Exploration of fragment-derived modulators of glycoside hydrolases

Eleni Makraki

Doctor of Philosophy

University of York

Chemistry

Jan 2020

Abstract

Previous work at York demonstrated that fragment molecules can increase the activity of the glycoside hydrolase, BtGH84. The initial aim of this project was to use fragment-based discovery methods to identify activators of several enzymes used in cellulose degradation where low activity is one of the limiting steps in the industrial process. This was successful for one enzyme, the fungal glycoside hydrolase, *Tr*Bgl2. The characterisation of the mechanism of activation for this enzyme is the main focus of this thesis.

A fragment screen of a library of 560 commercially available fragments using a kinetic assay identified a small molecule activator of *Tr*Bgl2. An analogue by catalogue approach and detailed kinetic analysis identified compounds that behaved as nonessential activators with up to a 2-fold increase in maximum activation. The compounds did not activate the related bacterial glycoside hydrolase *Cc*BglA demonstrating specificity. Interestingly, an analogue of the initial fragment inhibits both *Tr*Bgl2 and *Cc*BglA, apparently through a mixed-model mechanism.

Although it was not possible to determine crystal structures of activator binding to 55 kDa *Tr*Bgl2, solution NMR experiments demonstrated a specific binding site for the activator. A partial assignment of the NMR spectrum gave the identity of the amino acids at this site, allowing a model for *Tr*Bgl2 activation to be built. The activator binds at the entrance of the substrate binding site, stabilizing the enzyme-substrate complex.

List of contents

List of figures

List of tables

List of abbreviations

AC_{50} Compound concentration causing 50% activation

ADP   Adenosine diphosphate

A_{MAX} Maximum activation

AMPK   5' AMP-activated protein kinase

ATP   Adenosine triphosphate

BG   β-glucosidase

BRAF   v-raf murine sarcoma viral oncogene homolog B1

CARM1   Coactivator-associated arginine methyltransferase 1

CAZY   Carbohydrate-active enzymes database

CBH   Cellobiohydrolases

CcBglA   β-glucosidase from *Clostridium cellulovorans*

CgoX   Coproporphyrinogen oxidase

CR   Calorie restriction

CSPs   Chemical shift perturbations

CTR   C-terminal regulatory segment

Da   Dalton

DAPK1   Death associated protein kinase 1

DMSO   Dimethyl sulfoxide

DRAK2   Death-associated protein-related apoptotic kinase-2

DNA   Deoxyribonucleic acid

DSS   Sodium trimethylsilylpropanesulfonate

DTT   Dithiothreitol

D_2O Deuterium Oxide (heavy water)

EDTA   Ethylenediaminetetraacetic acid

EG   Endoglucanases

FBLD   Fragment-based ligand discovery

Fdl   Fluor-de-Lys

GK   Glucokinase

GDH   Glutamate dehydrogenase

GKA   Glucokinase activator

GKRP   Glucokinase regulatory protein

GlcNAc   N-acetylglucosamine

G6PD   Glucose-6-phosphate dehydrogenase

HCC   Human hepatocellular carcinoma

His   Histidine

HM   Hydrophobic motif

HSQC   Heteronuclear single quantum coherence

HTS	High-throughput screening
IC_{50}	Compound concentration causing 50% inhibition
IDE	Insulin-degrading enzyme
INEPT	Insensitive nuclei enhanced by polarization transfer
IPTG	Isopropyl β-D-1 thiogalactopyranoside
ICT	Isothermal titration calorimetry
k_{cat}	Catalytic rate constant
K_D	Dissociation constant
kDa	kilo-Dalton
K_m	Michaelis constant
LB	Lysogeny Broth
LC	Lignocellulosic
LOX	Lipoxygenase
MALS	Multi-Angle Light Scattering
MUG	4-methylumbelliferyl-β-D-glucopyranoside
NAD	Nicotinamide adenine dinucleotide
nM	Nanomolar
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect

NOESY	Nuclear overhauser effect spectroscopy
PAGE	Polyacrylamide-based discontinuous gel
PCL	Phospholipase C
PDPK1	Phosphoinositide- dependent protein kinase 1
PIF	PDPK1-interacting fragment
*p*NPG	*p*-Nitrophenyl-β-D-glucopyranoside
PPA	Porcine pancreatic alpha-amylase
RMSD	Root-mean-square deviation
SARs	Structure-activity relationship
SBD	Sirtuin binding domain
SDS PAGE	Sodium dodecyl sulfate
SEC	Size Exclusion Chromatography
SIRT	Sirtuins
STAC	Sirtuin-activating compounds
STD	Saturation transfer difference
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TrBgl2	β-glucosidase from *Trichoderma reesei*
TRIS	Tris(hydroxymethyl)aminomethane
TROSY	Transverse relaxation-optimised spectroscopy

TSA   Thermal shift assay

V_{max} Maximal enzyme velocity

WaterLOGSY   Water-Ligand Observed via Gradient Spectroscopy

µg   Microgram

µL   Microlitre

µM   Micromolar

Acknowledgments

First, I would like to thank my supervisor Professor Rod Hubbard for giving me the opportunity, the guidance and the support to do my PhD in YSBL. Many thanks go to John Darby for his invaluable support and continual practical assistance in the lab. I also would like to thank Alex Heyam for teaching me the basics of NMR and Jim Brannigan during his time in our group.

Many of the results obtained during this work would not have been possible without the assistance of our collaborators Gregg Siegal, Marta Carneiro and Eiso AB from ZoBio (Leiden, The Netherlands).

I want to special thank Sandra Greive, Paul Bond, Dan Wright and Imogen Breen for proofreading of my thesis.

There many more people I should mention here who helped and supported me during my PhD. My thanks go to all members of YSBL both past and present, for their help and kind friendship. I have enjoyed my time as part of YSBL.

An especially warm thanks to Hanna for being a good friend. Hanna has made my weekends in York so much fun.

Finally, I would like to thank my family, my friends and my partner for their love and support.

Declaration

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

Signature:

Name: Eleni Makraki

Date: 14/01/2020

Chapter 1: Introduction

1.1. Enzyme activation

The binding of a small organic molecule can affect an enzyme in two possible ways: it can either decrease (inhibit) or increase (activate) the activity of the enzyme. There are numerous examples where binding has evolved to inhibit activity in biological systems –such as direct inhibition by binding to the active site (for example Vemurafenib for BRAF kinase) [3] or indirect inhibition by disrupting a protein-protein interaction (for example Venetoclax for B-cell lymphoma 2 [4-6]). There are many enzymes whose inhibition can generate a therapeutic effect. Although there are some examples of inhibition through compounds that bind to an allosteric site (for example the allosteric inhibitor of BCR-ABL1 [7-8]) most drug discovery projects aim to generate inhibitors that bind directly to active sites and inhibit enzyme activity, typically with small organic molecules with a molecular weight (MW) of less than 500 Da.

The phenomenon of activation through non-covalent binding of small molecules is well established for receptors –that is proteins which increase biological activity through binding of a ligand, known as agonist action. Examples include the nuclear hormone receptors (such as the estrogen receptor [9]) and G-protein coupled receptors (such as the α-adrenergic receptor [10]) where binding of the ligand induces an active conformation.
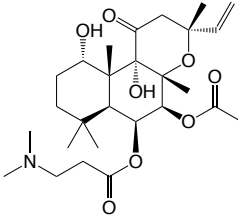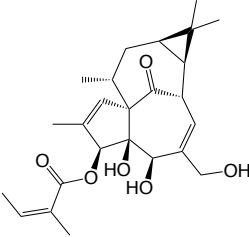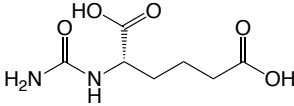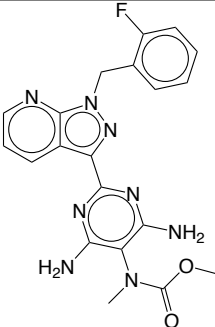
Covalent modification of an enzyme is a widely observed mechanism for modulating the biological activity of an enzyme, affecting cellular function.

The most widely described enzyme modification is the phosphorylation of amino acids such as serine, threonine or tyrosine at specific sites in an enzyme by selective protein kinases. This modification can alter the activity of an enzyme in various ways, for example by changing localization such as in the DAP kinase-related apoptosis-inducing kinase 2 (DRAK2) which changes between the nucleus and cytoplasm [11], by releasing inhibition by a domain such as in the phospholipase C (PLC)-γ isozyme [12] or by changing an enzyme's conformation to an active one such as in the MAP (mitogen-activated protein) kinase ERK2 (extracellular signal-regulated protein kinase) [13]. Enzyme modifications also include covalent modifications by carbohydrates (such as widespread addition of GlcNAc sugars to many intracellular proteins), which affect the overall activity of a cell.

Although, there are relatively few examples of naturally occurring small molecules that activate an enzyme through binding non-covalently, there have been some well characterized non-natural small molecules that lead to remarkable increase in enzyme activity. Enzyme activators usually act by binding to an allosteric pocket. An allosteric site is a site other than the orthosteric site of a protein where effectors can bind, often causing the enzyme to undergo a conformational change involving protein dynamics that may alter its catalytic or binding properties. It is thought that only a small increase in activation (as low as 10%) may be sufficient to elicit an effective therapeutic response whereas enzyme inhibitors usually require at least 90% inhibition to completely block a phenotype [14]. In addition,

allosteric sites are less likely to be conserved, increasing the opportunity for selectivity. Although enzyme inhibition is the most widely used approach in drug discovery, enzyme activation may provide new approaches for disease related studies. For example, there are some widely known marketed drugs that act as activators (Table 1.1).

Table 1.1. *Widely known marketed drugs as activators*

Drug	Target	Structure
Colforsin	Adenylate cyclase	
Ingenol	Protein kinase C	
Carbaglu	Carbamoyl phosphate synthetase	
Adempas	Guanylate cyclase	

1.2. Mechanisms of enzyme activation

A review by Zorn *et al.* [15] defined four general mechanisms (A1, A2, B1 and B2) of activation describing how the activity of an enzyme can be increased by a small molecule to a site other than the primary or orthosteric binding site. These are summarized in Figure 1.1. The classification proposed by Zorn *et al*. provides a general framework for classifying modes of activation. The main focus of the studies reported in this thesis is the A1 mechanism, where in a small molecule binds directly to the catalytic domain of the enzyme, thereby affecting some aspect of the enzyme kinetics.
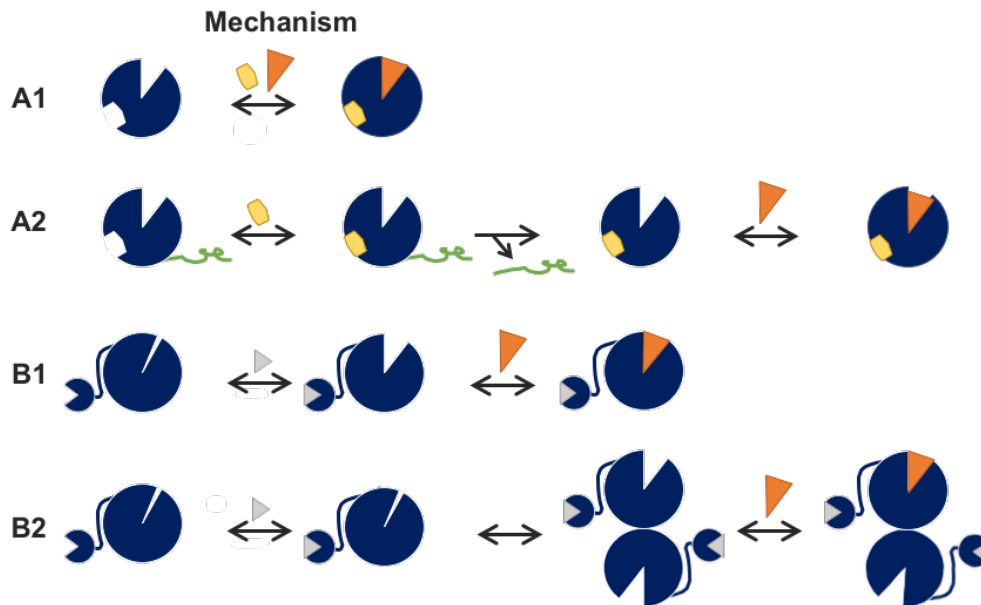
Figure 1.1. *Mechanisms of small molecule enzyme (orange square) activation [15]. The type A1 mechanism involves a direct binding interaction of the small-molecule activator (yellow) with an allosteric site on the enzyme's catalytic domain stabilizing an active conformation. The type A2 mechanism results from an irreversible post-translational modification. One example is proteolysis induced by binding of a small molecule to the catalytic domain of a protease, as in the case of MART$_{Vc}$ toxin activator. In a type B1 mechanism, the activator (yellow) binds to a regulatory subunit resulting in a conformational change in the catalytic domain. One example is AMP-activated protein kinase (AMPK). Type B2 mechanisms are where additional oligomerisation of the catalytic domain generates the active enzyme either by binding of small molecules to an allosteric site on the catalytic domain or by interaction of small molecules with a regulatory subunit. One example is RNAse L activators. The substrate is shown as an orange triangle.*

1.3. Examples of enzyme activation by small molecules

A search of the literature using the keywords "small molecules", "activators" and "enzyme activation", identified reports of small molecule activators of enzymes. These fall into two main classes depending on whether or not there is clear evidence that the ligands bind to the activated target. Table 1.2 summarises the papers published since 2012 where there is no clear evidence for binding of the ligands to the activated target.

Table 1.2. *Recent examples of enzyme activation through unidentified mechanism.*

Target	How activators found	Refs
Coactivator-associated arginine methyltransferase 1 (CARM1)	HTS	[16]
Insulin-degrading enzyme (IDE)	HTS	[17]
Porcine pancreatic alpha-amylase (PPA)	HTS of compounds with similar structure to natural metabolites	[18]
Glutamate dehydrogenase (GDH)	HTS	[19]

Target	How activators found	Refs
15-lipoxygenase (15-LOX)	Virtual screening against predicted allosteric sites	[20]
Coproporphyrinogen oxidase (CgoX)	HTS	[21-22]
Death associated protein kinase 1 (DAPK1)	HTS	[23]
Glucose-6-phosphate dehydrogenase (G6PD)	HTS	[24]

The second class of examples which is the focus of this thesis are those where there is evidence that the activation is through an A1 type mechanism –that is wherein the small molecule binds to an allosteric site on the catalytic domain of the enzyme to achieve activation. Some examples demonstrating this mechanism are discussed in detail below.

1.3.1. Glucokinase (GK)

GK is the mammalian enzyme that catalyzes the ATP-dependent phosphorylation of glucose to glucose-6-phosphate (G-6-P) [25]. In the liver, GK is retained in the nucleus by the glucokinase regulatory protein (GKRP) at low glucose levels whilst on increase of glucose concentration it is released into the cytoplasm where it promotes the conversion of glucose to glycogen [26-27]. In β-cells GK acts as a "glucose sensor" [28-30] and

inactivating mutations are associated with maturity-onset of type 2 diabetes while activating mutations are linked to improved glucose tolerance [31]. This suggests that activation of GK is a promising target for the treatment of type 2 diabetes.

A 120,000-compound library was screened for compounds that reverse inhibition by GKRP. This identified a glucokinase activator (GKA) that binds to an allosteric site of GK close to the active site without affecting GKRP. This led to the synthesis of a chemically optimised GKA (**1**) (Figure 1.2), which decreases the AC_{50} (the concentration of compound required to increase enzyme activity by 50%) of GK for glucose by approximately 4-fold and increases its V_{max} by 1.5-fold in a dose-dependent manner [29][32].
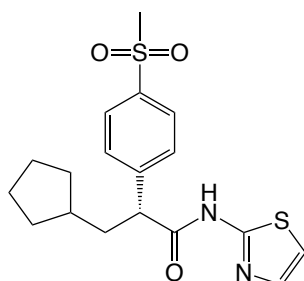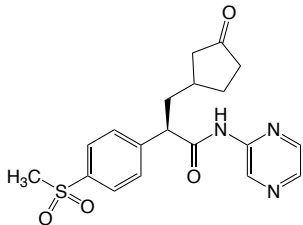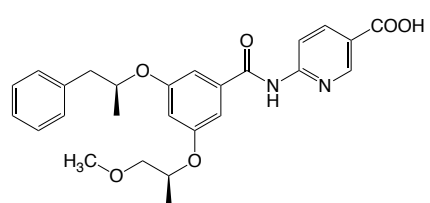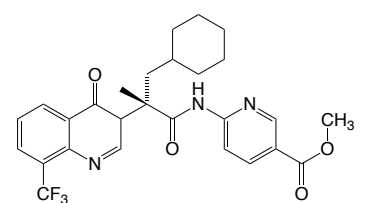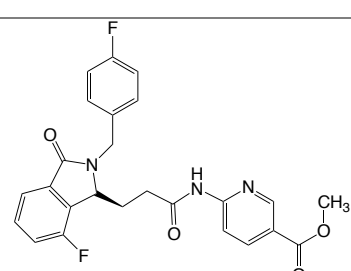
Figure 1.2. *Chemical structure of GK activator* ***1***.

In vivo studies in diabetic rodent models have shown that GKAs reduce blood glucose levels, increase plasma insulin and improve glucose tolerance [29]. Following this, a number of structurally unique GKAs (Table 1.3) [33-34] and patents have since been reported [35-48]. Some of these GKAs are in clinical trials in human patients with type 2 diabetes [49-50].

Table 1.3. *Main classes and chemical structures of GKAs.*

Class	Compound	Company
Carbon-centered		Hoffman-La Roche
Benzene- and pyridine-centred		Astra-Zeneca
Amino acid-based		Takeda
Pyrrolone-based		Johnson & Johnson

The crystal structures of active (in complex with glucose and a novel GKA (**2**)) and inactive (unbound) forms of GK provide an insight into the mechanism of GKAs [30]. GK is a monomeric enzyme consisting of a large and a small domain. Its single active site is located on the connection region between the domains [30]. Figure 1.3 shows the large conformational change in GK triggered by glucose binding, which result in a large rotation of its small domain [51]. This conformational change slowly interconverts

the enzyme from an ''open'' low affinity glucose form to a ''closed'' high-affinity glucose active form [30], which exposes an allosteric pocket that is located around 20 Å away from the active site in the low affinity form (Figure 1.3.A) [51]. This interconversion is affected by glucose concentration and liver specific GKA [52]. Thus, GKAs appear to bind to this allosteric site, thereby stabilising the ''closed'' active form of GK. Interestingly, the structural analysis of GK revealed that the GK activating mutations are clustered at the same allosteric site where GKAs interact [29-30].

More recently, it has been reported that some GKAs are able to function in a glucose-independent manner by promoting an active enzyme conformation similar to the effects produced by glucose binding [53].
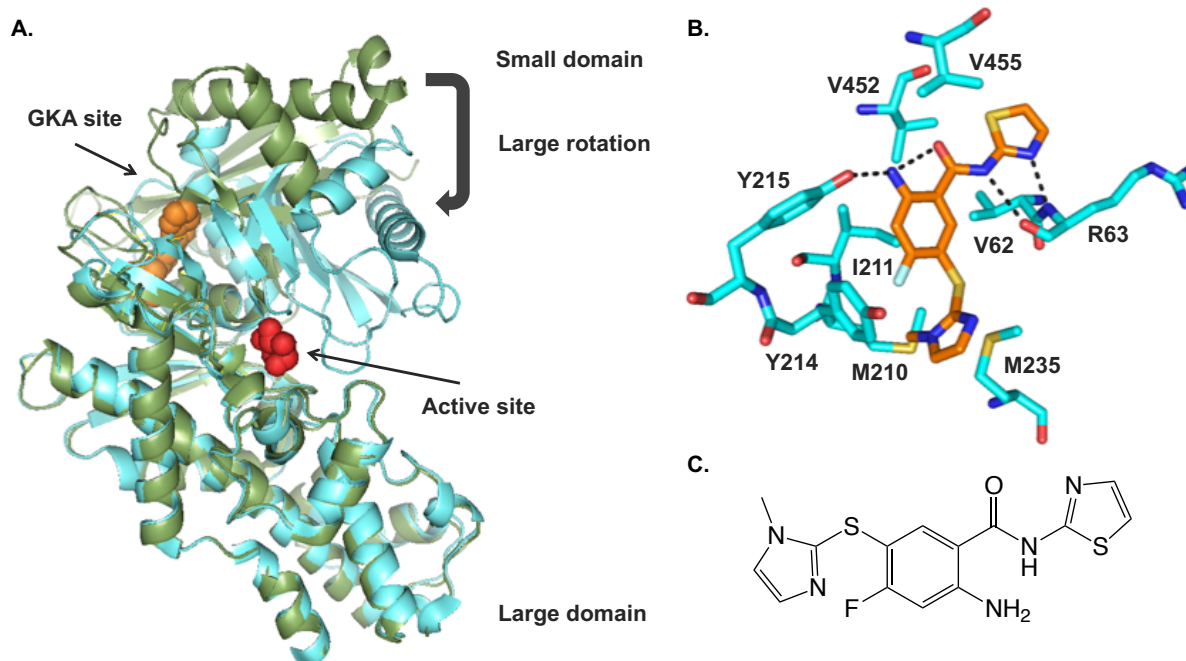
Figure 1.3. *A. Structural overlay of bound and unbound GK. The protein backbone is shown as ribbon diagram. GK bound to a novel GKA **2** and glucose (PDB ID 1V4S) is shown in cyan and the unbound GK (PDB ID 1V4T) is shown in green. Glucose and GKA **2** are shown as spheres in red and orange respectively. B. GKA allosteric site in closed form glucokinase. GKA forms hydrogen bonds with Arg63 (R63) and Tyr215 (Y215) and makes hydrophobic interactions with Val452 (V452) and Val455 (V455) in the small domain and with Val62 (V62), Met210 (M210), Ile211 (I211), Tyr214 (Y214), and Met235 (M235) in the large domain [30] (cyan stick model). Activator **2** is shown as lines in orange. Oxygen atoms are shown in red and nitrogen in blue. Hydrogen bonds are shown as black dashed lines. C. Chemical structure of **2**.*

1.3.2. Phosphoinositide-dependent protein kinase 1 (PDPK1)

Protein kinases catalyze the phosphorylation of specific sites on other

proteins to affect their activity, providing the most common mechanism for intracellular regulation. An example is the AGC protein kinase family where mutation that leads to deregulation has been linked to human diseases, such as diabetes and cancer [54-55]. AGC protein kinase family enzymes are activated by phosphorylation of a serine or threonine to stabilize the active conformation of the so-called activation loop within the kinase domain as well as phosphorylation of a residue in the hydrophobic motif (HM) Phe-Xaa-Xaa-Phe of the conserved region at the C-terminus of the catalytic domain. In some cases, the phosphorylated hydrophobic motif can function as a docking site that can recruit a HM-binding site in the catalytic domain of the protein kinase catalytic domain [56-57].

One member of the AGC protein kinase family is phosphoinositide-dependent protein kinase 1 (PDPK1). This enzyme phosphorylates the activation loop of other protein kinases from the AGC family, thereby activating them [56-58]. PDPK1 does not have a HM instead, it has a regulatory homologous pocket in its N-terminal lobe which is called the PIF (PDPK1-interacting fragment)-binding pocket, "PIF pocket". The PIF pocket recruits only phosphorylated HM of substrate protein kinases and allosterically activates itself through a process involving conformational change of its ATP binding site [58-60].

Engel *et al.* reported the first rational design of small compounds that mimic the phosphorylation-dependent docking interaction of the PIF-pocket with HM-containing substrates [61]. *In silico* screening of a 60,000-compound

library identified small-molecule compounds that possess features similar to those essential for the HM PIF-pocket interaction (such as Phe347 and Phe350) [61]. The hits that were selected from the computational screening were further evaluated *in vitro*.

Compound **3** (Figure 1.4) showed a 5-fold activation of PDPK1, with an AC_{50} of 34 µM [61]. This compound contains a negatively charged carboxylate that mimics the phosphate group of the phosphorylated HM, and two aromatic ring hydrophobic moieties that probably mimic the two phenylalanine residues on the HM of substrates [61]. The ability of **3** to activate PDPK1 relies on the specific interaction with the regulatory HM/PIF-pocket to induce the required conformational change [61].
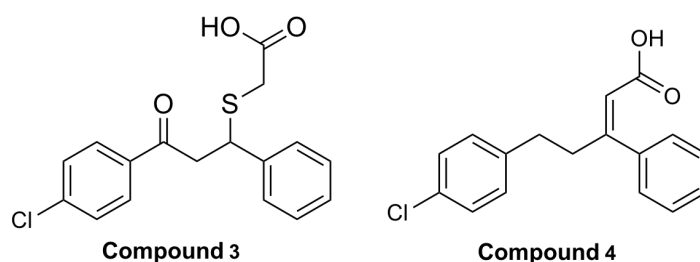
Figure 1.4. *Chemical structure of PDPK1 activators.*

Structure-activity relationships (SARs) guided further optimization of the initial compound **3** leading to improved compounds with AC_{50} in the range 2.8 to 41.3 µM and maximum activation (A_{max}) increased between 1.4 to 4.4-fold [62]. The AC_{50} values were consistent with the binding affinities that were derived from ITC experiments. Compound **4** was identified as the best compound with an AC_{50} value of 8 µM, showing a 4-fold increase in A_{max} as compared to the PDPK1 activity levels. The crystal structure of PDPK1 in complex with activator **4** and ATP gives an insight into the

mechanism of its activation [63] (Figure 1.5). The binding of the small-molecule activator results in modest structural changes local to the PIF-pocket producing local changes that lead to allosteric modifications in the ATP binding site and the activation loop (Figure 1.6) [63]. Furthermore, all the surface residues of the PIF-pocket appear to have decreased B-factors in the presence of activator **4**, indicating a dynamic rearrangement of the surface residues, thereby stabilising the bound activator [63].
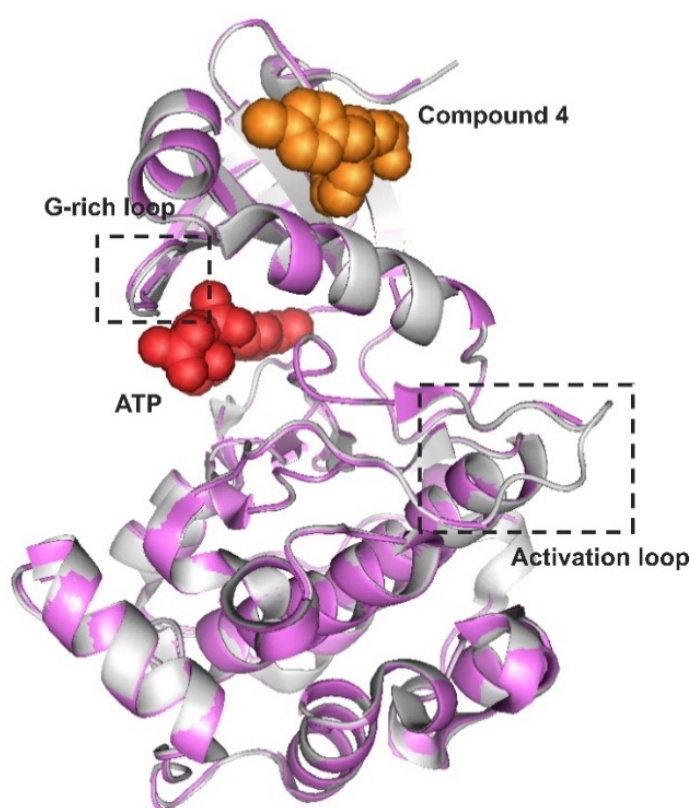
Figure 1.5. *Structural overlay of bound and unbound PDPK1. The protein backbone is shown as ribbon diagram. PDPK1 bound to **4** and ATP (PDB ID 3HRF) is shown in grey and **4**-unbound PDPK1 (PDB ID 3HRC) is shown in magenta. Compound **4** and ATP are shown as spheres in orange and red respectively.*
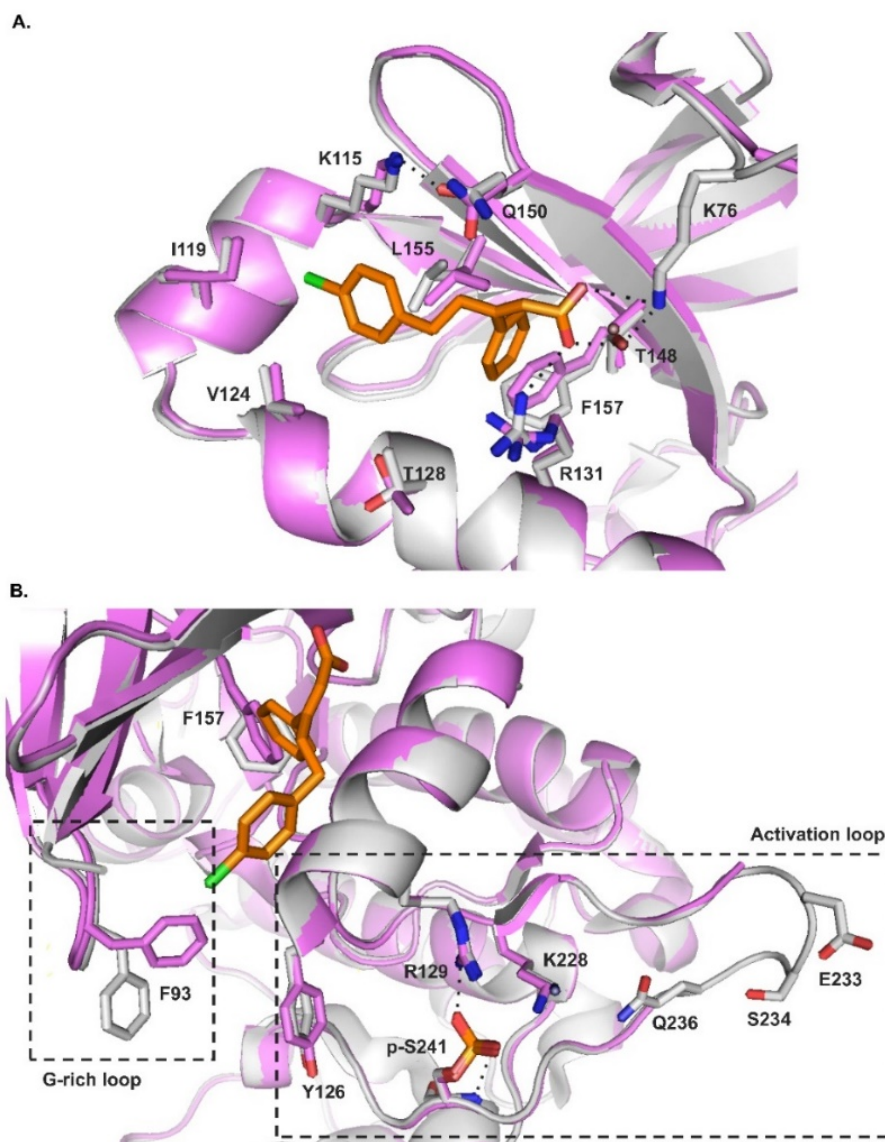
Figure 1.6. *Structural overlay of PDPK1 bound to **4** and ATP (PDB ID 3HRF) and PDK1 not bound to **4** (PDB ID 3HRC) is shown in grey and magenta respectively. A. Moderate conformational changes local to the PIF-pocket upon binding of **4**; Lys76 (K76) became ordered upon binding of **4**, Arg131 (R131) moved to interact with the carboxylate of **4** and Phe157 (F157) changed position to leave free space for **4** to bind. B. Allosteric conformational changes to the activation loop and glycine-rich loop upon binding of **4**. The activation loop became more ordered after **4** binding and Phe93 (F93) in the glycine-rich loop shifted. Compound **4** is shown as stick diagram in orange. Oxygen atoms are shown in red and nitrogen in blue. Hydrogen bonds are shown as black dashed lines.*

More recently, Wei *et al.* designed a novel binder compound **5** (Figure 1.7), through virtual docking of compounds into the PIF pocket of PDPK1.
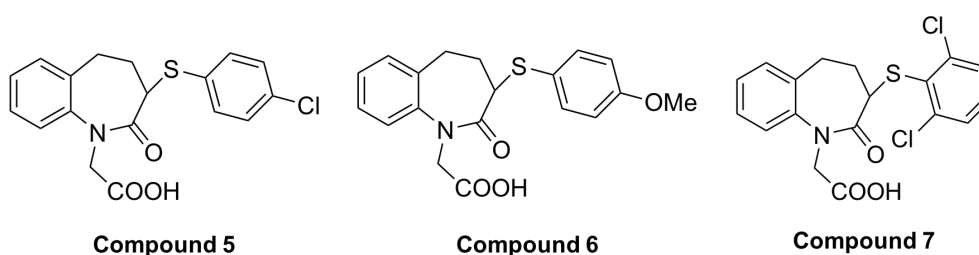
Figure 1.7. *Chemical structure of activator **5**, **6** and **7**.*

Molecular modeling showed that compound **5** binds to the PIF pocket through the hydrophobic interaction of its two aromatic groups with the PIF pocket and through a specific interaction between the tail substitute and the residues that form the phosphate-binding site of PDPK1. Compound **5** resulted in a 5-fold increase in binding affinity with an AC_{50} of 23 µM, compared to the previously reported compound **3** which was measured with an AC_{50} of 133 µM [64]. Compound **5** and two other benzodiazepine-2-ones (Compound **6** and **7**) were further tested in an *in vitro* enzymatic assay using a peptide substrate. Compounds **5**, **6** and **7** increased the activity of PDPK1 by 14-fold, 10-fold and 2-fold respectively at 25 µM. Compound **6** and **7** were measured with an AC_{50} of 53 and 45 µM respectively, which shows the absence of strong correlation between the activity levels of PDPK1 and the AC_{50} values. This suggests that small molecules interacting with the HM/PIF pocket may result in slightly different conformational changes which lead to differential activation of PDK1 [64].

1.3.3. Sirtuins (SIRT)

Sirtuins are nicotinamide adenine dinucleotide (NAD$^+$)-dependent deacetylases. Their primary biological function is to remove acetyl groups from modified lysine residues in histone proteins, thus releasing DNA for transcription. This regulation of transcription has been shown to influence aging in yeast; the *SIR2* gene is overexpressed in response to calorie restriction (CR) suppressing rDNA circle formation which results in slow ageing [65]. Mammals have seven sirtuin homologues (SIRT1-7). SIRT1, SIRT6 and SIRT7 are mainly located in the nucleus while SIRT2 is mainly cytosolic although nucleus functions have also been described. SIRT1 and SIRT6 are associated to ageing and stress responses. The most studied sirtuin, SIRT1, also links to CR while SIRT6 has been found to have low expression in many cancers. SIRT3, SIRT4 and SIRT5 are regulating mitochondrial proteins and processes. Sirtuins bind substrate polypeptide in a cleft between a Rossmann-fold and Zn^{2+}-binding domain allowing the substrate to react with the co-substrate NAD$^+$. This reaction releases nicotinamide and forms an alkylimidate intermediate which is then hydrolyzed [66].

SIRT1

Resveratrol (3,5, 4'-trihydroxystilbene), a polyphenolic plant metabolite, (Figure 1.8) was the most potent natural compound proposed as a SIRT1 activator that can mimic CR effects. Resveratrol was identified through a high-throughput screen (HTS) from a small molecule library that included

analogues of natural effectors such as ε-acetyl lysine, NAD$^+$, NAD$^+$ precursors, nucleotides and purinergic ligands [67]. A fluorimetric assay in which an acetyl lysine is linked to a fluorophore-labelled peptide substrate revealed that resveratrol decreases the Michaelis constant (K_m) of SIRT1 for an acetylated substrate 35-fold and NAD$^+$ over 5-fold without significant effect on V$_{max}$ [67].

Subsequent to the discovery of resveratrol, an HTS of a 290,000-compound library identified an imidazothiazole scaffold, chemically distinct from resveratrol, using a high-throuput in vitro fluorescence assay. This initial compound was optimized to give a second generation of SIRT1 activators such as compound **8**, **9**, and **10** (Figure 1.8) which provide better AC$_{50}$ values than resveratrol; resveratrol: AC$_{50}$ = 46.2 µM and A$_{max}$ = 201%, Compound **8**: AC$_{50}$ = 2.9 µM and A$_{max}$ = 447%, Compound **9**: AC$_{50}$ = 0.36 µM and A$_{max}$ = 296%, Compound **10**: AC$_{50}$ = 0.16 µM and A$_{max}$ = 781% [68]. None of the compounds were found to affect the V$_{max}$ or K_m for NAD$^+$ but all of them decreased the K_m for the acetylated peptide substrate. Isothermal titration calorimetry (ITC) showed that these compounds activate SIRT1 by binding to a SIRT1-peptide substrate complex. The binding of an acetylated peptide substrate seems to promote a more productive conformational change that results in the exposure of an allosteric site [68].

Resveratrol   **Compound 8**   **Compound 9**   **Compound 10**
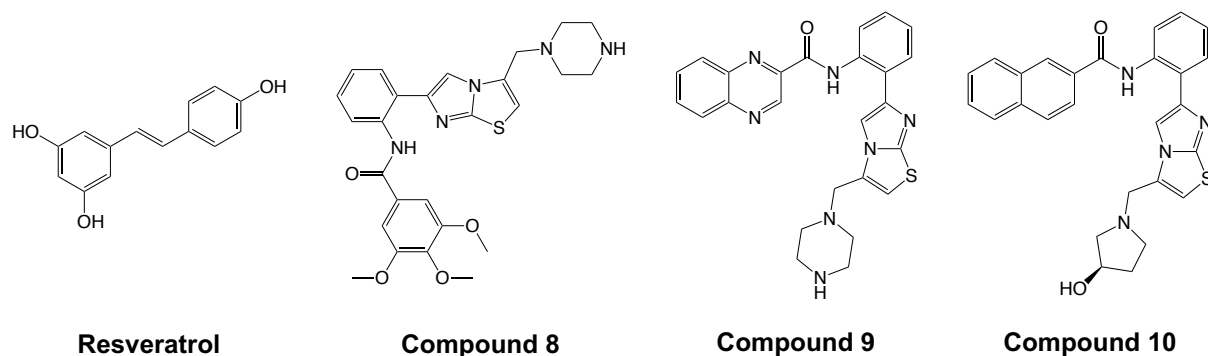
Figure 1.8. Chemical structure of SIRT1 activators.

