>P-GTP. The reaction products were run on denaturing (7 M urea) 18-20% polyacrylamide gel, which was then dried and imaged onto a photostimulable phosphor plate for detection (Personal Molecular Imager, Bio-Rad). Data were quantified using ImageQuant TL software version 8 (GE).

## 7.4 NMR

Samples (550  $\mu$ L) were prepared in protein storage buffer unless otherwise specified and contained 3-trimethyl-silyl-propane 1-sulfonic acid (DSS) (0.1 mM) and D<sub>2</sub>O (5-10%). Compounds were added from neat stocks or 200 mM stocks in deuterated DMSO. The sample temperature was 298 K. NMR data were collected on a 700 MHz Bruker Avance Neo spectrometer, equipped with a 5 mm triple-resonance N<sub>2</sub>-cooled cryoprobe.

### 7.4.1 Compound quality controls

Fragments were diluted to 500  $\mu$ M into QC buffer (NaP<sub>i</sub> pH 7.80 mM, D<sub>2</sub>O 10%, DSS 100  $\mu$ M). Proton spectra were recorded with water suppression by excitation sculpting (Hwang & Shaka, 1995) with 16-32 scans, a spectral width of 16 ppm, an acquisition duration of 2 s, and a recycle delay of 8 s. All analysis was done using Bruker TopSpin software version 3.5.6 or 4.0.2. Fragments were evaluated as ‘passed’ if their spectra were identical to their original spectra recorded when their stock solutions were first prepared.

### 7.4.2 Ligand-observed experiments

All samples contained purified protein (15-20  $\mu$ M) exchanged into fresh buffer: either NaP<sub>i</sub> or Tris or HEPES at pH 7.5, NaCl (50 mM), MgCl<sub>2</sub> (10 mM), DTT (1 mM), DSS (0.1 mM), D<sub>2</sub>O (5%), compound (0.5-1 mM), and DMSO-d<sub>6</sub> (0.25-1.5%). For WaterLOGSY experiments, the ephogsygpn pulse program (Dalvit *et al.*, 2000) was used with a mixing time of 1.5 s. For saturation transfer difference (STD) experiments, the protein saturation frequency peak was set to 0.5 ppm downfield of the transmitter frequency and a gaussian pulse shape was used. For Carr-Purcell-Meiboom-Gills (CPMG) experiments, the filter delay was set to 1 s. Fragment screening against full-length primase was conducted at the Crick Institute (London, UK) using a 700 MHz Bruker Avance III HD equipped with a cryo-probe and a SampleJet 96 well plate array autosampler. The screening buffer consisted of NaP<sub>i</sub> (25 mM) pH 7.5, NaCl (45 mM), MgCl<sub>2</sub> (10 mM), DSS (0.1 mM), D<sub>2</sub>O (5%), compound (0.5 mM), and DMSO-d<sub>6</sub> (1.5%).

### 7.4.3 Protein-observed experiments

Data for the ZBD were collected on a 500 MHz Bruker Avance I equipped with a room temperature BBI probe. For the ZBD, the hsqcetgpsi pulse program with water suppression was used (Schleucher *et al.*, 1994; Kay *et al.*, 1992), with 96 scans, 2048 and 512 complex points per experiment, and an F1 offset and spectral width of 120 and 32 ppm respectively. For the HBD,  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra were recorded using the SOFAST experiment (Schanda *et al.*, 2005) with 8 scans, 2048 and 256 complex points per experiment, and an F1 offset and spectral width of 120 and 40 ppm respectively.  $^1\text{H}$ - $^{15}\text{N}$ - $^{13}\text{C}$  HNC0 spectra were recorded using the hncogpwg3d pulse program (Kay *et al.*, 1994; Grzesiek & Bax, 1992), with 8 scans, 5% or 15% non-uniform sampling, 2048, 128, and 256 complex points per experiment, F1 offset and spectral width of 174 and 12 ppm respectively, and F2 offset and spectral width of 118 and 28 ppm respectively. Data were processed using Topspin and analysed using CcpNmr Analysis v2 (Vranken *et al.*, 2005). The position of the main DSS peak was used as an internal zero-frequency reference and published  $\delta$  values were used for extension to indirect dimensions (Markley *et al.*, 1998).

### 7.5 MD simulations

Fragment parametrisation was done using Gaussian09 with the HF/6-31G\* basis set (Hariharan & Pople, 1973; Francl *et al.*, 1982) for calculation of electrostatic potentials. Further work was done using several programs within the Amber14 or AmberTools17 program suites (Case *et al.*, 2014). Antechamber was used for normalisation of Gaussian output into restrained electrostatic potentials (Bayly *et al.*, 1993), and missing bonding/angles parameters were added manually. Homology modelling of the *E. coli* ZBD (residues 2-101) was done in MOE version 2016.0802 based on the crystal structure of the *G. stearothermophilus* homologue (Pan & Wigley, 2000). Parameters for the zinc ion were added manually based on reported values for zinc in CHC2-type zinc binding proteins. Systems were prepared by placing one protein and one ligand molecule in different orientations at least 5 Å away from each other. Leap was used to solvate these models in explicit TIP3P water in an octahedron with at least 4 Å distance between any non-solvent atom and the box boundaries.  $\text{Na}^+$  and  $\text{Cl}^-$  ions were added to effective concentrations of 33-44 mM to neutralise the system. Proteins were described using the AMBER ff14SB forcefield (Maier *et al.*, 2015). Systems were energy-minimised then heated from 100-300 K in four steps of

200 ps at constant volume, then vacuums were removed in one step of 1 ns at constant pressure of 1 atmosphere. Temperature was kept at 300 K employing Langevin dynamics with a collision frequency of  $4 \text{ ps}^{-1}$  (Izaguirre *et al.*, 2001). Simulations were run for 500 ns for each system (8 in total, 4 per protein) at constant volume, using SHAKE (Ryckaert *et al.*, 1977) to allow for time steps of 2 fs, writing information only once per 40 ps to reduce calculation times. All simulations were run using the particle mesh Ewald (PME) method (Darden *et al.*, 1993) (direct space sum cut-off 9 Å) on NVIDIA Tesla K80 graphics processing units (Salomon-Ferrer *et al.*, 2013) on the Minotauro cluster of the Barcelona Supercomputing Center. Calculation times varied between 20-100 ns per day depending on the number of atoms in the system and the number of available nodes. Trajectory analyses were done using cpptraj (Roe & Cheatham, 2013).

## 7.6 SEC-MALS

All SEC-MALS experiments were conducted by the University of York Department of Biology Biotechnology Facility. Experiments were conducted on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu HPLC system. Work was conducted at room temperature ( $20 \pm 2 \text{ }^\circ\text{C}$ ). The column (Superdex 75 or 200 analytical grade) was equilibrated with at least 2 column volumes of protein storage buffer before use and flow was continued at the working flow rate until baselines for UV, light scattering and refractive index detectors were all stable. Sample injection volume was 100  $\mu\text{L}$ . Shimadzu LC Solutions software was used to control the HPLC and Astra V software for the HELEOS-II and rEX detectors. Data were analysed using the Astra V software. Molecular weights were estimated using the Zimm fit method with degree 1. A control sample of bovine serum albumin (BSA) was run to correct for changes in instrument calibration and  $dn/dc$  values in the buffer used. The molecular weight of BSA was taken as 66,400 Da.

## 7.7 Mass spectrometry

All MS experiments were conducted by the University of York Department of Biology Biotechnology Facility Metabolomics & Proteomics centre. Samples were diluted 1 in 10 with aqueous 50% (v:v) acetonitrile containing 1% (v:v) formic acid before infusing at  $3 \mu\text{Lmin}^{-1}$  into a Bruker maXis-HD qTOF mass spectrometer using

electrospray ionisation. Positive ESI-MS spectra were averaged over a 1 min acquisition before maximum entropy deconvolution. Deconvoluted mass spectra were peak picked to monoisotopic masses using the SNAP algorithm (repetitive building blocks C4.9384; N1.3577; O1.4773; S0.0415; H7.7583), requiring a minimum signal to noise ratio and minimum percent relative intensity of 10. Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control and DataAnalysis, Bruker Daltonics).

## 7.8 Crystallography

All crystallisations were set up using the sitting drop vapour diffusion method. 200 nL drops with equal volumes protein:precipitant were prepared using a Mosquito liquid handler (SPT Labtech). Purified RPD (aa 111-433 or 111-429) (7-15 mgmL<sup>-1</sup>) was subjected to several broad crystallisation screens that were commercially available. A crystal cluster that grew from an impurity in bicine pH 9 (0.1 M) + PEG6000 (20%) (JCSG+ screen, Molecular Dimensions) was used for seeding in the same condition, which yielded a single crystal in the I222 space group ( $a/b/c = 69.8/73.3/129.2 \text{ \AA}$ ). Optimisation of the condition included varying the pH between 8.5-9.0, PEG concentration up to 24%, addition of ethylene glycol up to 20%, and varying the protein:reservoir volume ratios. In all these trials, however, crystals only grew in the P212121 space group again ( $a/b/c = 38/57/135-148 \text{ \AA}$ ) where the c axis length varied depending on soaking conditions. Soaking was performed by adding mother liquor mixed with up to 50% compound stock (200 mM in DMSO for fragments, or 100 mM in water for nucleotides) to the crystal drops. Bigger crystals often cracked upon addition of compound. Slow increase of compound concentrations was found to be the best way to preserve crystals. After soaking overnight, crystals were frozen directly into liquid nitrogen.

All datasets were collected at the Diamond Light Source. Some crystals were centred manually to avoid irradiating cracks which would give multiple diffraction patterns. Data were collected in 0.1° slices up to 220° with no beam attenuation. Data processing was done in CCP4i2 and consisted of automatic integration, scaling and merging using the XDS 3dii pipeline, data reduction in Pointless/Aimless, molecular replacement in Molrep using a previously solved apo model with a matching c axis, and refinement in Refmac and Coot. Where possible, the same FreeR set derived from

an apo dataset was used for all compound-soaked datasets to avoid biases. Datasets that were almost completely isomorphous to available apo models were processed using the automated ligand solution workflow within CCP4i2. Resolution cut-offs for refinement were based on anisotropic CC1/2 statistics. A small number of crystals showed signs of radiation damage as evidenced from increased  $R_{\text{merge}}$  values of later images. These images were then excluded while still preserving at least 90° of data. To judge the presence of difference density that could accommodate the soaked molecule, datasets were subjected to one round of refinement and manual model building. Statistics for a representative dataset (soaked in 10% DMSO overnight) after one round of automatic and manual refinement are given in Table 7-2.

Table 7-2. Data collection and refinement statistics for primase RPD (111-433) soaked in 10% DMSO.

<b>Data collection</b>	
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
a, b, c (Å)	38.00, 57.95, 145.42
Resolution (Å)	1.95 (2.00-1.95)
$R_{\text{merge}}$ (%)	9.2 (24.4)
$I / \sigma(I)$	9.7 (0.9)
CC1/2 (%)	99.8 (36.0)
Completeness (%)	100 (100)
Redundancy	8.0 (8.3)
<b>Refinement</b>	
Resolution (Å)	1.95
Number of unique reflections	24356 (1718)
$R_{\text{work}} / R_{\text{free}}$ (%)	22.5 / 27.6
Number of protein atoms	2538
Number of waters	43
B-factors (Å <sup>2</sup> )	
Protein	48.95
Water	47.71
RMS bond length deviation (Å)	0.01
RMS bond angle deviation (°)	1.41
Ramachandran (favoured/allowed/disallowed)	311/5/2

## 7.9 Synthesis of diazirine analogues and UV crosslinking experiments

The author of this manuscript gratefully acknowledges the contributions of Dr. James D. Firth to results presented in section 5.2, *i.e.* the design and synthesis of compounds D1, D2, D3, D4, and other derivatives of various replisome fragment inhibitors. All

compounds and intermediate reagents were characterised by infrared spectroscopy,  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectroscopy, and high-resolution ESI-MS.

Compounds (0.5 mM in NMR QC buffer unless otherwise specified) and reaction mixtures were irradiated with UV-B light (peak 302 nm, range 280-350 nm) in a CL-1000 ultraviolet crosslinker ( $1\text{ mJcm}^{-2}$ ) (UVP) at room temperature. For incubation times longer than 10 minutes, irradiation resulted in an increase in chamber temperature from 22 °C to a maximum of 28 °C. Samples were kept in open transparent containers during irradiation. After irradiation, samples were either immediately used in subsequent experiments (replication assays, NMR) or stored at 4 °C until further use (MS).

### **7.10 Click reactions**

Alkyne-tagged diazirine-protein mixtures were used directly after UV crosslinking. To each mixture were added to final concentrations:  $\text{CuSO}_4$  (aqueous stock) 1.5 mM, tris((1-benzyl-4-triazolyl)methyl)amine (TBTA) (DMSO stock) 1.5 mM, TAMRA-azide dye (DMSO stock) 30  $\mu\text{M}$ , ascorbic acid (fresh aqueous stock) 1.5 mM. Click reaction mixtures were incubated at 40 °C for 1 h, then diluted with Laemmli sample buffer and directly used for gel electrophoresis.

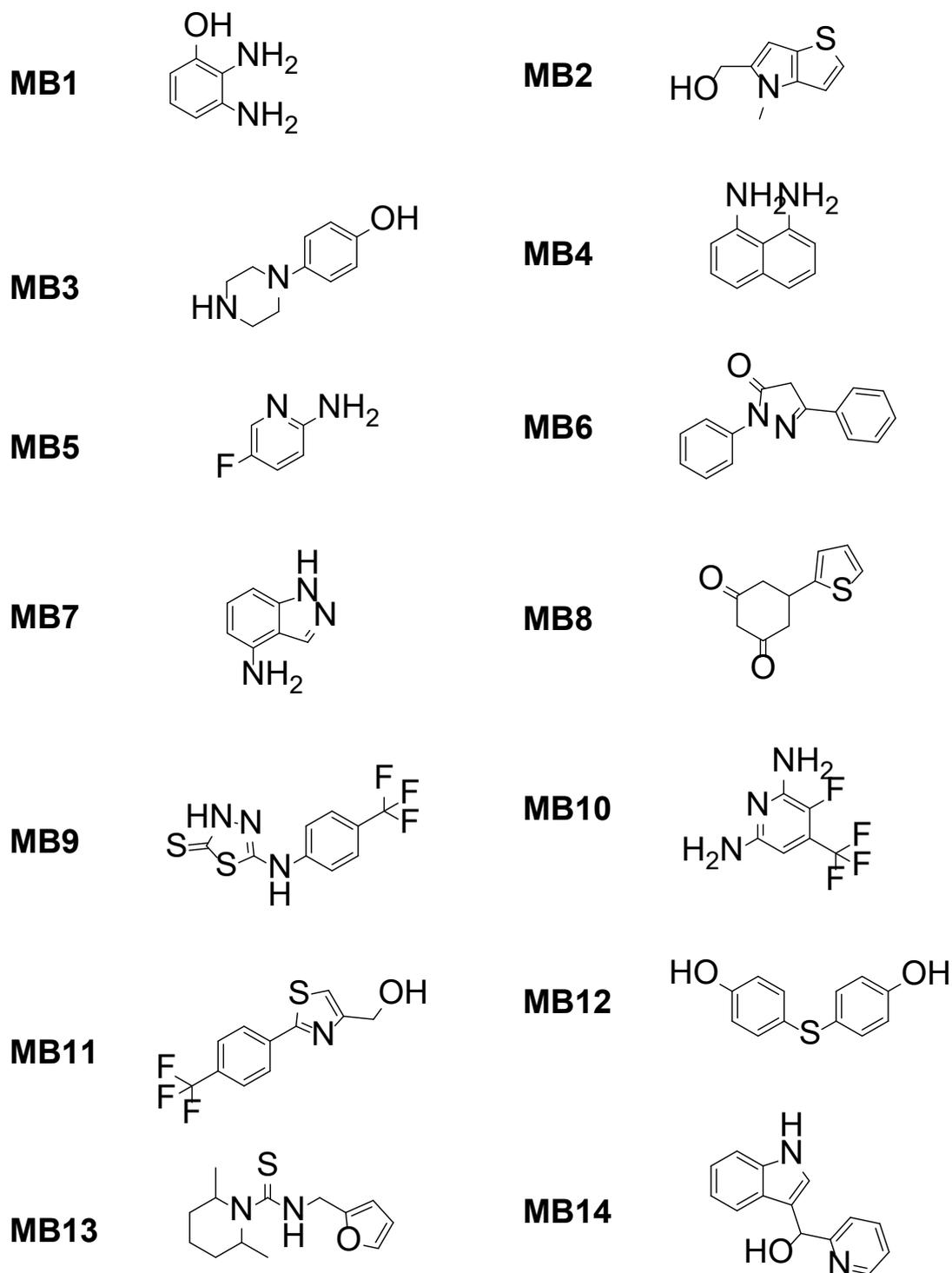
### **7.11 Covalent fragment screening**

Electrophilic fragments were provided by the Keserú group (Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary) as 100 mM DMSO stock solutions. For storage periods longer than one month, these fragments were stored in argon atmosphere at -20 °C. All replication reaction methods and materials were identical to those used in previous experiments (see section 7.2) (BSA and DTT were omitted from all screening reactions).