However, a number of reports have claimed that the activation of SIRT1 seen for resveratrol and imidazothiazoles is an *in vitro* artefact due to the intrinsic fluorescence of the compounds interfering with the assay reagents [69-70]. Following this, new studies reported that if the fluorescent moieties on the substrates are replaced by hydrophobic residues, this results in SIRT1 activation [71-73], which suggests that the fluorophore-modified peptide substrate "Fluor-de-Lys" 1 (FdL1) possibly mimics natural hydrophobic amino acids that occur on native sirtuin substrates.

More recent work offers a better understanding of the molecular basis of SIRT1 activation. Dai *et al.* has designed an engineered human SIRT1 (Mini-hSIRT1) which includes the minimum structural elements for the activity of SIRT1: (1) the main catalytic domain, (2) the N-terminal STAC (Sirtuin-activating compounds) binding domain (termed SBD) that is located before the main catalytic domain in full-length hSIRT1 and (3) the remote human C-terminal regulatory segment (CTR), which enhances the stability and activity of the truncated hSIRT1 [20]. The crystal structure of Mini-hSIRT1 showed that the conserved catalytic domain of all sirtuins [74] comprise a

large Rossmann-fold lobe, a small zinc-binding lobe, the three-helical N-terminal SBD and the β-strands C-terminal CTR peptide [20]. The activator was further shown to interact with specific hydrophobic residues (Leu206, Thr209, Pro211, Pro212, Leu215, Thr219, Ile223, Ile227) that are located on the helix-turn-helix (H2-T-H3) motif within the SBD and form a shallow hydrophobic surface with a deeper off-center hydrophobic pocket (Figure 1.9) [20]. This explains the requirement reported in previous studies for specific hydrophobic motifs in allosteric activators for human SIRT1 [71-73]. After modeling, it was speculated that the N-terminal SBD domain swings through the flexible linker towards the active site resulting in a direct contact between activator and substrate [20].
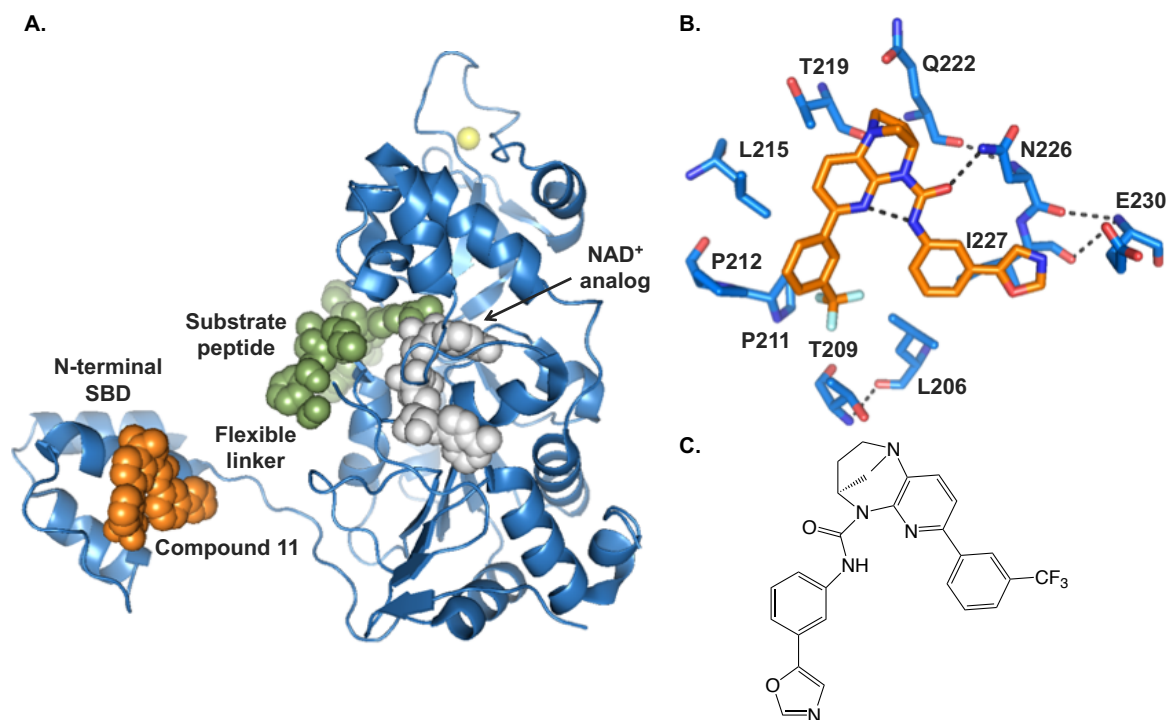
Figure 1.9. *Structure of Mini-hSIRT1 in complex with **11** and substrate peptide. A. Structure of Mini-hSIRT1 in complex with **11** and substrate peptide is shown as ribbon diagram (PDB ID 4ZZJ). Compound **11** is shown in orange and substrate peptide in green as spheres. The zinc ion is shown as yellow sphere. B. hSIRT1-binding site of **11** with the interacting hydrophobic residues shown as stick diagram in blue. Compound **11** is shown as stick diagram in orange. Oxygen atoms are shown in red and nitrogen in blue. Hydrogen bonds are shown as black dashed lines. C. Chemical structure of **11**.*

This structure determination of a hSIRT1-activator complex provides the proof of direct allosteric activation of hSiRT1 by small molecule activators, however the detailed mechanism of SIRT1 activation is still elusive.

SIRT3 and SIRT5

SIRT5 activation by resveratrol was also detected against the artificial FdL1 substrate, resulting in approximately 2.5-fold activation at 0.2 mM

resveratrol concentration. A resveratrol natural metabolite, piceatannol (Figure 1.10), which carries an additional hydroxyl group was also found to stimulate SIRT3 displaying an AC_{50} of 0.07±0.02 mM.
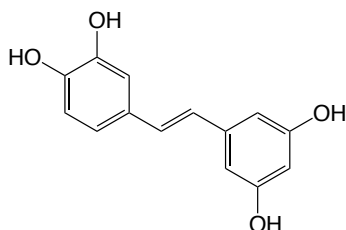
Figure 1.10. *Chemical structure of piceatannol.*

The crystal structure of human SIRT5 with FdL1 and resveratrol revealed a direct contact of resveratrol to the substrate (Figure 1.11). According to this, the C-terminal FdL1 fluorophore is positioned in a hydrophobic tunnel made by Tyr255 (Y255) and Pro256 (P256). Resveratrol is located in direct contact with this FdL1 fluorophore while the A-ring of resveratrol binds between Thr278 (T278) and Thr279 (T279) of a loop on one side of FdL1, and between Gly72 (G72) and Ala73 (A73) on the other side of it (Figure 1.11B). The B-ring of resveratrol binds between Arg71(R71) and Gly72 (G72) while it interacts in a perpendicular direction with the FdL1 fluorophore (Figure 1.11). Binding of the activator seems to close the active site entrance, thereby trapping the bound substrate and resulting in a more suitable substrate orientation for the subsequent binding and reaction [75].
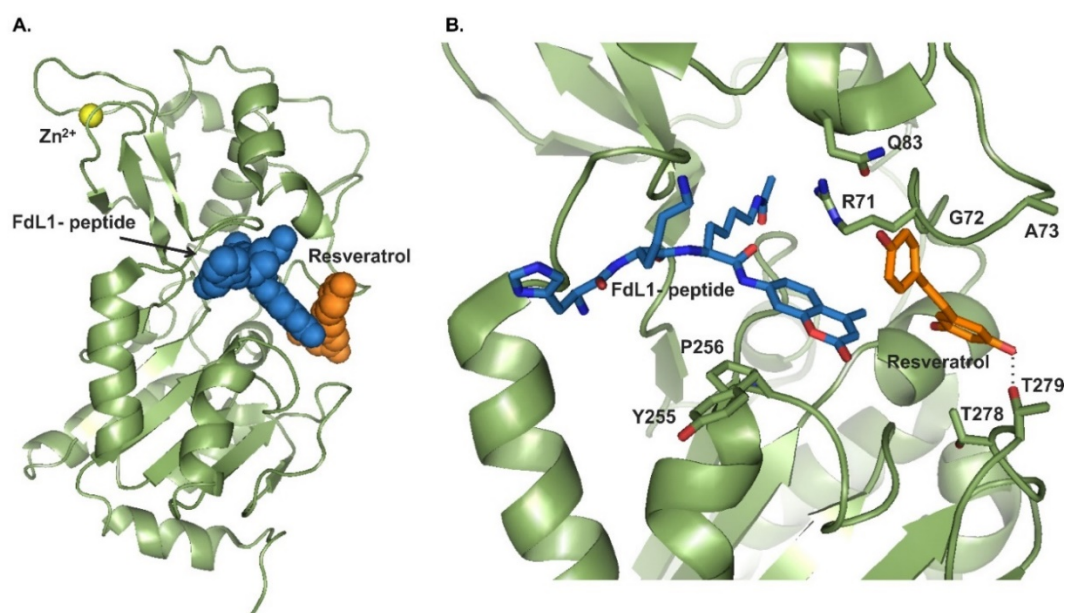
Figure 1.11. *Crystal structure of SIRT5 in complex with the FdL1-peptide and resveratrol. A. Overall structure of the SIRT5/FdL1/resveratrol complex (PDB ID 4HDA) is shown as ribbon diagram. FdL1-peptide and resveratrol are shown as spheres in blue and orange respectively. The zinc ion is shown as a yellow sphere. B. FdL1-peptide and the resveratrol-binding site with the interacting residues shown as stick diagram. Resveratrol and FdL1-peptide are shown as blue and orange stick diagrams respectively. Oxygen atoms are shown in red and nitrogen in blue. Hydrogen bonds are shown as black dashed lines.*

The crystal structure of SIRT3 in complex with FdL1-peptide and piceatannol showed direct contact between FdL1-peptide and piceatannol similar to that displayed in the SIRT5/FdL1/resveratrol complex [75] and thus is not discussed in detail here.

SIRT6

A weak molecule activator (**12**) for SIRT6 was determined through a docking screen and subsequent FdL assays after efforts to identify isoform-

specific sirtuin inhibitors [76]. A more potent compound (**15**) was identified

from a group of activator **12** derivatives, yielding a dose-dependent

enhancement in SIRT6 activity with an AC_{50} of 38 µM and a maximum 2-

fold activation [77]. Due to the limited solubility of **15**, definitive

calculations of K_m and V_{max} values could not be performed. The crystal

structures of SIRT6 in complex with derivatives of **12** (compound **13**, **14**

and **15**) were determined revealing the same binding site. The SIRT6

complex with ADP-ribose and the most potent activator **15** is described

below as a representative example of the interactions between these

compounds and SIRT6 [77].

Compound 12   **Compound 13**   **Compound 14**   **Compound 15**
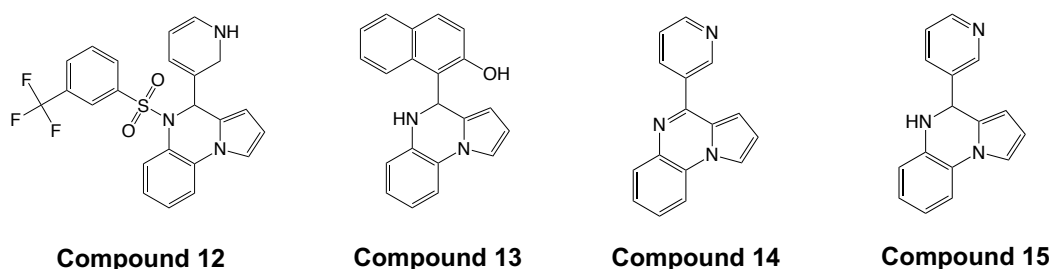
Figure 1.12. *Chemical structures of SIRT6 activators*

Compound **15** is located in a hydrophobic pocket at the exit of the SIRT6

active site that is formed by residues Phe64/82/86 (F64/82/86), Ile61

(I61), Pro62(P62), Val115(V115) and Met136/157(M136/157). The bottom

of the pyrrolo[1,2-*a*] quinoxaline moiety of **15** is exposed to solvent (Figure

1.13). The pyridine nitrogen atom could potentially make a polar interaction

and hydrogen bond to the Pro62(P62) backbone oxygen (Figure 1.13B)

[77].
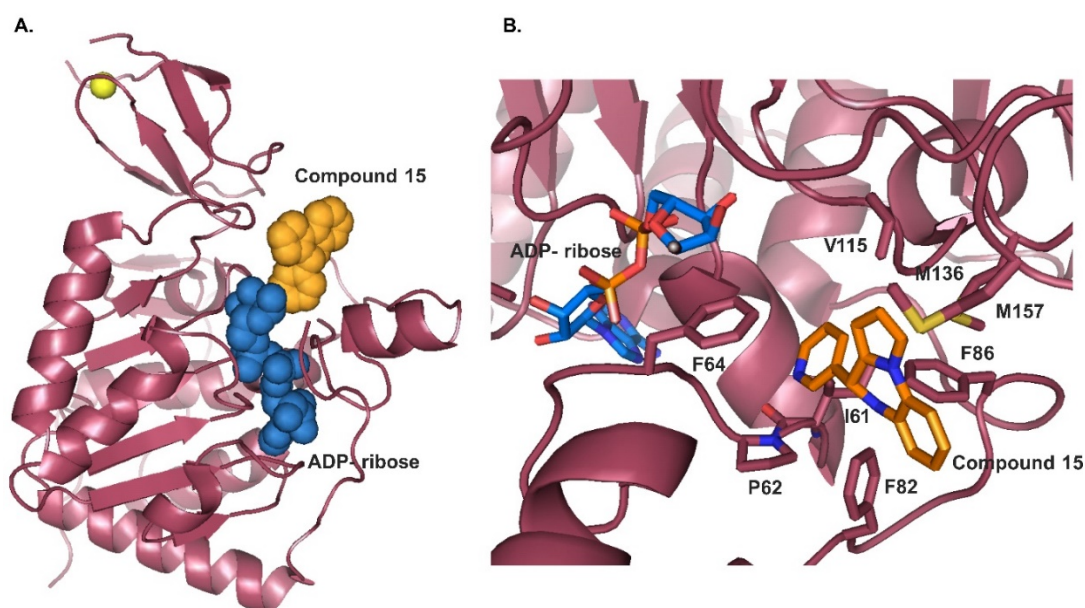
Figure 1.13. *Crystal structure of human SIRT6 in complex with ADP-ribose and*

15. *A. Overall structure of the SIRT6/ADP-ribose/**15** complex (PDB ID 5MF6) is*

*shown as ribbon diagram. Compound **15** and ADP-peptide are shown as spheres*

in orange and blue respectively. The Zinc ion is shown as yellow sphere. B.

*Compound **15**-binding site with the interacting residues is shown as stick diagram.*

*Compound **15** and ADP-ribose are shown in orange and blue stick diagram*

respectively. Oxygen atoms are shown in red and nitrogen in blue.

The discovery of compound **15** was followed by the identification of **16** and

17 after virtual screening of more than 5,000,000 compounds on the

predicted allosteric site of SIRT6 around Phe82 (F82) and Phe86 (F86) [78].

Kinetic characterization of the compounds using an acetylated peptide

(RHKK-ac-AMC) as substrate showed an activation with AC_{50} values of

approximately 174 and 218 µM respectively. Chemical optimization of the

initial hits resulted in selective SIRT6 activator **18** and **19** with improved

AC_{50} values of approximately 10 and 6 µM respectively and significantly

increased activity by more than 22-fold at 100 µM. Compound **18** and **19**

were found to decrease the K_m by 3.2-fold and 7.8-fold and strongly increase the k_{cat} by 7.8-fold and 41-fold respectively for the acetylated substrate and NAD$^+$ [78].
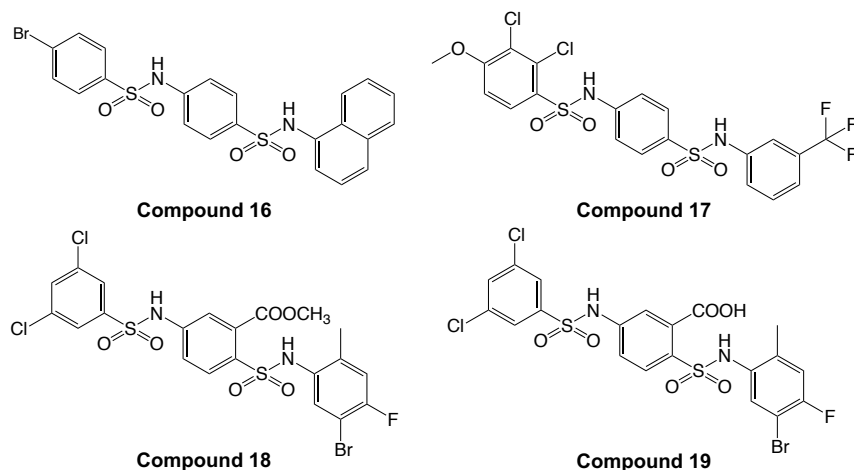
Figure 1.14. *More chemical structures of SIRT6 activators.*

Both compounds **18** and **19** were found to efficiently increase the deacetylation activity of SIRT6 on another physiologically acetylated H3K9 peptide (KQTARK-ac-STGGWW, H3K9ac) and on the H3K9ac peptide, which lacks two tryptophan residues at the C-terminus (KQTARK-ac-STGG) to exclude the possibility of nonspecific effects [78]. Cellular SIRT6 activation has also been observed through decreasing H3K9ac levels and inhibiting the proliferation of human hepatocellular carcinoma (HCC) by the addition of **10** [78].

The crystal structure of SIRT6, H3K9-Myr peptide-ADPR (a product analog of deacetylation) with **19** revealed that **19** is located in the same hydrophobic pocket as **15** in a location other than the substrate-binding site (Figure 1.13) [78]. Compounds **18** and **19** were found to stabilize a more active form of SIRT6 through a mechanism similar to that of SIRT6

allosteric compound **15** (Figure 1.13).

As described above, several sirtuin isoforms (SIRT1, SIRT3, SIRT5, SIRT6) can be activated by direct activator-substrate contacts and mostly through hydrophobic interactions between SIRT6 amino acids and the activators. To date, only activators of SIRT1 have been assessed through clinical trials but results from these trials suggest that compounds with better pharmacokinetic and tolerability profiles are needed.

1.4. Characterising enzyme activation

Michaelis-Menten kinetic analysis can be used to assess enzyme activators although it doesn't apply as well as in the case of enzyme inhibitors [15]. Usually, reversible activators follow Michaelis-Menten kinetics. The ordered binding of the substrate to the enzyme following by activator binding is analogous to uncompetitive inhibition and represents the simplest activation mechanism (eq 1).

$$
\begin{array}{c}
\mathrm{E} + \mathrm{S} \underset{}{\overset{K_s}{\rightleftharpoons}} \mathrm{ES} \overset{k_2}{\rightarrow} \mathrm{E} + \mathrm{P}_1 + \mathrm{P}_2 \\
K_A \big\updownarrow \mathrm{I A I} \\
\mathrm{AES} \overset{\beta k_2}{\rightarrow} \mathrm{E} + \mathrm{A} + \mathrm{P}_1 + \mathrm{P}_2
\end{array}
\tag{1}
$$
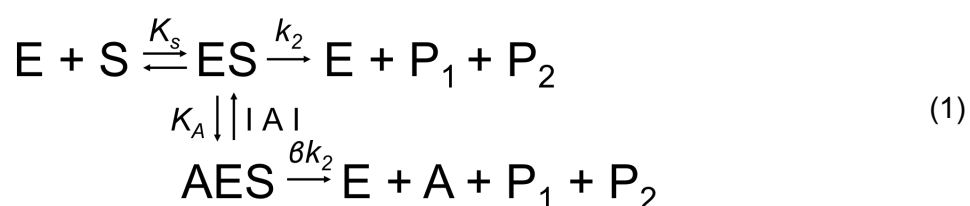
where E is the enzyme, S is the substrate, P_1 and P_2 are the products, A is the activator, ES is an enzyme-substrate complex, K_s is the dissociation constant for the ES complex, K_A is the dissociation constant for the AES complex to A+ES, k_2 is the rate constant and β is a constant.

The nonessential activator model (eq 2) which is similar to the proposed reversible mixed inhibition mechanism [79] is represented schematically as

$$
\begin{array}{ccccc}
\text{E} + \text{S} & \underset{}{\overset{K_s}{\rightleftharpoons}} & \text{ES} & \overset{k_2}{\longrightarrow} & \text{E} + \text{P}_1 + \text{P}_2 \\
+ & & + & & \\
\text{A} & & \text{A} & & \\
\updownarrow K_A & & \updownarrow \alpha K_A & & \\
\text{EA} + \text{S} & \underset{}{\overset{\alpha K_s}{\rightleftharpoons}} & \text{AES} & \overset{\beta k_2}{\longrightarrow} & \text{E} + \text{A} + \text{P}_1 + \text{P}_2
\end{array}
\tag{2}
$$
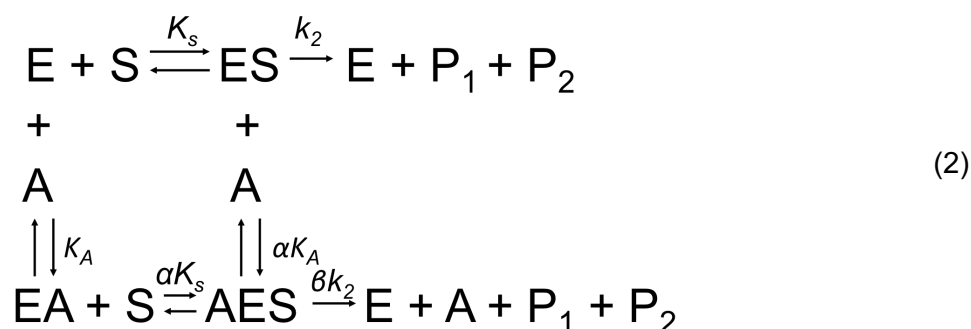
The kinetic constants k_{cat} (k_{cat} = catalytic rate constant), K_m (K_m = Michaelis constant) and V_{max} are used to characterise the potency of an enzyme activator. V_{max} and K_m values are determined by fitting a plot of initial velocity versus substrate concentration. Consequently, the k_{cat} value is measured using the following equation:

$$k_{cat} = V_{max} / [E]_i$$

Where $[E]_i$ is the initial enzyme concentration and V_{max} is the maximum velocity.

There are two important measures of activation; the maximum activation obtained following addition of the activator (A_{max}) and the concentration of the activator required to half-increase the maximum activation of the enzyme (AC_{50}). A_{max} values in combination with structural information of a series of activators bound to an enzyme often helps to get more insight about the relationship between the chemical structures of compounds and the increased activation observed. The AC_{50} value correlates well with a compound's binding affinity for the enzyme, but there is an important caveat for activators that facilitate irreversible modification (this kind of activator is not described here) [15].

1.5. Discovery of small molecules that activate enzymes

High-Throughput Screening (HTS) is the most widely used approach for small molecule discovery in large pharmaceutical organisations. This approach uses robotic automation and liquid handling devices to test typically up to tens of thousands of compounds per day in an assay. As well as considerable amounts of equipment, HTS has two main requirements –a compound library and a robust assay that can be performed in miniaturised plate format (usually between 96 to 1536 samples per plate). Organisations that perform these assays typically invest large sums of resources in preparation and curation of the compound library which can vary from 100s of thousands to millions of compounds in the largest organisations. Screening large numbers of modulators produces a number of hits that is lower in small molecule activator screens in comparison to inhibitor screens. HTS libraries are typically selected to focus on randomly picked compounds as seen with GK or natural metabolites and precursors related to the target such as seen with SIRT1. The primary role of HTS is to detect lead compounds which are characterized and are then optimized [15].

Metal contaminants in stock compounds need to be regularly checked as they are known to cause the most false positives in activator discovery screening approaches.

Rational design approaches that mimic endogenous interactions looking for allosteric sites on enzymes have also proven successful as seen with PDPK1 [61].

1.6. Previous work

The York laboratory has previously demonstrated the use of fragments to identify the first small-molecule activator **20**, 4-ethoxyquinazoline, discovered for any glycoside hydrolase, which was in this case the bacterial O-GlcNAc hydrolase (OGA) from *Bacteroides thetaiotaomicron*, BtGH84 [73]. BtGH84 removes the post-translational attachment of N-acetylglucosamine (GlcNAc) to the hydroxy groups of serine and threonine residues. The small molecule activator (**20**) (Figure 1.15) was identified from a fragment screen of 1,000 compounds in the presence of the known O-GlcNAcase inhibitor PUGNAc. *In vitro* kinetic characterization of the compound using 4-methylumbelliferyl-N-acetyl glucosamine (4MU-GlcNAc) and p-nitrophenyl-N-acetyl-glucosamine (pNP-GlcNAc) as substrate showed an activation with an AC_{50} value of 3.5 mM. The activator was found to decrease the K_m value from 1.7 to 0.87 mM and increases the k_{cat}/K_m values from 28000 to 54000 M^{-1} s^{-1} following the nonessential type of activation.
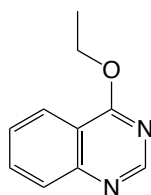
Figure 1.15. *Chemical structure of GlcNAc hydrolase activator* **20**.

Further optimization of the initial compound using the analogue-by catalogue approach resulted in compounds with better AC_{50} values and significantly improved k_{cat}/K_m values with factors of around 8-fold (Table 1.4.); these compounds also had a small effect on the V_{max} [73].

Table 1.4. *Activator data from a 4MU-GlcNAc cleavage assay.*

Compound	AC_{50} [µM]	A_{max} [%]	K_m [µM]	k_{cat}/K_m [M^{-1} s^{-1}]
-	-	-	1700	29000
20	575	284	809	92000
21	259	371	416	210000
22	274	329	643	125000
23	702	234	649	77500

The crystal structure of BtGH84 provides more details about the mechanism of activation (Figure 1.16) [73]. BtGH84 is observed in either an "open" or "closed" form depending on the interactions made by any ligands. More specifically, the loops that present the catalytic and substrate binding residues shift between two distinct conformational states (Figure 1.16.B). The activator binds directly to the catalytic domain in close proximity to the active site of the enzyme. Activator **20** is located on top of Tyr137 (Y137) and forms a hydrogen bond with Arg347 (R347). The binding of BtGH84 activators occurs with the "loop closed" form of the enzyme, where the Tyr137 (Y137) residue can form a hydrogen bond with catalytic Asp243 (D243), thus stabilising it, and is potentially enhanced by the π-stacking

interaction with the activator. In the "loop open" structure the loop containing the catalytic residues Asp242 (D242) and Asp243 (D243) is shifted away from the active site and the hydrogen bond between Tyr137 (Y137) and Asp243 (D243) breaks (Figure 1.16.B) [73].
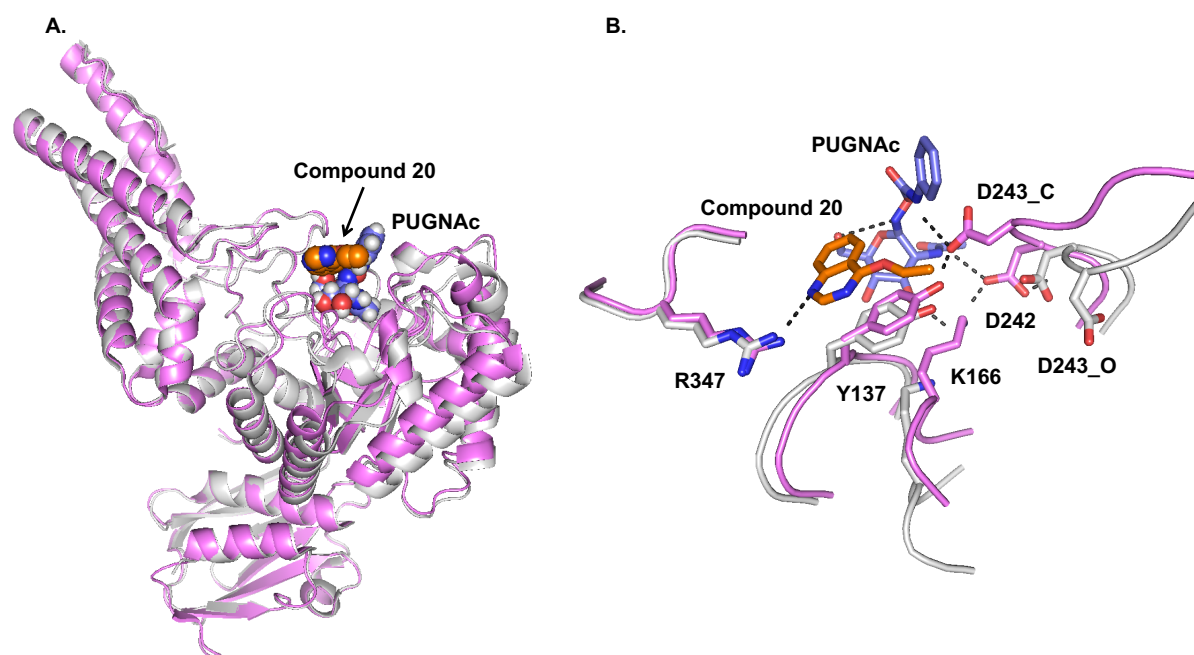
Figure 1.16. *A. Structural overlay of bound and unbound structure BtGH84. The protein backbone is shown as ribbon diagram. BtGH84 bound to 20 and PUGNAc (PDB ID 4UR9) is shown in magenta and unbound BtGH84 (PDB ID 2W4X) is shown in grey. Compound 20 and PUGNAc are shown as spheres with carbon atoms of 20 in orange and of PUGNAC in blue. B. The binding sites of 20 and PUGNAc on BtGH84. The protein backbone of the catalytic loop is shown schematically with key active site residues in stick representation. The ligands are shown as stick diagrams with carbon atoms of 20 in orange and of PUGNAC in blue. Hydrogen bonds are shown as black dashed lines.*

Following this study, the small molecule activator was covalently tethered to the enzyme; Tyr550 was mutated to Cys and an activated fragment

designed and synthesized inducing a 35-fold increase in catalytic activity (k_{cat}/K_m). The successful attachment was confirmed by mass spectroscopy and determination of the crystal structure of the covalently modified BtGH84 [80].

1.7. Fragment based ligand discovery (FBLD)

FBLD has emerged over the last two decades and is now well established as an approach for the discovery of novel lead compounds. There are many examples of compounds derived from fragments that are in clinical trials, Vemurafenid and Venetoclax being two examples (Figure 1.17) that are now being used to treat patients with lymphocytic leukaemia and metastatic melanoma respectively.

Vemurafenib   **Venetoclax**
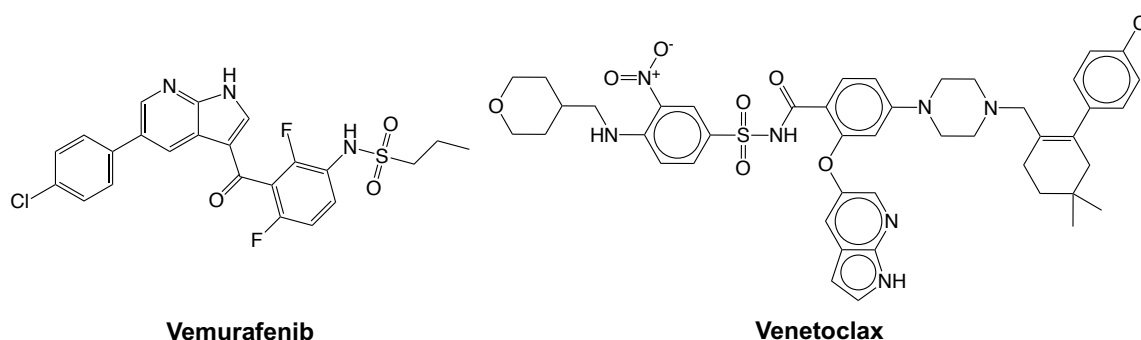
Figure 1.17. *Chemical structures of Vemurafenib and Venetoclax.*

The key feature of FBLD is to find a low molecular weight molecule (typically between 120-250 Da) by screening a small library (usually 1000-2000 fragments) that is big enough to probe interactions in the target protein but also small enough to minimize unfavourable interactions. Focusing on lower molecular weight compounds allows a significant region of chemical space to be explored using fewer compounds and avoiding the need to

perform very large high-throughput screens. Fragments bind with very weak affinity (dissociation constant for complex, K_D) requiring high compound concentrations during screening and the use of a biophysical method that is sensitive enough to reliably detect interactions in the low millimolar (mM) range. The most robust screening method is ligand-observed NMR which has a dynamic range between 10 mM and 100 nM and which does not require protein labelling and can be used for proteins of any molecular weight. A small fragment hit can be further optimised to larger, higher affinity lead compounds. There are three main strategies for the development of fragments: growing, merging and linking. In the growing approach, which is the most widely used, a fragment hit either provides a central scaffold with which to search a database of accessible compounds or is used as a basis to grow more effective ligands, by structure-guided ligand design. In the linking approach is to link together two fragments that bind in adjacent sites while in the merging approach, binding mode information of fragments combined with data from literature, HTS hits or virtual screening are used to synthesize new scaffolds.

FBLD has been widely used to identify enzyme inhibitors and protein-protein interactions. However, very little has been reported on the use of FBLD to identify activators of enzyme activity; notable exceptions being the activator of BtGH84 described above [73]. For this artificial system wherein in mechanistic insight was obtained, the mechanism of activation requires that the activator concentration is in excess of the enzyme concentration. Darby *et* al. demonstrated that this limitation can be overcome through

developing a rational covalent strategy to design structure guided covalent modifications at the site identified from fragment screening leading to a modified enzyme with significantly increased activity [80].

Therefore, FBLD could be an alternative to traditional genetic approaches to identify compounds that affect the conformational changes occurring with the large number of enzymes that perform catalysis through mechanisms involving conformational changes. There are many studies wherein in conditions such as pH, solvents and immobilisation were altered to affect the activity of industrially important enzymes, but no such studies have been reported regarding enzyme activators. Indeed, non-covalent activators may not be commercially feasible due to the cost of the activators and the probable need to separate activator from the products. However, covalent attachment could address this issue by providing the foundation for a combined FBLD and tethering approach as a general strategy for enzyme optimisation (Figure 1.18). Previous work on subtilisin *Bacillus lentus* demonstrated activation by introducing a cysteine in the substrate binding site [81-82]. A different tethering approach includes screening of disulphide-containing compounds that react with natural or introduced cysteine residues as seen with a kinase [83-84]. Although, these strategies show that covalent attachment of a molecule to an enzyme can result in increased activity, prior knowledge of structure to rational design such modifications is required.

This approach may have general applicability for improving the activity of enzymes but also for developing probes to study the impact of modulating enzymes in cell biology.
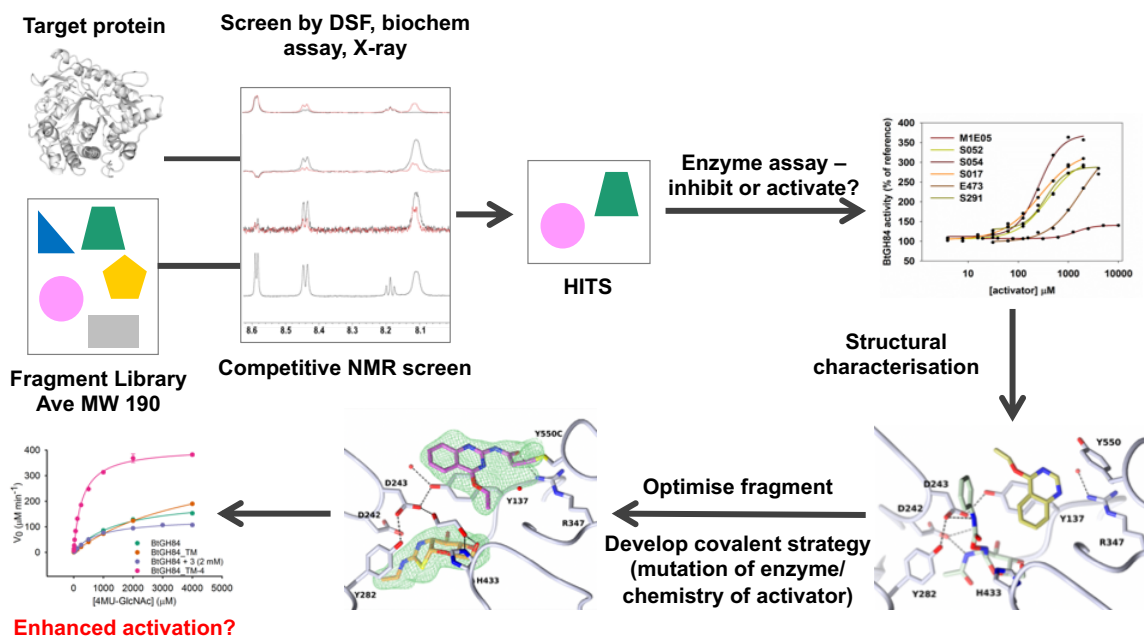
Figure 1.18. *The FBLD process for discovery of activators.*

1.8. Aims of the project

This project is to use FBLD methods to explore the activation by small molecules of carbohydrate processing enzymes used in key industrial applications. The idea arose from previous research described above on the enzyme BtGH84 which can be considered a model system. This provided the experience to identify the requirements of an enzyme suitable for identification of activators using fragment based lead discovery (FBLD): catalytically active proteins that can be sourced easily and in large quantities, for example by being produced through bacterial over-expression, the availability of a high resolution crystal structure of the enzyme, the availability of an activity assay suitable for screening the enzyme's activity against ligands in high concentrations, and the availability of specific inhibitors for assay validation and competition in screening.