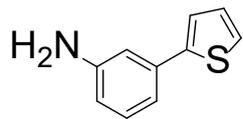


## Appendix

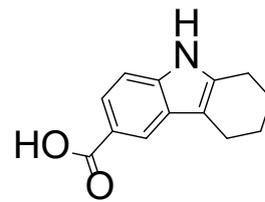
### Appendix A – Structures of 25 replisome screen fragment hits



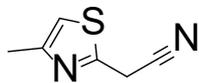
**MB15**



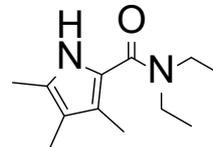
**MB16**



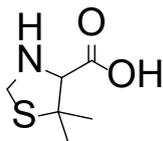
**MB17**



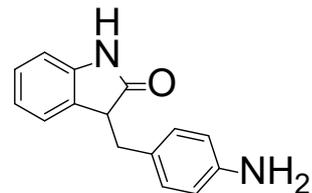
**MB18**



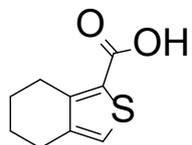
**MB19**



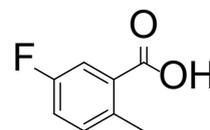
**MB20**



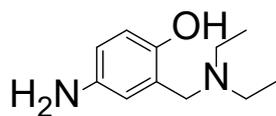
**MB25**



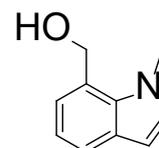
**MB33**



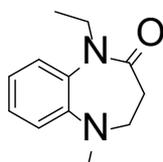
**MB41**



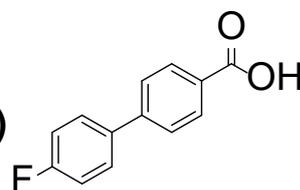
**MB63**



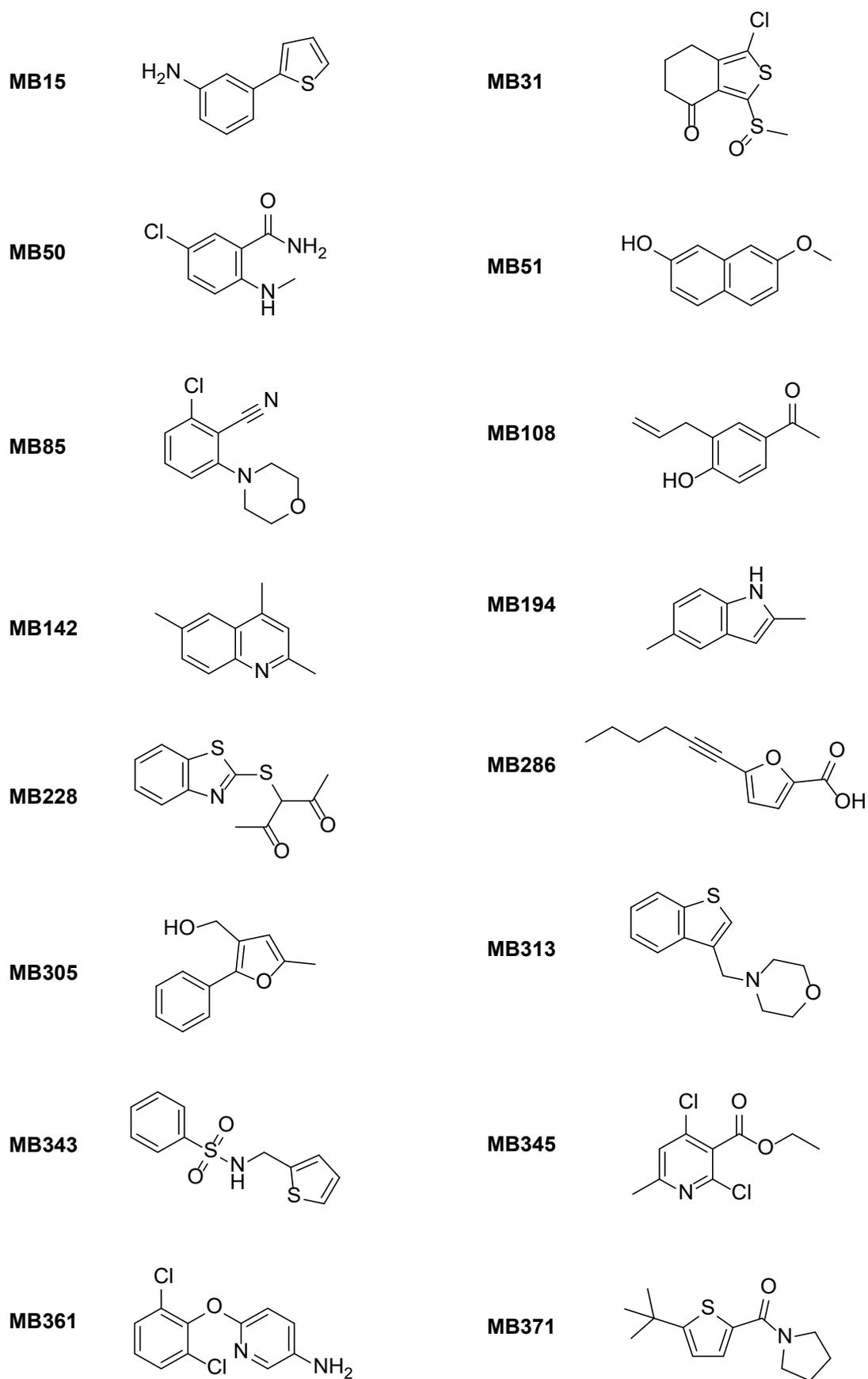
**MB99**

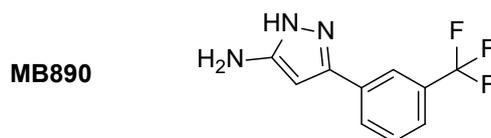
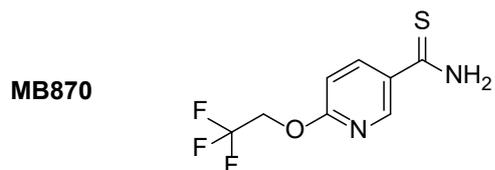
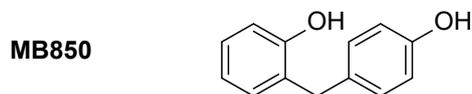
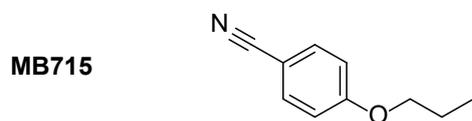
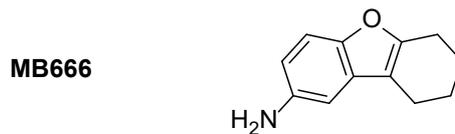
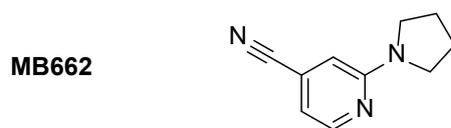
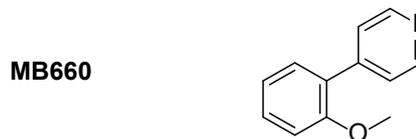
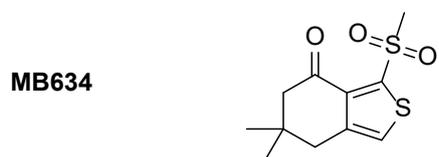
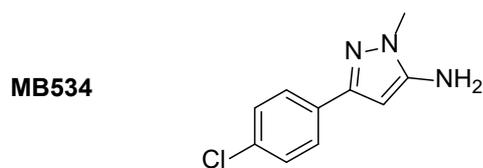
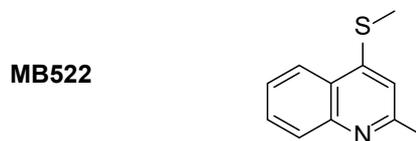
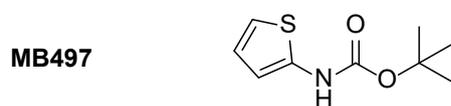
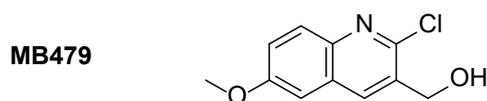
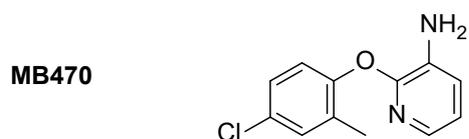
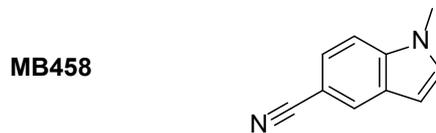
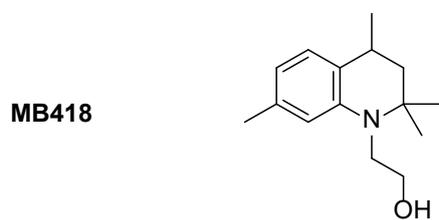
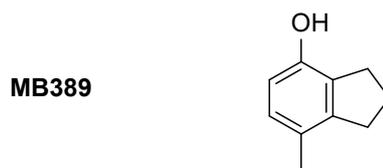
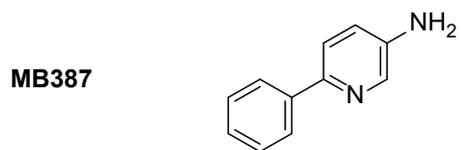


**F5 (control fragment)**



Appendix B – Structures of 35 primase NMR screen fragment hits





## Abbreviations

aa – amino acid

ApCpp – adenosine 5'-( $\alpha,\beta$ -methylene)triphosphate

bp – base pair

BSA – bovine serum albumin

CSP – chemical shift perturbation

CTD – carboxy-terminal domain

DMSO – dimethylsulfoxide

dsDNA – double-stranded DNA

DSS – 3-trimethyl-silyl-propane 1-sulfonic acid

DTT – dithiothreitol

FBLD – Fragment-based lead discovery

FL – full-length

GpCpp – guanosine 5'-( $\alpha,\beta$ -methylene)triphosphate

GppNHp – guanosine 5'-( $\beta,\gamma$ -imido)triphosphate

HA – heavy atom

HBD – helicase-binding domain

HSQC – heteronuclear single quantum coherence

HTS – high-throughput screening

IC<sub>50</sub> – 50% inhibitory concentration

ITC – isothermal titration calorimetry

K<sub>D</sub> – equilibrium dissociation constant

LE – ligand efficiency

LO – ligand-observed

MD – molecular dynamics

MS – mass spectrometry

NMR – nuclear magnetic resonance

NTD – amino-terminal domain

PO – protein-observed

pol III HE – polymerase III holoenzyme

ppGpp – guanosine 3',5'-bispyrophosphate

PPI – protein-protein interaction

rmsd – root mean square deviation

RPD – RNA polymerase domain

SAR – structure-activity relationships

SEC-MALS – size exclusion chromatography-multi-angle light scattering

SPR – surface plasmon resonance

SSB – Single-stranded DNA-binding protein

ssDNA – single-stranded DNA

TSA – thermal shift assay

ZBD – zinc-binding domain

## References

- Aakre, C. D., Phung, T. N., Huang, D. & Laub, M. T. (2013). *Mol. Cell.* **52**, 617–628.
- Ábrányi-Balogh, P., Petri, L., Imre, T., Szijj, P., Scarpino, A., Hrast, M., Mitrović, A., Fonovič, U. P., Németh, K., Barreteau, H., Roper, D. I., Horváti, K., Ferenczy, G. G., Kos, J., Ilaš, J., Gobec, S. & Keserű, G. M. (2018). *Eur. J. Med. Chem.* **160**, 94–107.
- Aiello, D., Barnes, M. H., Biswas, E. E., Biswas, S. B., Gu, S., Williams, J. D., Bowlin, T. L. & Moir, D. T. (2009). *Bioorg. Med. Chem.* **17**, 4466–4476.
- Alberts, B. M., Barry, J., Bedinger, P., Formosa, T., Jongeneel, C. V. & Kreuzer, K. N. (1983). *Cold Spring Harb. Symp. Quant. Biol.* **47**, 655–668.
- Allen, K. N., Bellamacina, C. R., Ding, X., Jeffery, C. J., Mattos, C., Petsko, G. A. & Ringe, D. (1996). *J. Phys. Chem.* **100**, 2605–2611.
- Alvarez-Garcia, D. & Barril, X. (2014). *J. Med. Chem.* **57**, 8530–8539.
- Aravind, L. (1998). *Nucleic Acids Res.* **26**, 4205–4213.
- Arias-Palomo, E., O’Shea, V. L., Hood, I. V. & Berger, J. M. (2013). *Cell.* **153**, 438–448.
- Arias-Palomo, E., Puri, N., O’Shea Murray, V. L., Yan, Q. & Berger, J. M. (2019). *Mol. Cell.* **74**, 173–184.
- Arrowsmith, C. H., Audia, J. E., Austin, C., Baell, J., Bennett, J., Blagg, J., Bountra, C., Brennan, P. E., Brown, P. J., Bunnage, M. E., Buser-Doepner, C., Campbell, R. M., Carter, A. J., Cohen, P., Copeland, R. A., Cravatt, B., Dahlin, J. L., Dhanak, D., Edwards, A. M., Frederiksen, M., Frye, S. V, Gray, N., Grimshaw, C. E., Hepworth, D., Howe, T., Huber, K. V. M., Jin, J., Knapp, S., Kotz, J. D., Kruger, R. G., Lowe, D., Mader, M. M., Marsden, B., Mueller-Fahrnow, A., Müller, S., O’Hagan, R. C., Overington, J. P., Owen, D. R., Rosenberg, S. H., Ross, R., Roth, B., Schapira, M., Schreiber, S. L., Shoichet, B., Sundström, M., Superti-Furga, G., Taunton, J., Toledo-Sherman, L., Walpole, C., Walters, M. A., Willson, T. M., Workman, P., Young, R. N. & Zuercher, W. J. (2015). *Nat. Chem. Biol.* **11**, 536–541.

- Auffinger, P., Cheatham, T. E. & Vaiana, A. C. (2007). *J. Chem. Theory Comput.* **3**, 1851–1859.
- Ayotte, Y., Bilodeau, F., Descoteaux, A. & LaPlante, S. R. (2018). *ChemMedChem.* **13**, 1377–1386.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. (2006). *Mol. Syst. Biol.* **2**, 2006–0008.
- Baell, J. B. & Holloway, G. A. (2010). *J. Med. Chem.* **53**, 2719–2740.
- Bailey, S., Eliason, W. K. & Steitz, T. A. (2007). *Science.* **318**, 459–463.
- Ballanco, J. & Mansfield, M. L. (2011). *PLoS One.* **6**, e18881.
- Barros, T., Guenther, J., Kelch, B., Anaya, J., Prabhakar, A., O'Donnell, M., Kuriyan, J. & Lamers, M. H. (2013). *BMC Struct. Biol.* **13**, 8.
- Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. (1993). *J. Phys. Chem.* **97**, 10269–10280.
- Benkert, P., Biasini, M. & Schwede, T. (2011). *Bioinformatics.* **27**, 343–350.
- Blinkova, A., Hervas, C., Stukenberg, P. T., Onrust, R., O'Donnell, M. E. & Walker, J. R. (1993). *J. Bacteriol.* **175**, 6018–6027.
- Bloom, L. B., Chen, X., Fygenson, D. K., Turner, J., O'Donnell, M. & Goodman, M. F. (1997). *J. Biol. Chem.* **272**, 27919–27930.
- Bollag, G., Tsai, J., Zhang, J., Zhang, C., Ibrahim, P., Nolop, K. & Hirth, P. (2012). *Nat Rev Drug Discov.* **11**, 873–886.
- Bottaro, S. & Lindorff-Larsen, K. (2018). *Science.* **361**, 355–360.
- Bressanin, D., Stefan, A., Piaz, F. D., Cianchetta, S., Reggiani, L. & Hochkoepler, A. (2009). *Biochim. Biophys. Acta - Proteins Proteomics.* **1794**, 1606–1615.
- Bunnage, M. E., Chekler, E. L. P. & Jones, L. H. (2013). *Nat. Chem. Biol.* **9**, 195–199.
- Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham, T., Darden, T., Duke, R., Gohlke, H., Goetz, A., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossváry, I., Kovalenko, A., Lee, T., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K., Paesani, F., Roe, D., Roitberg, A., Sagui, C., Salomon-