The degradation of cellulose is an important and challenging area for industrial biotechnology. The first-generation biofuels and sugar production from corn starch industries are one of the largest markets worldwide for industrial enzymes. Three enzymes used in these processes were chosen for investigation: Two cellulose degrading enzymes (β-glucosidase from *Trichoderma reesei* (*Tr*Bgl2) and β-glucosidase from *Clostridium cellulovorans* (*Cc*BglA) and a "Termamyl-like" *Geobacillus stearothermophilus* α-amylase.

The aims of this project are to identify small molecules that activate these enzymes, and to characterise the kinetics, mechanism of action, and any

changes to substrate and product profiles. A longer-term aim (not realised in this thesis) was to design covalent strategies in order to attach the small molecules to the enzymes for constitutive enzyme activation.

In this thesis, work towards the identification and validation of a series of novel activators of the fungal *Tr*Bgl2 is described in Chapter 2. Chapter 3 describes characterisation of binding of the compounds to *Tr*Bgl2 using a number of biophysical techniques. Chapter 4 describes attempts to characterise the structure and interaction of the activator with the enzyme. It was not possible to obtain a crystal structure of activator bound to the enzyme so advanced NMR experiments were used to generate insight into the mechanism of activation. Chapter 5 summarises work on other enzymes that were investigated for enzyme activation and Chapter 6 is a discussion and conclusion of the results of the thesis. A number of established biomolecular and biophysical techniques were used in these studies. The relevant section in Chapter 7 (Materials and Methods) provides a summary of the principles, limitations and information generated for each of the techniques.

Chapter 2: Identifying and validating activators of the fungal glycosyl hydrolase, TrBgl2

2.1. Introduction

The depletion of fossil fuel in combination with the increasing demand for energy worldwide has instigated research on alternative and sustainable energy sources. Lignocellulosic (LC) biomass such as wood, agricultural residues and dedicated energy crops is abundant and available at low cost and has received considerable attention as the most promising alternative, renewable source for second-generation biofuel production such as cellulosic ethanol [85]. This thesis contains studies to identify activators of a number of enzymes that can break down cellulose.

LC biomass, the structural backbone of all plant cell walls, is composed mainly of cellulose (40-50%), in combination with hemicellulose (25-30%) and lignin (15-20%) [86]. Cellulose is a linear polymer formed by D-glucose units linked by β-1,4-glycosidic bonds that are further stabilised by hydrogen bonds (Figure 2.1). Cellulose linear chains arrange themselves to form rod-like structures of 3-4 nm diameter called microfibrils, which contains highly ordered crystalline structures [85]. Hemicellulose is an amorphous branched or linear heteropolymer of 200-400 units of different hexoses (C6 sugars), pentoses (C5 sugars) and uronic acids. Lignin is a crosslinked polymer of the three phenyl propane units *p*-coumaryl, coniferyl and sinapyl alcohol with an amorphous structure [85].
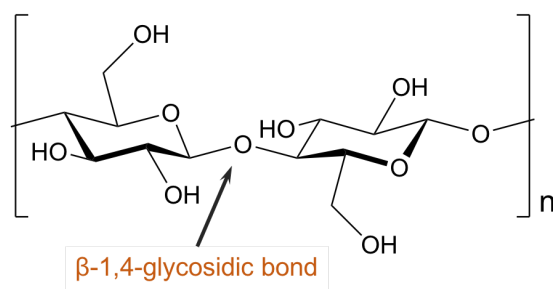
Figure 2.1. *Schematic representation of cellulose structure and its breakdown to D-glucose.*

A mixture of enzymes that together are known as cellulases catalyse cellulose degradation and comprise three categories of enzymes; endoglucanases (EG) (EC 3.2.1.4), exoglucanases or cellobiohydrolases (CBH) (EC 3.2.1.91) and β-glucosidases (BG) (EC 3.2.1.21) [85][87]. Endoglucanases cleave the internal β-1,4-glycosidic bonds of cellulose microfibrils releasing small fragments. Subsequently, exoglucanases or cellobiohydrolases (CBH) act on the reducing and non-reducing ends resulting in short chain cello-oligosaccharides such as cellobiose, which are hydrolysed into glucose by the action of β-glucosidases (Figure 2.2) [85][88].

Biotechnology companies, such as Novozymes, have already developed cellulolytic enzyme cocktails for LC biomass degradation. Breakdown of LC requires pre-treatment of plant material to disrupt the cell wall structure to improve access of enzymes to breakdown the cellulose. This is not straightforward due to the heterogeneity and the tight three-dimensional structure that is created by the main components of LC; lignin, cellulose and hemicellulose. The cost of such breakdown has so far limited the

amount of LC biomass entering the global energy market. In addition, fluctuations in oil prices (political events as well as fracking in the US) has recently limited investment. However, the increasing environmental concern over burning of fossil fuels may eventually lead to renewed focus on LC biomass conversion. Once the LC biomass is pre-treated, the accumulation of cellobiose is regarded as the main limiting step in degradation due to the low efficiency of β-glucosidase, resulting in endoglucanase and exoglucanase inhibition [85][89-93]. Exogenous supplementation of β-glucosidase has been used in some case as an alternative to overcome this issue [94-95]. Therefore, β-glucosidases with improved catalytic activity could address one of the main limiting factors in the production of second-generation biofuels.
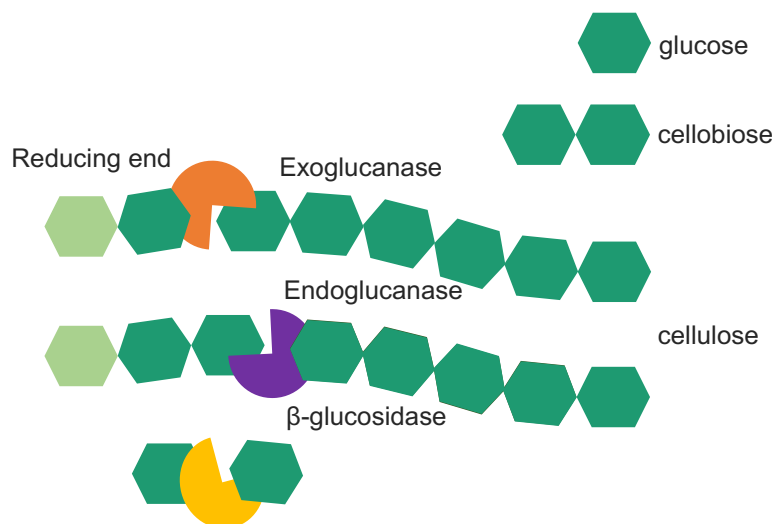
Figure 2.2. Cellulose degradation. Three classes of enzymes are working synergistically for the hydrolysis of cellulose: EG act on the internal β-1,4-glycosidic bonds releasing small fragments, subsequently CBH act on the reducing and non-reducing ends resulting in short chain cello-oligosaccharides and then BG hydrolase soluble short cello-oligosaccharides to glucose units [85].

2.1.1. Glycosyl hydrolases mechanism

Glycosyl hydrolases occur in all kingdoms of life and play crucial roles in many biological pathways, such as degradation of carbohydrates for nutrient uptake [96-97], cellular signalling and defence against pathogen attack [98-100]. Glycosyl hydrolases hydrolyse the glycosidic bond between carbohydrates or between a carbohydrate and a non-carbohydrate moiety. The hydrolysis of the glycosidic bond takes place via general acid catalysis by two critical residues: a general acid (proton donor) and a nucleophile/base. Hydrolysis occurs with either an overall retention, or an overall inversion of the anomeric configuration depending on the spatial position of these critical residues [101]. In both mechanisms, the position of the general acid (proton donor) is identical however the nucleophile/base differs between mechanisms. In retaining enzymes, the base is located near to the sugar anomeric carbon whereas in inverting enzymes, there is a water molecule between the nucleophile/base and the sugar. This results in an average distance between the two catalytic residues of approximately 5.5 and 10 Å in retaining and inverting enzymes respectively (Figure 2.3) [102]. The foundations of these widely adopted mechanistic models of glycosyl hydrolases were first proposed by Koshland [1].
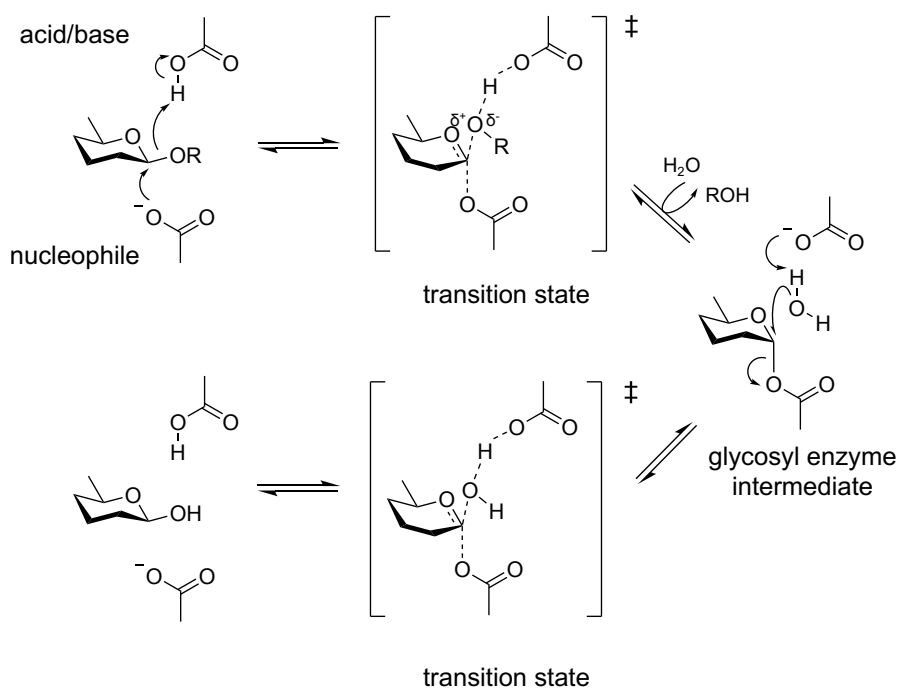
Figure 2.3. *Retaining mechanism of glycosyl hydrolases as first proposed by Koshland [1].*

2.1.2. The fungal glycosyl hydrolase Trichoderma reseei, TrBgl2

Trichoderma reseei produces large amounts of thermostable cellulolytic enzymes, which make them attractive targets for industrial applications [103]. As such, they have been widely studied in previous years [104-105]. Genetically modified β-glucosidases from *Trichoderma reesei* have been commercialised and are included in cellulolytic enzyme cocktails that have been developed by companies such as Novozymes (http://www.novozymes.com/) and Genencor(http://www.genencor.com/) [105]. *Tr*Bgl2 belongs to the β-retaining glycoside hydrolase family 1, according to the classification of Carbohydrate-Active enzymes (CAZY) (http://www.cazy.org/) [106-108]. The structure and biochemical

properties of apo *Tr*Bgl2 have been elucidated by Jeng [2]. *Tr*Bgl2 adopts a $(\alpha/\beta)_8$-TIM barrel-like fold as a typical member of the GH1 enzyme (Figure 2.4). The Protein Interfaces, Surfaces and Assemblies (PISA) server predicts that the protein does not share any specific interactions that promote oligomerisation. Nevertheless, four protein molecules of *Tr*Bgl2 crystallised in an asymmetric unit. There is no evidence that the tetramer is functionally relevant [2].
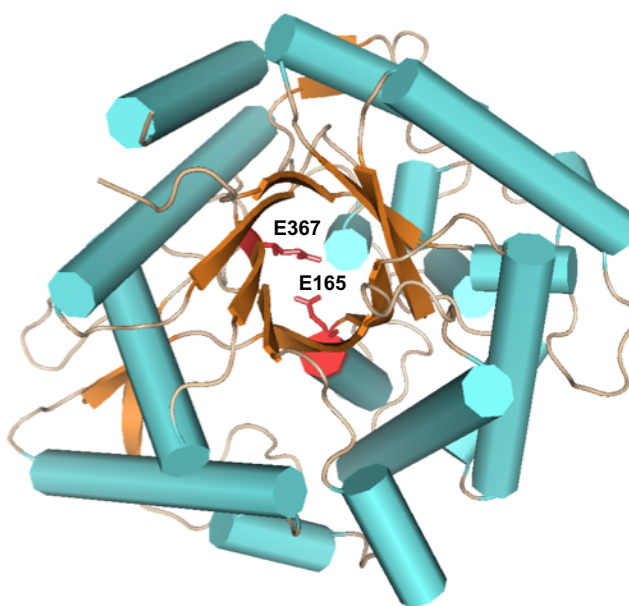
Figure 2.4. *Cartoon representation of the TrBgl2 structure (PDB ID 3AHY). The catalytic acid/proton donor, Glu165 (E165) and the catalytic nucleophile/base, Glu367 (E367) are shown as red stick diagram. The helices are shown in cyan, whereas strands in orange, loops in wheat.*

TrBgl2   TT →β1→   αααα α1   ααα η1   ααααααα α2   TT   TT   ααα η2   αααα

```
              β1          α1           η1          α2                  η2
TrBgl2   TT  ───►    ℓℓℓℓ         ℓℓℓ      ℓℓℓℓℓℓℓ          TT        TT  ℓℓℓ  ℓℓℓℓ
         1        10          20          30          40          50          60
         .        .           .           .           .           .           .
TrBgl2   MLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNRTAED

            α3            β2        α4            α5                        β3
TrBgl2   ℓℓℓℓℓℓℓ     ───►      ℓℓℓℓ      TT      ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ        ───►
              70          80          90          100         110         120
              .           .           .           .           .           .
TrBgl2   IALLKSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFITLFHW

            α6      η3            α7              β4        α8            T
TrBgl2   ℓℓℓℓℓℓℓ   ℓℓℓ     ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ   ───►     ℓℓℓℓℓℓℓℓℓ
              130         140         150         160         170         180
              .           .           .           .           .           .
TrBgl2   DLPEGLHQRYGGLLNRTEFPLDFENYARVMFRALPKVRNWITFNEPLCSAIPGYGSTFA

         T              α9                      TT      β5        β6          ℓℓℓℓ
TrBgl2      ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ        ───►      ───►
              190         200         210         220         230         240
              .           .           .           .           .           .
TrBgl2   PGRQSTSEPWTVGHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDAADPADK

            α10           α11           α12     η4       α13           β7
TrBgl2   ℓℓℓℓℓℓℓℓℓ   ℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓℓℓ       ℓℓℓℓℓℓℓ     ───►     ─
              250         260         270         280         290         300
              .           .           .           .           .           .
TrBgl2   EAAERRLEFFTAWFADPIYLGDYPASMRKQLGDRLPTFTPEERALVHGSNDFYGMNHYTS

            β8          TT         β9        TT       TT       α14
TrBgl2   ───►                    ───►                    ℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
              310         320         330         340         350         360
              .           .           .           .           .           .
TrBgl2   NYIRHRSSPASADDTVGNVDVLFTNKQGNCIGPETQSPWLRPCAAGFRDFLVWISKRYGY

            β10           η5    α15        α16                    β11
TrBgl2   ───►          ℓℓℓ  ℓℓℓℓℓ   ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ       ───►
              370         380         390         400         410         420
              .           .           .           .           .           .
TrBgl2   PPIYVTENGTSIKGESDLPKEKILEDDFRVKYYNEYIRAMVTAVELDGVNVKGYFAWSLM

            η6          β12        β13       α17
TrBgl2   ℓℓℓ          ───►      ───►      ℓℓℓℓℓℓℓℓℓℓℓℓ
              430         440         450         460
              .           .           .           .
TrBgl2   DNFEWADGYVTRFGVTYVDYENGQKRFPKKSAKSLKPLFDELIAAA
```

Figure 2.5. *The secondary structure elements of TrBgl2. Numbering of the principal secondary structure elements underlines the (α/β)₈-TIM barrel structure; additional secondary elements are indicated by letters. The 3¹⁰-helices are indicated by the character η [2]. Figure generated with ESPript 3.0 (http://espript.ibcp.fr/ESPript/ESPript/) [109].*

The active site of the enzyme is in a 15-20 Å deep slot-like cleft and is surrounded by negatively charged residues. The catalytic acid/proton donor is Glu165, located on the TXNEP motif at the end of β-strand 4 (where X is a hydrophobic amino acid) (Figure 2.5), while the catalytic nucleophile/base is Glu367 for *Tr*Bgl2. Two *cis*-peptide bonds were found, Ala180-Pro181 and Trp417-Ser418, which their presence is very common for the glycosyl hydrolase family 1 [110].

A TRIS molecule was found to bind in different orientations in the active site of *Tr*Bgl2 which forms five direct hydrogen bonds with three residues of *Tr*Bgl2 (Glu165 (E165), Glu367 (E367) and Glu424 (E424)) resulting in slight inhibition of the enzyme (Figure 2.6) [2]. TRIS-based buffers are reported as extremely potent glycosidase inhibitors. The positively charged amino-center of TRIS partially mimics the positive transition state charge of the reaction while the various flexible alkyl hydroxyl moieties mimic the oxygen atoms of the sugar substrates. This results in such tight binding that it has proven difficult to remove [111-112]. In this project, TRIS was only present in the purification scheme (Section 7.5.1) and crystallisation conditions (Section 7.9.1) due to the difficulties faced to crystalize *Tr*Bgl2 when TRIS was replaced by HEPES. The rest of the experiments were performed in the absence of TRIS to avoid TRIS binding at the active site which could potentially displace low affinity ligand binding during the library screening.
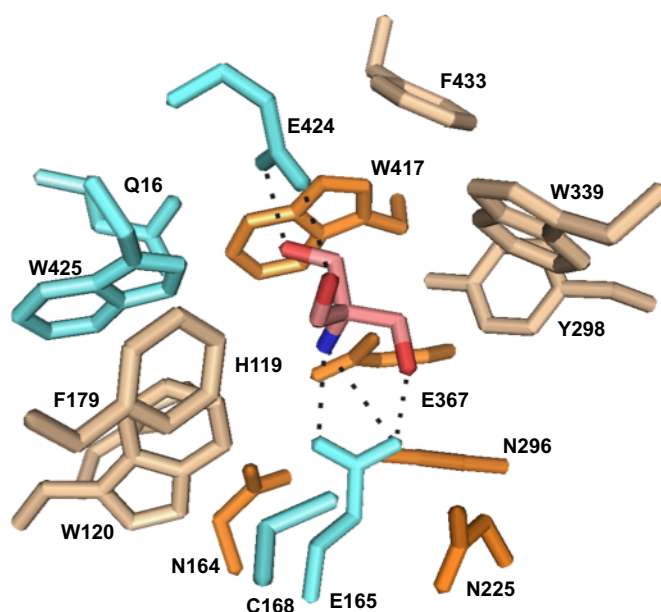
Figure 2.6. *TrBgl2 active site (PDB ID 3AHY) is shown as stick diagram. TRIS molecule is shown as pink stick diagram. The helices are shown in cyan, whereas strands in orange, loops in wheat. Oxygen atoms are shown in red and nitrogen in blue. The hydrogen bonds are shown as black dashed lines.*

2.2. Cloning and protein expression

Trbgl2 gene was cloned into the pET-YSBLIC3C plasmid under the control of T7 promoter as described in section 7.2. Figure 2.7 shows the construct pET-YSBLIC3C_*Trbgl2* which incorporated a His$_6$ cleavable extension on the N-terminus.

DNA sequencing confirmed the success of the cloning process. The construct pET-YSBLIC3C_*Trbgl2* was transformed into the *E. coli* BL21 DE3 bacterial strain and *Tr*Bgl2 was successfully overexpressed (as described in Section 7.4.1) in soluble form, present in the supernatant after centrifugation. The molecular weight of the recombinant protein with the His$_6$ cleavable extension on the N-terminus is 55 kDa.
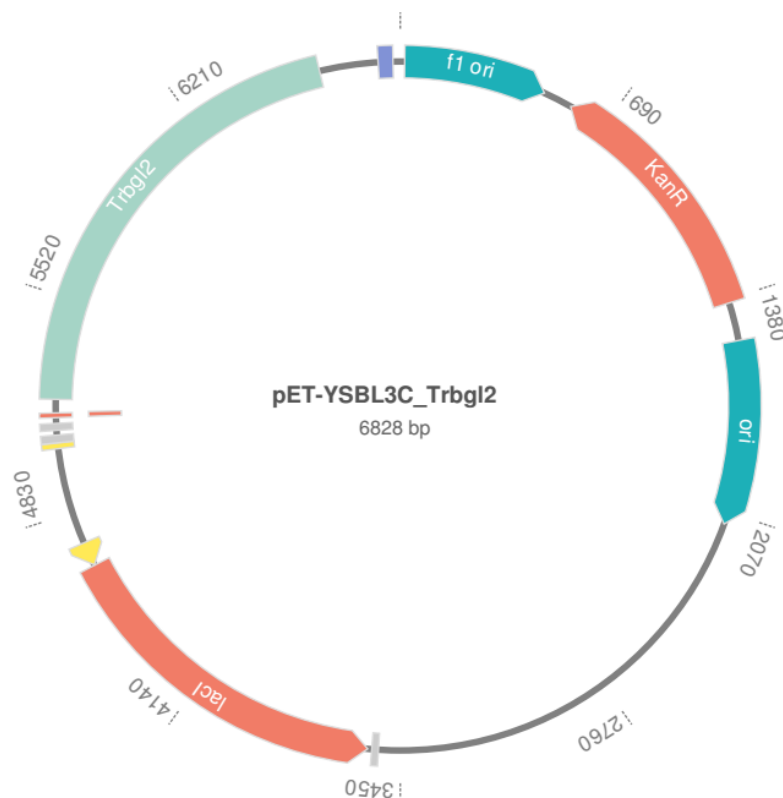
Figure 2.7. *Plasmid map of pET-YSBLIC3C with Trbgl2 cloned into the YSBL-LIC3C site. The ORF containing a gene encoding TrBgl2 is under control of the T7 promoter and begins transcription with an N-terminal 3C-cleavage His$_6$ affinity tag. The plasmid map generated using SnapGene® Viewer.*

2.3. Protein production and characterisation

The protein purification protocol (Section 7.5.1) that was used for *Tr*Bgl2 yields approximately 60 mg/L in optimal conditions. It was isolated in two chromatographic steps comprising a Ni^{2+} affinity capture of the His$_6$-tagged protein and subsequent size-exclusion (S200) chromatographic separation with purity more than 90% as determined by SDS-PAGE. *Tr*Bgl2 protein is eluted in TRIS buffer as one population at a volume that was consistent with the protein being monomeric (Figure 2.8).
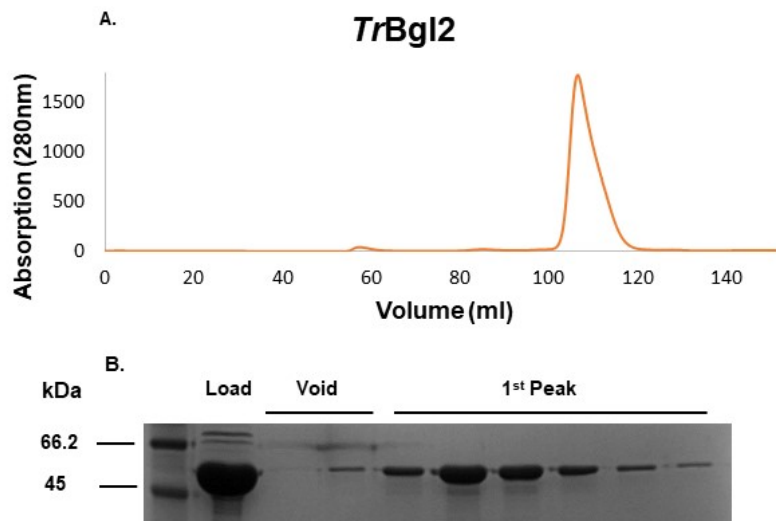
Figure 2.8. *S200 column TrBgl2 eluate. A. S200 elution peaks detected by UV absorbance at 280 nm, obtained after 2 mL injection of TrBgl2 at 17 mg/mL concentration. B. SDS PAGE of the Sephacryl S200 column TrBgl2 eluate. Samples were electrophoresed on a 12% polyacrylamide gel under denaturing and reducing conditions.*
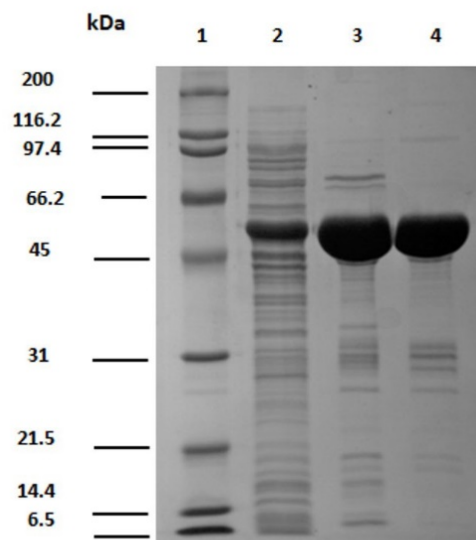
Figure 2.9. *SDS PAGE of the purified TrBgl2. 1. Broad molecular weight marker, 2. Cell culture crude extract, 3. Ni-NTA agarose column eluate, 4. Sephacryl S200 column eluate. Samples were electrophoresed on a 12% polyacrylamide gel under denaturing and reducing conditions.*

2.4. Oligomerisation studies

Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS) analysis was carried out to investigate the oligomeric state of the recombinant *Tr*Bgl2. SEC-MALS measures absolute molar mass and size of molecules in solution using the intensity and the light scattered by the sample at a range of angles (Section 7.7) [113]. The analysis revealed the presence of one main population of *Tr*Bgl2 which corresponds to an average mass of approximately 53 kDa (Figure 2.8). This mass was consistent with the monomeric form of purified *Tr*Bgl2 in solution.
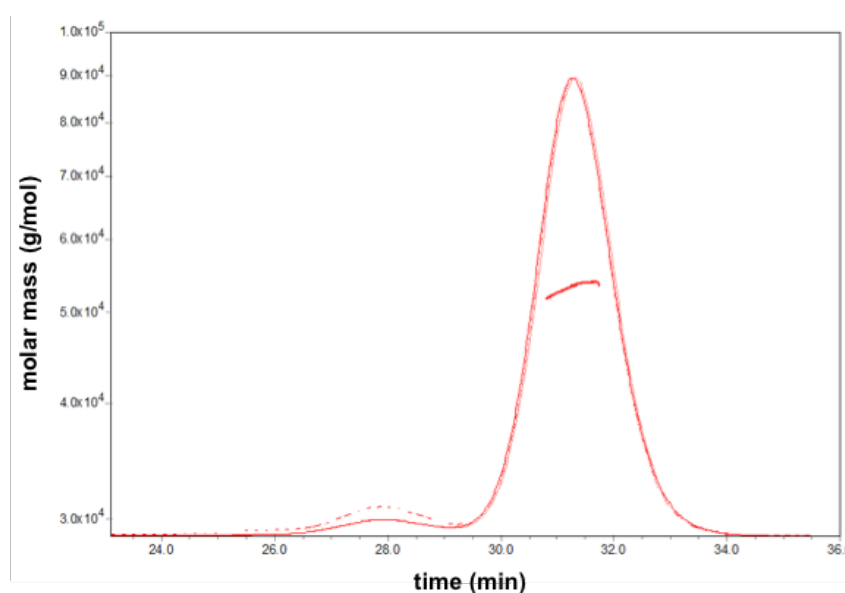
Figure 2.10. *SEC-MALS chromatogram of the elution of TrBgl2 from a S200 10/300 analytical column. The chromatogram displays the UV at 280 nm (dotted lines), light scattering at 90° angle, LS, (solid line), refractive index, RI, (dashed line) together with the molar mass of each peak calculated by MALS.*

2.5. Crystal structures

2.5.1. Apo TrBgl2

*Tr*Bgl2 protein was screened against commercially purchased screens (Table 7.1) in a 96-well plate format, using sitting drop vapour diffusion methodology. Crystals of apo-*Tr*Bgl2 were obtained in the Index Screen (Hampton). The initial conditions for protein crystallisation were optimised and the best results were obtained with 35% (w/v) PEG 3350 and 0.25 M NaCl in 0.1 M TRIS pH 8.5 at 293 K (Section 7.8). The final crystal size was obtained in 3 days.
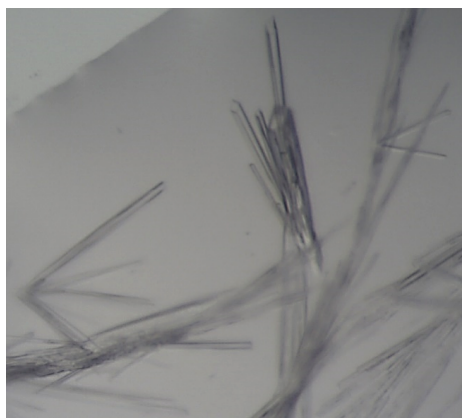
Figure 2.11. Crystals of TrBgl2 protein.

The *apo* structure of *Tr*Bgl2 was determined and refined at a resolution of 1.8 Å and contained four protein and TRIS molecules in the asymmetric unit (details of data collection, structure determination and refinement are in section 7.9.2). Data-processing and refinement statistics are summarized in Table 2.1. The crystal structures of the current *apo Tr*Bgl2 and the previously determined *apo Tr*Bgl2 (PDB ID 3AHY) were superposed. This superposition gives an overall RMSD of 0.39 Å and on the active site

residues a RMSD of 0.17 Å which shows that the *apo Tr*Bgl2 structure obtained is almost identical to the already determined *apo Tr*Bgl2 structure (PDB ID 3AHY).

Table 2.1. *Data collection and refinement statistics for apoTrBgl2.*

X-ray-source	Diamond I04
Wavelength [A]	0.976250
Space group	P 1 21 1
a, b, c (Å)	93,04, 103.78, 94.76
α, β, γ (^{o})	90.00, 105.56, 90.00
Resolution range (Å)	74.76-1.83
Completeness (%)	100 (100) *
Multiplicity	4.1 (3.8)
Total reflections	653030 (66163)
Unique reflections	160317 (15913)
$<I\sigma(I)>$	8.1 (2.3)
R_{merge}	0.089 (0.518)
R_{pim}	0.078 (0.455)
$CC^{1/2}$	0.990 (0.680)
R_{cryst}	0.192
R_{free}	0.238
Bond [Å]	0.0103
Angle [^{o}]	1.440

*Values in parentheses are for highest-resolution shell.

2.5.2. TrBgl2 complexed with isofagomine

Crystals of *Tr*Bgl2 liganded with the known inhibitor for any glycoside hydrolase, isofagomine (Figure 2.12) were obtained after co-crystallisation in the same conditions as apo *Tr*Bgl2. This novel crystal structure of the *Tr*Bgl2/isofagomine complex was determined at 1.9 Å resolution and contained four protein molecules in the asymmetric unit.

The structure was solved using the previously determined structure of apo *Tr*Bgl2 (Section 7.9.2). The electron density of the Fourier difference map, calculated by including all refined atoms except isofagomine, showed all atoms of the isofagomine which has displaced the TRIS molecule seen in the apo-structure. A total of 5 residues of *Tr*Bgl2 (Gln16, His119, Glu367, Glu424 and Trp425) form seven direct hydrogen bonds with isofagomine (Figure 2.12). Data-processing and refinement statistics are summarized in Table 2.2.
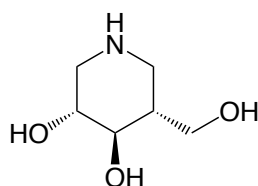
Figure 2.12. *Chemical structure of isofagomine.*

Table 2.2. *Data collection and refinement statistics for TrBgl2 complexed with isofagomine.*

X-ray-source	Diamond I04
Wavelength [A]	0.979499
Space group	P 1 21 1
a, b, c (Å)	93.003, 103.532, 94.370
α, β, γ (°)	90.00, 105.285, 90.00
Resolution range (Å)	74.46-1.93
Completeness (%)	100(99.9)
Multiplicity	4.2(4.2)
Total reflections	568096 (57558)
Unique reflections	135630 (13462)
$<I\sigma(I)>$	5.7(1.1)
R_{merge}	0.123(1.327)
R_{pim}	0.105(0.128)
$CC^{1/2}$	0.959(0.385)
R_{cryst}	0.20
R_{free}	0.25
Bond [Å]	0.01
Angle [°]	1.643

*Values in parentheses are for highest-resolution shell.
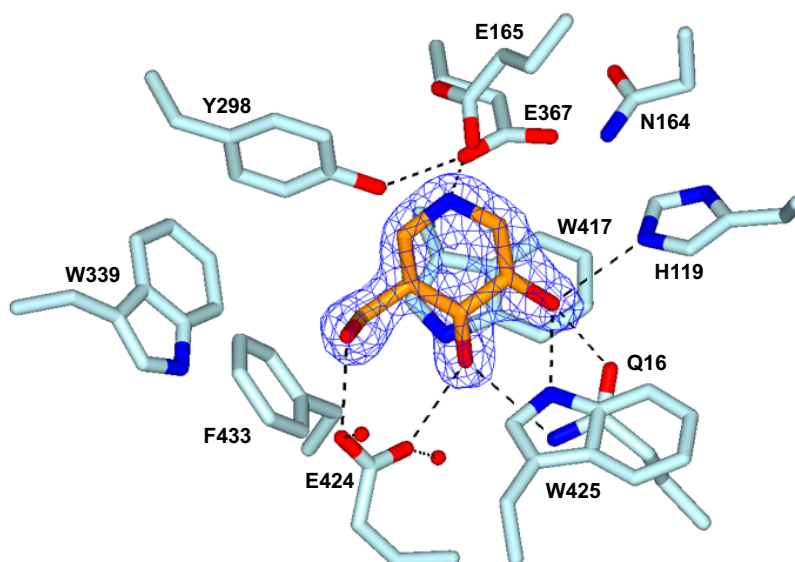
Figure 2.13. *Isofagomine binding site on TrBgl2. The active site is shown with key active site residues as stick diagram. The isofagomine-bound structure is shown as stick diagram, with carbon atoms in orange with the SA-F_O-F_C omit map contoured at 2.3σ r.m.s. The isofagomine is bound within TrBgl2 active site and forms H-bond to Glu367 (E367), His119 (H119), Gln16 (Q16), Trp 425 (W425) and Glu424 (E424). Oxygen atoms are shown in red and nitrogen in blue. Hydrogen bonds are shown as black dashed lines. The figure was created in CCP4MG [114].*

2.6. Assay development

The kinetic properties of *Tr*Bgl2 were investigated by continuous assay using synthetic *p*-Nitrophenyl-β-D-glucopyranoside (*p*NPG) and 4-methylumbelliferyl-β-D-glucopyranoside (MUG) as substrate at the optimal conditions (40°C, pH 6.0) -the previously determined pH and temperature optimum for enzymatic activity (Section 7.6) [2]. The Michaelis-Menten parameters were calculated (*p*NPG: k_{cat} = 34±4.5 s^{-1}, K_m = 1.73±0.69 mM, MUG: k_{cat} = 24±0.4 s^{-1}, K_m = 0.46±0.02 mM) (Figure 2.13.) through direct

fit to the Michaelis-Menten equation and are displayed in Table 2.3. *Tr*Bgl2 shows higher affinity and catalytic efficiency toward the latter substrate, MUG, which is consistent with the results obtained previously by Jeng [2].

Table 2.3. *Catalytic activity of TrBgl2*

p-Nitrophenyl-β-ᴅ-glucopyranoside (*p*NPG)			4-methylumbelliferyl-β-ᴅ-glucopyranoside (MUG)		
k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})
34±4.5	1.73±0.69	19.68±6.54	24±0.4	0.46±0.02	53±19.8
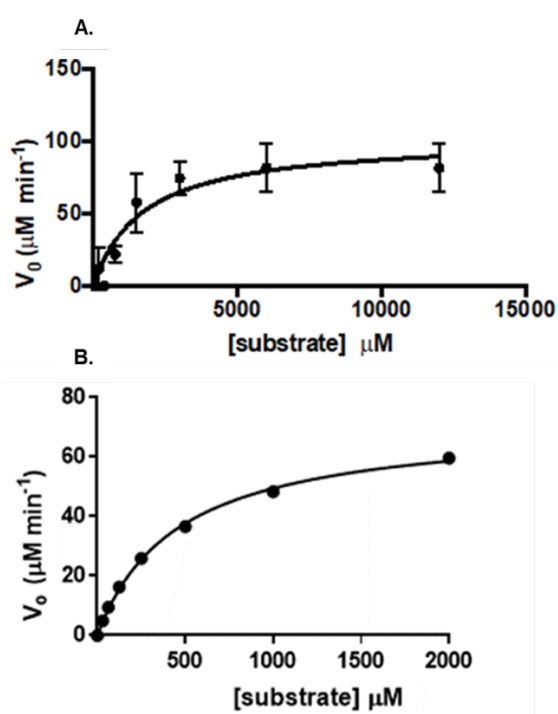
Figure 2.14. *Michaelis-Menten plots for TrBgl2. A. pNPG cleavage assay Michaelis-Menten plot for TrBgl2 (100 nM) and B. MUG cleavage assay Michaelis-Menten plot for TrBgl2 (100 nM).*

The inhibition efficiency of *Tr*Bgl2 by the known glycosyl hydrolase inhibitor isofagomine was assessed by continuous assay of hydrolysis of MUG at the optimal conditions (40°C, pH 6.0) (Figure 2.15) [115-116].

Isofagomine showed an inhibition of *Tr*Bgl2 with an IC_{50} value of 0.53±0.12 µM (Figure 2.15; IC_{50} =compound concentration causing 50% inhibition). Isofagomine can thus be used to categorize any hit compound as either competitive or non-competitive with isofagomine.
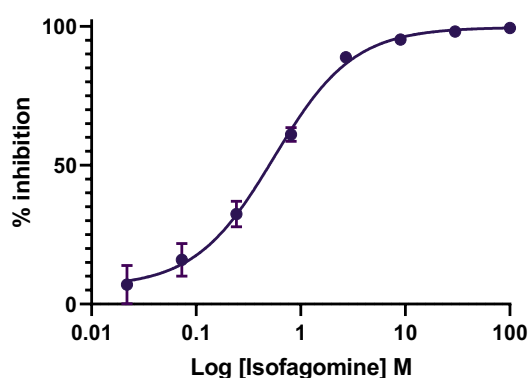
Figure 2.15. *Inhibition of TrBgl2 by isofagomine.*

2.7. Fragment screening

The York fragment library at the time of these experiments consisted of 570 compounds (MW less than 250 Da). This was screened in the *Tr*Bgl2 enzymatic assay using MUG as substrate at the optimal conditions of the enzyme (40°C, pH 6.0). The MUG activity-based assay was previously validated with the well-known inhibitor for any glycoside hydrolase, isofagomine to show that the assay is robust enough for compound screening (Section 2.6).

In the initial screen of 570 compounds, 65 fragments inhibited *Tr*Bgl2 activity (Appendix E) and 5 fragments enhanced the ability of *Tr*Bgl2 protein to cleave MUG. Only 1 (**21**) (Figure 2.16) of the 5 fragments was reconfirmed as activator after repeating the MUG activity-based assay at different compound concentrations and running negative controls in order to reject any false positives due to possible fluorescent compounds. Interestingly, when one of the initial hits (**31**) (Appendix D) was synthesized in-house by James Firth or when different chemical reagents such as 1 mM EDTA and 1 mM reducing agents (DTT, TCEP) were added in the presence of purchased compound were found to be inactive suggesting possible metal contamination.
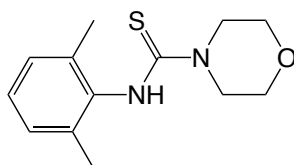
Figure 2.16. *Chemical structure of the initial hit (**21**)*

Kinetic characterization of the hit compound (**21**), using 500 µM MUG as a substrate and 100 nM *Tr*Bgl2 at the optimal conditions (40°C, pH 6.0), yielded measurable AC_{50} values and showed an approximately 160% maximum activation (Figure 2.17).

Rational exploration of commercially available analogues of **21** through an "analogue-by-catalogue" approach (with the help of Dr James Firth) was used to identify compounds that could have greater impact on the maximum rate of *Tr*Bgl2, better solubility and chemotypes that are essential for the activation.
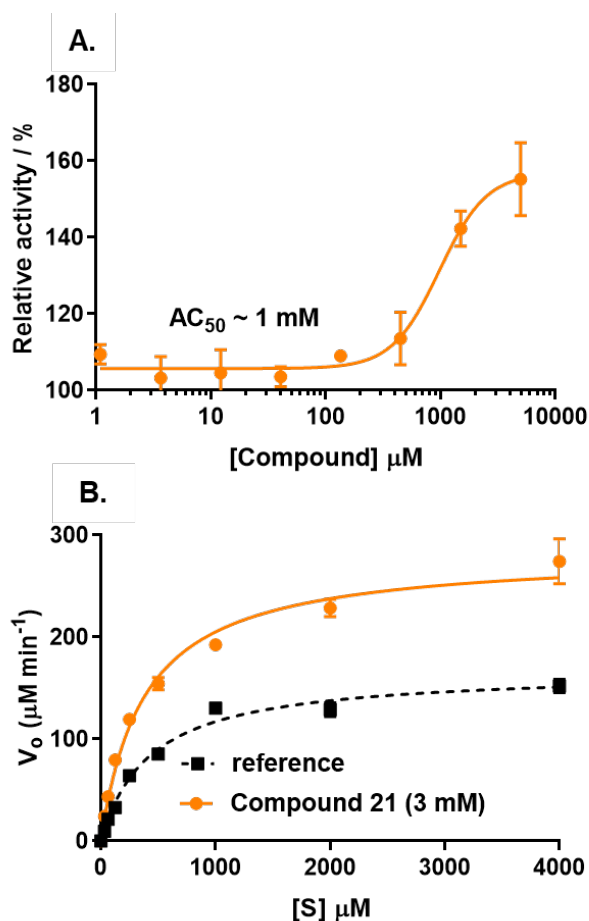
Figure 2.17. *A. MUG cleavage assay AC_{50} curve for **21**. Substrate concentration = 500 µM. B. MUG cleavage assay Michaelis-Menten plot for TrBgl2 in the absence and presence of **21**. [S] = substrate concentration.*

Replacement of the thiourea group by a thio-amide or replacement of sulphur by oxygen yielded an inactive analogue. Other available compounds that retained thiourea but had a methyl group meta or para to thiourea instead of ortho to thiourea were also inactive. On the other hand, as summarised in Table 2.4, chemical groups other than methyl group (chrolo, fluoro, methoxy) ortho to thiourea yielded active compounds. Attempts to alter the morpholino moiety were more successful and led to more active compounds.

Table 2.4. *Activator data from a MUG cleavage assay.*

Compound	Chemical structure	AC$_{50}$ [µM]	A$_{max}$ [%]
22		>5000[1]	~120[1]
23		~1800[1]	~150[1]
24		>5000[1]	~160[1]
25		>5000[2]	~140[2]
26		>5000[2]	~160[2]
27		~1500[2]	~170[2]
28		~1600[1]	~130[1]
29		>5000[1]	~140[1]

[1] Maximum activation obtained with a titration of compound from 30 µM to 3.8 mM (substrate conc. = 500µM)
[2] Maximum activation obtained with a titration of compound from 63 µM to 8 mM (substrate conc. = 500µM)
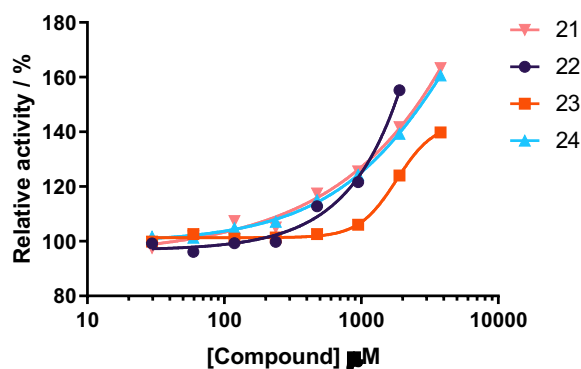
Figure 2.18. *MUG cleavage assay AC$_{50}$ curve for the hit compounds. AC$_{50}$ values obtained with a titration of compounds from 30 µM to 3.8 mM (substrate concentration = 500 μM). The optimal temperature (40°C) and the optimal pH of TrBgl2 (pH 6.0) were used in the assay.*
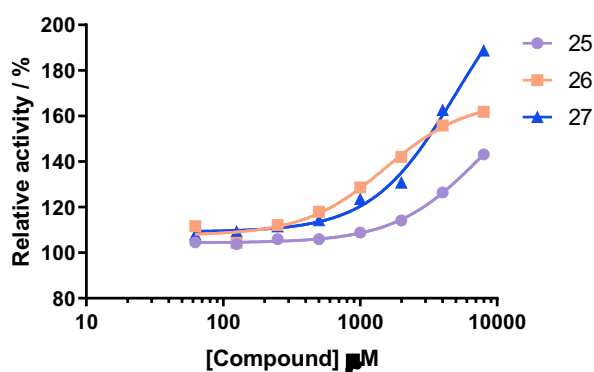
Figure 2.19. *MUG cleavage assay AC$_{50}$ curve for the hit compounds. AC$_{50}$ values obtained with a titration of compounds from 63 µM to 8 mM (substrate concentration = 500 μM). The optimal temperature (40°C) and the optimal pH of TrBgl2 (pH 6.0) were used in the assay.*

Methyl groups ortho to thiourea causes the aryl ring to twist out of the plane with the thiourea to limit steric clashes between the thiourea and the ortho groups. Groups other than methyl groups in the ortho position also result in this twist while substituents in the meta or para positions do not cause this to happen. This twist of aryl ring out of the plane with thiourea

might be essential for the activation and could explain why compounds that retained thiourea but had a group meta or para to thiourea instead of ortho to thiourea appeared to be inactive.

Compound **27** retains the important thiourea and the methyl group ortho to thiourea generating a better activator of *Tr*Bgl2 with improved maximum activation and much better solubility (up to 16 mM) (Table 2.4). The properties of this most potent activator (**27**) will be characterised below.

A series of kinetics experiments were performed with covariation of substrate and compound **27** concentration to study the mode of activator **27** interaction with *Tr*Bgl2. Substrate titrations were performed at four concentrations of **27** that confer 0, 25, 50 and 75% activation when measured at $[S] = K_m$ (Figure 2.21). The resulting data were fitted to nonessential reversible activator kinetic model, similar to reversible mixed inhibition model in which the compound displays binding affinity for both the free enzyme and the enzyme-substrate complex.

$$E + S + A \underset{}{\overset{K_M}{\rightleftharpoons}} ES + A \xrightarrow{k_{cat}} P + E + A$$

$$\big\updownarrow K_A \qquad\qquad \big\updownarrow \alpha K_A$$

$$EA + S \underset{}{\overset{\alpha K_M}{\rightleftharpoons}} ESA \xrightarrow{\beta k_{cat}} P + E + A$$
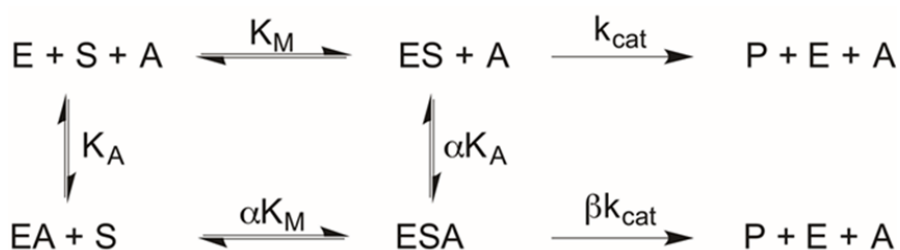
Figure 2.20. *Nonessential reversible activator kinetic model. E = enzyme; S = substrate; A = activator; P = product [73].*

In nonessential type activation model, the dissociation constant of the substrate and activator are modified in the presence of activator and

substrate respectively by the constant a. In addition, the k_{cat} of the enzyme in the presence of activator is modified by the constant β (Figure 2.20). The modifier a and β were calculated as 0.64 and 1.5 respectively, which indicate a decreased dissociation constant for the MUG substrate and increased catalytic rate in the presence of **27**.
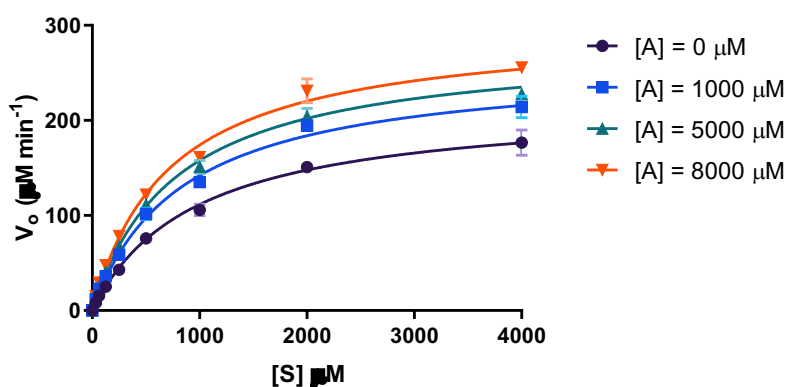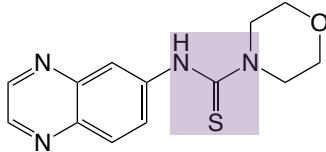
Figure 2.21. *Covariation curve fits to the non-essential activator model for activator **27**. [A] = compound concentration.*

Interestingly, one of the fragment analogues (Table 2.5) appears to inhibit *Tr*Bgl2.

Table 2.5. *Inhibition data from a MUG cleavage assay.*

Compound	Chemical structure	IC$_{50}$ [μM]	Maximum inhibition [%]
30		~300[2]	~95[2]

[2] Maximum activation obtained with a titration of compound from 63 μM to 8 mM (substrate conc. = 500μM)

A series of kinetics experiments were performed with covariation of substrate and compound **30** concentration. Substrate titrations were

performed at four concentrations of **30** that confer 0, 25, 50 and 75% inhibition when measured at [S] = K_m (Figure 2.23).
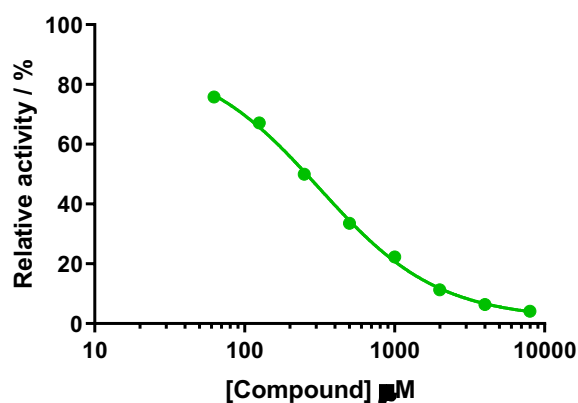
Figure 2.22. *MUG cleavage assay IC$_{50}$ curve for inhibitor **30**. IC$_{50}$ values obtained with a titration of compounds from 63 μM to 8 mM (substrate concentration = 500 μM). The optimal temperature (40°C) and the optimal pH of TrBgl2 (pH 6.0) were used in the assay.*
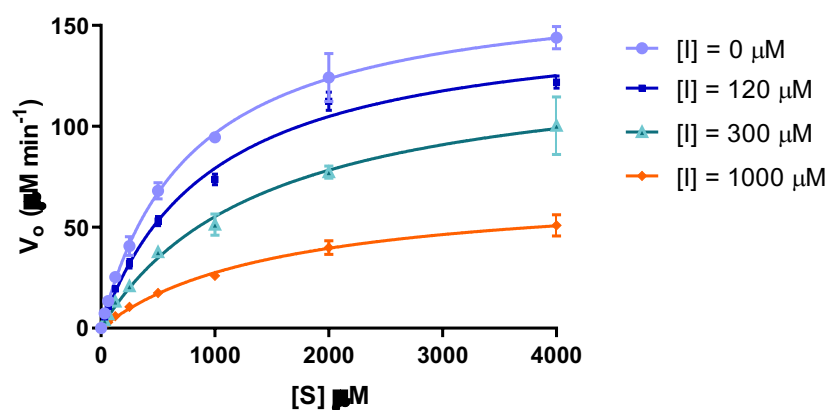
Figure 2.23. *Covariation curve fits to the reverse mixed inhibition model for inhibitor **30**. [I] = inhibitor concentration*

The resulting data were consistent with reversible mixed inhibition model. In this situation, two dissociation constants are defined, one for the binary enzyme-inhibitor complex (K_I) and one for the ternary ESI complex (αk_I).

The modifier α was calculated as 3.65 after covariation analysis for compound **30** which indicates an increased dissociation constant for the MUG substrate in the presence of **30**.

2.8. Effect of chemical reagents

There have been a number of reports (unpublished) [117] that changes in enzyme activity in high concentration screening can be non-specific due to contaminants in the ligand sample (such as traces of metals left over from synthesis), oxidation of the enzyme or due to aggregation of the ligand.
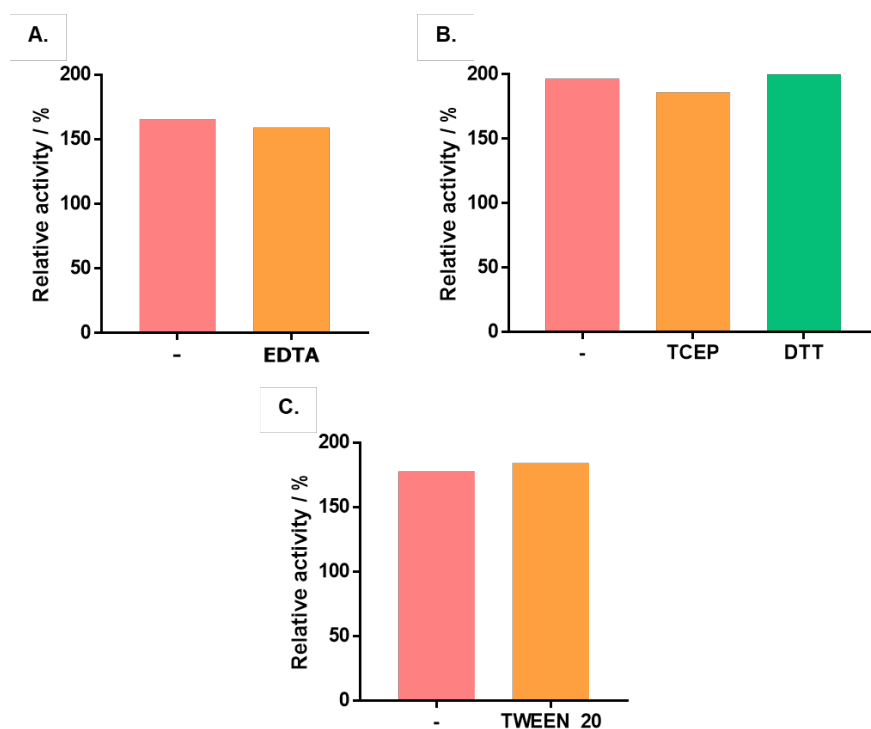
Figure 2.24. *Effect of chemical reagents on the TrBgl2 activity with the hit compound (**21**). A. Effect on the TrBgl2 activity in the presence of 1 mM EDTA, B. Effect on the TrBgl2 activity in the presence of 1 mM TCEP and DTT, C. Effect on the TrBgl2 activity in the presence of 0.1% Tween-20. The optimal temperature (40ºC) and the optimal pH of TrBgl2 (pH 6.0) were used in the assay.*

A series of experiments were therefore performed to investigate the effect of different chemical reagents. *Tr*Bgl2 with the hit compound **21** was assayed in the presence of 1 mM EDTA, 1 mM reducing agents (DTT, TCEP) or 0.1% Tween-20 as shown in Figure 2.24. No significant change in the activator activity can be observed indicating that the change in activity observed is due to specific binding of the activator.

2.9. Specificity

Chapter 5 will describe the work on a closely related homology enzyme, bacterial β-glycosidase *Clostridium cellulovorans*, *Cc*BglA (Figure 5.1). The activators showed no activity on this enzyme which demonstrates that the effect is specific. Interestingly, the inhibitor **30** inhibits both enzymes.

2.10. Conclusion

Expression, purification and kinetic analysis of *Tr*Bgl2 was carried out successfully following the published work by Jeng *et al* [2]. *Tr*Bgl2 does not show any oligomeric pattern after SEC- MALS analysis. The crystal structure of apo *Tr*Bgl2 and the novel structure of *Tr*Bgl2 in complex with the known inhibitor, isofagomine, was determined at 1.8 and 1.9 Å resolution, respectively and both contained four proteins in the asymmetric unit. In the structure of apo *Tr*Bgl2 a TRIS molecule was found to bind at the active site. This TRIS molecule was displaced by the inhibitor in the crystal structure of *Tr*Bgl2 in complex with the known inhibitor, isofagomine. *Tr*Bgl2 showed higher affinity and catalytic efficiency for the MUG substrate compared to *p*NPG confirming previous work [2].

In the initial screen of the York 570 fragment library, one compound appeared to enhance the ability of *Tr*Bgl2 to cleave the substrate (MUG). A series of experiments investigated the effect of different chemical reagents such as reducing agent and detergent. No significant change in the activator activity can be observed indicating that the change in activity observed is due to specific binding of the activator. The effect is selective as no such activation was observed for the homologous enzyme *Cc*BglA (see Chapter 5 for details). Characterisation of commercially available analogues of the initial hit allowed the identification of activators that have greater impact on the maximum rate of *Tr*Bgl2 yielding up to 170% increased activity and solubility; the AC_{50} ranged from 1500 to more than 5000 mM.

Interestingly, one of the fragment analogues of the initial hit displayed inhibition of *Tr*Bgl2 and also of *Cc*BglA.

The kinetics of *Tr*Bgl2 activation are consistent with a nonessential activation model similar to reverse inhibition. According to this model, the activator binds to both free enzyme and enzyme- substrate complex and the affinity of the substrate to the enzyme is increased in the presence of the activator leading to an enhanced catalytic rate via the enzyme-substrate- activator complex.

Biophysical characterisation is necessary to further investigate and confirm the binding of the hit compound to the *Tr*Bgl2. This is described in the following chapter (Chapter 3).

Chapter 3: Characterisation of compound binding to TrBgl2

The previous chapter describes the identification of compounds that bind to the enzyme *Tr*Bgl2 and increase its enzymatic activity. This chapter describes the studies to characterise the binding of the most potent activator to the enzyme, using NMR spectroscopy and thermal shift assay (TSA). The chapter begins with a description of the generation and initial characterisation of the catalytically inactive E367Q mutant of *Tr*Bgl2 that can be used to study the possible role of the substrate (MUG) to activator binding. NMR spectroscopy and TSA are then performed to confirm the binding of the hit compound and investigate the role of the substrate in ligand binding.

3.1. Inactive mutant, E367Q TrBgl2

A Gibson Assembly Cloning Kit (NEB) was used to mutate the catalytic nucleophile/base glutamic acid, Glu (E367) to glutamine, Gln (Q) in the previously produced plasmid pET-YSBLIC3C_*Tr*bgl2 (Figure 2.4).

DNA sequencing confirmed the success of the mutagenesis and was followed by transformation of the construct into the *E. coli* BL21 DE3 bacterial expression strain. The E367Q *Tr*Bgl2 inactive mutant was successfully overexpressed in soluble form (Section 7.4.1) and purified (Section 7.5.1) in the same manner as the wild type *Tr*Bgl2 yielding

approximately 60 mg/L with a purity greater than 90% as determined by SDS-PAGE.

The activity of the E367Q TrBgl2 mutant was tested in a kinetic assay using MUG as the substrate at the optimal conditions (40°C, pH 6.0) (Section 7.8). The mutant appeared to be completely inactive (Figure 3.1).
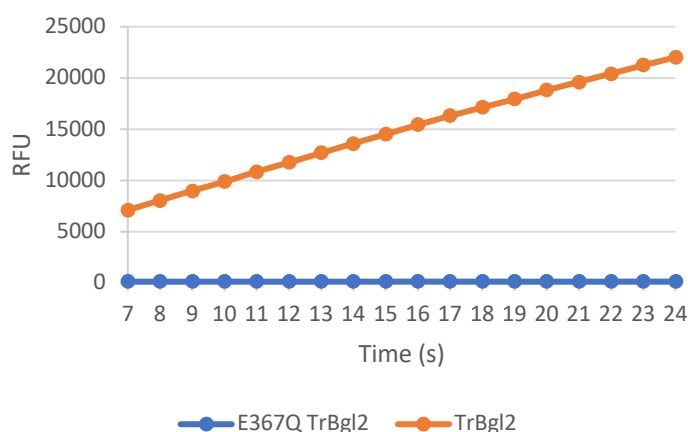
Figure 3.1. *Plot of Relative Fluorescent Units (RFU) over time for TrBgl2 in blue and E367Q TrBgl2 in orange.*

3.2. ^{1}H NMR spectroscopy

The binding of ligands to proteins can be studied in solution using NMR spectroscopy. The details of the physics of NMR are not described here but a summary of the principles of the technique is as follows [118].

The principle behind NMR is that certain nuclei (such as ^{1}H, ^{2}H, ^{13}C, ^{15}N) have a magnetic spin which can be converted from one spin state to another by absorbance of radiofrequency radiation. The frequency at which radiation is absorbed is dependent on the magnitude of the applied magnetic field, the type of atom and the covalent and non-covalent

environment of the atom. Therefore, information about the nucleus chemical environment can be derived from its resonant frequency. The chemical shift is defined by the frequency of the resonance relative to a reference compound in a magnetic field. Chemical shift is expressed in parts per million (ppm) and is independent of the spectrometer frequency.

$$\text{Chemical shift } (\delta) = \frac{\text{frequency of signal} - \text{frequency of reference}}{\text{spectrometer frequency}} \times 10^6$$

J-coupling (or scalar coupling) is a through shared electrons interaction between two nuclei in which the spin of one nucleus perturbs the spins of the second spin. This effect can cause the so-called "splitting" of the NMR signal for each type of nucleus in multiple peaks. In addition, energy can also be transferred through space from one population of spin-active nuclei (such as ^{1}H, ^{13}C and ^{15}N) to another that is close in space (known as the Nuclear Overhauser Effect, NOE).

As described in Section 7.11., these different properties can be explored through experiments using pulse sequences. The 1-dimensional (1D) NMR spectrum shows a signal at a given resonance frequency for each nucleus. A combination of pulse sequences can generate a 2-dimensional (2D) spectrum, where each axis reflects the chemical shift of the two nuclei whose relationship is being explored.

Figure 3.2 shows 1D ^{1}H NMR of *Tr*Bgl2 and the typical distribution of peaks expected in a folded protein sample. The amide proton region of the spectrum is well dispersed and contains distinctive amide peaks. The

presence of peaks at around 0.0 ppm (upfield) indicates folded methyl groups packing against aromatic amino acids in the hydrophobic core (Figure 3.2) [119-120].
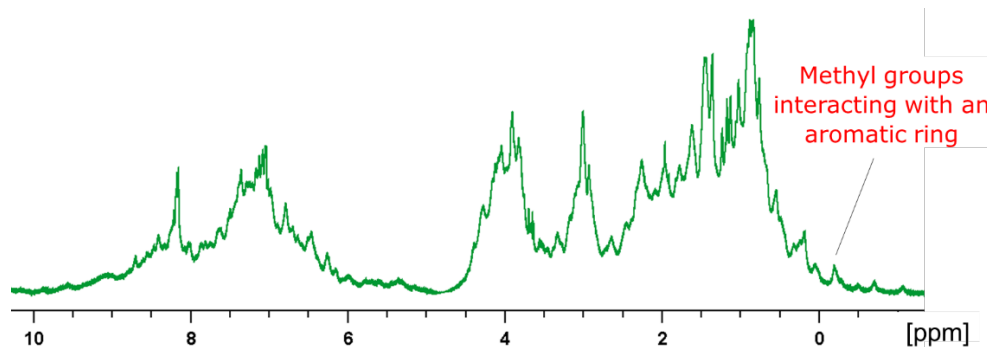
Figure 3.2. *1D ^{1}H NMR spectrum of TrBgl2. Spectrum acquired with 100 µM protein in D$_2$O.*

The typical ^{1}H 1D spectrum of ligands contains many fewer peaks, as can be seen in Figure 3.3. A number of NMR experiments can be carried out where changes in this 1D spectrum can report on the ligand binding to a protein.
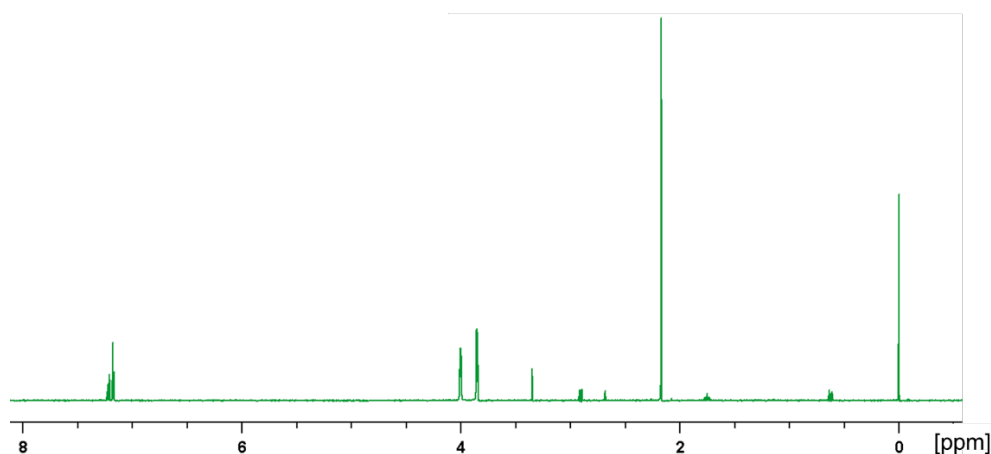
Figure 3.3. *1D ^{1}H NMR spectrum of the initial hit (**21**).*

All NMR spectra of this section were acquired in a phosphate buffer at pH 6.0. These conditions avoid the strong interference peaks generated in

protonated buffers and is sufficiently acidic to slow the exchange of some amide protons so that they are visible in the spectra.

3.2.1. Characterising the initial hit (21)

Compound aggregation has been implicated in giving false positive ("promiscuous" compounds) in fragment screening [121]. Therefore, a simple NMR assay, using dilution experiments, was used to detect possible formation of aggregates of **21**. This strategy is effective as aggregates are sensitive to changes in concentration and structural changes in the aggregate and its environment can be detected by ^{1}H NMR spectroscopy. Aggregating compounds can be identified by "unusual" features in the ^{1}H NMR spectrum, which appear with increasing concentration. For compounds that have "unusual" features, the resonances are broad as expected for slow-tumbling aggregate in solution, and there are changes in number, shape, and chemical shifts. On the other hand, for soluble compounds that have "normal" features, the resonances are sharp and only the resonance intensities are expected to change as a function of concentration [121].
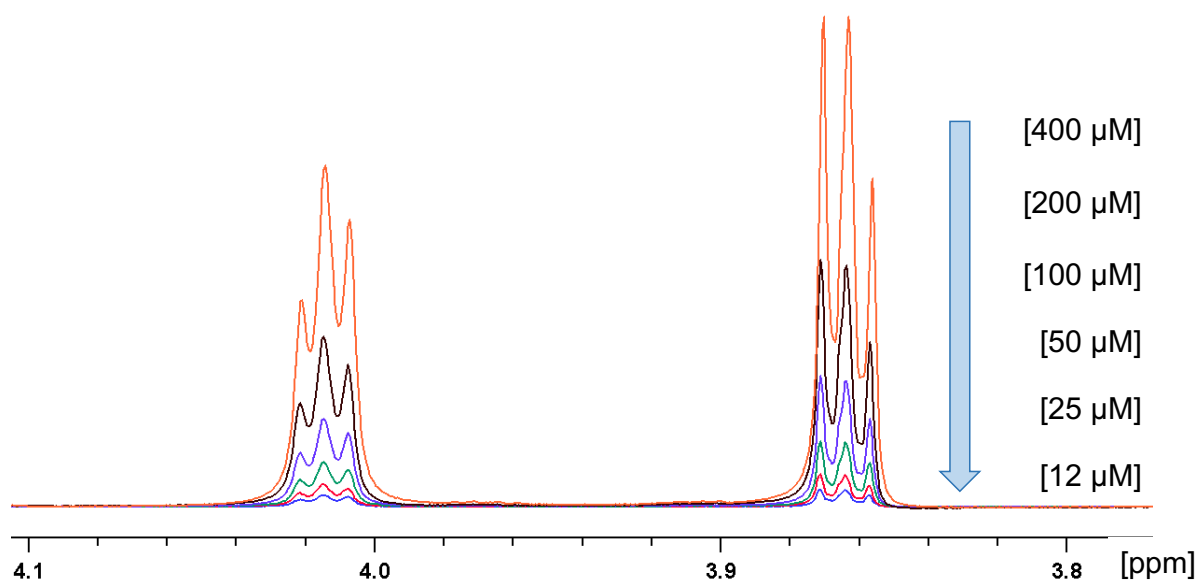
Figure 3.4. *NMR aggregation assay. A series of ^{1}H NMR spectra are superimposed from various concentrations of **21** (dilutions from 400 to 12 µM). Spectra were collected under identical conditions.*

NMR resonances of **21** are sharp at all concentrations and they do not shift left or right. In addition, there is no change in the number and shape of the resonances indicating a non-aggregating compound. However, for very large precipitates and solids that tumble too slowly, resonances are not expected, thus resulting in "NMR-invisible" large aggregates. To verify the presence of "NMR-invisible" aggregates, peak intensities were plotted as a function of **21** concentration. Resonance intensities correlated well with the nominal and measured concentrations which confirms that **21** is a soluble compound [121].

3.2.2. ^{1}H STD NMR

In the saturation transfer difference (STD) NMR experiment, radiofrequency pulses are applied to selected protein resonances (usually

the hydrophobic core); this magnetization is transferred to any bound ligand during the timescale of the experiment leading to partial saturation and a decrease in the intensity of those ligand resonances in the *on*-resonance spectrum. A second *off*-resonance spectrum is subsequently recorded with the selective excitation moved to a point where no protein or ligand signals are detectable (usually around 30 ppm). Subtraction of the *on*-resonance from the *off*-resonance spectrum results in the STD NMR spectrum which shows only ligand resonances that are in close contact with the protein binding site (Section 7.11.1).

Compound **27** and **30** are soluble under the conditions of the NMR experiment. Their ^{1}H NMR spectra show narrow lines and no signal overlap and therefore the STD approach can be utilized as a tool to characterize the binding of these fragments to the protein.

The STD NMR spectra of compound **27** and **30** were acquired at a ligand: *Tr*Bgl2 ratio of 50 and 25 respectively (Figure 3.5-3.6).
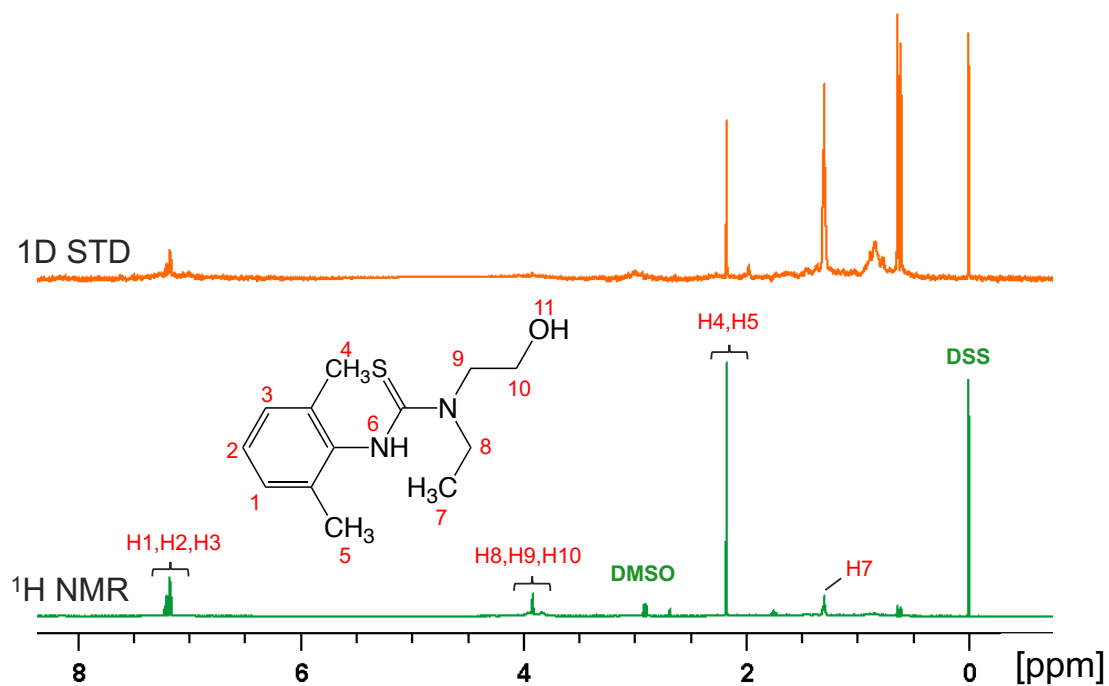
Figure 3.5. STD NMR spectrum of **27** (1 mM) with TrBgl2 (20 µM) shown in orange and ¹H reference spectrum shown in green (500 µM). Spectra were collected under identical conditions.
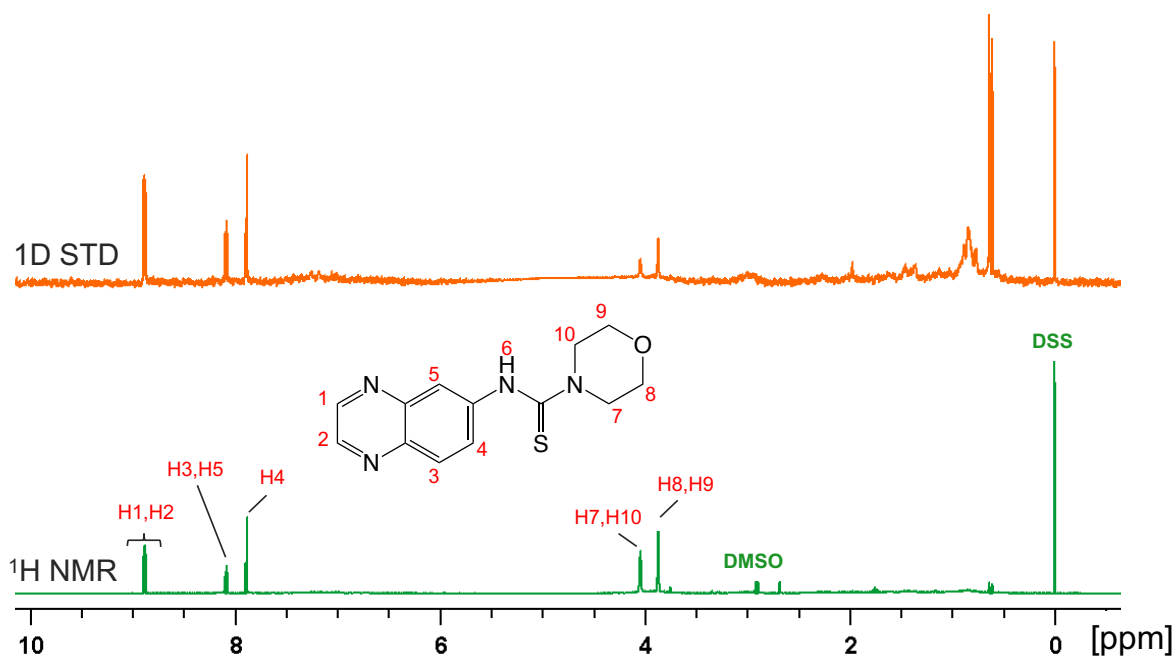
Figure 3.6. STD NMR spectrum of **30** (500 µM) with TrBgl2 (20 µM) shown in orange and ¹H reference spectrum shown in green (500 µM). Spectra were collected under identical conditions.