- Ferrer, R., Seabra, G., Simmerling, C., Smith, W., Swails, J., Walker, R., Wang, J., Wolf, R., Wu, X. & Kollman, P. (2014).
- Catazaro, J., Periago, J., Shortridge, M. D., Worley, B., Kirchner, A., Powers, R. & Griep, M. A. (2017). *Biochemistry*. **56**, 932–943.
- Chen, I.-J. & Hubbard, R. E. (2009). *J. Comput. Aided. Mol. Des.* **23**, 603–620.
- Chilingaryan, Z., Headey, S., Lo, A., Xu, Z.-Q., Otting, G., Dixon, N., Scanlon, M. & Oakley, A. (2018). *Antibiotics*. **7**, 14.
- Congreve, M., Carr, R., Murray, C. & Jhoti, H. (2003). *Drug Discov Today*. **8**, 876–877.
- Corn, J. E., Pease, P. J., Hura, G. L. & Berger, J. M. (2005). *Mol. Cell*. **20**, 391–401.
- Corn, J. E., Pelton, J. G. & Berger, J. M. (2008). *Nat. Struct. Mol. Biol.* **15**, 163–169.
- Coyle, J. & Walser, R. (2020). *SLAS Discov. Adv. Sci. Drug Discov.* **25**, 471–490.
- Dallmann, H. G. (2000). *J. Biol. Chem.* **275**, 15512–15519.
- Dalvit, C., Caronni, D., Mongelli, N., Veronesi, M. & Vulpetti, A. (2006). *Curr. Drug Discov. Technol.* **3**, 115–124.
- Dalvit, C., Fogliatto, G., Stewart, A., Veronesi, M. & Stockman, B. (2001). *J. Biomol. NMR*. **21**, 349–359.
- Dalvit, C., Pevarello, P., Tatò, M., Veronesi, M., Vulpetti, A. & Sundström, M. (2000). *J. Biomol. NMR*. **18**, 65–68.
- Danielson, U. (2009). *Curr. Top. Med. Chem.* **9**, 1725–1735.
- Darby, J. F., Atobe, M., Firth, J. D., Bond, P., Davies, G. J., O’Brien, P. & Hubbard, R. E. (2017). *Chem. Sci.* **8**, 7772–7779.
- Darden, T., York, D. & Pedersen, L. (1993). *J. Chem. Phys.* **98**, 10089–10092.
- Davies, J. (2006). *J. Ind. Microbiol. Biotechnol.* **33**, 496–499.
- Dervyn, E., Suski, C., Daniel, R., Bruand, C., Chapuis, J., Errington, J., Jannièrè, L. & Ehrlich, S. D. (2001). *Science*. **294**, 1716–1719.
- Dietrich, M., Pedro, L., Garcia, J., Pons, M., Huttener, M., Paytubi, S., Madrid, C. &

- Juarez, A. (2014). *J. Bacteriol.* **196**, 1102–1112.
- Erlanson, D. A. (2019). Practical Fragments Blog - fragment-finding methods in 2019.
- Erlanson, D. A., Davis, B. J. & Jahnke, W. (2019). *Cell Chem. Biol.* **26**, 9–15.
- Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. (2016). *Nat. Rev. Drug Discov.* **15**, 605–619.
- Erlanson, D. A., Wells, J. A. & Braisted, A. C. (2004). *Annu. Rev. Biophys. Biomol. Struct.* **33**, 199–223.
- Fabio Lapenta, Silva, A. M., Brandimarti, R., Lanzi, M., Gratani, F. L., Gonzalez, P. V., Peticarari, S. & Hochkoepller, A. (2016). *PLoS One.* **11**, e0157207.
- Ferenczy, G. G. & Keserü, G. M. (2016). *Fragm. Drug Discov. Lessons Outlook.* 75–98.
- Fernandez-Leiro, R., Conrad, J., Scheres, S. H. W. & Lamers, M. H. (2015). *Elife.* **4**, 1–16.
- Fernandez-Leiro, R., Conrad, J., Yang, J.-C., Freund, S. M. V, Scheres, S. H. W. & Lamers, M. H. (2017). *Nat. Struct. Mol. Biol.*
- Firth, J. D. & O'Brien, P. (2018). *Chemical and Biological Synthesis: Enabling Approaches for Understanding Biology*, Vol. pp. 74–113.
- Flower, A. M. & McHenry, C. S. (1990). *Proc. Natl. Acad. Sci.* **87**, 3713–3717.
- Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., DeFrees, D. J. & Pople, J. A. (1982). *J. Chem. Phys.* **77**, 3654–3665.
- Frick, D. N. & Richardson, C. C. (2001). *Annu. Rev. Biochem.* **70**, 39–80.
- Galletto, R., Maillard, R., Jezewska, M. J. & Bujalowski, W. (2004). *Biochemistry.* **43**, 10988–11001.
- Galmozzi, A., Parker, C. G., Kok, B. P., Cravatt, B. F. & Saez, E. (2018). Vol. pp. 115–127.
- Gao, D. & McHenry, C. S. (2001). *J. Biol. Chem.* **276**, 4441–4446.
- Gao, Y., Cui, Y., Fox, T., Lin, S., Wang, H., de Val, N., Zhou, Z. H. & Yang, W.

- (2019). *Science*. **363**, eaav7003.
- Georgescu, R. E., Langston, L., Yao, N. Y., Yurieva, O., Zhang, D., Finkelstein, J., Agarwal, T. & O'Donnell, M. E. (2014). *Nat. Struct. Mol. Biol.* **21**, 664–670.
- Gerry, C. J. & Schreiber, S. L. (2018). *Nat. Rev. Drug Discov.* **17**, 333–352.
- Ghose, A. K., Viswanadhan, V. N. & Wendoloski, J. J. (1998). *J. Phys. Chem. A.* **102**, 3762–3772.
- Godson, G. N., Schoenich, J., Sun, W. & Mustaev, A. A. (2000). *Biochemistry.* **39**, 332–339.
- Goodford, P. J. (1985). *J Med Chem.* **28**, 849–857.
- Gossert, A. D. & Jahnke, W. (2016). *Prog Nucl Magn Reson Spectrosc.* **97**, 82–125.
- Graham, J. E., Mariani, K. J. & Kowalczykowski, S. C. (2017). *Cell.* **169**, 1201–1213.
- Grant, E. K., Fallon, D. J., Eberl, H. C., Fantom, K. G. M., Zappacosta, F., Messenger, C., Tomkinson, N. C. O. & Bush, J. T. (2019). *Angew. Chemie Int. Ed.* **58**, 17322–17327.
- Griep, M. A., Blood, S., Larson, M. A., Koepsell, S. A. & Hinrichs, S. H. (2007). *Bioorg. Med. Chem.* **15**, 7203–7208.
- Grzesiek, S. & Bax, A. (1992). *J. Magn. Reson.* **96**, 432–440.
- Hajduk, P. J., Olejniczak, E. T. & Fesik, S. W. (1997). *J. Am. Chem. Soc.* **119**, 12257–12261.
- Hall, R. J., Mortenson, P. N. & Murray, C. W. (2014). *Prog. Biophys. Mol. Biol.* **116**, 82–91.
- Hallenbeck, K. K., Turner, D. M., Renslo, A. R. & Arkin, M. R. (2017). *Curr. Top. Med. Chem.* **17**, 4–15.
- Hamdan, S., Carr, P. D., Brown, S. E., Ollis, D. L. & Dixon, N. E. (2002). *Structure.* **10**, 535–546.
- Hamdan, S. M., Loparo, J. J., Takahashi, M., Richardson, C. C. & van Oijen, A. M. (2009). *Nature.* **457**, 336–339.