Evidence of fragment binding was observed in the STD NMR with a clear signal intensity for both compound **27** and **30** (Figure 3.5-3.6). Addition of the known glycosyl hydrolase inhibitor, isofagomine gave no change in intensity for either compounds. This allows categorisation of the hit compounds as being non-competitive with isofagomine (Figure 3.7-3.8).
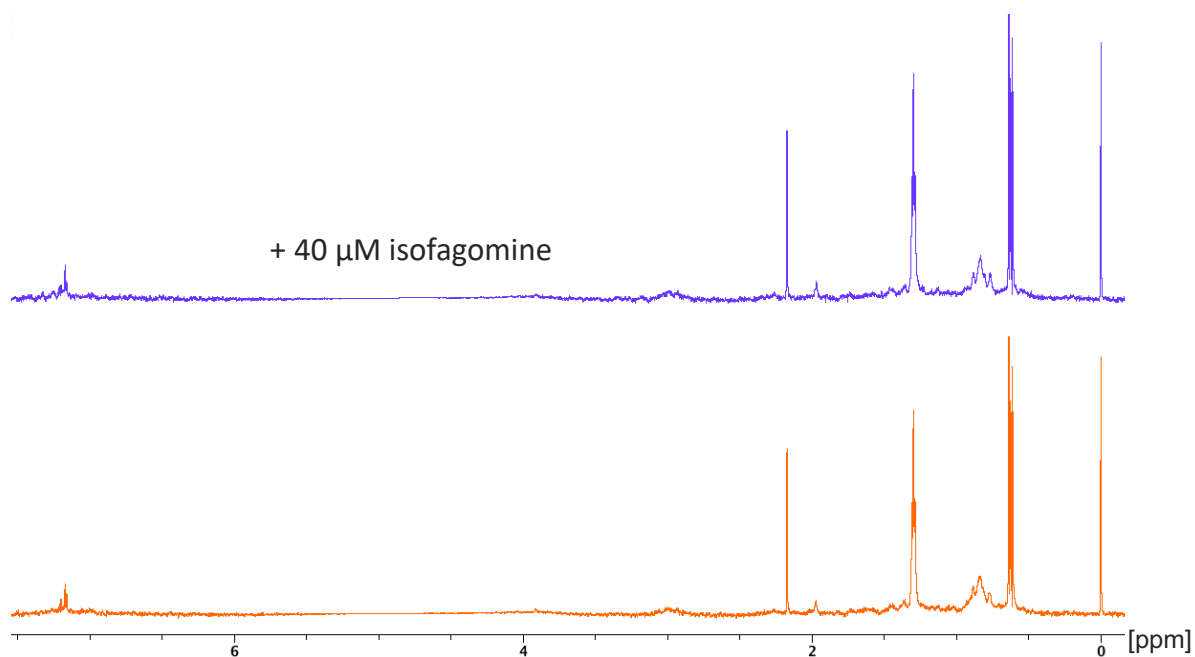
Figure 3.7. *STD NMR spectrum of* **27** *(500 µM) with TrBgl2 (20 µM) shown in orange and following addition of isofagomine (40 µM) shown in purple where several peaks appear to be the same in intensity following addition of isofagomine. Spectra were collected under identical conditions.*
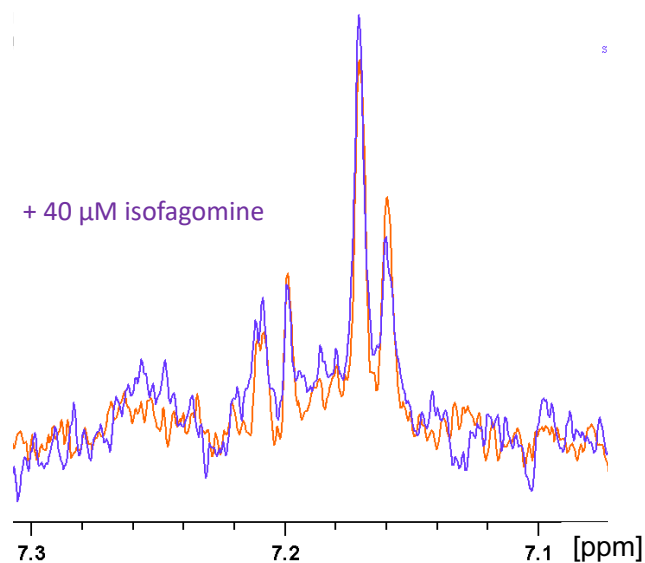
Figure 3.8. *STD NMR spectrum of **27** (500 μM) with TrBgl2 (20 μM) shown in orange and following addition of isofagomine (40 μM) shown in purple at 7.1-7.3 ppm where several peaks appear to be the same in intensity following addition of isofagomine. Spectra were collected under identical conditions.*

3.2.3. ^{1}H STD NMR in addition of the substrate

The possible role of substrate in the binding of **21**-like compounds to TrBgl2 was also investigated with STD NMR using the inactive E367Q TrBgl2 mutant that was previously generated. The ability of the E367Q TrBgl2 mutant to still bind the MUG substrate was confirmed by an STD NMR experiment (Figure 3.9).

The STD-NMR spectra of the mixture of **27** with E367Q TrBgl2, MUG with E367Q TrBgl2 and both **27** and the MUG substrate with E367Q TrBgl2 were acquired under identical conditions and are shown in Figure 3.10.
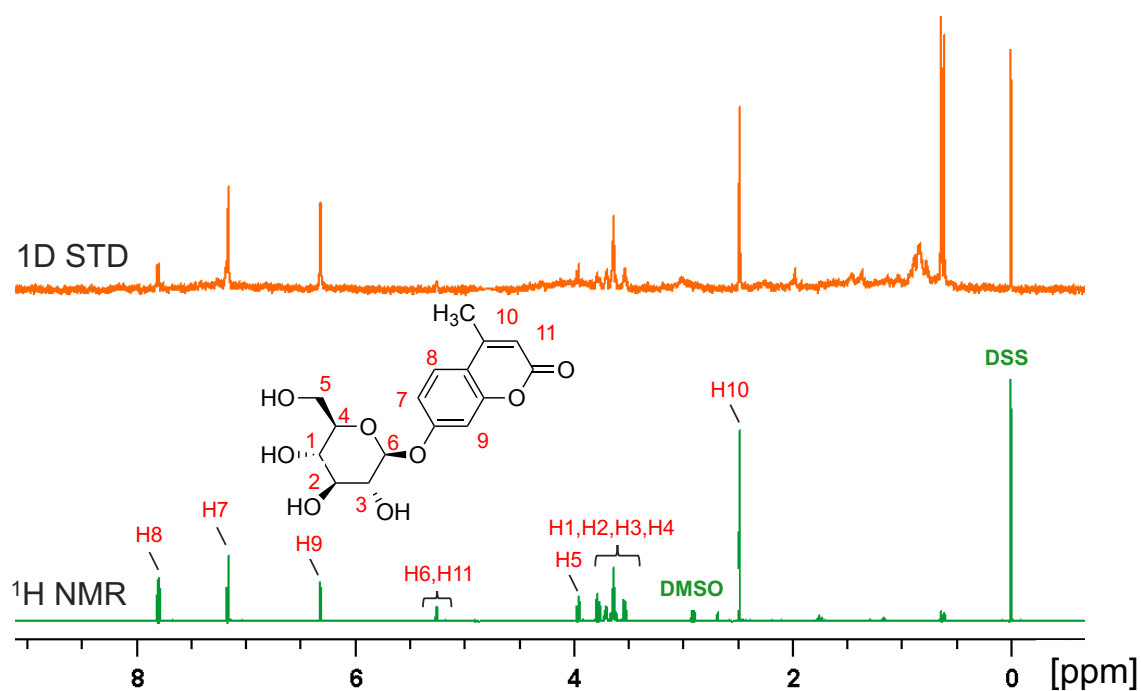
Figure 3.9. *STD NMR spectrum of MUG (500 μM) with E367Q TrBgl2 (20 μM) shown in orange and ^{1}H reference spectrum shown in green (500 μM). Spectra were collected under identical conditions.*
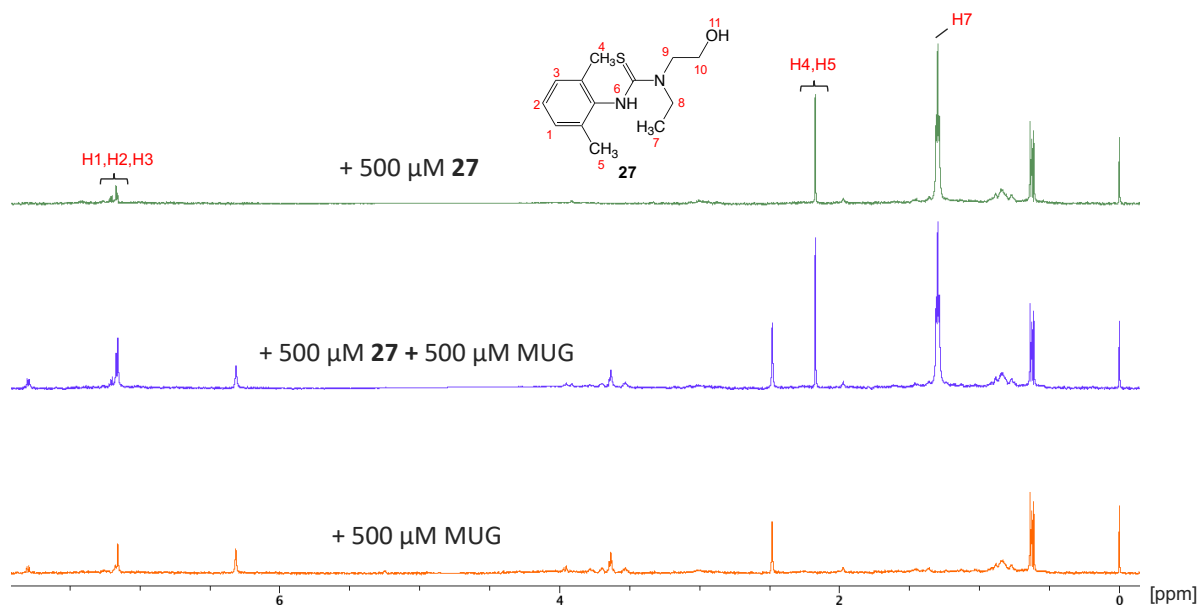
Figure 3.10. *STD NMR spectrum of **27** (500 μM) with E367Q TrBgl2 (20 μM) shown in green, MUG (500 μM) with E367Q TrBgl2 (20 μM) in orange and both **27** (500 μM) and MUG (500 μM) with E367Q TrBgl2 (20 μM) in purple. Spectra were collected under identical conditions.*

In the E367Q *Tr*Bgl2/**27**/MUG spectrum, all the STD signals of compound **27** and MUG are enhanced compared to E367Q *Tr*Bgl2/**27** and E367Q *Tr*Bgl2/MUG spectrums (Figure 3.10-3.13). This observation indicates that binding to **27** and the MUG substrate were stronger when both were present in full accordance with the kinetic experiments which suggest a decreased dissociation constant for the **27** and MUG in the presence of MUG and **27** respectively (Figure 2.21). These data suggest a significant substrate effect in activator binding to *Tr*Bgl2 through a nonessential reversible activator kinetic model as described in Section 2.7.
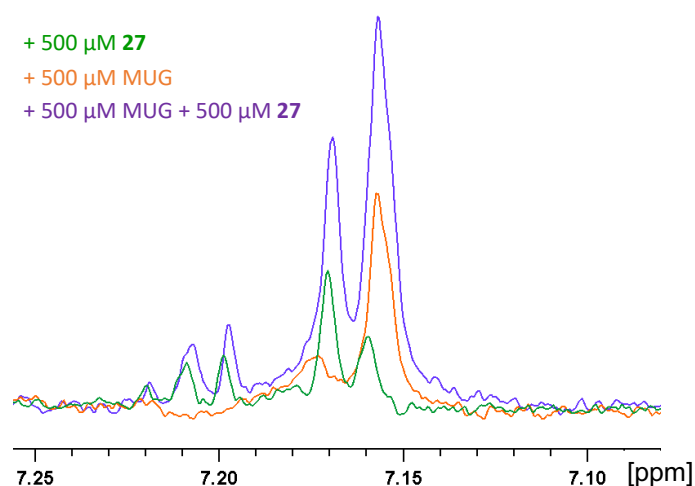
Figure 3.11. *STD NMR spectrum of **27** (500 μM) with E367Q TrBgl2 (20 μM) shown in green, MUG (500 μM) with E367Q TrBgl2 (20 μM) in orange and both **27** (500 μM) and MUG (500 μM) with E367Q TrBgl2 (20 μM) in purple at 7.10-7.25 ppm where the peaks for both **27** and the substrate, MUG were enhanced when both were present. Spectra were collected under identical conditions.*
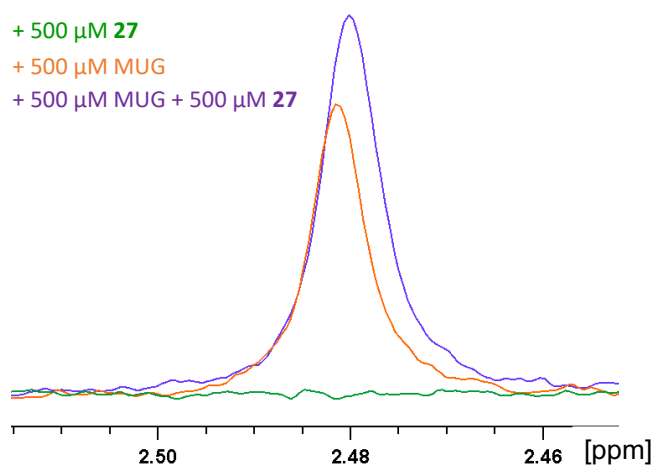
Figure 3.12. STD NMR spectrum of **27** (500 µM) with E367Q TrBgl2 (20 µM) shown in green, MUG (500 µM) with E367Q TrBgl2 (20 µM) in orange and both **27** (500 µM) and MUG (500 µM) with E367Q TrBgl2 (20 µM) in purple at 2.46-2.50 ppm where the peak for MUG was enhanced following addition of **27**. Spectra were collected under identical conditions.
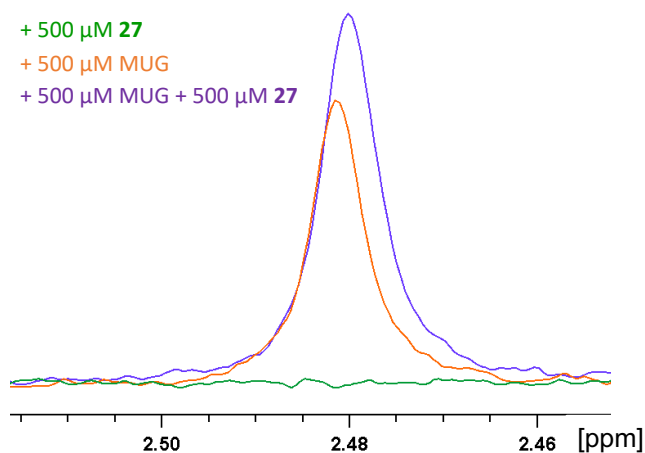
Figure 3.13. STD NMR spectrum of **27** (500 µM) with E367Q TrBgl2 (20 µM) shown in green, MUG (500 µM) with E367Q TrBgl2 (20 µM) in orange and both **27** (500 µM) and MUG (500 µM) with E367Q TrBgl2 (20 µM) in purple at 2.16-2.18 ppm where the peak for **27** was enhanced following addition of MUG. Spectra were collected under identical conditions.

3.2.4. Using WaterLOGSY NMR to characterise ligand binding

WaterLOGSY is another ligand-observed NMR experiment in which radiation is transferred from bulk water molecules to a ligand. Free ligands experience fast tumbling which results in a positive NOE (negative WaterLOGSY signal). By contrast, protein-bound ligands experience negative NOE effects (positive WaterLOGSY signal) because they are tumbling more slowly. Thefore, the observed WaterLOGSY signs are opposite for signals of free and protein-bound ligands which allows discrimination of binders and non-binders [122-123].

The WaterLOGSY spectra of the mixture of **30** with E367Q *Tr*Bgl2, MUG with E367Q *Tr*Bgl2 and both **30** and MUG with E367Q *Tr*Bgl2 were acquired at identical conditions and are shown in Figure 3.14.

The WaterLOGSY spectrum of **30** with E367Q *Tr*Bgl2 shows positive signals for the resonances H1, H2, H3, H4 and H5 which represent the typical observation of a bound-ligand. The transfer of NOE from the protein to the bound ligand depends on distance and flexibility. Directly bound atoms experience the most efficient NOE built up, while the NOE transfer is slower and of lower magnitude for more distant and flexible regions. This explains why the resonances of the H7, H8, H9, H10 behave more like free ligand peaks exhibiting negative signal. This behaviour could indicate that this region of the ligand is accessible to bulk water because it may not be completely buried in the binding site but protrudes from the protein.
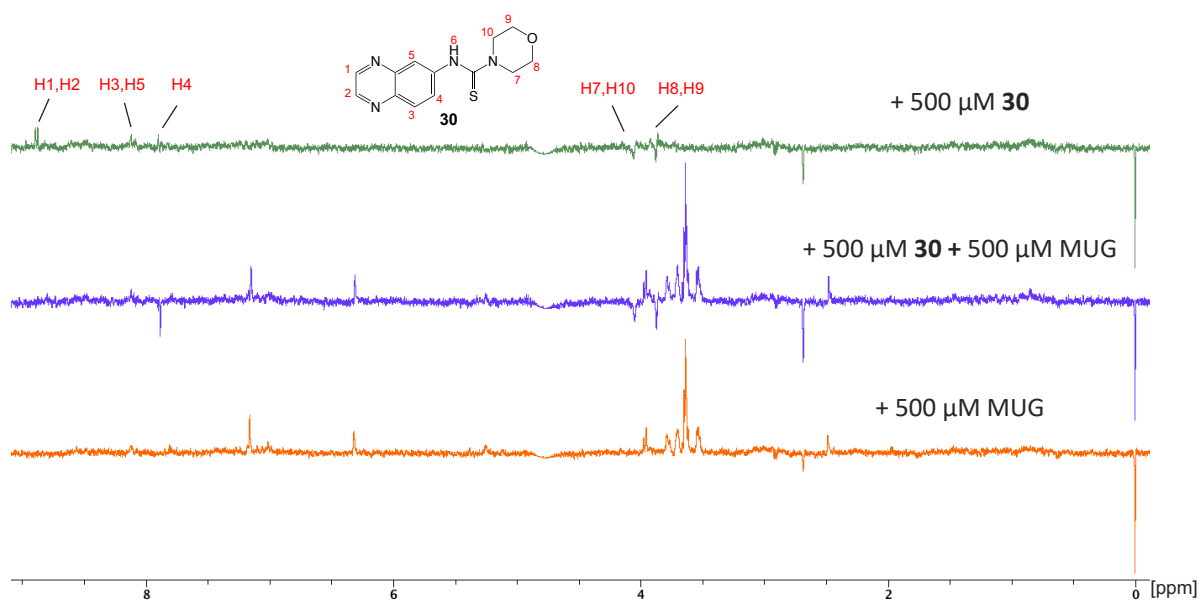
Figure 3.14. *WaterLOGSY spectrum of MUG (500 μM) with E367Q TrBgl2 (20 μM) shown in orange, 30 (500 μM) with E367Q TrBgl2 (20 μM) shown in green and both MUG (500 μM) and 30 (500 μM) with E367Q TrBgl2 (20 μM) shown in purple. Spectra were collected under identical conditions.*

In the E367Q *Tr*Bgl2/**30**/MUG spectrum, the signals for the resonances H1, H2, H4, H7, H8, H9 and H10 of compound **30** were degraded compared to that for E367Q *Tr*Bgl2/**30**, whereas the signals for the resonances H3 and H5 seem to be not affected (Figure 3.14-3.15). It's likely that the addition of MUG exposes the most of compound's groups to bulk water while H3 and H5 are still buried in the protein. This observation indicates weaker compound **30** binding in the presence of MUG in full accordance with the kinetic experiments performed which show an increased dissociation constant for the **30** in the presence of MUG (Figure 2.23).
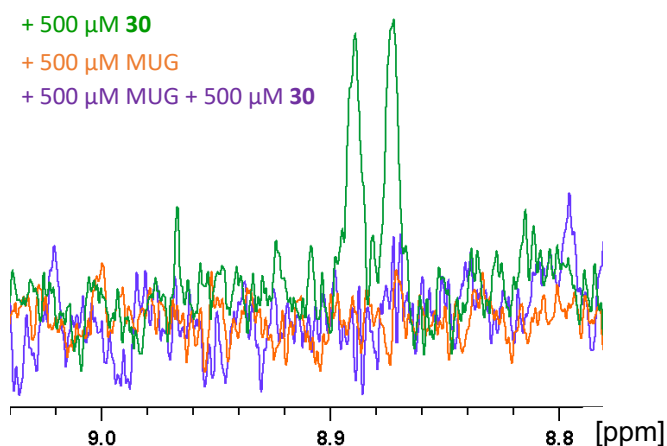
Figure 3.15. *WaterLOGSY spectrum of MUG (500 µM) with E367Q TrBgl2 (20 µM) shown in orange, **30** (500 µM) with E367Q TrBgl2 (20 µM) shown in green and both MUG (500 µM) and **30** (500 µM) with E367Q TrBgl2 (20 µM) shown in purple at 8.8-9 ppm where the peak for **30** was degraded following addition of MUG substrate. Spectra were collected under identical conditions.*

3.3. Thermal shift assay (TSA)

TSA can be used to detect ligand binding which changes the thermostability of the protein. This approach uses an environmentally sensitive dye, SYPRO orange, which binds non-specifically to hydrophobic surfaces and water quenches its fluorescence. By gradually increasing the temperature, the protein unfolds, and the exposed hydrophobic surfaces bind to the dye, increasing the fluorescence by excluding the quenching water to provide an estimation of the T_m of the protein.

The influence of fragment binding on E367Q *Tr*Bgl2 stability was assessed by using a TSA, which defined the melting temperature (T_m) of E367Q *Tr*Bgl2 with and without the substrate, MUG (Figure 3.16).

Both fragments, compound **27** and **30**, have a destabilising effect on apo E367Q *Tr*Bgl2 by approximately 1 and 4°C respectively. This may be a contributing factor to the difficulty in obtaining co-crystal structures of these fragments with apo *Tr*Bgl2.

In contrast, MUG has a stabilising effect on E367Q *Tr*Bgl2/**27** complex whereas the addition of MUG does not affect the significantly large destabilising effect on E367Q *Tr*Bgl2 caused by **30** (Figure 3.16).
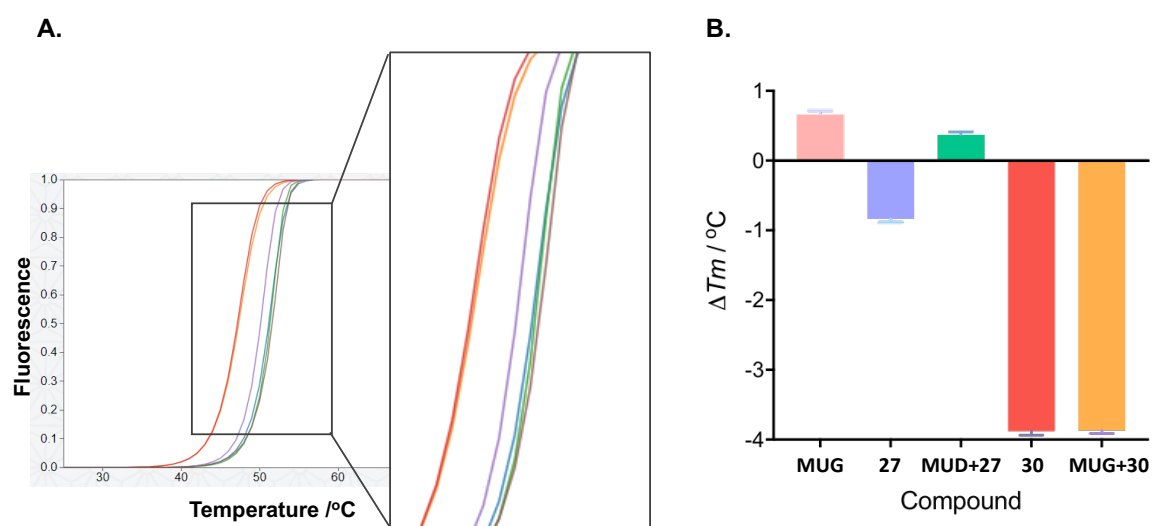
Figure 3.16. *TSA data. A. DSF curves for TrBgl2 in the presence of **27** (8 mM) in purple, **27** (8 mM) and MUG (4 mM) in green, **30** (8 mM) in red, **30** (8 mM) and MUG (4 mM) in orange compared to TrBgl2 in presence of DMSO in blue and TrBgl2 in presence of MUG (4 mM) in pink. B. Maximum shift in melting temperature from the reference (ΔT_m).*

3.4. Conclusion

A ligand-observed NMR experiment (STD NMR) was used to confirm the binding of the hit compounds to *Tr*Bgl2. These experiments were repeated following addition of the known inhibitor (isofagomine) allowing the categorisation of the hit compounds as non-competitive with isofagomine.

In order to investigate the possible role of substrate to compound binding, an inactive mutant E367Q *Tr*Bgl2 was generated for further ligand-observed NMR and TSA experiments. Substrate and activator binding are enhanced in the presence of activator and substrate respectively. On the other hand, the fragment inhibitor displayed reduced affinity to *Tr*Bgl2 in the presence of the substrate. In addition, TSA experiments show that the substrate causes a stabilising effect on the enzyme-activator complex increasing slightly the melting temperature of the enzyme while the binding of substrate on enzyme-inhibitor complex significantly destabilises it reducing the melting temperature of the enzyme by approximately 4 degrees. These results are in full agreement with the kinetic experiments performed in Section 2.7 and suggest that these series of analogues are allosteric modulators of *Tr*Bgl2. They can either allow the enzyme/substrate complex to adopt a productive conformation increasing the activity of the enzyme or a non-productive conformation inhibiting the enzyme.

Structural insight into the site at which the activator binds is necessary in order to investigate the mechanism of the activation. As no X-ray structure of the *Tr*Bgl2/activator complex is available, more sophisticated NMR

experiments are needed which require isotopic labelling of *Tr*Bgl2 and are described in the following chapter.

Chapter 4: Structural characterisation of compound binding to TrBgl2

The previous chapter described characterisation of the binding of the most potent activator compound to *Tr*Bgl2 using NMR spectroscopy and TSA. This included generation of the catalytically inactive E367Q mutant of *Tr*Bgl2 which is used to study the role of the substrate (MUG) in activator binding. This chapter describes experiments that attempt to obtain structural insight into the binding site of the activator using protein crystallography, NMR spectroscopy and modelling. The objective was to obtain a model of the mechanism of activation of *Tr*Bgl2.

4.1. Attempted ligand complex formation

A number of attempts were made to determine the crystal structure of *Tr*Bgl2 complexed with an activator. Compound **27** was chosen for the crystallisation attempts as it appears to be the most soluble at high compound concentrations compared to the other analogues.

The first attempt was to soak 10, 20 and 30 mM final concentration of activator **27**, into apo *Tr*Bgl2 crystals followed by 2-4 h and overnight incubations. This proved unsuccessful. Seeding strategies were explored by adding microcrystals of apo *Tr*Bgl2 to the mixtures of *Tr*Bgl2/**27** in solution in the crystallisation conditions that previously proved successful for crystallisation of apo *Tr*Bgl2 (Section 7.9.1). Tiny amounts of powdered activator were also added directly into apo *Tr*Bgl2 crystals letting the

activator dissolve slowly. Diffraction data obtained under these conditions produced identical models with no observable density corresponding to Compound **27**. Despite sustained attempts at co-crystallisation, involving the rescreening (Table 7.1) of *Tr*Bgl2/**27**, E367Q *Tr*Bgl2/MUG/**27** and *Tr*Bgl2/isofagomine/**27** mixtures containing saturated substrate (MUG)(8 mM) and isofagomine (1 mM), it was not possible to identify conditions which gave successful growth of crystals with activator bound. The final concentration of **27** used in these experiments ranged from 20 to 30 mM.

It appears that DMSO inhibits crystal growth, as co-crystallisation of a *Tr*Bgl2/DMSO mixture did not generate crystals in conditions that otherwise did give crystals. Therefore, additional attempts were made after dissolving **27** at 200 mM in ethanol and methanol. These also proved unsuccessful.

In addition to DMSO, it was also seen that TRIS binds tightly in the active site of apo *Tr*Bgl2 as it mimics the positive transition state charge of the reaction. As TRIS is present in the crystallisation medium, this presumably further reduces the chance of success for soaking crystals with weakly binding (1.5 mM) compound **27**. Rescreening of apo *Tr*Bgl2 for crystallisation in buffers that did not contain TRIS was not successful (Table 7.1).

4.2. ^{15}N labelled protein production and purification

*Tr*Bgl2 was expressed from the pET-YSBL3C_*Trbgl2* construct (Section 2.2) in M9 minimal media supplemented with ^{15}N NH$_4$Cl (Section 7.4.2). The protein was purified according to the previously used protocol for *Tr*Bgl2 (Section 2.3) yielding approximately 50 mg/L culture medium.

4.3. 2D ^{1}H ^{15}N Transverse relaxation-optimised spectroscopy (TROSY) NMR

NMR spectroscopy was used to validate the binding of compound **27** to *Tr*Bgl2. The size of the protein (55 KDa) means that the usual ^{1}H ^{15}N HSQC experiments were not possible as the fast transverse relaxation times (*T2)* of large molecules leads to line broadening such that the spectra cannot be interpreted. An alternate experiment is to use 2D ^{1}H ^{15}N Transverse relaxation-optimised spectroscopy (TROSY) NMR which allows studies of larger proteins (>30 kDa) [124]. TROSY NMR results in a correlation of a protein's amide proton and nitrogen resonances generating one peak for each backbone amide. Ligand binding will induce chemical shifts of the protein's ^{1}H and ^{15}N nuclei (Section 7.11.3).

4.3.1. 2D ^{1}H ^{15}N TROSY NMR of apo TrBgl2

Figure 4.1 shows the 2D ^{1}H ^{15}N TROSY spectrum of apo *Tr*Bgl2 at 120 µM in 50 mM TRIS buffer containing 100 mM NaCl and 10% D$_2$O at pH 8.0. The 2D ^{1}H ^{15}N TROSY spectrum of apo *Tr*Bgl2 was also recorded in 50 mM phosphate buffer containing 25 mM NaCl and 10% D$_2$O at pH 6.0 in order

to be consistent with the buffer used in the kinetics experiments (Section 2.7), however the spectra quality was poor, with several peaks being missing at pH 6.0 that were visible at pH 8.0. Although more amide peaks were expected to be in intermediate exchange with bulk solvent at pH 8.0, it seems that the protein is less stable in the absence of TRIS and at lower salt concentration at pH 6.0 resulting in broadening of the signal beyond detection [125]. Therefore, the generally accepted rule that lower pH gives better NH spectra seems to not be the case for *Tr*Bgl2. The quality of apo *Tr*Bgl2 spectra obtained at pH 8.0, which includes mainly separate and distinguishable peaks is unusually good for such a large protein. Each residue that has an amide proton attached to a nitrogen in the peptide bond is visible except for proline. The sidechains of asparagine and glutamine residues are observed to yield two peaks at the same nitrogen but different hydrogen values.

The ^{1}H ^{15}N TROSY spectrum of apo *Tr*Bgl2 contains 403 peaks (including sidechains) out of the 529 peaks predicted from the sequence of the *Tr*Bgl2 sample (Figure 4.1).
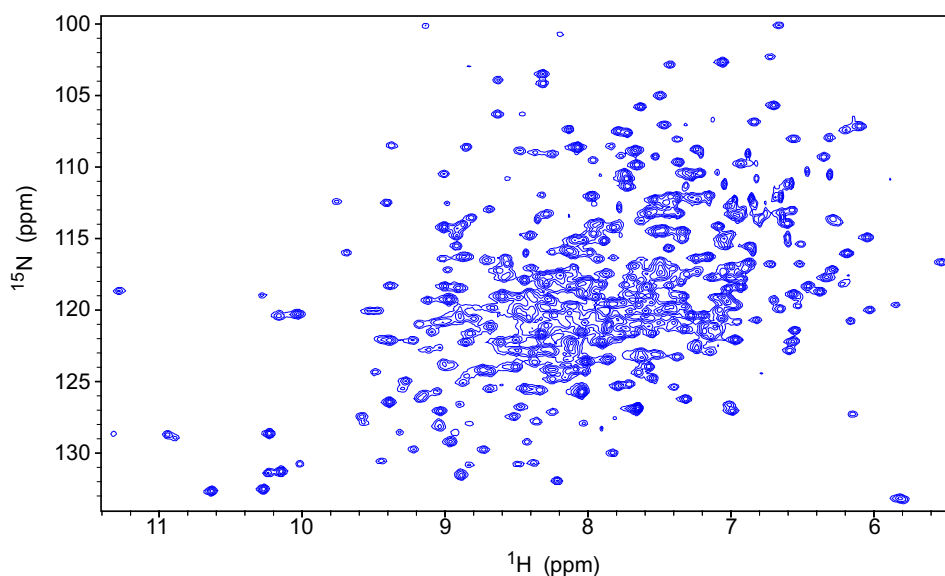
Figure 4.1. ^1H ^{15}N *TROSY spectrum (700 MHz, 298 K) of apo TrBgl2 at 120 µM.*

The ^1H ^{15}N TROSY spectra were unchanged following protein storage at 298 K for several (~5) days. This demonstrates that the sample is sufficiently stable for long multi-dimensional NMR experiments which can be used for studies of *Tr*Bgl2 in complex with the compound and substrate to investigate the mechanism of activation and the possible role of the substrate in compound binding.

4.3.2. 2D ^1H ^{15}N TROSY NMR of TrBgl2 with activator 27

The activity of **27** on *Tr*Bgl2 was first tested in a kinetic assay using MUG as the substrate in 50 mM TRIS buffer at pH 8.0. Compound **27** showed similar activity on *Tr*Bgl2 in both phosphate and TRIS buffer at pH 6.0 and 8.0, respectively (Figure 4.2).
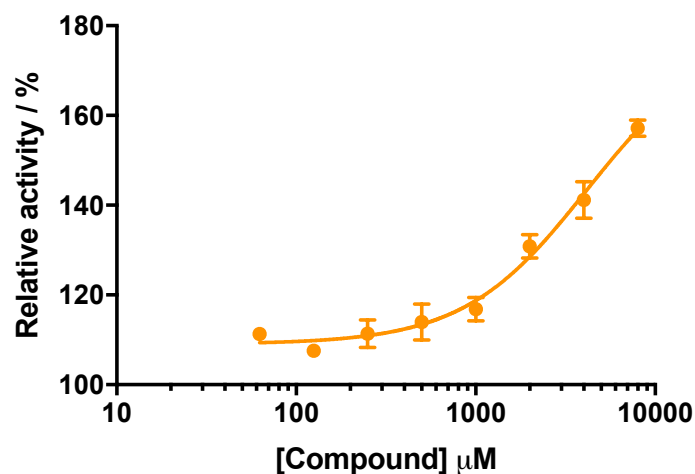
Figure 4.2. *MUG cleavage assay AC_{50} curve for compound **27** at pH 8.0. AC_{50} values obtained with a titration of compound from 63 µM to 8 mM (substrate concentration = 500 µM). The optimal temperature ($40^{o}C$) was used in the assay.*

Figure 4.3 shows the TROSY spectrum of apo *Tr*Bgl2 at 120 µM (red), *Tr*Bgl2 at 120 µM with **27** at 8 mM (light purple) and *Tr*Bgl2 at 120 µM with **27** at 16 mM (dark purple) in 50 mM TRIS buffer containing 100 mM NaCl and 10% D_2O at pH 8.0. The peaks of all spectra are very well dispersed, characteristic of a fully folded protein. Following addition of **27** at 8 mM and 16 mM some changes in amide chemical shifts across the protein were observed, clearly indicating a specific binding event and not any denaturation or other nonspecific effect (Figure 4.3-4.4).
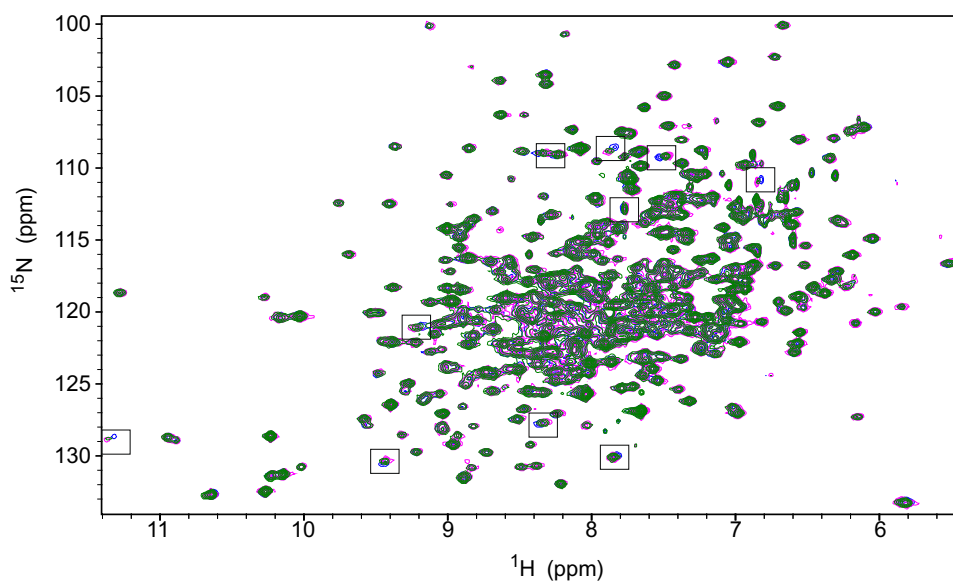
Figure 4.3. ^{1}H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo TrBgl2 (blue) overlaid with spectra of TrBgl2 with **27** at 8 mM (green) and 16 mM (magenta) concentration. Spectra collected under identical conditions.
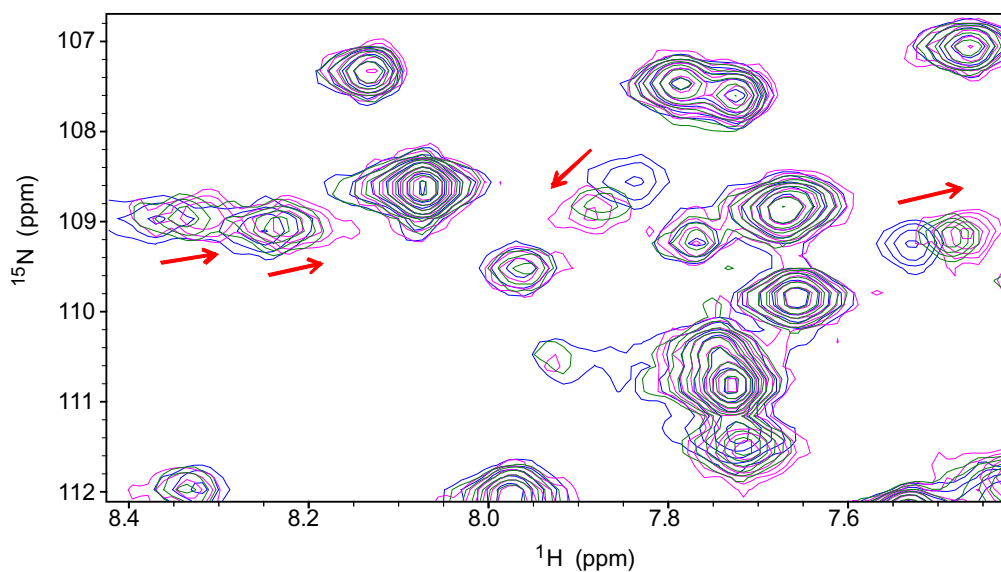
Figure 4.4. Expanded panel of ^{1}H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo TrBgl2 (blue) overlaid with spectra of TrBgl2 with **27** at 8 mM (green) and 16 mM (magenta) concentration. Red arrows indicate the direction of chemical shifts on increasing ligand concentration. Spectra collected under identical conditions.

4.3.3. 2D ^1H ^{15}N TROSY NMR of TrBgl2 with inhibitor 30

Figure 4.5 shows the TROSY spectra of apo *Tr*Bgl2 at 120 μM (blue) and *Tr*Bgl2 at 120 μM with **30** at 8 mM (red) in 50 mM TRIS buffer containing 100 mM NaCl and 10% D_2O at pH 8.0. The peaks of all spectra are very well dispersed, characteristic of a fully folded protein. Following addition of **30** at 8 mM, smaller changes in amide chemical shifts were observed than those observed on addition of **27** indicating that a binding event had taken place. It was not possible to measure the TROSY spectra of *Tr*Bgl2 with **30** at concentration higher than 8 mM due to the limited solubility of the compound.
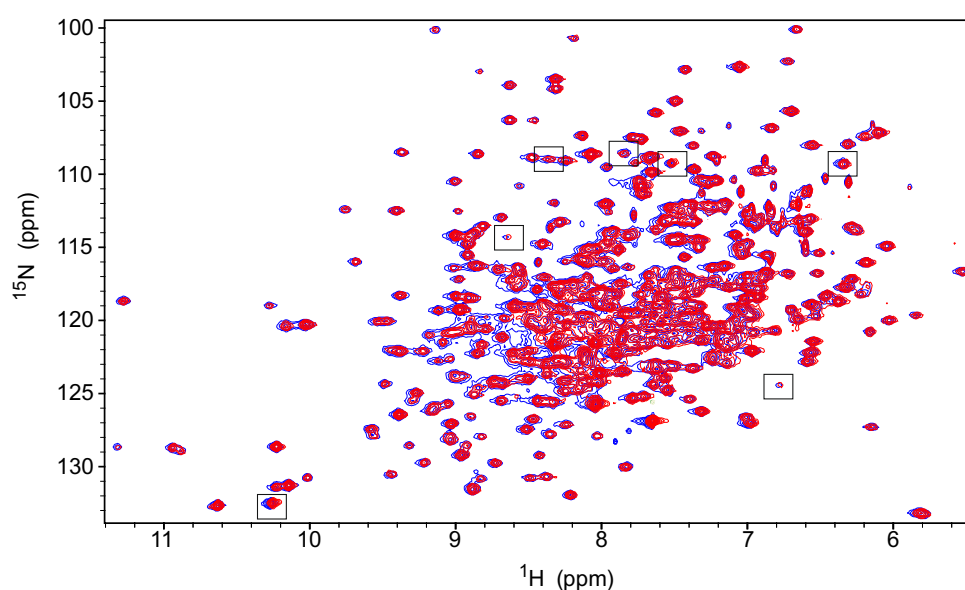
Figure 4.5. *^1H ^{15}N TROSY spectrum (700 MHz, 298 K) of TrBgl2 (blue) overlaid with spectra of TrBgl2 with **30** at 8 mM (red) concentration. Spectra collected under identical conditions.*

4.3.4. 2D ^1H ^{15}N TROSY NMR of E367Q TrBgl2

The possible role of substrate in fragment binding was further investigated by TROSY NMR. Figure 4.6 shows the TROSY spectra of apo E367Q TrBgl2 at 120 µM in 50 mM TRIS buffer containing 100 mM NaCl and 10% D_2O at pH 8.0.
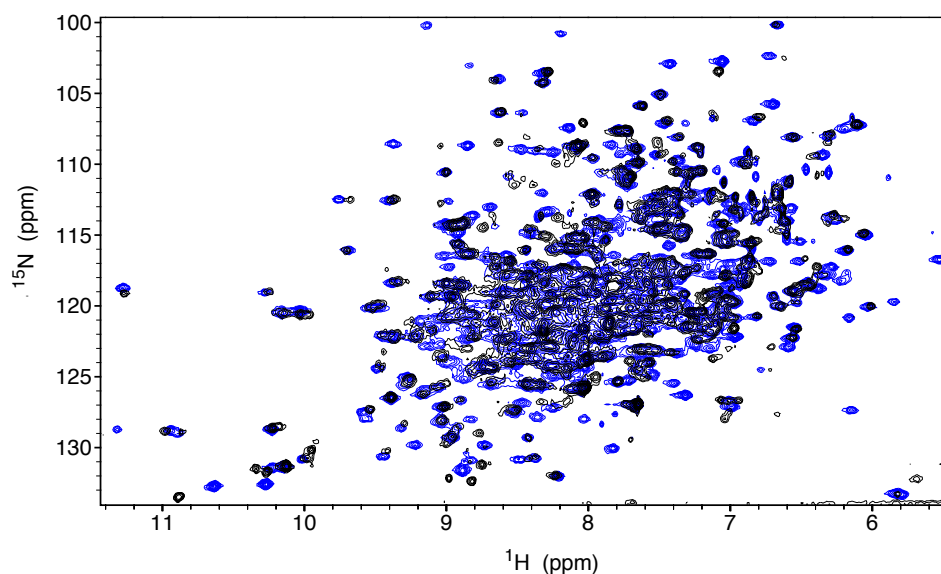
Figure 4.6. ^1H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo E367Q TrBgl2 (black) overlaid with spectra of apo TrBgl2 (blue) at 120 µM.

Although the peaks of the spectra are very well dispersed, characteristic of a fully folded protein, the resonances are broader than observed in the wild type TrBgl2 spectra (Figure 4.6). This may be due to internal dynamic and/or conformational changes of TrBgl2 caused by the inactive mutation introduced to the wild type protein. A single amino acid substitution can influence protein structure stability and binding affinity with ligands. Although, the ability of the E367Q TrBgl2 mutant to still bind the substrate (MUG) has been previously confirmed by ligand observed NMR (Section 3.2.3) and TSA (Section 3.3), such protein changes induced by the E367Q

120

mutation could result in different effect of **27** and MUG substrate binding between the wild type and E367Q-mutated protein.

Comparison of the spectrum when **27** was added at 8 mM concentration to the apo E367Q *Tr*Bgl2 spectrum (Figure 4.7), several resonances are missing. This may be due to potential changes in dynamics as discussed above. Although few changes in chemical shifts were observed, there were three peaks which have similar chemical shifts in both wild type *Tr*Bgl2 and inactive mutant E367Q *Tr*Bgl2 indicating that **27** binds to the same site to both wild type and inactive protein.
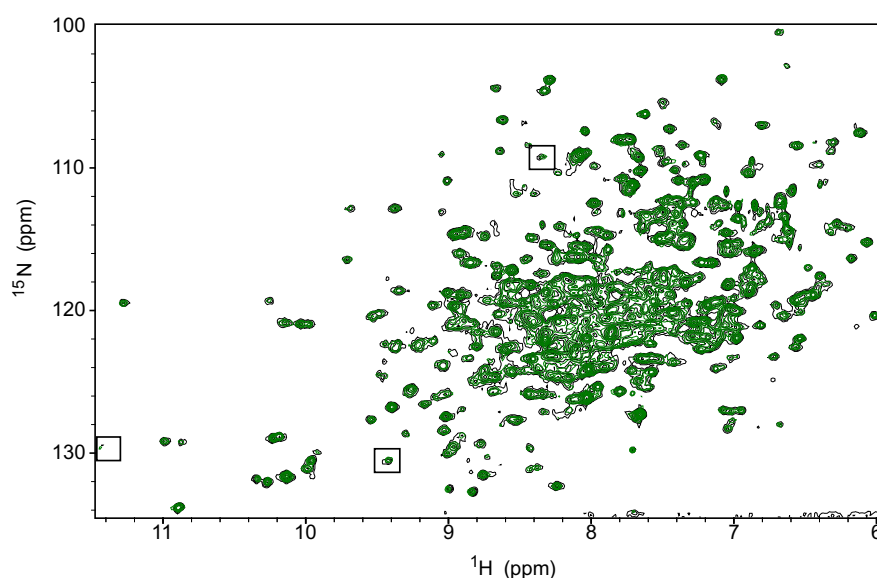
Figure 4.7. ^1H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo E367Q TrBgl2 (black) overlaid with spectra of E367Q TrBgl2 with **27** at 8 mM concentration (green). Spectra collected under identical conditions.

The addition of MUG at 5 mM resulted in several changes in amide chemical shifts and signal disappear across the E367Q *Tr*Bgl2 consistent with tight substrate binding (Figure 4.8).
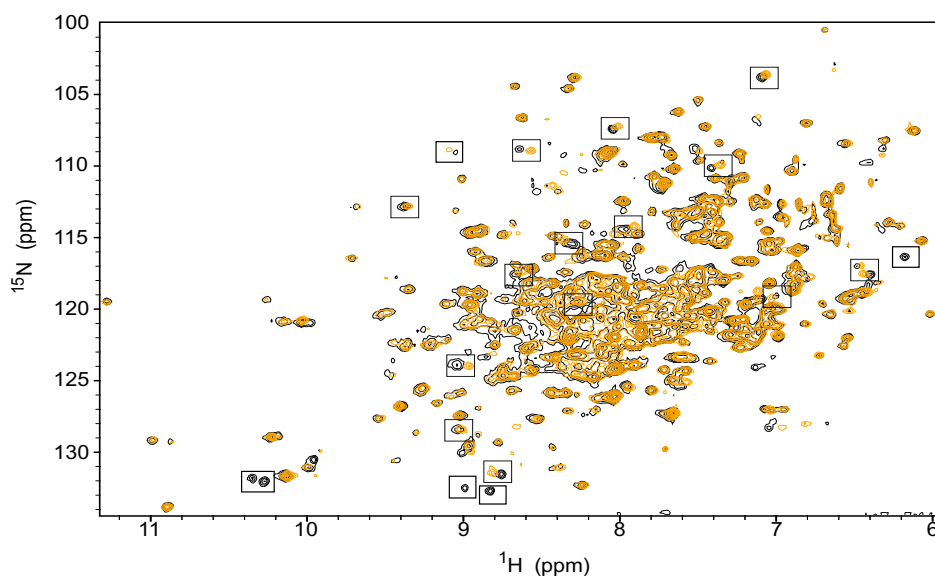
Figure 4.8. *^{1}H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo E367Q TrBgl2 (black) overlaid with spectra of E367Q TrBgl2 with MUG at 5 mM concentration (orange). Spectra collected under identical conditions.*
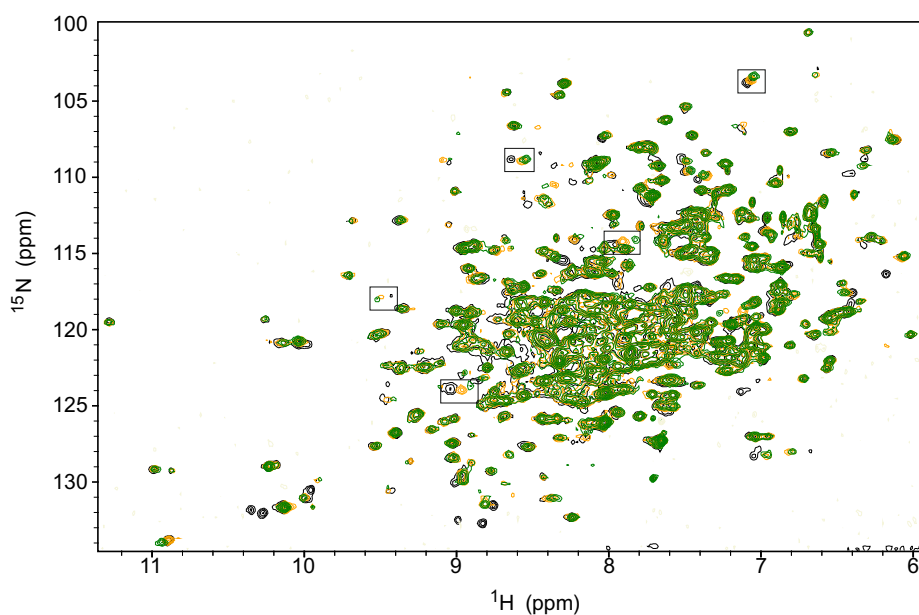
Figure 4.9. *^{1}H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo E367Q TrBgl2 (black) overlaid with spectra of E367Q TrBgl2 with MUG at 5 mM concentration (orange) and spectra of E367Q TrBgl2 bound with both MUG and **27** at 5 and 8 mM concentration respectively (green). Spectra collected under identical conditions.*

Following the addition of **27** at 8 mM, the same substrate specific chemical shifts were seen indicating that the interaction site remains the same in the ternary complex. Moreover, the magnitude of the substrate specific chemical shifts that are not affected by **27** binding to the apo protein is larger in the ternary complex, compared to the binary E367Q *Tr*Bgl2/MUG complex, suggesting an increase in the amount of enzyme that is in complex with the substrate (Figure 4.7). Therefore, considering that the conditions and concentration of MUG were kept identical, these results also indicate higher substrate affinity to the enzyme in the presence of the activator **27** (Figure 4.9-4.10).
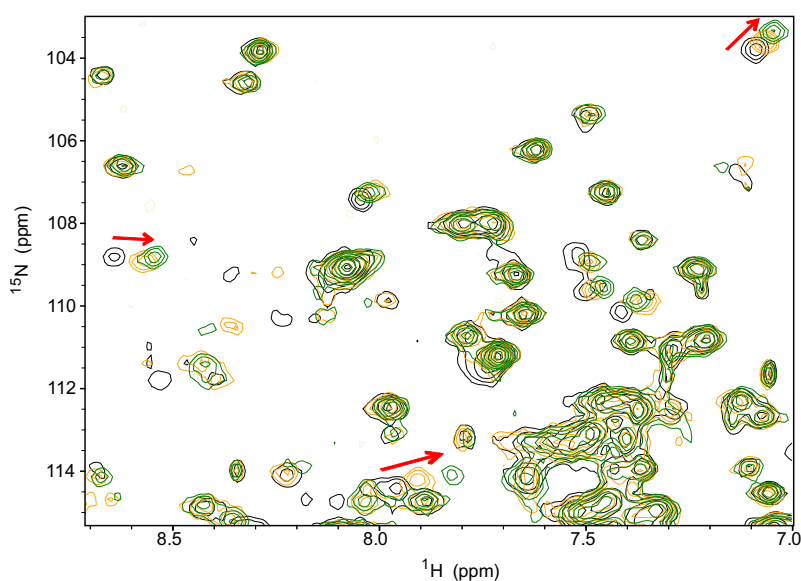
Figure 4.10. *Expanded panel of ^{1}H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo E367Q TrBgl2 (black) overlaid with spectra of E367Q TrBgl2 with MUG at 5 mM concentration (orange) and spectra of E367Q TrBgl2 bound with both MUG and 27 at 5 and 8 mM concentration respectively (green). Red arrows indicate the direction of chemical shifts. Spectra collected under identical conditions.*
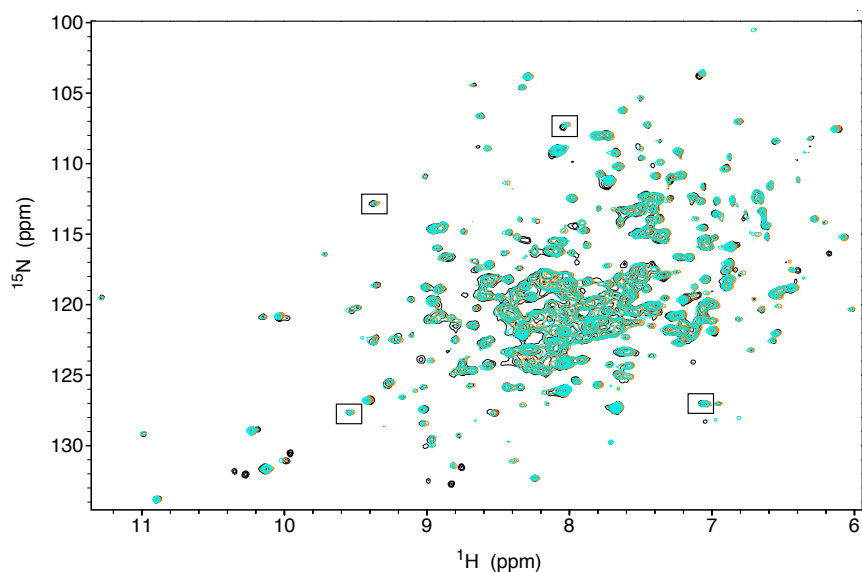
Figure 4.11. *¹H ¹⁵N TROSY spectrum (700 MHz, 298 K) of apo E367Q TrBgl2 (black) overlaid with spectra of E367Q TrBgl2 with MUG at 5 mM concentration (orange) and spectra of E367Q TrBgl2 bound with both MUG and **30** at 5 and 4 mM concentration respectively (cyan). Spectra collected under identical conditions.*
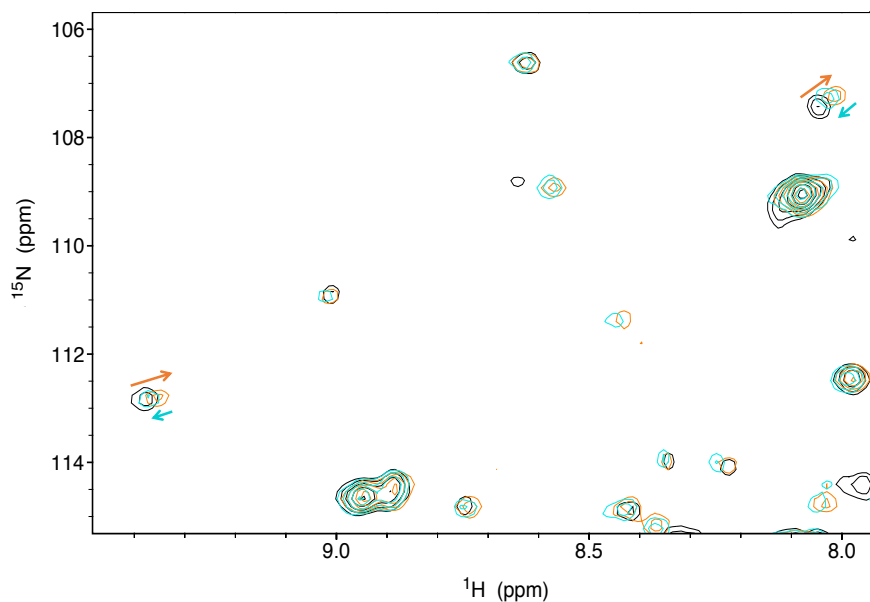
Figure 4.12. *Expanded panel of ¹H ¹⁵N TROSY spectrum (700 MHz, 298 K) of apo E367Q TrBgl2 (black) overlaid with spectra of E367Q TrBgl2 with MUG at 5 mM concentration (orange) and the spectra of E367Q TrBgl2 bound with both MUG at 5 mM and **30** at 4 mM (red). Spectra collected under identical conditions.*

On the other hand, following the addition of **30** at 4 mM, the substrate specific chemical shifts were smaller suggesting lower substrate affinity to the enzyme in the presence of the fragment inhibitor **30** (Figure 4.11-4.12).

Both observations are in full agreement with the kinetic experiments performed which indicate a decreased and increased dissociation constant for MUG in the presence of **27** and **30** respectively (Section 2.7).

4.3.5. Constant time (CT) ^{1}H ^{13}C HSQC of IIe (I), Val (V), and Leu (L) methyl sidechains

TROSY spectroscopy is the most common approach used for large proteins to determine chemical shift perturbations (CSPs) caused by ligand binding. Backbone TROSY CSPs are limited by the fact that most ligand-protein interactions are mediated by sidechains that are distant from the detected backbone amides. Methyl groups are frequently located in hydrophobic cores and at protein-protein interfaces and NOEs between methyl protons provide valuable structural restraints [126-127]. Methyl resonances are also usually intense since there are three protons per methyl group and quite narrow due to rapid rotation around their three-fold axis so that even for high molecular weight systems methyl spectra are of good quality [126-127]. Selective protonation of methyl groups allows protons to be retained in key positions of a structure to allow study of specific features of structure or dynamics. Methyl groups could also serve as probes of molecular interactions and additional chemical shift studies [127]. To investigate

ligand-IVL methyl side-chains interactions, therefore, an IVL methyl protonated sample of deuterated TrBgl2 (Section 7.4.4) was used.

A CT ^{1}H ^{13}C HSQC experiment is the carbon equivalent of the ^{1}H ^{15}N HSQC measuring the correlation of the H-C resonances in the protein (clear description in Section 7.11.4). Ligand binding induces chemical shifts of the protein's H-C nuclei. Figure 4.13 shows the CT ^{1}H ^{13}C HSQC spectra of a methyl protonated {I(δ1 only), L(^{13}CH$_3$,^{12}CD$_3$), V(^{13}CH$_3$,^{12}CD$_3$)} U-[^{15}N, ^{13}C, ^{2}H] sample of apo TrBgl2 at 250 µM (green) in 50 mM TRIS buffer containing 100 mM NaCl and 10% D$_2$O at pH 8.0.
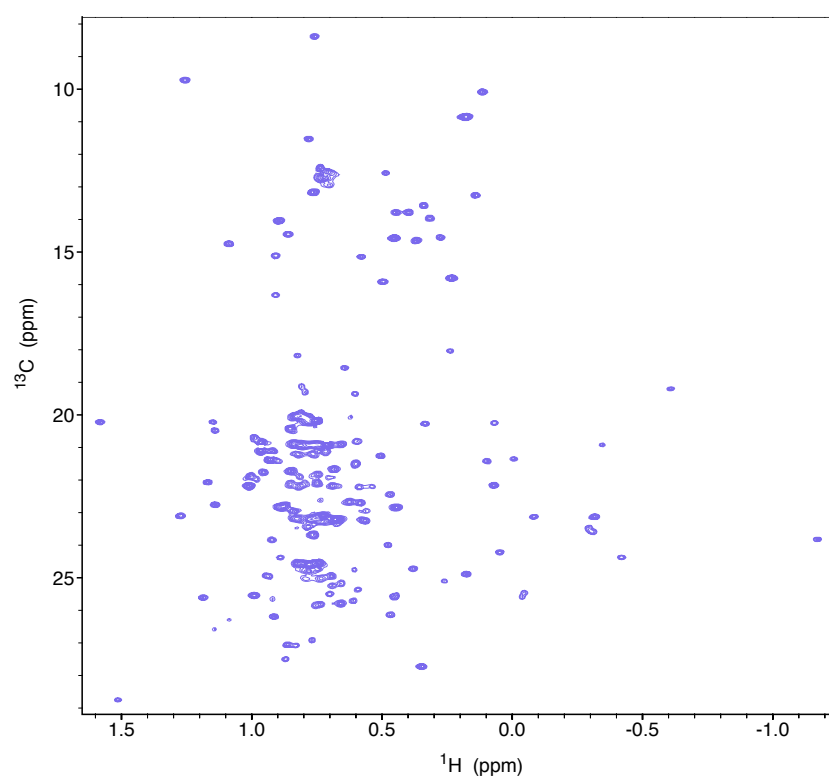
Figure 4.13. *CT ^{1}H ^{13}C HSQC spectrum of an IVL methyl protonated sample (700 MHz, 298 K) of apo TrBgl2 at 250 µM.*
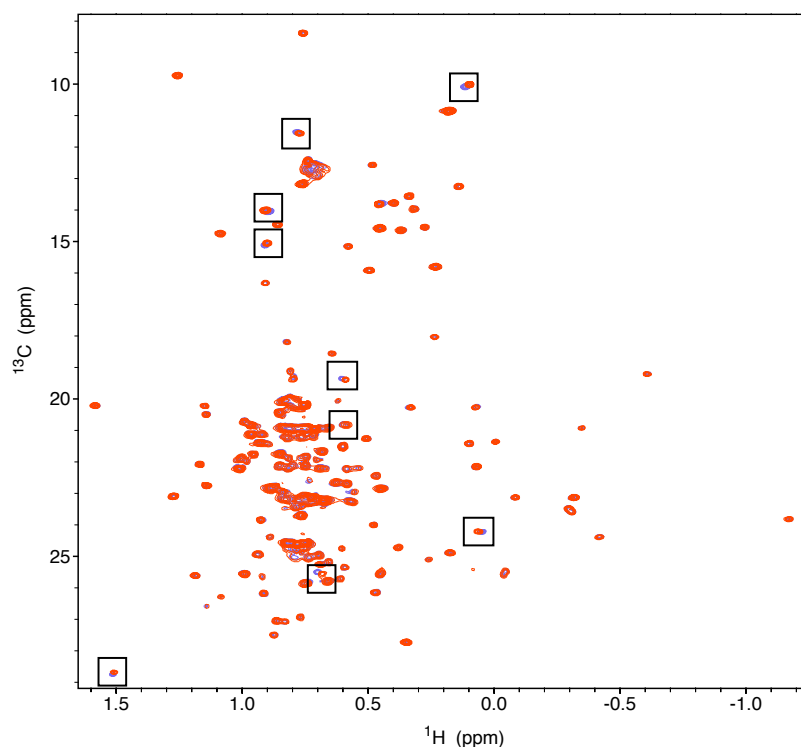
Figure 4.14. CT ^1H ^{13}C HSQC spectrum of an IVL methyl protonated sample (700 MHz, 298 K) of apo TrBgl2 (purple) overlaid with spectra of TrBgl2 with **27** at 8 mM (red). Spectra collected under identical conditions.

Following addition of **27** at 8 mM concentration, significant chemical shift perturbations for several protein methyl groups were observed confirming binding of **27** to the enzyme (Figure 4.14).

4.4. Resonance assignment

The NMR spectra presented so far in this chapter show that only a few chemical shifts are affected by binding of the activator, consistent with the activator binding to a specific site on the enzyme. If the resonance assignment of TrBgl2 is available, mapping of chemical shifts onto the available apo structure of TrBgl2 can provide information about the activator's binding site.

Historically, the complete assignment of resonances is usually only possible for proteins of less than 30 kDa [127-129]. *Tr*Bgl2 has a molecular weight of 55 kDa, which is usually considered too large for analysis by many NMR methods. However, the relatively good quality of the spectra observed above (Figure 4.1) using the TROSY-based pulse sequences and the apparent stability of the protein considering its size suggested that it could be possible, through challenging, to obtain partial assignments for *Tr*Bgl2.

Replacement of the non-exchangeable protons by deuterons while retaining ^{1}H occupancy at backbone amide positions has helped assignment through spin diffusion suppression and reduction in relaxation rates of ^{13}C and ^{15}N spins. Due to the efficient distribution of magnetization through the spin system of dipolar coupled protons, the linewidths of the NMR signals increase with molecular weight thereby lowering signal-to-noise. The use of a deuterated protein reduces the number of proton signals and therefore, eliminate many relaxation pathways and improve the signal-to-noise of NMR spectra [130]. Deuteration also decreases the dipolar interaction between ^{13}C or ^{15}N and the bound proton spin which mainly causes relaxation. The relaxation times of ^{13}C and ^{15}N spins are then increased which results in smaller linewidths and higher signal-to-noise [131].

Protein samples were therefore generated with uniform ^{15}N, ^{13}C, ^{2}H isotopic backbone labelling and methyl protonated $\{$I(δ1 only), L(^{13}CH$_3$, ^{12}CD$_3$), V(^{13}CH$_3$, ^{12}CD$_3$$\}$ U-[^{15}N, ^{13}C, ^{2}H] sidechain labelling (Section 7.4.4). This

protein was then used in experiments that correlate resonances via both bond and space interactions between the nuclei [126] (an overview of several experiments used for assignment of proteins by NMR can be found in [132]). This section outlines the approach taken to assign the resonances observed for apo TrBgl2.

4.4.1. Backbone assignments

Backbone sequential assignments of TrBg2 were made using a set of seven 3D experiments; HNCO, HN(CA)CO, HNCA, HN(CO)CA, HN(CO)CACB, HNCACB and HN(CA)CB. In these experiments, different nuclei are correlated on the basis of via bond scalar couplings (a fuller description is provided in Appendix A) [132]. The nuclei enclosed in parentheses above were not frequency labelled during the pulse sequence. Each of these experiments expands the 2D ^{1}H ^{15}N TROSY into a third ^{13}C dimension, thus allowing correlation of ^{13}C resonances to the amide backbone resonances identified in the TROSY spectra.

Strips of the seven 3D spectra at a specific ^{1}H ^{15}N resonance TROSY peak were displayed and compared together in pairs (Figure 4.15). The HNCACB and HN(CO)CACB spectra show the resonances of Cα and Cβ correlated to a specific amide resonance. The HNCACB spectrum contains the Cα and Cβ shifts of the current (*i*) and previous (*i-1*) residues as contrasted with the HN(CO)CACB spectrum, which only has peaks for Cα and Cβ of the *i-1* residue (Figure 4.15). To compliment this, HNCA, HN(CO)CA and HN(CA)CB experiments were recorded allowing further assignment. The HNCA and

HN(CO)CA spectra contains only the Cα shifts compare to HNCACB and HN(CO)CACB spectra while the HN(CA)CB spectra contains the Cβ shifts of the current (*i*) and previous (*i-1*) residue. These experiments were useful in the cases where Cα and Cβ resonances were not visible in the HNCACB and HN(CO)CACB spectra. By an analogous method, the C' resonances can be assigned to amide frequencies using the HNCO with the HN(CA)CO experiments which show the C' shifts of the previous (*i-1*) and current (*i*) residues respectively. The HNCO and HNCA–type spectra are in general more sensitive than the HNCACB spectra, however they provide less information about the amino acid type than the Cβ chemical shift [132].

Figure 4.15. ^{13}C ^1H strips from 3D HN(CO)CACB (first strip of each amino acid) and HNCACB (second strip of each amino acid) spectrum (700MHz, 298 K) of ^2H, ^{15}N, ^{13}C TrBgl2.

The strips were taken at ^{15}N chemical shifts of residues 127-130 and centered on the corresponding ^{1}H chemical shifts (indicated at the bottom of each strip). The sequence-specific assignment is indicated at the top of each strip by the one-letter amino acid code and the sequence number. The sequential and intra-residual connectivities are indicated by dashed lines. ^{13}Cα and ^{13}Cβ peaks are coloured in blue and black, respectively.

Chemical shifts of each backbone experiment were submitted to the automated resonance assignment algorithm FLYA to determine as far as possible sequential assignment of *Tr*Bgl2 peaks on the basis of peak lists from through-bond experiments [133]. The initial FLYA run assigned approximately 50% of the backbone to a high degree of confidence. The assigned peaks were inspected manually by visually correlating *i* peaks to *(i-1)* peaks of the following spin system and adjusted in areas of uncertainty. The assigned peaks and regions of sequence that caused difficulty in assignment by FLYA were inspected manually.

The sidechain resonances of tryptophans are assigned by combined analysis of the backbone experiments and the ^{15}N-resolved NOESY and ^{13}C-resolved NOESY spectra. The ^{13}C-resolved NOESY experiment did not yield highly resolved spectra with good sensitivity, preventing assignments for tryptophan (W) sidechains in *Tr*Bgl2.

Recombinant *Tr*Bgl2 consists of 488 amino acids (including N-terminal His$_6$ cleavable tag). In total, 265 backbone amide resonances from 459 non-proline residues were sequentially assigned (Figure 4.17). Some regions of

the sequence were unable to be assigned due to missing peaks, spectral overlap, many spin systems with similar chemical shifts, or a combination of all these factors (Figure 4.16).

```
-21 MGSSHHHHHHSSGLEVLFQGPA M L P K D F Q W G F A T A A Y Q I E G A
 21 V D Q D G R G P S I W D T F C A Q P G K I A D G S S G V T A C D S Y N
 56 R T A E D I A L L K S L G A K S Y R F S I S W S R I I P E G G R G D A V
 92 N Q A G I D H Y V K F V D D L L D A G I T P F I T L F H W D L P E G L H
128 Q R Y G G L L N R T E F P L D F E N Y A R V M F R A L P K V R N W I T F
164 N E P L C S A I P G Y G S G T F A P G R Q S T S E P W T V G H N I L V A
200 H G R A V K A Y R D D F K P A S G D G Q I G I V L N G D F T Y P W D A
235 A D P A D K E A A E R R L E F F T A W F A D P I Y L G D Y P A S M R K Q
271 L G D R L P T F T P E E R A L V H G S N D F Y G M N H Y T S N Y I R H R
307 S S P A S A D D T V G N V D V L F T N K Q G N C I G P E T Q S P W L R
342 P C A A G F R D F L V W I S K R Y G Y P P I Y V T E N G T S I K G E S D
378 L P K E K I L E D D F R V K Y Y N E Y I R A M V T A V E L D G V N V K G
414 Y F A W S L M D N F E W A D G Y V T R F G V T Y V D Y E N G Q K R F P
449 K K S A K S L K P L F D E L I A A A
```

Figure 4.16. *Sequence of TrBgl2 used for NMR studies. Tag residues are underlined, and the assigned sequence is coloured green.*

Figure 4.17. *2D ^{1}H ^{15}N TROSY spectrum (700 MHz, 298 K) of ^{15}N-labeled TrBgl2. Backbone resonance peaks are indicated by the one-letter amino acid code and the sequence number is shown in black.*

*Tr*Bgl2 secondary structure motifs were predicted by the TALOS-N software (Figure 4.18) [134]. The software indicated that the overall secondary structures are in agreement with the X-ray structure of *Tr*Bgl2 (PDB ID 3AHY) [2].

Figure 4.18. *Secondary structure prediction of TrBgl2 analyzed with TALOS-N using the assigned chemical shifts compared to the secondary structure of the X-ray structure of TrBgl2. Top: colored bars (red and blue bars indicate α-helix and β-strands respectively) show the secondary structure type predicted by TALOS-N. The bar height represents the prediction confidence. The black line shows the predicted S^2 order parameter, a measure of flexibility. Bottom: the secondary structure of TrBgl2 as determined by X-ray crystallography (PDB ID 3AHY) [2].*

4.4.2. IIe (I), Val (V), and Leu (L) Methyl sidechain assignment

As most Cα, Cβ and C' assignments are known during backbone assignment, methyl resonances can be assigned using COSY-type magnetisation transfers from the methyl carbon and proton to already assigned atoms [129]. This approach makes use of a biosynthetic precursor that leads to labelling of Leu (L) and Val (V) at only a single methyl position

134

resulting in linear arrangement of coupled carbons for these residues [126][129].

Data was then added from experiments that relay magnetisation from methyl carbon and proton spins of Ile (I), Leu (L) and Val (V) residues to aliphatic carbons (HMCM(CG)CBCA) and from methyl groups to carbonyls and back (Val-HMCM(CBCA)CO and Ile, Leu-HMCM(CGCBCA)CO) (clear description in Appendix A) [126][135]. Further, additional methyl-methyl NOESY, ^{15}N-resolved NOESY and ^{13}C-resolved NOESY experiments (clear description in Appendix A) were performed, in which the different nuclei are correlated via space interactions. The known crystal structure of apo *Tr*Bgl2 with respect to the already assigned backbone peak lists was also analysed to allow more complete assignment of the expected resonances. Rerunning the automated resonance assignment algorithm (FLYA) on the combined data resulted in an improved output. The assigned peaks were inspected manually.

Methyl side-chains were assigned by identifying the corresponding Cα, Cβ and C' resonances in the relevant amide strip of the HNCA, HNCACB and HN(CA)CO spectra, respectively. Further confirmation of the assignment was performed by comparison with the through space 3D CH$_3$-CH$_3$ NOESY, ^{15}N-resolved NOESY and ^{13}C-resolved NOESY spectra.