- Hann, M. M., Leach, A. R. & Harper, G. (2001). *J Chem Inf Comput Sci.* **41**, 856–864.
- Hariharan, P. C. & Pople, J. A. (1973). *Theor. Chim. Acta.* **28**, 213–222.
- Harvey, M. J. & De Fabritiis, G. (2012). *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 734–742.
- Hassan, M., Brown, R. D., Varma-O'Brien, S. & Rogers, D. (2006). *Mol. Divers.* **10**, 283–299.
- Hermann, J. C., Chen, Y., Wartchow, C., Menke, J., Gao, L., Gleason, S. K., Haynes, N.-E., Scott, N., Petersen, A., Gabriel, S., Vu, B., George, K. M., Narayanan, A., Li, S. H., Qian, H., Beatini, N., Niu, L. & Gan, Q.-F. (2013). *ACS Med. Chem. Lett.* **4**, 197–200.
- Horne, J. E., Walko, M., Calabrese, A. N., Levenstein, M. A., Brockwell, D. J., Kapur, N., Wilson, A. J. & Radford, S. E. (2018). *Angew. Chemie Int. Ed.* **57**, 16688–16692.
- Hou, C., Biswas, T. & Tsodikov, O. V. (2018). *Biochemistry.* **57**, 2084–2093.
- Hurwitz, J. & Wickner, S. (1974). *Proc. Natl. Acad. Sci.* **71**, 6–10.
- Hwang, T. L. & Shaka, A. J. (1995). *J. Magn. Reson. Ser. A.* **112**, 275–279.
- Ichihara, O., Barker, J., Law, R. J. & Whittaker, M. (2011). *Mol. Inform.* **30**, 298–306.
- Ilic, S., Akabayov, S. R., Arthanari, H., Wagner, G., Richardson, C. C., Akabayov, B., Silver, L. L., Lewis, K., Marians, K. J., Yao, N. Y., O'Donnell, M., Frick, D. N., Richardson, C. C., Arezi, B., Kuchta, R. D., Payne, D. J., Gwynn, M. N., Holmes, D. J., Pompliano, D. L., Baker, M., Lepre, C. A., Peng, J. W., Lepre, C. A., Fejzo, J., Abdul-Manan, N., Moore, J. M., Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W., Jhoti, H., Venkitakrishnan, R. P., Benard, O., Max, M., Markley, J. L., Assadi-Porter, F. M., Scott, D. E., Coyne, A. G., Hudson, S. A., Abell, C., Wasko, M. J., Pellegrine, K. A., Madura, J. D., Surratt, C. K., Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., Coleman, R. G., Morris, G. M., Kato, M., Ito, T., Wagner, G., Richardson, C. C., Ellenberger, T., Huber, H. E., Tabor, S., Richardson, C. C., Notarnicola, S. M., Mulcahy, H. L., Lee, J., Richardson, C. C., Nakai, H., Richardson, C. C., Frick, D. N., Baradaran, K.,

- Richardson, C. C., Kusakabe, T., Richardson, C. C., Mendelman, L. V., Richardson, C. C., Sklenar, V., Piotto, M., Leppik, R., Saudek, V., Piotto, M., Saudek, V., Sklenar, V., Mayer, M., Meyer, B., Delaglio, F., Johnson, B. A., Blevins, R. A., Hyberts, S. G., Milbradt, A. G., Wagner, A. B., Arthanari, H., Wagner, G., Delaglio, F., Corn, J. E., Pease, P. J., Hura, G. L., Berger, J. M., Rymer, R. U., Wagstaff, J. L., Taylor, S. L. & Howard, M. J. (2016). *Sci. Rep.* **6**, 36322.
- Indiani, C. & O'Donnell, M. (2006). *Nat. Rev. Mol. Cell Biol.* **7**, 751–761.
- Itsathitphaisarn, O., Wing, R. A., Eliason, W. K., Wang, J. & Steitz, T. A. (2012). *Cell*. **151**, 267–277.
- Izaguirre, J. A., Catarello, D. P., Wozniak, J. M. & Skeel, R. D. (2001). *J. Chem. Phys.* **114**, 2090–2098.
- Jackman, J. E., Gott, J. M. & Gray, M. W. (2012). *RNA*. **18**, 886–899.
- Jencks, W. P. (1975). *Adv. Enzymol. Relat. Areas Mol. Biol.* **43**, 219–410.
- Jencks, W. P. (1981). *Proc Natl Acad Sci U S A*. **78**, 4046–4050.
- Jergic, S., Horan, N. P., Elshenawy, M. M., Mason, C. E., Urathamakul, T., Ozawa, K., Robinson, A., Goudsmits, J. M. H., Wang, Y., Pan, X., Beck, J. L., van Oijen, A. M., Huber, T., Hamdan, S. M. & Dixon, N. E. (2013). *EMBO J.* **32**, 1322–1333.
- Jeruzalmi, D., O'Donnell, M. & Kuriyan, J. (2001). *Cell*. **106**, 429–441.
- Jeruzalmi, D., Yurieva, O., Zhao, Y., Young, M., Stewart, J., Hingorani, M., O'Donnell, M. & Kuriyan, J. (2001). *Cell*. **106**, 417–428.
- Jhoti, H., Williams, G., Rees, D. C. & Murray, C. W. (2013). *Nat. Rev. Drug Discov.* **12**, 644–644.
- Johnson, S. K., Bhattacharyya, S. & Griep, M. A. (2000). *Biochemistry*. **39**, 736–744.
- Jonsson, U., Fagerstam, L., Ivarsson, B., Johnsson, B., Karlsson, R., Lundh, K., Lofas, S., Persson, B., Roos, H. & Ronnberg, I. (1991). *Biotechniques*. **11**, 620–627.
- Kathman, S. G. & Statsyuk, A. V. (2016). *Medchemcomm.* **7**, 576–585.

- Kathman, S. G., Xu, Z. & Statsyuk, A. V. (2014). *J. Med. Chem.* **57**, 4969–4974.
- Kay, L. E., Xu, G. Y. & Yamazaki, T. (1994). *J. Magn. Reson. Ser. A.* **109**, 129–133.
- Kay, L., Keifer, P. & Saarinen, T. (1992). *J. Am. Chem. Soc.* **114**, 10663–10665.
- Keck, J. L., Roche, D. D., Lynch, A. S. & Berger, J. M. (2000). *Science.* **287**, 2482–2486.
- Keeley, A., Petri, L., Ábrányi-Balogh, P. & Keserű, G. M. (2020). *Drug Discov. Today.* **25**, 983–996.
- Keserű, G. M., Erlanson, D. A., Ferenczy, G. G., Hann, M. M., Murray, C. W. & Pickett, S. D. (2016). *J Med Chem.*
- Kirsch, P., Hartman, A. M., Hirsch, A. K. H. & Empting, M. (2019). *Molecules.* **24**, 4309.
- Kling, A., Lukat, P., Almeida, D. V., Bauer, A., Fontaine, E., Sordello, S., Zaburannyi, N., Herrmann, J., Wenzel, S. C., König, C., Ammerman, N. C., Barrio, M. B., Borchers, K., Bordon-Pallier, F., Bronstrup, M., Courtemanche, G., Gerlitz, M., Geslin, M., Hammann, P., Heinz, D. W., Hoffmann, H., Klieber, S., Kohlmann, M., Kurz, M., Lair, C., Matter, H., Nuermberger, E., Tyagi, S., Fraisse, L., Grosset, J. H., Lagrange, S. & Muller, R. (2015). *Science.* **348**, 1106–1112.
- Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., Xia, B., Beglov, D. & Vajda, S. (2015). *Nat Protoc.* **10**, 733–755.
- Lamers, M. H., Georgescu, R. E., Lee, S.-G., O'Donnell, M. & Kuriyan, J. (2006). *Cell.* **126**, 881–892.
- Lancy, E. D., Lifshits, M. R., Kehres, D. G. & Maurer, R. (1989). *J. Bacteriol.* **171**, 5572–5580.
- Larson, M. A., Griep, M. A., Bressani, R., Chintakayala, K., Soultanas, P. & Hinrichs, S. H. (2010). *Nucleic Acids Res.* **38**, 7167–7178.
- Leach, A. R. & Hann, M. M. (2011). *Curr. Opin. Chem. Biol.* **15**, 489–496.
- Lee, S.-J. & Richardson, C. C. (2011). *Curr. Opin. Chem. Biol.* **15**, 580–586.
- Leeson, P. D. & Springthorpe, B. (2007). *Nat. Rev. Drug Discov.* **6**, 881–890.

- Leu, F. P., Hingorani, M. M., Turner, J. & O'Donnell, M. (2000). *J. Biol. Chem.* **275**, 34609–34618.
- Li, B., Pai, R., Aiello, D., Di, M., Barnes, M. H., Peet, N. P., Bowlin, T. L. & Moir, D. T. (2013). *Bioorg. Med. Chem. Lett.* **23**, 3481–3486.
- Lin, N. S.-C. & Pratt, D. (1972). *J. Mol. Biol.* **72**, 37–49.
- Linker, S. M., Magarkar, A., Köfinger, J., Hummer, G. & Seeliger, D. (2019). *J. Chem. Theory Comput.* **15**, 4974–4981.
- Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. (2001). *Adv. Drug Deliv. Rev.* **46**, 3–26.
- Liu, B., Eliason, W. K. & Steitz, T. A. (2013). *Nat. Commun.* **4**, 2495.
- Liu, B., Lin, J. & Steitz, T. A. (2013). *Structure.* **21**, 658–664.
- Lo, M.-C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M. & Ellestad, G. (2004). *Anal. Biochem.* **332**, 153–159.
- Lonsdale, R. & Ward, R. A. (2018). *Chem. Soc. Rev.* **47**, 3816–3830.
- Lu, D., Bernstein, D. a, Satyshur, K. a & Keck, J. L. (2010). *Proc. Natl. Acad. Sci. U. S. A.* **107**, 633–638.
- Maciąg, M., Kochanowska, M., Łyżeń, R., Węgrzyn, G. & Szalewska-Pałasz, A. (2010). *Plasmid.* **63**, 61–67.
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E. & Simmerling, C. (2015). *J. Chem. Theory Comput.* **11**, 3696–3713.
- Makowska-Grzyska, M. & Kaguni, J. M. (2010). *Mol. Cell.* **37**, 90–101.
- Marceau, A. H., Bahng, S., Massoni, S. C., George, N. P., Sandler, S. J., Marians, K. J. & Keck, J. L. (2011). *EMBO J.* **30**, 4236–4247.
- Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kaptein, R., Sykes, B. D., Wright, P. E. & Wüthrich, K. (1998). *J. Biomol. NMR.* **12**, 1–23.
- Mattos, C. & Ringe, D. (1996). *Nat. Biotechnol.* **14**, 595–599.
- Mayer, M. & Meyer, B. (1999). *Angew. Chemie Int. Ed.* **38**, 1784–1788.

- McGlynn, P. & Guy, C. P. (2008). *J. Mol. Biol.* **381**, 249–255.
- McGovern, S. L., Helfand, B. T., Feng, B. & Shoichet, B. K. (2003). *J. Med. Chem.* **46**, 4265–4272.
- McHenry, C. & Kornberg, A. (1977). *J. Biol. Chem.* **252**, 6478–6484.
- McHenry, C. S. (1982). *J. Biol. Chem.* **257**, 2657–2663.
- Meiboom, S. & Gill, D. (1958). *Rev. Sci. Instrum.* **29**, 688–691.
- Mendes, F., Groessl, M., Nazarov, A. A., Tsybin, Y. O., Sava, G., Santos, I., Dyson, P. J. & Casini, A. (2011). *J. Med. Chem.* **54**, 2196–2206.
- Miller, R. M., Paavilainen, V. O., Krishnan, S., Serafimova, I. M. & Taunton, J. (2013). *J. Am. Chem. Soc.* **135**, 5298–5301.
- Miranker, A. & Karplus, M. (1991). *Proteins.* **11**, 29–34.
- Miyaki, M., Murata, I., Osabe, M. & Ono, T. (1977). *Biochem. Biophys. Res. Commun.* **77**, 854–860.
- Mulcair, M. D., Schaeffer, P. M., Oakley, A. J., Cross, H. F., Neylon, C., Hill, T. M. & Dixon, N. E. (2006). *Cell.* **125**, 1309–1319.
- Murale, D. P., Hong, S. C., Haque, M. M. & Lee, J.-S. (2016). *Proteome Sci.* **15**, 14.
- Murray, C. W., Newell, D. R. & Angibaud, P. (2019). *Medchemcomm.* **10**, 1509–1511.
- Naue, N., Beerbaum, M., Bogutzki, A., Schmieder, P. & Curth, U. (2013). *Nucleic Acids Res.* **41**, 4507–4517.
- O’Neill, J. (2016). Tackling Drug-Resistant Infections Globally: FINAL REPORT AND RECOMMENDATIONS.
- O’Reilly, M., Cleasby, A., Davies, T. G., Hall, R. J., Ludlow, R. F., Murray, C. W., Tisi, D. & Jhoti, H. (2019). *Drug Discov. Today.* **24**, 1081–1086.
- Oakley, A. J., Loscha, K. V., Schaeffer, P. M., Liepinsh, E., Pintacuda, G., Wilce, M. C. J., Otting, G. & Dixon, N. E. (2005). *J. Biol. Chem.* **280**, 11495–11504.
- Oakley, A. J., Prosselkov, P., Wijffels, G., Beck, J. L., Wilce, M. C. J. & Dixon, N. E. (2003). *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59**, 1192–1199.

- Painter, R. E., Adam, G. C., Arocho, M., DiNunzio, E., Donald, R. G. K., Dorso, K., Genilloud, O., Gill, C., Goetz, M., Hairston, N. N., Murgolo, N., Nare, B., Olsen, D. B., Powles, M., Racine, F., Su, J., Vicente, F., Wisniewski, D., Xiao, L., Hammond, M. & Young, K. (2015). *Chem. Biol.* **22**, 1362–1373.
- Pan, H. & Wigley, D. B. (2000). *Structure.* **8**, 231–239.
- Pandey, P., Verma, V., Gautam, G., Kumari, N., Dhar, S. K. & Gourinath, S. (2017). *FEBS Lett.* **591**, 2311–2322.
- Pansegrau, W. & Lanka, E. (1992). *Nucleic Acids Res.* **20**, 4931–4931.
- Parker, C. G., Galmozzi, A., Wang, Y., Correia, B. E., Sasaki, K., Joslyn, C. M., Kim, A. S., Cavallaro, C. L., Lawrence, R. M., Johnson, S. R., Narvaiza, I., Saez, E. & Cravatt, B. F. (2017). *Cell.* **168**, 527–541 e29.
- Parker, C. G. & Pratt, M. R. (2020). *Cell.* **180**, 605–632.
- Patterson, D. M., Nazarova, L. A. & Prescher, J. A. (2014). *ACS Chem. Biol.* **9**, 592–605.
- Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. (2007). *Nat. Rev. Drug Discov.* **6**, 29–40.
- Petros, A. M., Dinges, J., Augeri, D. J., Baumeister, S. A., Betebenner, D. A., Bures, M. G., Elmore, S. W., Hajduk, P. J., Joseph, M. K., Landis, S. K., Nettesheim, D. G., Rosenberg, S. H., Shen, W., Thomas, S., Wang, X., Zanze, I., Zhang, H. & Fesik, S. W. (2006). *J. Med. Chem.* **49**, 656–663.
- Pinheiro, S., Soteras, I., Gelpí, J. L., Dehez, F., Chipot, C., Luque, F. J. & Curutchet, C. (2017). *Phys. Chem. Chem. Phys.* **19**, 9849–9861.
- Pratt, D. & Erdahl, W. S. (1968). *J. Mol. Biol.* **37**, 181–200.
- Presolski, S. I., Hong, V. P. & Finn, M. G. (2011). *Curr. Protoc. Chem. Biol.* **3**, 153–162.
- Rannou, O., Le Chatelier, E., Larson, M. A., Nouri, H., Dalmais, B., Laughton, C., Janniere, L. & Soultanas, P. (2013). *Nucleic Acids Res.* **41**, 5303–5320.
- Renaud, J.-P., Chung, C., Danielson, U. H., Egner, U., Hennig, M., Hubbard, R. E. & Nar, H. (2016). *Nat. Rev. Drug Discov.* **15**, 679–698.

- Roe, D. R. & Cheatham, T. E. (2013). *J. Chem. Theory Comput.* **9**, 3084–3095.
- Roope, L. S. J., Smith, R. D., Pouwels, K. B., Buchanan, J., Abel, L., Eibich, P., Butler, C. C., Tan, P. S., Walker, A. S., Robotham, J. V. & Wordsworth, S. (2019). *Science*. **364**, eaau4679.
- Roychowdhury, A., Szymanski, M. R., Jezewska, M. J. & Bujalowski, W. (2009). *Biochemistry*. **48**, 6730–6746.
- Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. . (1977). *J. Comput. Phys.* **23**, 327–341.
- Rymer, R. U., Solorio, F. A., Tehranchi, A. K., Chu, C., Corn, J. E., Keck, J. L., Wang, J. D. & Berger, J. M. (2012). *Structure*. **20**, 1478–1489.
- Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. (2013). *J. Chem. Theory Comput.* **9**, 3878–3888.
- San Martin, M. C., Stamford, N. P. J., Dammerova, N., Dixon, N. E. & Carazo, J. M. (1995). *J. Struct. Biol.* **114**, 167–176.
- Sanders, G. M., Dallmann, H. G. & McHenry, C. S. (2010). *Mol. Cell*. **37**, 273–281.
- Sandler, S. J., Samra, H. S. & Clark, A. J. (1996). *Genetics*. **143**, 5–13.
- Sauer, W. H. B. & Schwarz, M. K. (2003). *J. Chem. Inf. Comput. Sci.* **43**, 987–1003.
- Savvides, S. N., Raghunathan, S., Fütterer, K., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2004). *Protein Sci.* **13**, 1942–1947.
- Scarpino, A., Ferenczy, G. G. & Keserü, G. M. (2018). *J. Chem. Inf. Model.* **58**, 1441–1458.
- Schanda, P., Kupče, Ě. & Brutscher, B. (2005). *J. Biomol. NMR*. **33**, 199–211.
- Schenone, M., Dančik, V., Wagner, B. K. & Clemons, P. A. (2013). *Nat. Chem. Biol.* **9**, 232–240.
- Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sørensen, O. W. & Griesinger, C. (1994). *J. Biomol. NMR*. **4**, 301–306.
- Schmitz, C. & Bonvin, A. M. J. J. (2011). *J. Biomol. NMR*. **50**, 263–266.