In total, approximately 55% of methyl sidechain resonance peaks were assigned (Figure 4.19).

Figure 4.19. ^1H ^{13}C HSQC spectrum (700 MHz, 298 K) of an IVL methyl protonated sample of apo TrBgl2. Methyl sidechain resonances are indicated by the one-letter amino acid code and the sequence number is shown in black.

4.5. Binding site mapping

CSPs do not completely correlate with proximity to the ligand-binding site as CSPs can also result from ligand-induced conformational changes. Nevertheless, when ligand binding does not cause significant conformational changes, most of the largest CSPs are caused by the ligand contacting atoms within or around the ligand binding site. However, resonances that are shifted by nearby aromatic groups undergo large CSPs even when distant from the ligand binding site [136-137].

A high percentage of backbone assignments of a protein is required to locate the binding site of a ligand based on CSPs. This allows correlation of the CSPs to amino acids within the protein sequence. When the 3D structure is available, as is the case for *Tr*Bgl2, CSPs can be mapped to the surface of the protein structure and clusters of shifts in continuous structural regions of the protein can be used to predict ligand binding sites.

The assigned backbone CSPs in the TROSY spectrum and the IVL methyl chemical shifts in the ^{1}H ^{13}C HSQC after compound **27** titration were mapped onto the known crystal structure of *Tr*Bgl2 using the resonance peaks determined for the protein.

A scaled Pythagoras equation was used to calculate backbone CSP values ($\Delta\delta$) for each resonance following addition of **27**. The typical range of shifts is significantly different for ^{1}H, ^{13}C and ^{15}N nuclei which means the importance of changes in one dimension would be overstated without the

adjustment. Therefore, a scaling factor (*c*) of 0.1 and 0.2 was used to balance the contributions of ^{1}H-^{15}N and ^{1}H-^{13}C respectively to the magnitude of the perturbation [137]. CSPs were quantified using an adjusted chemical shift value (Δδ) of the following equation:

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (c\Delta\delta_{N/C})^2}$$

Standard practice was used to decide which shifts are large enough to be considered indicators of **27** interaction to *Tr*Bgl2. The standard deviation (SD) and average (AVG) of the shift changes were first calculated and residues for which the shift change is greater than the sum of these two values (SD, AVG) were identified as statistically significant shifts [138]. The value of 0.012 ppm and 0.01 ppm was calculated as the cut-off for the backbone and methyl chemical shifts, respectively.
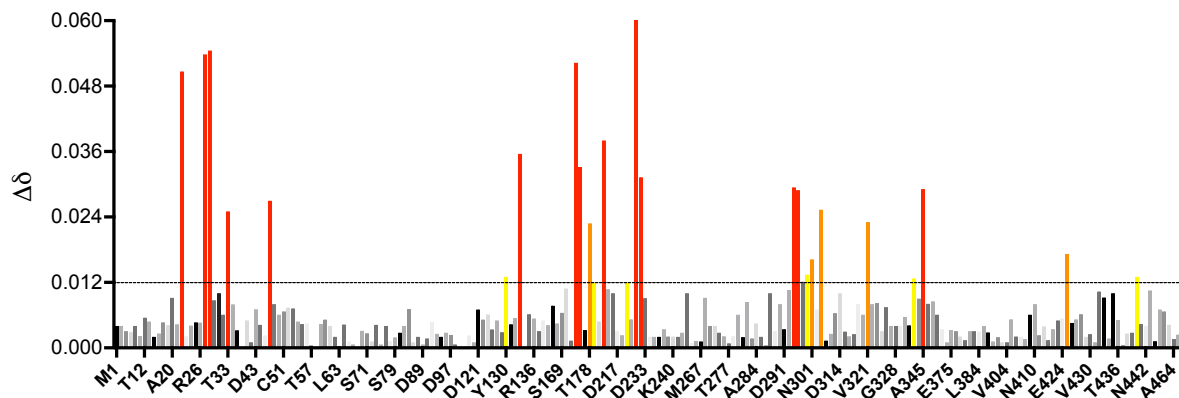
Figure 4.20. *A histogram showing the backbone chemical shift value (Δ) as a function of the TrBgl2 sequence and across the titration with **27** against apo TrBgl2. The threshold value for Δδ used as a criterion for significance (0.012) is marked using dashed line. Backbone CSPs corresponding to residues with Δδ > 0.025 are coloured red, Δδ values 0.012-0.025 are shaded in a gradient yellow-orange-red.*

CSP: 🟥 > 🟧 > 🟨   CSP: 🟥 > 🟧 > 🟨
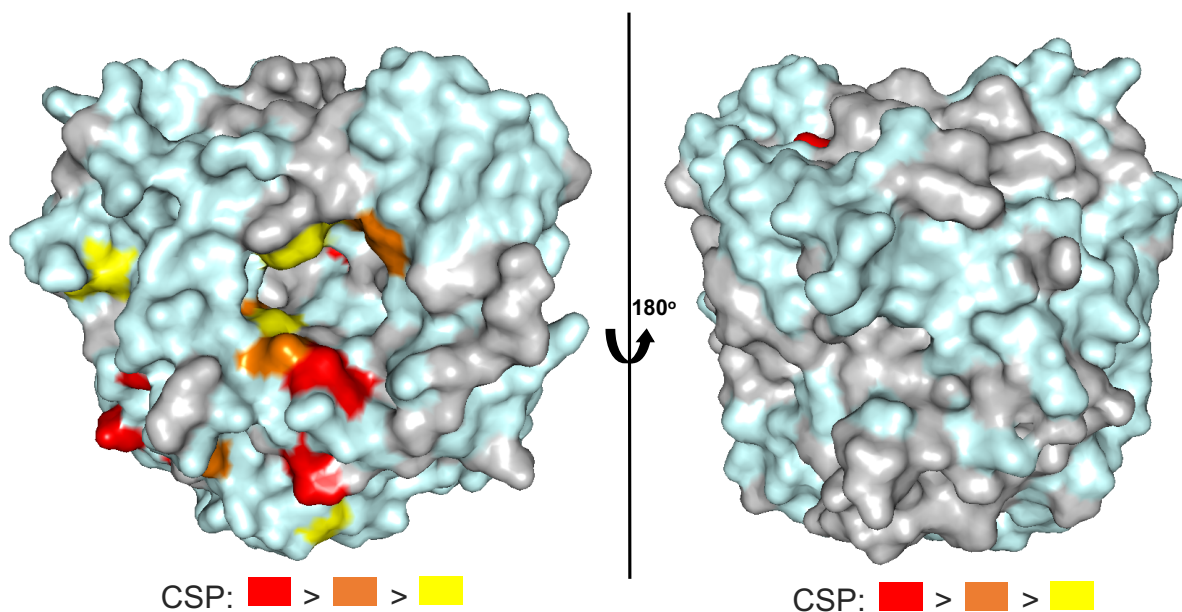
Figure 4.21. *Amide chemical shift perturbations caused by the presence of **27** mapped to the protein surface of the crystal structure of TrBgl2. Backbone CSPs corresponding to residues with Δδ > 0.025 are coloured red, Δδ values 0.012-0.025 are shaded in a gradient yellow-orange-red. Unassigned residue are coloured grey. The figure was created in PyMOL.*
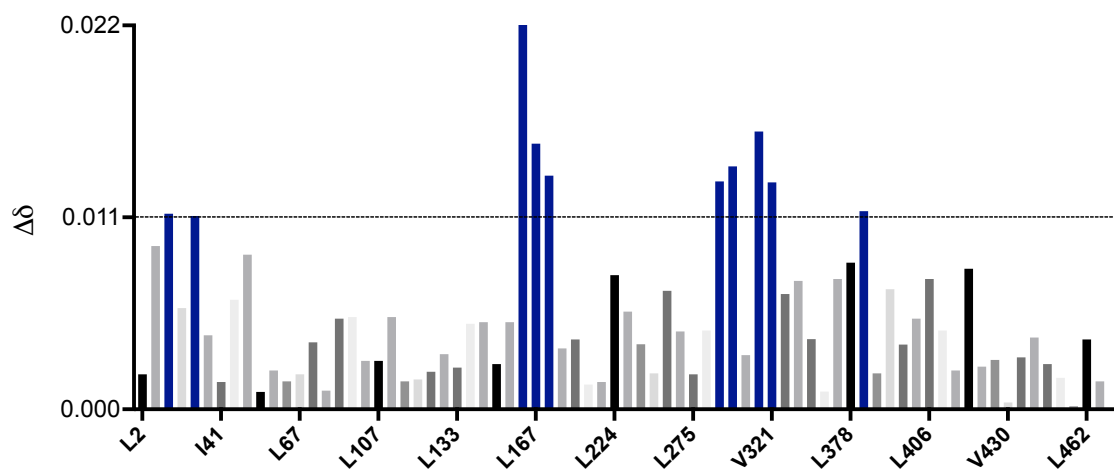
Figure 4.22. *A histogram showing the methyl chemical shift value (Δ) as a function of the TrBgl2 sequence and across the titration with **27** against apo TrBgl2. The threshold value for Δδ used as a criterion for significance (0.011) is marked using dashed line.*

Table 4.1. *List of assigned CSPs after compound **27** titration.*

Backbone TROSY CSPs	D22, G27, S29, T33, S46, Y130, L133, G175, S176, T178, F179, R183, L224, G226, F228, M295, N296, S300, N301, I303, V321, W339, A345, W425, Y440
¹H ¹³C HSQC of IVL methyl CSPs	I17, V21, L167, I171, I303, V316, V321
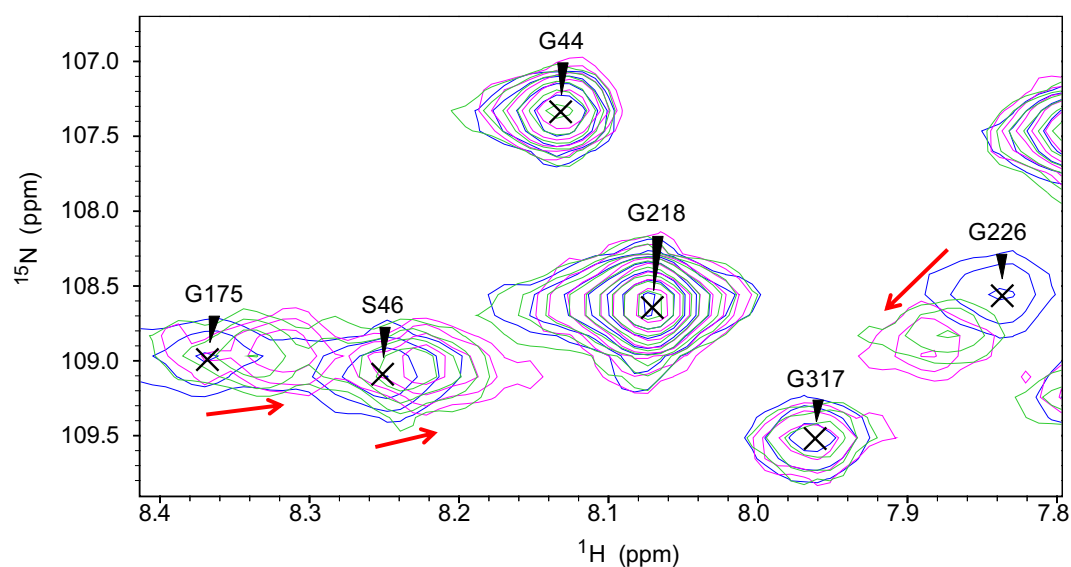
Figure 4.23. *A small region of ^1H ^{15}N TROSY spectrum (700 MHz, 298 K) of apo TrBgl2 (blue) overlaid with spectra of TrBgl2 with **27** at 8 mM (green) and 16 mM (magenta) concentration. The figure shows three backbone CSPs (S46, G175, G226) which correspond to residues with Δδ > 0.025. Red arrows indicate the direction of chemical shifts on increasing ligand concentration. Spectra collected under identical conditions.*
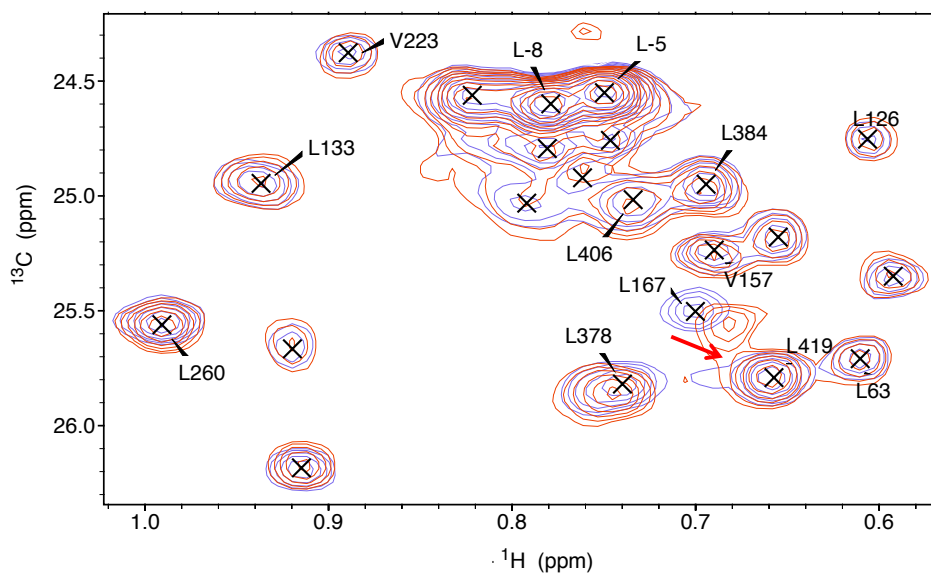
Figure 4.24. *A small region of ^1H ^{13}C HSQC spectrum (700 MHz, 298 K) of an IVL methyl protonated sample of apo TrBgl2 (purple) overlaid with spectra of TrBgl2 with **27** at 8 mM (red) concentration. The figure shows one methyl CSPs (L167) which corresponds to residue with Δδ > 0.011. Red arrows indicate the direction of chemical shifts on increasing ligand concentration. Spectra collected under identical conditions.*

It should be highlighted that the degree of saturation affects the absolute CSP measured. Ligands that bind in a well-defined orientation result in larger CSPs compare to ligands that bind in a less defined orientation [138]. In this system, the larger shift in Δδ is only 0.06 ppm. Due to the very weak binding of **27,** it was not possible to use the ligand's saturating concentration. Therefore, the observed CSPs represent only a proportion of the bound state. It is likely that the ligand adopts multiple bound conformations and the shifts are averaged which results in smaller CSPs [138].

Mapping CSPs of **27** to the *Tr*Bgl2 structure indicated the entrance of the active site as binding site (Figure 4.25). Although, most of the CSPs are located at the entrance of the active site, residues at the surrounding part of the active site are also significantly affected, potentially indicating that conformational changes are induced in the protein upon binding of activator **27**.
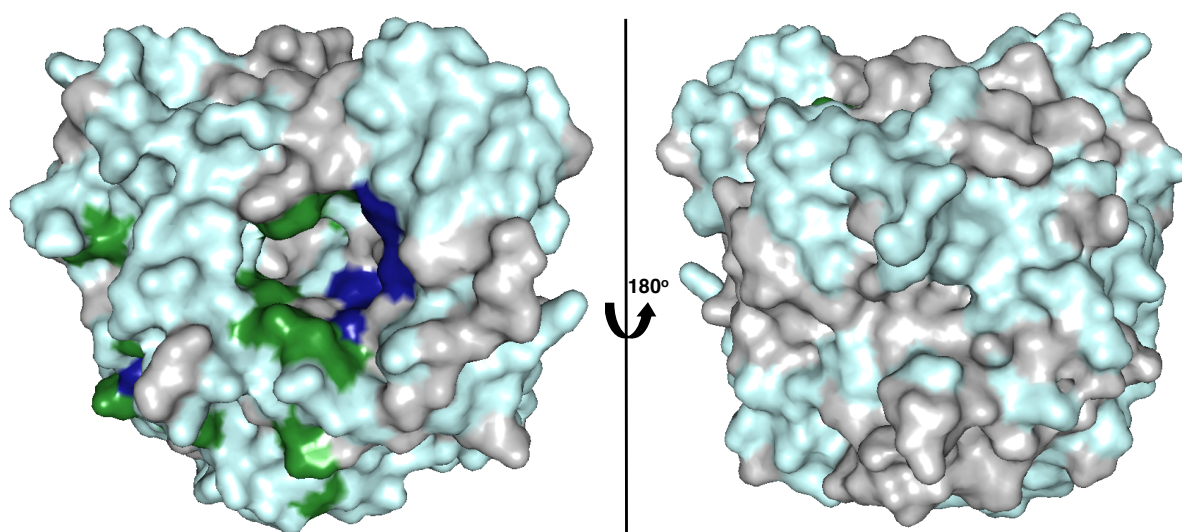
Figure 4.25. *Amide and IVL methyl chemical shift perturbations caused by the presence of **27** mapped to the protein surface of the crystal structure of TrBgl2. Backbone CSPs are coloured green and IVL methyl CSPs are coloured blue. Unassigned residues are coloured grey. The figure was created in PyMOL.*

The backbone CSPs (G27, S46, D52, H127, L133, S169, G175, G226, W232, F228, G288, G294, M295, N296, S300, N301, T335, W339, L340, I372, W425, F433) in the TROSY spectrum after adding inhibitor **30** were also mapped onto the known crystal structure of *Tr*Bgl2 (Appendix C; Figure C.1). The pattern of mapped CSPs of **30** is quite similar to that of **27** indicating that the two compounds bind in the same binding site.

4.6. Molecular docking

The known 3D structure of *Tr*Bgl2 was used for molecular docking studies using HADDOCK-based semi-rigid, data-driven docking to investigate structurally the mechanism of the activation of *Tr*Bgl2 by activator **27** [139, 140]. Although, the observed CSPs reflect only a fraction of the bound enzyme, docking is still informative as the CSPs are not used quantitatively but qualitatively.

HADDOCK experiments were performed using the available apo structure of *Tr*Bgl2 (PDB ID 3AHY), the chemical structure of compound **27** and the larger CSPs located within the *Tr*Bgl2 active site (backbone CSPs: G175, S176, T178, F179, G226, M295, N296, N301, W339 and W425; IVL methyl group CPSs: L167, I171, V316 and V321) as these residues are more likely to directly interact with the ligand.

The lowest energy cluster (model 1) exhibited a HADDOCK score of -45.8 and a minimum restraint violation energy of 77.4 kcal/mol, while the next lowest energy ligand cluster (model 2) exhibited a HADDOCK score and a minimum restraint violation energy of -40.1 and 85.9 kcal/mol respectively.
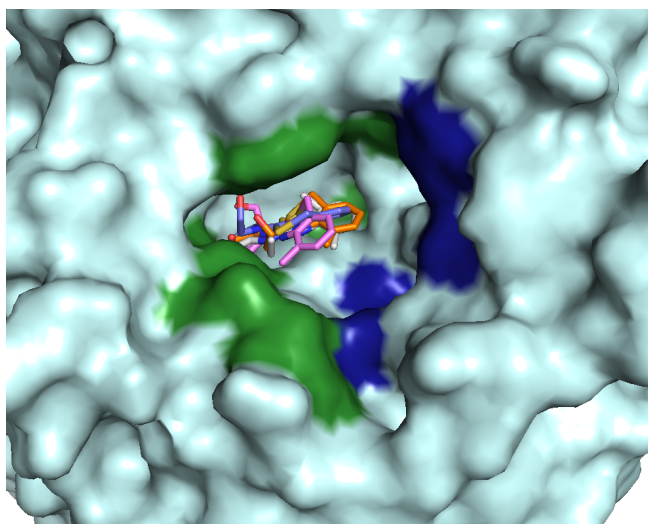
Figure 4.26. *Overlays of the lowest energy HADDOCK model 1 structures of **27** docked with TrBgl2 using the backbone amide and methyl CSPs interpreted to locate at the TrBgl2's active site. The figure was created in PyMOL Backbone (G175, S176, T178, F179, G226, M295, N296, N301, W339, W425) and IVL methyl group (L167, I171, V316, V321) CSPs used in HADDOCK are shown in green and blue respectively. The figure was created in PyMOL.*

HADDOCK model 1 structures were superimposed on the crystal structure of a close homologue of *Tr*Bgl2, termite *Neotermes koshunensis* glycosyl hydrolase (*Nk*Bgl-E193D) in complex with this enzyme's substrate, *p*-Nitrophenyl-β-D-glucopyranoside (*p*NPG) [2]. When observing the superposed structures, the substrate of *Nk*Bgl-E193D, *p*NPG, overlaps with the determined binding site for compound **27**. However, the kinetic (Section 2.7), NMR (Section 3.2.1-4.3.4) and TSA experiments (Section 3.3) performed have shown that **27** binds to both free enzyme and enzyme/MUG complex suggesting that it binds to an allosteric binding site. Notably, substrate and activator binding are enhanced following addition of the activator and MUG respectively stabilising the *Tr*Bgl2/**27**/MUG complex.
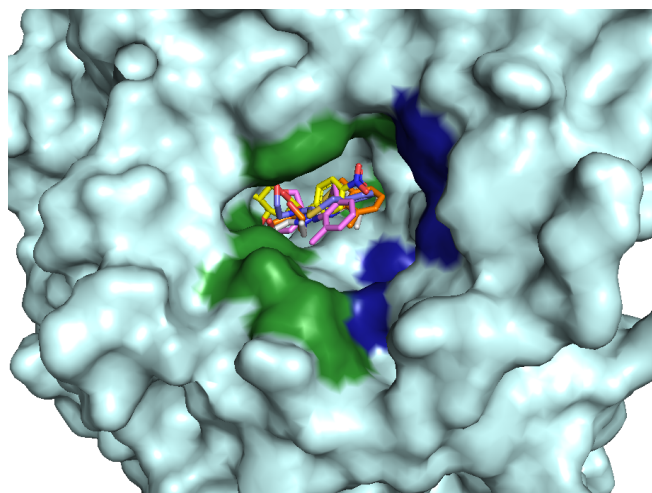
Figure 4.27. *Structural overlay of the lowest energy HADDOCK model 1 structures with the close homologue of TrBgl2, NkBgl-E193D bound to pNPG, (PDB ID 3AI0) in yellow. Backbone's (G175, S176, T178, F179, G226, M295, N296, N301, W339, W425) and IVL methyl group's (L167, I171, V316, V321) CSPs used in HADDOCK are shown in green and blue respectively. The figure was created in PyMOL.*

Due to a possible substrate role in the compound binding to *Tr*Bgl2, we sought to generate a model of the *Tr*Bgl2/MUG complex, and then *Tr*Bgl2/MUG/**27** complex after docking **27** into the *Tr*Bgl2/MUG model using HADDOCK.

Although attempts were made to determine the crystal structure of E367Q *Tr*Bgl2 complexed to MUG were not successful, the crystal structure of *Nk*Bgl-E193D in complex with *p*NPG as determined by Jeng is available [2]. *Nk*Bgl and *Tr*Bgl2 share a highly conserved active site. After binding of the substrate (*p*NPG) on *Nk*Bgl-E193D significant changes in the structure occurred at the sidechains of residues Glu/Asp193, Asp253, Trp374, Trp444, Asn449 and Glu451 (Appendix B). The substrate *p*NPG was

observed to form direct hydrogen bonds to Gln45, His148, Asn255, Glu402, Trp444, Glu451 and Trp452 of *Nk*Bgl-E193D corresponding to residues Gln16, His119, Asp227, Glu367, Trp417, Glu424 and Trp425 respectively of *Tr*Bgl2 after structural alignment performed between the homologues (Figure 4.28).
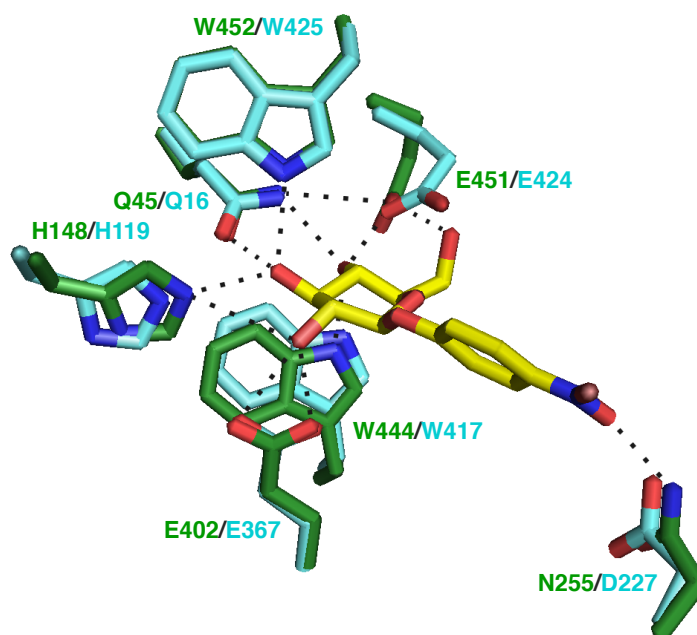
Figure 4.28. *Structural overlay of TrBgl2 and the homologue NkBgl-E193D bound to pNPG active site is shown as stick diagram. TrBgl2 (PDB ID 3AHY) is shown in cyan and NkBgl-E193D bound to pNPG (PDB ID 3AI0) is shown in green. pNPG is shown as yellow stick diagram. Hydrogen bonds are shown as black dashed lines.*

First, HADDOCK was performed using the apo structure of *Tr*Bgl2 (PDB ID 3AHY), the chemical structure of the substrate (MUG) and the amino acids of *Tr*Bgl2 defined (Gln16, His119, Glu367, Trp417, Glu424, Trp425) after analysis of structural overlay of *Tr*Bgl2 and *Nt*Bgl2 bound to *p*NPG (PDB ID 3AI0) (Figure 4.28). The amino acid, Asp227 that directly interacts with the nitrophenyl part of *p*NPG substrate, was excluded.

The unique *Tr*Bgl2/MUG model 3 exhibited a HADDOCK score of -60 and a minimum restraint violation energy of 1.1 kcal/mol (Figure 4.29).
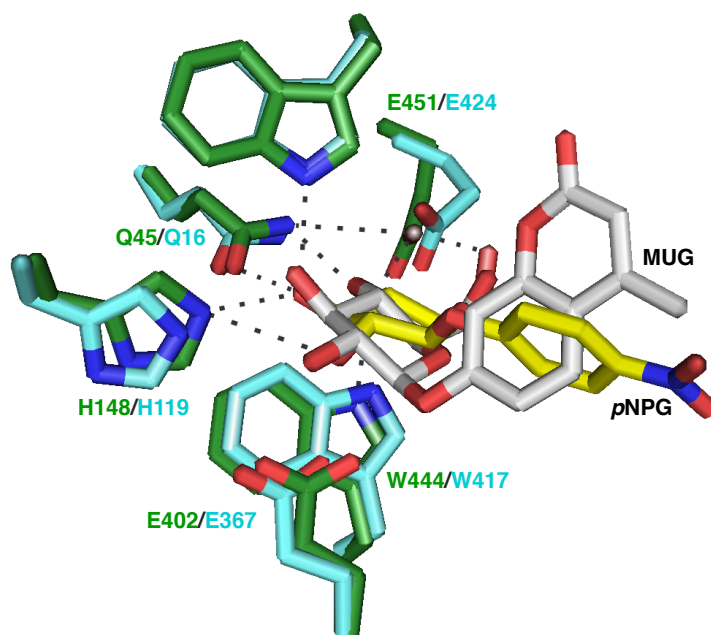
Figure 4.29. *Structural overlay of TrBgl2 bound to MUG (Model 3) and the homologue NkBgl-E193D bound to pNPG active site is shown as stick diagram. TrBgl2 bound to MUG (Model 3) is shown in cyan and NkBgl-E193D bound to pNPG (PDB ID 3AI0) is shown in green. MUG is shown in white and pNPG is shown in yellow, both as stick diagrams. Hydrogen bonds are shown as black dashed lines.*

Next, HADDOCK was used to generate a model of compound **27** docked into the *Tr*Bgl2/MUG cluster using model 3, the chemical structure of **27** and the CSPs interpreted belonging to residues located at the *Tr*Bgl2 active site (backbone CSPs: G175, S176, T178, F179, G226, M295, N296, N301, W339 and W425; IVL methyl group CPSs: L167, I171, V316 and V321).

The lowest energy cluster (Model 4) exhibited a HADDOCK score of -19.5 and a minimum restraint violation energy of 35.4 kcal/mol, whereas the next ligand cluster (Model 5) exhibited a HADDOCK score and minimal

restraint violation energy of -8.1 and 72.1 kcal/mol, respectively. Both models reveal the binding site of **27** at the substrate entrance of *Tr*Bgl2, confirming the experiments described above, and suggests a possible direct activator **27**/MUG contact (Figure 4.30).
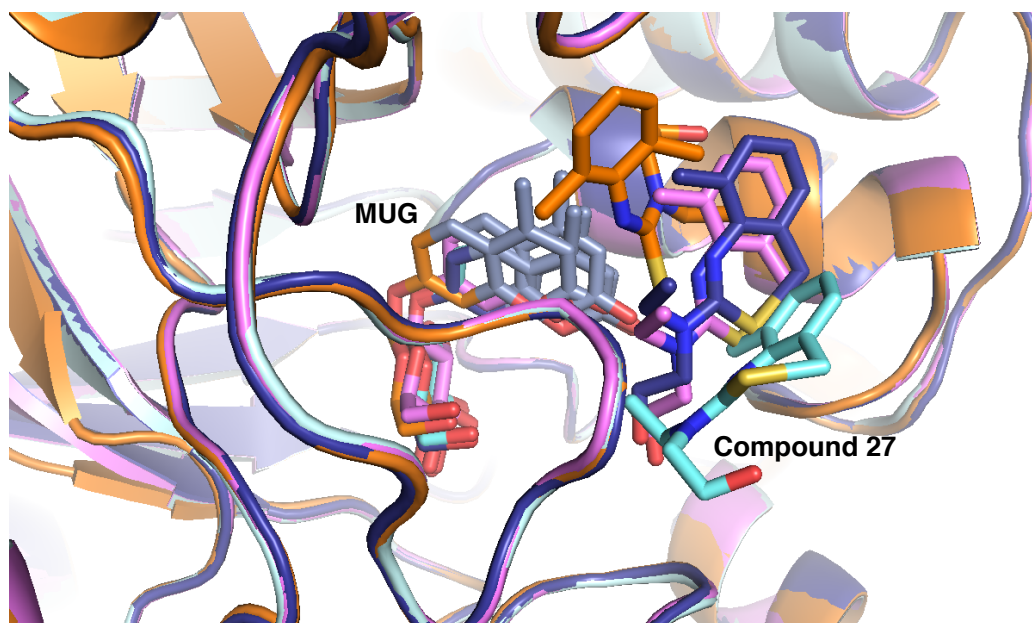
Figure 4.30. *Overlays of the lowest energy HADDOCK model structures of **27** docked with TrBgl2/MUG complex using the backbone amide and methyl CSPs interpreted to locate at the TrBgl2's active site. The figure was created in PyMOL.*

Although, mapping of CSPs and molecular docking provides information about the activator **27** binding site, determination of the precise orientation of ligand and intermolecular contacts is beyond the capabilities of CSP analysis. In principle, intermolecular NOE experiments should be able to detect interactions between the IVL methyl groups within the activator binding site and **27** that are within NOE distance (usually 5 Å). Although in this case, only a few IVL methyl groups of the active site could be assigned it was possible that some interactions may be observed.

Once binding of **27** to *Tr*Bgl2 was saturated, a 2D NOESY (Bruker pulse sequence: noesygpphwgx2) and a 3D ^{13}C NOESY-HSQC (Bruker pulse sequence: hsqcgpnowgx33d) filtering out ^{13}C- and ^{15}N in the direct dimension were recorded [141-142] by Dr Alex Heyam (University of York) in collaboration with ZoBio B.V. (Leiden, The Netherlands). Unfortunately, intermolecular NOEs between the ligand resonances and IVL methyl groups as well as backbone amide protons of *Tr*Bgl2 were not detected. Although intermolecular NOE approaches are quite powerful, they suffer from several limitations regarding the requirement for many intermolecular NOE contacts and do not always work well with weakly binding ligands [143].

4.7. Conclusion

This chapter has described the various experiments used to obtain structural information on the binding of activator **27** to *Tr*Bgl2 to get insight into the mechanism of activation. A number of crystallisation attempts were proven unsuccessful. Extensive multi-dimensional NMR experiments were therefore carried out on isotopically labelled *Tr*Bgl2.

The initial good quality spectra of TROSY NMR of apo *Tr*Bgl2 laid the path for further NMR studies. Following the addition of activator **27** and inhibitor **30**, changes in amide chemical shifts across the protein occurred which clearly indicated a specific binding event. The possible role of the substrate MUG in fragment binding was also investigated by TROSY NMR measurements with inactive E367Q *Tr*Bgl2 mutant. Notably, substrate (MUG) specific chemical shifts were further enhanced following addition of **27** indicating possible higher substrate affinity to the enzyme in the presence of activator **27**. On the other hand, following the addition of inhibitor **30**, the substrate-specific chemical shifts were reduced suggesting lower substrate affinity to *Tr*Bgl2 in the presence of this fragment inhibitor. Both observations demonstrate the important role of substrate in fragment binding and are in agreement with the kinetic (Section 2.7) and TSA (Section 3.3.) experiments.

In the second part of this chapter, approximately 56% of backbone resonances and 40% of methyl resonances of apo *Tr*Bgl2 were successfully assigned. Although *Tr*Bgl2 was too large for full structure determination by

NMR, the assignments obtained were enough to elucidate the binding site of activator **27**. Mapping of the observed CSPs of activator **27** onto the known apo structure of *Tr*Bgl2 indicated the entrance of the active site as a possible binding site. Although, most of the CSPs are located at the substrate entrance of *Tr*Bgl2, residues at the surrounding part of active site are also significantly affected indicating that conformational changes may be induced in *Tr*Bgl2 upon binding of **27**. A previous study revealed that mutations of amino acids located in the entrance of *Tr*Bgl2's active site result in a greater increase in catalytic efficiency of the enzyme compared to amino acids located much closer to the active site [144]. Interestingly, some of those amino acids (L167, T178, F179) are the same ones identified to cause the greater CSPs following the addition of **27** to *Tr*Bgl2. The CSPs of fragment inhibitor (**30**) were also mapped onto the apo structure of *Tr*Bgl2 revealing that both compound analogues bind in the same binding site.

Molecular docking using HADDOCK was then used to generate a model of the *Tr*Bgl2/substrate/activator complex using the ability of HADDOCK to include CSPs as ambiguous restraints. The resulting model confirms the binding site of activator **27** at the entrance of *Tr*Bgl2's active site and suggests possible direct activator-substrate contacts. Frequently, activator binding is accompanied by conformational changes in a protein. Indeed, although most of the CSPs observed are located at the entrance of *Tr*Bgl2's active site, residues close to the active site are also significantly affected,

indicating that the protein may undergo conformational changes upon compound binding.

After substrate binding, compound **27** was hypothesized to attach to the entrance of active site thereby blocking the substrate exit and stabilizing the enzyme-substrate complex. Depending on the fit between the substrate (MUG) and a **21**-like compound, the substrate either properly fits in the active site acting as an improved substrate and adopting a productive conformation which leads to enhanced *Tr*Bgl2 activity (**27**) or obtains a less favor fit adopting a non-productive conformation which leads to *Tr*Bgl2 inhibition (**30**).

Chapter 5: Other enzymes investigated

5.1. The bacterial glycosyl hydrolase, CcBglA

5.1.1. Introduction

The bacterium *Clostridium cellulovorans* produces a thermostable cellulolytic enzyme, *Cc*BglA, homologous to the fungal *Tr*Bgl2 (Figure 5.1) but which has higher activity [2]. *Cc*BglA also belongs to the β-retaining glycoside hydrolase family 1 (CAZY) and consists of a typical $(\alpha/\beta)_8$-TIM barrel-like fold. The structure and biochemical properties of *Cc*BglA were determined by Jeng in 2011 [2]. Although, the protein does not share any specific interactions that promote oligomerisation according to the Protein Interfaces, Surfaces and Assemblies (PISA) server, four protein molecules of *Cc*BglA crystallised in an asymmetric unit [2].