- Seitz, H., Weigel, C. & Messer, W. (2000). *Mol. Microbiol.* **37**, 1270–1279.
- Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M. & Keck, J. L. (2008). *Crit. Rev. Biochem. Mol. Biol.* **43**, 289–318.
- Shi, Y. & von Itzstein, M. (2019). *Molecules.* **24**, 2838.
- Shimizu, M., Noguchi, Y., Sakiyama, Y., Kawakami, H., Katayama, T. & Takada, S. (2016). *Proc. Natl. Acad. Sci.* **113**, E8021–E8030.
- Shishmarev, D., Wang, Y., Mason, C. E., Su, X.-C., Oakley, A. J., Graham, B., Huber, T., Dixon, N. E. & Otting, G. (2014). *Nucleic Acids Res.* **42**, 2750–2757.
- Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1996). *Science.* **274**, 1531–1534.
- Silver, L. L. (2013). *Antibiotics*, Vol. pp. 23–52. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.
- Simonetta, K. R., Kazmirski, S. L., Goedken, E. R., Cantor, A. J., Kelch, B. A., McNally, R., Seyedin, S. N., Makino, D. L., O'Donnell, M. & Kuriyan, J. (2009). *Cell.* **137**, 659–671.
- Śledź, P. & Caflisch, A. (2018). *Curr. Opin. Struct. Biol.* **48**, 93–102.
- Slee, A. M., Wuonola, M. A., McRipley, R. J., Zajac, I., Zawada, M. J., Bartholomew, P. T., Gregory, W. A. & Forbes, M. (1987). *Antimicrob. Agents Chemother.* **31**, 1791–1797.
- Souers, A. J., Levenson, J. D., Boghaert, E. R., Ackler, S. L., Catron, N. D., Chen, J., Dayton, B. D., Ding, H., Enschede, S. H., Fairbrother, W. J., Huang, D. C., Hymowitz, S. G., Jin, S., Khaw, S. L., Kovar, P. J., Lam, L. T., Lee, J., Maecker, H. L., Marsh, K. C., Mason, K. D., Mitten, M. J., Nimmer, P. M., Oleksijew, A., Park, C. H., Park, C. M., Phillips, D. C., Roberts, A. W., Sampath, D., Seymour, J. F., Smith, M. L., Sullivan, G. M., Tahir, S. K., Tse, C., Wendt, M. D., Xiao, Y., Xue, J. C., Zhang, H., Humerickhouse, R. A., Rosenberg, S. H. & Elmore, S. W. (2013). *Nat. Med.* **19**, 202–208.
- Soylu, I. & Marino, S. M. (2017). *Bioinformatics.* **33**, 2395–2396.
- Srivatsan, A. & Wang, J. D. (2008). *Curr. Opin. Microbiol.* **11**, 100–105.

- Stano, N. M., Chen, J. & McHenry, C. S. (2006). *Nat. Struct. Mol. Biol.* **13**, 458–459.
- Su, X.-C., Schaeffer, P. M., Loscha, K. V., Gan, P. H. P., Dixon, N. E. & Otting, G. (2006). *FEBS J.* **273**, 4997–5009.
- Swart, J. R. & Griep, M. A. (1995). *Biochemistry.* **34**, 16097–16106.
- Syson, K., Thirlway, J., Hounslow, A. M., Soutanas, P. & Waltho, J. P. (2005). *Structure.* **13**, 609–616.
- Taft-Benz, S. A. & Schaaper, R. M. (2004). *J. Bacteriol.* **186**, 2774–2780.
- Tautermann, C. S., Binder, F., Büttner, F. H., Eickmeier, C., Fiegen, D., Gross, U., Grundl, M. A., Heilker, R., Hobson, S., Hoerer, S., Luippold, A., Mack, V., Montel, F., Peters, S., Bhattacharya, S., Vaidehi, N., Schnapp, G., Thamm, S. & Zeeb, M. (2019). *J. Med. Chem.* **62**, 306–316.
- Tommasi, R., Brown, D. G., Walkup, G. K., Manchester, J. I. & Miller, A. A. (2015). *Nat. Rev. Drug Discov.* **14**, 529–542.
- Tougu, K. & Mariani, K. J. (1996). *J. Biol. Chem.* **271**, 21391–21397.
- Tougu, K., Peng, H. & Mariani, K. J. (1994). *J. Biol. Chem.* **269**, 4675–4682.
- Tsuchihashi, Z. & Kornberg, A. (1990). *Proc. Natl. Acad. Sci.* **87**, 2516–2520.
- Tyrell, R. M. (1973). *Photochem. Photobiol.* **17**, 69–73.
- Velten, M., McGovern, S., Marsin, S., Ehrlich, S. D., Noirot, P. & Polard, P. (2003). *Mol. Cell.* **11**, 1009–1020.
- Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. (2005). *Proteins Struct. Funct. Bioinforma.* **59**, 687–696.
- Wang, G., Klein, M. G., Tokonzaba, E., Zhang, Y., Holden, L. G. & Chen, X. S. (2008). *Nat. Struct. Mol. Biol.* **15**, 94–100.
- Wang, Y., Dix, M. M., Bianco, G., Remsberg, J. R., Lee, H.-Y., Kalocsay, M., Gygi, S. P., Forli, S., Vite, G., Lawrence, R. M., Parker, C. G. & Cravatt, B. F. (2019). *Nat. Chem.* **11**, 1113–1123.
- Xu, L. & Mariani, K. J. (2000). *J. Biol. Chem.* **275**, 8196–8205.

- Yao, N., Turner, J., Kelman, Z., Stukenberg, P. T., Dean, F., Shechter, D., Pan, Z.-Q., Hurwitz, J. & O'Donnell, M. (1996). *Genes to Cells*. **1**, 101–113.
- Yeeles, J. T. P., Poli, J., Marians, K. J. & Pasero, P. (2013). *Cold Spring Harb. Perspect. Biol.* **5**, a012815.
- Yin, Z., Kelso, M. J., Beck, J. L. & Oakley, A. J. (2013). *J. Med. Chem.* **56**, 8665–8673.
- Yin, Z., Whittell, L. R., Wang, Y., Jergic, S., Liu, M., Harry, E. J., Dixon, N. E., Beck, J. L., Kelso, M. J. & Oakley, A. J. (2014). *J. Med. Chem.* **57**, 2799–2806.
- Yoda, K. & Okazaki, T. (1991). *MGG Mol. Gen. Genet.* **227**, 1–8.
- Yuzhakov, A., Kelman, Z. & O'donnell, M. (1999). *Cell*. **96**, 153–163.
- Zhang, C., Ibrahim, P. N., Zhang, J., Burton, E. A., Habets, G., Zhang, Y., Powell, B., West, B. L., Matusow, B., Tsang, G., Shellooe, R., Carias, H., Nguyen, H., Marimuthu, A., Zhang, K. Y. J., Oh, A., Bremer, R., Hurt, C. R., Artis, D. R., Wu, G., Nespi, M., Spevak, W., Lin, P., Nolop, K., Hirth, P., Tesch, G. H. & Bollag, G. (2013). *Proc. Natl. Acad. Sci.* **110**, 5689–5694.
- Zhang, X., Gong, Z., Li, J. & Lu, T. (2015). *J. Chem. Inf. Model.* **55**, 2138–2153.
- Zhou, Y., Luo, H., Liu, Z., Yang, M., Pang, X., Sun, F. & Wang, G. (2017). *Sci. Rep.* **7**, 659.