The active site of the enzyme forms a deep slot-like cleft and is surrounded by negatively charged residues. *Cc*BglA and *Tr*Bgl2 share a highly conserved active site whereas there is more sequence divergence in the outer regions that surround the opening to the active site (Figure 5.2). The catalytic acid/base for *Cc*BglA is Glu166, located on the TXNEP sequence motif at the end of β-strand 4 (where X is a hydrophobic amino acid), while the catalytic nucleophile is Glu352, located on the I/VTENG sequence motif at the end of β-strand 7. Two *cis*-peptide bonds were also found at Ala181-Pro182 and Trp399-Ser400 [110]. As discussed, TRIS-based buffers are potent inhibitors of glycosidases (Section 2.1.2) [111]. TRIS appeared to

inhibit the activity of *Cc*BglA although it was not detected in the active site

of the enzyme's crystal structure [2].

```
                          β1         α1           η1          α2           η2
CcBglA          TT   ─────▶  ℓℓℓℓ      ℓℓℓ     ℓℓℓℓℓℓ    TT  ℓℓℓ          ℓℓℓℓ
                1         10        20        30        40        50        60
CcBglA   MEKLRFPKDFIFGTATAAYQIEGAYKEDEKGESIWDRFSHIPGNVAKMHNGDIACDHYHR
TrBgl2   ....MLPKDFQWGFATAAYQIEGAVDQDGRGPSIWDTFCAQPGKIADGSSGVTACDSYNR

                α3              β2        α4                    α5              β3
CcBglA   ℓℓℓℓℓℓℓℓℓℓ      ─────▶     ℓℓℓℓ            ℓℓℓℓℓℓℓℓℓℓℓℓℓℓ      ───
                       70        80        90       100       110
CcBglA   YKEDVQLLKSLGIKSYRFSIAWPRIFPKGF..GEINQKGIQFYRDLIDELIKNDIEPAIT
TrBgl2   TAEDIALLKSLGAKSYRFSISWSRIIPEGGRGDAVNQAGIDHYVKFVDDLLDAGITPFIT

                     α6       η3          α7            β4       α8
CcBglA   ─────▶    ℓℓℓℓ   ℓℓℓ  ℓ.ℓℓℓℓℓℓℓℓ      ─────▶   ℓℓℓℓℓℓℓℓ
               120       130       140       150       160       170
CcBglA   IYHWDLPQKLQD.IGGWANP.QVADYYVDYANLLFREFGDRVKTWITHNEPWVASYLGYA
TrBgl2   LFHWDLPEGLHQRYGGLLNRTEFPLDFENYARVMFRAL.PKVRNWITFNEPLCSAIPGYG

                             α9                        β5         β6
CcBglA         TT  ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ         ─────▶   ─────▶
               180       190       200       210       220       230
CcBglA   LGVHAPGIKDMKMALLAAHNILLSHFKAVKAYRELEQ....DGQIGITLNLSTCYSNSA.
TrBgl2   SGTFAPGRQSTSEPWTVGHNILVAHGRAVKAYRDDFKPASGDGQIGIVLNGDFTYPWDA

                     α10             α11           α12
CcBglA   ℓℓℓℓℓℓℓℓℓℓℓℓ    ℓℓℓℓℓℓℓ      ℓℓℓℓℓℓ           ───
               240       250       260       270       280       290
CcBglA   DEEDIAAHRSDGWNNRWFLDAALKGTYPEDMIKIFSDTNIMPELPKELFTEVFETSDFL
TrBgl2   DPADKEAAERRLEFFTAWFADPIYLGDYPASMRKQLG..DRLPTFTPEERALVHGSNDFY

          β7        β8              η4  β9            α13
CcBglA   ─▶        ─────▶      TT  ℓℓℓ ─▶      TT     ℓℓℓℓℓℓℓℓℓ
               300       310       320       330
CcBglA   GINYYTRQVVKNN..............SEAFIGAESVAMDNPKTEMG.WEIYPQGLYDLLT
TrBgl2   GMNHYTSNYIRHRSSPASADDTVGNVDVLFTNKQGNCI.GPETQSPWLRPCAAGFRDFLV

          β10                  TT        α14                     ───
CcBglA   ℓℓℓℓℓℓ   ─────▶              ℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓℓ
               340       350       360       370       380       390
CcBglA   RIHRDYGNIDLYITENGAAF..NDMVNRDGKVEDENRLDYLYTHFAAALSAIE.AGVPLK
TrBgl2   WISKRYGYPPIYVTENGTSIKGESDLPKEKILEDDFRVKYYNEYIRAMVTAVELDGVNVK

          β11         η5        β12   TT.T  β13       α15
CcBglA   ─────▶    ℓℓℓℓℓℓ      ─────▶        ─────▶  ℓℓℓℓℓℓℓℓℓℓℓ
               400       410       420       430       440
CcBglA   GYYIWSFMDNFEWAEGYEKRFGIVHVNYK.TQERTIKKSAYWYKELIERSNK..
TrBgl2   GYFAWSLMDNFEWADGYVTRFGVTYVDYENGQKRFPKKSAKSLKPLFDELIAAA
```

Figure 5.1. *Amino acid sequence alignment of TrBgl2 and CcBglA enzymes. The secondary structure elements of CcBglA are indicated. Numbering of the principal*

secondary structure elements underlines the (α/β)₈-TIM barrel structure; additional secondary elements are indicated by letters. The 3^{10}-helices are indicated by the character η [2]. Sequence alignment generated using the T-coffee service (http://tcoffee.crg.cat/apps/tcoffee/index.html). Figure generated with ESPript 3.0 (http://espript.ibcp.fr/ESPript/ESPript/) where identical residues are highlighted in red and similar residues are written with black bold characters and boxed in yellow [109]. The residue numbering is that for CcBglA.
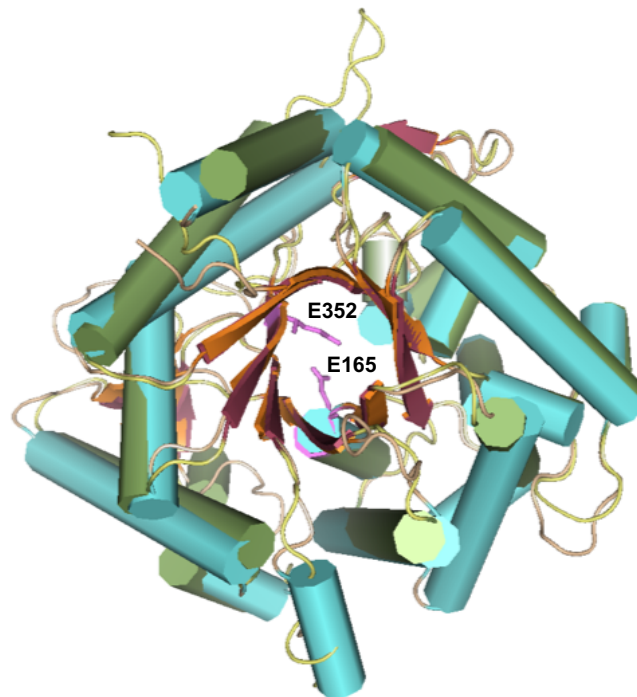
Figure 5.2. *Cartoon representation of the structural alignment of TrBgl2 (PDB ID 3AHY) and CcBglA (PDB ID 3AHX). The catalytic acid/proton donor, E166 and the catalytic nucleophile/base, E352 are shown as stick diagram in magenta. The helices of TrBgl2 and CcBglA are shown in cyan and green, strands in orange and red and loops in wheat and yellow respectively.*

5.1.2. Cloning and protein expression

The *CcbglA* gene was cloned into the pET-YSBLIC3C plasmid under control of the T7 promoter as described in section 7.2. Figure 5.3 shows the

construct pET-YSBLIC3C_*CcbglA* which includes the His$_6$ cleavable extension on the N-terminus.
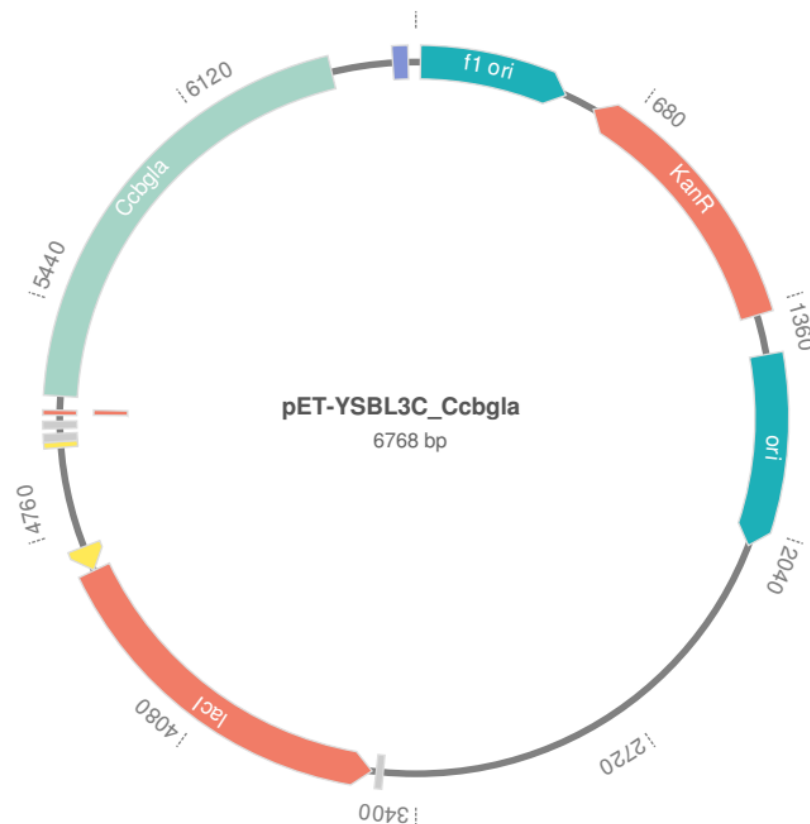
Figure 5.3. *Plasmid map of pET-YSBLIC3C with Ccbgla cloned into the YSBL-LIC3C site. The ORF containing a gene encoding CcBglA is under control of the T7 promoter and begins transcription with an N-terminal 3C-cleavage His$_6$ affinity tag. The plasmid map was generated using SnapGene® Viewer.*

DNA was sent for sequencing to confirm that the desired sequence successfully inserted into the vector. Plasmid DNA was transformed into the *E. coli* BL21 DE3 bacterial strain and *Cc*BglA was overexpressed (as described in Section 7.4) in soluble form.

The recombinant protein with the His$_6$ cleavable tag on the N-terminus has a molecular weight of 54 kDa.

5.1.3. Protein production and characterisation

The protein purification protocol (Section 7.5.1) that was used for CcBglA yields approximately 25 mg/L in optimal conditions. Protein was isolated in two chromatographic steps comprising of a Ni^{2+} affinity capture of the His_6-tagged protein and subsequent size-exclusion (S200) chromatographic separation with purity more than 90% as determined by SDS-PAGE (Figure 5.4). The CcBglA protein eluted in TRIS buffer as two populations, which was consistent with the protein being tetrameric (1st peak) and monomeric (2nd peak).
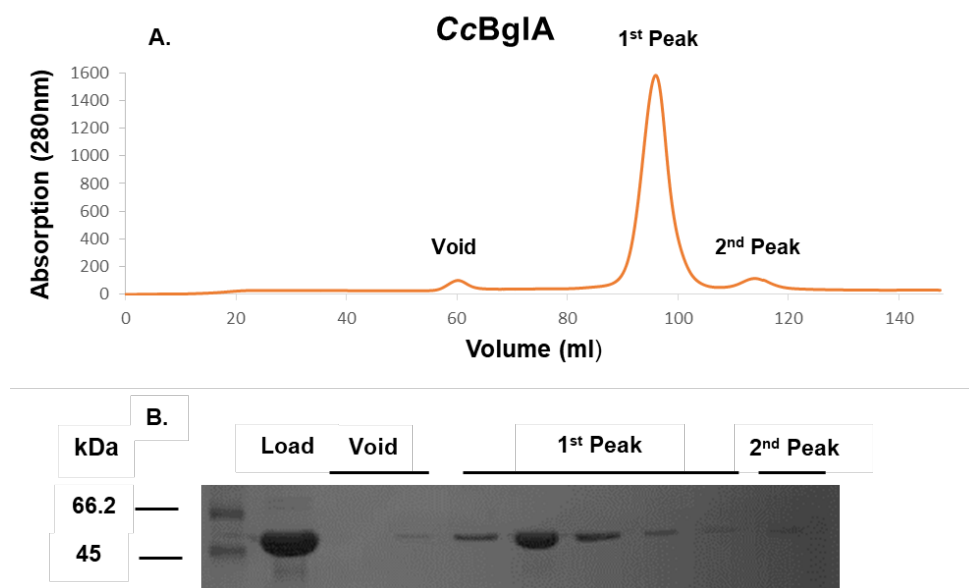
Figure 5.4. *S200 column CcBglA eluate. A. S200 elution peaks detected by UV absorbance at 280 nm, obtained after 2 mL injection of CcBglA at 10 mg/mL concentration, B. SDS PAGE of the Sephacryl S200 column CcBglA eluate. Samples were electrophoresed on a 12% polyacrylamide gel under denaturing and reducing conditions.*
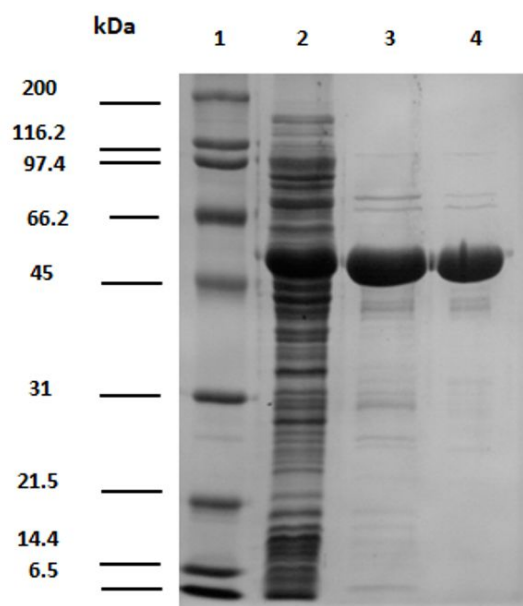
Figure 5.5. *SDS PAGE of the purified CcBglA. 1. Broad molecular weight marker, 2. Cell culture crude extract, 3. Ni-NTA agarose column eluate, 4. Sephacryl S200 column eluate. Samples were electrophoresed on a 12% polyacrylamide gel under denaturing and reducing conditions.*

5.1.4. Oligomerisation studies

SEC-MALS was used to investigate the oligomeric state of the recombinant *Cc*BglA. The analysis revealed the presence of one main population of the *Cc*BglA which corresponds to an average mass of approximately 216 kDa and 208 kDa after cleavage of the N-terminus His$_6$ tag (Figure 5.6). These masses were consistent with tetrameric forms of purified *Cc*BglA in solution and indicate that the oligomeric state was not affected by the presence of His$_6$ tag on the N-terminus.
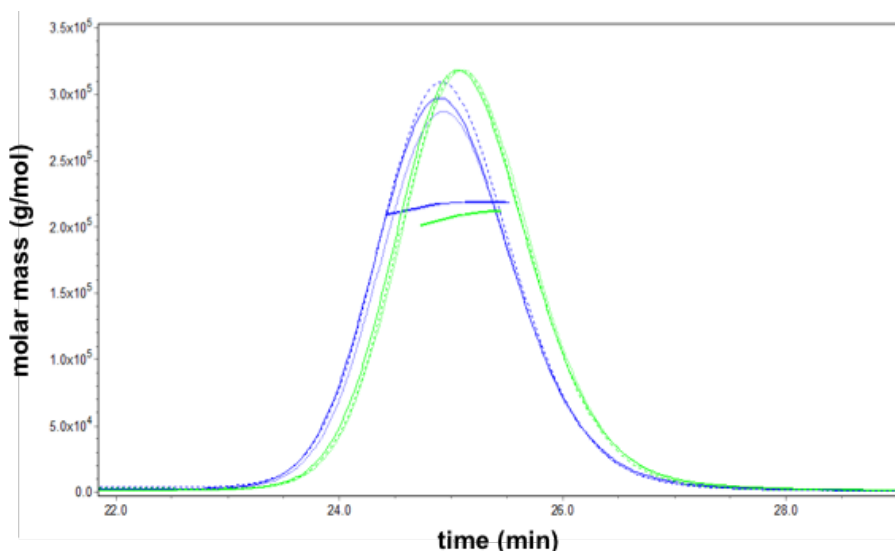
Figure 5.6. *SEC-MALS chromatograms of the elution of CcBglA from a S200 10/300 analytical column. SEC-MALS analysis of CcBglA before (blue) and after (green) cleavage of the N-terminus Hi$_6$. The chromatogram displays the UV at 280 nm (dotted lines), light scattering at a 90° angle, LS, (solid line), refractive index, RI, (dashed line) together with the molar mass of each peak calculated by MALS.*

5.1.5. Assay development

The kinetic properties of *Cc*BglA were investigated by a continuous assay using the same assay as for *Tr*Bgl2 (Section 2.6), with *p*NPG and MUG as a substrate. For *Cc*BglA, the assay was run at 45°C and pH 6.0 the previously reported pH and temperature optimum for activity of this enzyme [2] (Section 7.8). The Michaelis-Menten parameters were calculated (*p*NPG: k_{cat} = 251.7±14.44 s^{-1}, K_m = 0.2±0.04 mM, MUG**:** k_{cat} = 199±4.8 s^{-1}, K_m = 0.28±0.02 mM) (Figure 5.7) through direct fit to the Michaelis-Menten equation and are displayed in Table 5.1. *Cc*BglA shows higher affinity and catalytic efficiency toward the first substrate, *p*NPG, which is consistent with the results obtained previously by Jeng *et al.* [2].

Table 5.1. *Catalytic activity of CcBglA.*

pNPG			MUG		
k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1} mM^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (s^{-1}mM^1)
251.7±14.44	0.2±0.04	1266±349.72	199±4.8	0.28±0.02	714±234
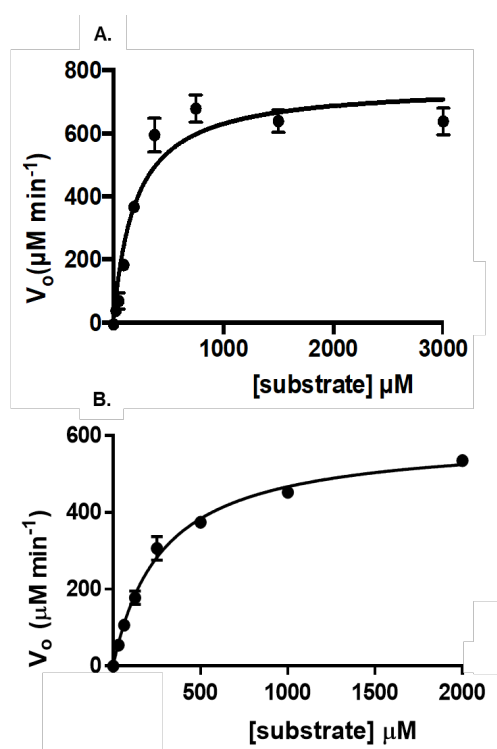
Figure 5.7. *Michaelis-Menten plots for CcBglA. A. pNPG cleavage assay Michaelis-Menten plot for CcBglA (100 nM) and B. MUG cleavage assay Michaelis-Menten plot for CcBglA (100 nM).*

The inhibition efficiency of CcBglA by the known glycosyl hydrolase inhibitor isofagomine (Figure 5.8) [115-116] was assessed by a continuous assay of hydrolysis of MUG at the optimal conditions (45°C, pH 6.0) (Section 7.8). Isofagomine showed inhibition of CcBglA with an IC$_{50}$ value of 2.3±0.3 µM.

Isofagomine can thus be used to categorize any hit compound as either competitive or non-competitive with isofagomine.
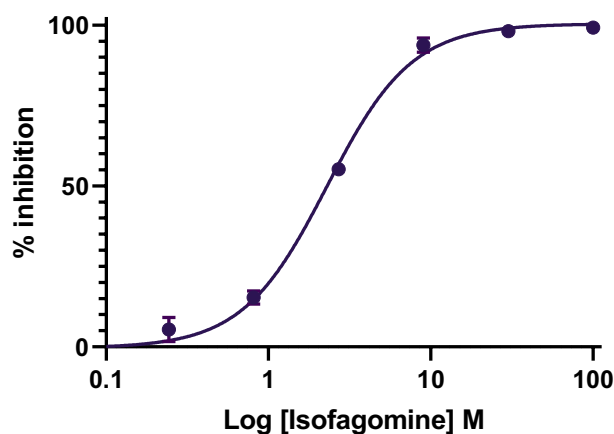
Figure 5.8. *Inhibition of CcBglA by isofagomine.*

5.1.6. Crystal structures

Crystals of *Cc*BglA were obtained by repeating the published conditions [2] using the sitting drop vapour diffusion methodology in a 24-well plate format. Crystal optimisation was performed through variation of these published conditions after removing the His$_6$ tag attached to the protein and the best results were obtained with 22% (w/v) PEG 3350, 0.3 M LiSO$_4$ in 0.1 M Hepes pH 7.5 at 293 K (Section 7.9). The triangular prism crystals of *Cc*BglA grew within 14 days (Figure 5.9) but when tested on the Diamond beamline, diffracted only to 6 Å.

Following this, *Cc*BglA was screened against commercially purchased screens (Table 7.1) in a 96-well plate format, using sitting drop vapour diffusion methodology. Crystallization hits were obtained and were subjected to fine-screening in larger scale 24-well sitting drop format to

test and improve crystal size, reproducibility and diffraction quality. None of the attempts resulted in crystals suitable for diffraction.
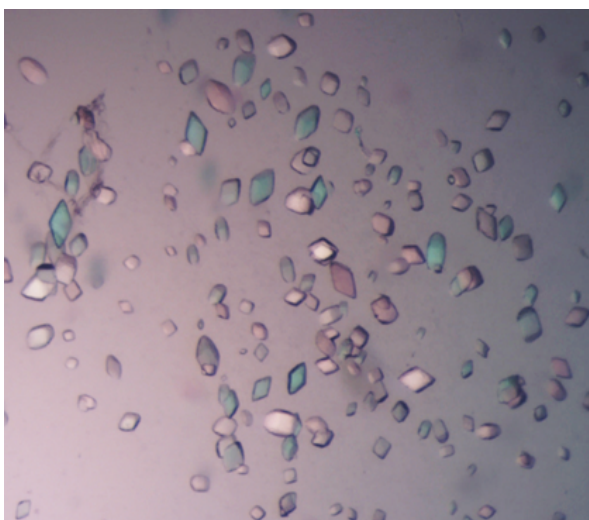
Figure 5.9. *CcBglA crystals obtained after optimisation of the published conditions [2].*

5.1.7. Fragment screening

The York fragment library at the time of these experiments consisted of 570 compounds. The inhibition of *Cc*BglA by isofagomine confirmed that the MUG activity-based assay for *Cc*BglA was robust enough for compound screening, as seen for *Tr*Bgl2 (Section 2.6). Fragment screening was performed at the optimal conditions of the enzyme (45°C, pH 6.0) (Section 7.8). In the initial screen of 570 compounds, 54 fragments inhibited *Cc*BglA activity and 8 fragments enhanced the ability of *Cc*BglA protein to cleave MUG. However, none of the hits were confirmed as an activator in repeats of the MUG activity-based assay at different compound concentrations and after control experiments to identify any false positives due to possible interference from innate compound fluorescence.

5.1.8. Conclusion

The enzyme, *Cc*BglA, was successfully expressed and purified and an enzyme assay established and validated following the published work by Jeng *et al* [2]. SEC-MALS analysis revealed the presence of one main population at a molecular weight consistent with a tetrameric form of purified *Cc*BglA. *Cc*BglA showed higher affinity and catalytic efficiency for the *p*NPG substrate compared to MUG, in full agreement with previous published studies on this enzyme [2]. Crystallisation using published conditions and screening of new conditions did not generate crystals which diffracted adequately for structure determination.

A screen of the York fragment library of 570 compounds in the enzyme assay did not identify any fragments which increased the activity of *Cc*BglA to cleave MUG substrate.

5.2. The α-amylase, TVB146

5.2.1. Introduction

α-amylases (E.C.3.2.1.1) are enzymes that hydrolyse the internal α-1,4-glycosidic bond in starch into polymers composed of glucose units. Amylases constitute one of the largest industrial enzyme markets occupying an approximately 25% share worldwide. They have proved to be extremely useful in most industrial processes, such as in the starch industry for the production of fructose and glucose syrups, in the fuel industry for the production of first-generation bio-ethanol, as well as in the detergent, food, paper and pharmaceutical industries.

The variant "Termamyl-like" α-amylase from *Geobacillus stearothermophilus* (TVB146) belongs to the GH13 family according to CAZY sequence-based classification system. This is a stabilised variant of the widely used subgrouping of industrial GH13 "Termamyl" α-amylase from *Bacillus licheniformis*. The mechanism of the GH13 family enzymes is well defined; they hydrolyse the α-1,4-glycosidic bonds with net retention of anomeric configuration through a covalent glycosyl-enzyme intermediate. TVB146 α-amylase features a two-residue deletion (Δ181-182), which results in a much tighter turn, less exposed loop and a single-point substitution (N193F). The Asp232, Glu262 and Asp329 residues of TVB146 act as the catalytic nucleophile, Bronsted acid/base and "helper" residue respectively.

The determination of the structure of TVB146 revealed the typical "Termamyl" α-amylase three-domain assembly [145]. It comprises domains A, B and C. Domain A is a $(\beta/a)_8$-barrel which is located at the core of the structure. Domain B consists of three antiparallel β-sheets, one four stranded sheet and a short two-stranded sheet. Domain C displays a globular structure and consists of an eight-stranded β-sheet in which four strands are folded in a Greek-key motif. TVB146 contains the unusual Ca^{2+}-Na^+-Ca^{2+} triad at the A/B interface as do all the "Termamyl" α-amylases [145].
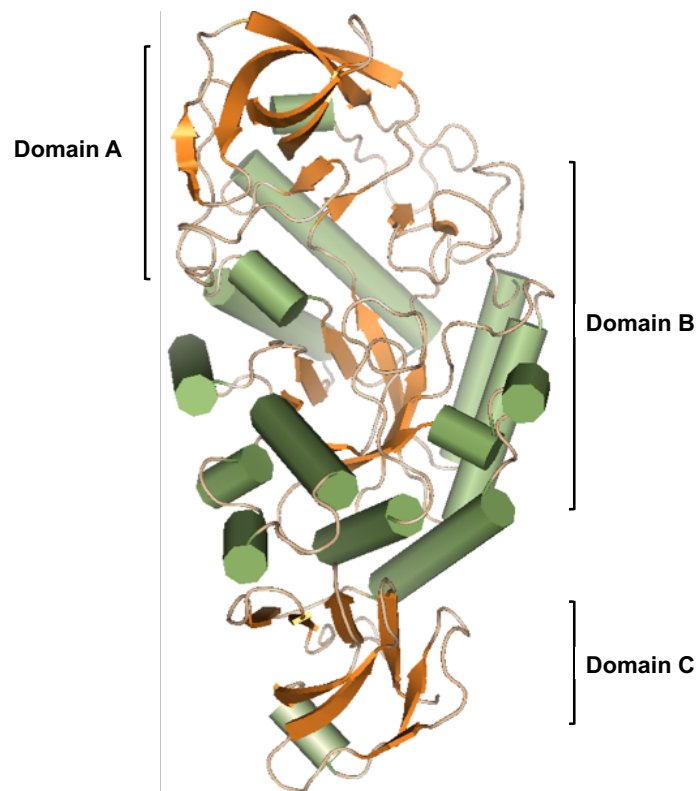
Figure 5.10. *Cartoon representation of the TVB146 amylase structure (PDB ID 4UZU). The helices are shown in green, strands in orange and loops in wheat.*

5.2.2. Cloning and protein expression

The *a-amylase* gene was cloned into the pET-YSBLIC3C plasmid under the control of the T7 promoter as described in section 7.2. Figure 5.11 shows the construct pET-YSBLIC3C_*a-amylase,* which incorporates the His$_6$ cleavable extension on the N-terminus.
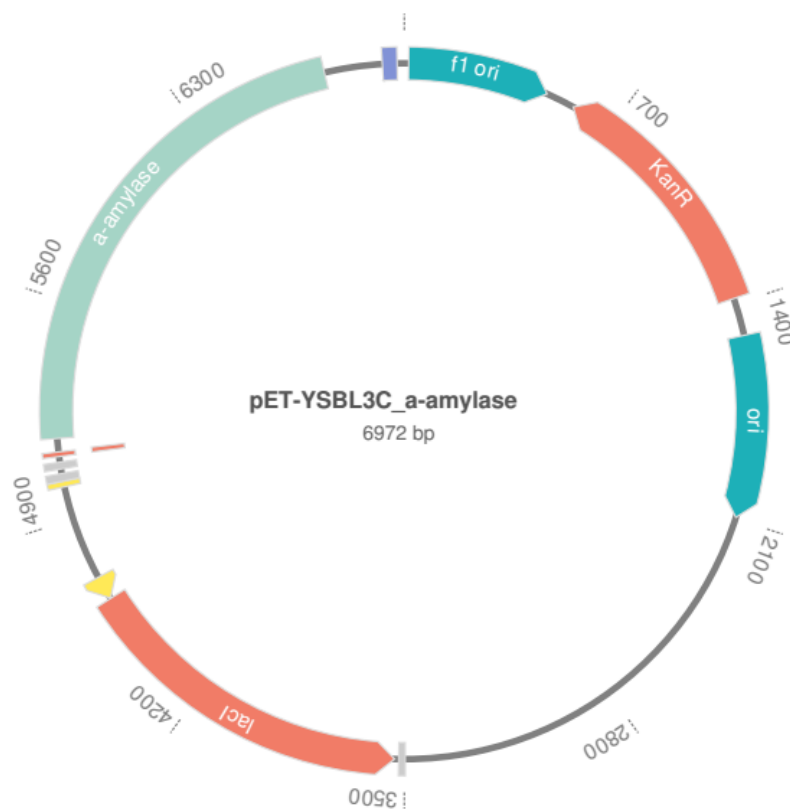
Figure 5.11. *Plasmid map of pET-YSBLIC3C with a-amylase cloned into the YSBL-LIC3C site. The ORF containing a gene encoding a-amylase is under control of the T7 promoter and begins transcription with an N-terminal 3C-cleavage His$_6$ affinity tag. The plasmid map was generated using SnapGene® Viewer.*

DNA sequencing confirmed the success of the cloning process and was followed by the transformation of the construct pET-YSBLIC3C_*a-amylase* into the *E. coli* BL21 DE3 bacterial strain. Unfortunately, the TVB146 variant

was only obtained in insoluble form, present in the pellet after centrifugation, despite screening under various conditions of temperature (16-37°C) and IPTG concentration (0.2-1 mM).

The molecular weight of the recombinant protein with the His$_6$ cleavable extension on the N-terminus is 61 kDa.

5.2.3. Conclusion

TVB146 α-amylase variant from *Geobacillus stearothermophilous* was not soluble after expression in *E. coli* heterologous expression system despite testing under various conditions. This is an industrial α-amylase variant that has been previously expressed successfully in *Bacillus licheniformis* as many other industrial α-amylases [145]. This change of expression system probably leads to incorrect protein folding resulting in an insoluble form of TVB146.

Chapter 6: General conclusions

Cellulose degrading enzymes are of central interest and extensive use in industrial biotechnology through their important role in the production of first-generation biofuels. These enzymes are used in enzymatic cocktails to degrade LC biomass but, as discussed, the low activity of glycoside hydrolases is regarded as the main limiting step in the process. Therefore, glycoside hydrolases with improved catalytic activity are urgently required to improve the economic feasibility for biofuel production. Novel glycoside hydrolases with excellent thermostability such as the fungal glycoside hydrolase, *Tr*Bgl2, have been the subject of considerable research interest.

Although, the majority of drug discovery projects aim to discover enzyme inhibitors, there are several successful examples of enzyme activation by small molecules. This suggested that the discovery of *Tr*Bgl2 activators could be possible. In the work reported here, FBLD methods were used to identify small molecule activators for improving the activity of *Tr*Bgl2. Throughout this work a robust biochemical assay was efficiently used for the initial screen of the York 570 fragment library. Despite the good performance of the assay, false-positive results and the requirement of significant quantities of compounds were an issue and finally only one fragment, compound **21**, was successfully identified as an activator of *Tr*Bgl2 for further investigation. Metal contaminants in stock compounds need to be regularly monitored since these are the cause of most false positives in activator discovery screening approaches. Contrary to this,

ligand observed NMR experiments and NMR CSPs have proved more precise and reliable to characterise and confirm the binding of small molecules to TrBgl2. Moreover, biophysical approaches are preferred for weak-binding compounds because of their improved sensitivity. However, NMR spectroscopy requires sufficiently concentrated samples which are not always available due to low expression yields of the protein of interest.

Rational exploitation of the commercially available fragment analogues of the initial hit, resulted in the identification of activators with greater impact on the maximum rate of TrBgl2 and with much better solubility. It is important to point out that the maximum activation calculated reflects the mechanism of activation which is possibly due to conformational changes or stabilisation of some aspect of the mechanism. This is different from the AC_{50} value that is calculated which is related to the affinity of the compound for the binding site of ligand. As discussed, thiourea and chemical groups ortho to thiourea appear to be essential for this enzyme activation. Groups larger than methyl at the ortho position to thiourea twist the aryl ring out of the plane with the thiourea which could explain why other compound that had a group meta or para instead of ortho to thiourea were inactive. Interestingly, one of the fragment analogues (**30**) appeared to be a potent TrBgl2 inhibitor when retaining the thiourea group in the absence of any chemical group ortho to thiourea.

Initially, this work intended to utilise X-ray crystallography to characterise the activator/protein/substrate complexes after library screening and

characterisation of the binding to the hit compounds. Despite considerable effort, it proved impossible to grow crystals of these complexes. However, a collaboration enabled the use of advanced NMR experiments determine and map the binding site of the activator identified. Although large proteins (>30 kDa) usually precludes analysis by NMR, selective labelling demonstrated that good quality spectra can be obtained identifying a sufficient number of restraints for a small molecule ligand bound to a protein. An extensive series of NMR experiments and data analyses were then used to make a partial assignment of the spectrum for *Tr*Bgl2 which was sufficient to allow the identification of the activator binding site. The CSPs observed in the TROSY spectrum are not necessarily correlated with proximity to the ligand-binding site as they can also result from ligand-induced conformational changes. The number of TROSY CSPs from nuclei in the protein backbone are limited as most ligand-protein interactions are mediated by sidechains which are distant from the detected backbone amides. Therefore, selective labelling of I, V, and L residues was also performed to provide more information about the interaction of the bound ligand with the methyl sidechains interactions of these amino acids, although the assignments for these groups where not as good compared to the backbone. Although methyl groups have favourable relaxation properties resulting in good quality spectra even for large proteins, one limitation is the requirement for methyl groups within the ligand binding pocket which is not always the case. The effort invested to assign as much of the backbone and IVL methyl group resonances for *Tr*Bgl2 as possible,

followed by mapping the assigned CSPs determined and molecular docking experiments allowed the investigation of activator-enzyme interactions that provided very interesting data and insight into the mechanism of activation. Another limitation is that complete assignment will fail for large proteins due to issues involving spectra overlap, resolution and poor sensitivity [126][129]. However, partial assignment in combination with available structural data may be sufficient for mapping the assigned CSPs as seen in this work. While it is true that these techniques are not as accurate as obtaining high resolution X-ray crystal structures of the activator-protein complexes, they have proven extremely useful in the cases where crystals are not available or weakly binding fragments do not give rise to electron density in the crystal structures. Nevertheless, such advanced NMR experiments are a rather expensive approach which require isotopic labelling of the protein, long experimental times and NMR expertise.

Although, mapping of CSPs and molecular docking was an alternative procedure to crystallography that provide some information about the activator binding site, the precise orientation of ligand and intermolecular contacts are well beyond the capabilities of CSPs analysis. Therefore, a set of intermolecular NOE experiments were performed where NOEs exchanged with the ligand to determine the accurate orientation of the bound activator, but this was not successful. Such approaches could provide valuable restrains to perform guided docking of small molecules. Although, intermolecular NOEs approaches are quite powerful and can provide the precise data about which protein nuclei are in contact with the compound,

they suffer from several limitations regarding the requirement for many intermolecular NOE contacts and the pattern of CSPs induced by ligand binding which might not be robust, is used by the resonance assignment [143].

Mapping of observed CSPs for activator **27** onto the known crystal structure of *Tr*Bgl2 in combination with molecular docking including CSPs as ambiguous restraints, provided a useful model when crystallography was not an option for defining the substrate entrance to the active site as the activator binding site. Although, the most numerous and largest chemical shifts were observed at the entrance of *Tr*Bgl2's active site, shifts for residues surrounding the active site were also observed indicating activator-induced conformational changes. The CSPs identified for fragment inhibitor **30** were also mapped onto the apo structure of *Tr*Bgl2 revealing the same binding site for both fragment analogues. Such observations allowed us to speculate that depending on the fit between the substrate, MUG and **21**-like compound, the substrate is properly fitted in the active site acting as an improved substrate adopting a productive conformation which leads to enhanced *Tr*Bgl2 activity (**27**) or the substrate obtains a less favour fit adopting a non-productive conformation which leads to *Tr*Bgl2 inhibition (**30**). The described mechanism is in fact similar to the model for sirtuin activation, which suggests that resveratrol binds close to the active site entrance and then traps the bound substrate resulting in a more suitable orientation of the substrate [75]. In summary, our results indicate a mechanism enabling **21**-like compounds to directly modulate *Tr*Bgl2.

The activators identified in this thesis could be the starting point for future attempts to design higher affinity compounds that induce higher rates of _Tr_Bgl2 activation while retaining the same binding site. The long-term aim that was not realised in this work is to develop covalent strategies to cross-link the activator to the enzyme increasing the economic feasibility and purity of the enzyme products. The activator is modified in order to tether covalently to a cysteine introduced into the enzyme. This is a strategy that has been already successfully applied to BtGH84, generating a constitutively-activated enzyme with significantly enhanced activity [80]. Although, NMR provided a solid platform for the initial steps towards a challenging aim, more precise characterisation of _Tr_Bgl2/activator/substrate interactions are required.

As already discussed, FBLD has been widely used to identify enzyme inhibitors. This work expands the use of FBLD and small molecule activation to industrially relevant biocatalysts and has laid the path for FBLD use as an alternative to traditional protein engineering approaches affecting the conformational changes that many enzymes perform during catalysis. Directed evolution is the most widely used protein engineering approach, whereby random mutagenesis is performed by error-prone PCR to generate large libraries of enzyme variants. However, it can only explore the chemistries through the genetic code [80].

Although other enzymes were also explored for enzyme activation such as the close bacterial homologue of _Tr_Bgl2, _Cc_BglA and the "Termamyl-like"

Geobacillus stearothermophilus α-amylase (TVB146), these were not successful. In the initial screen of the York library, none of the fragments appeared to enhance the activity of *Cc*BglA. Alternative approaches such as mimicking the activator binding site identified through genetic mutations could be considered for *Cc*BglA given the highly conserved active site that the *Tr*Bgl2 and *Cc*BglA enzymes share. In addition, *Cc*BglA exhibits higher activity and thermostability compared to *Tr*Bgl2 and therefore, *Cc*BglA is a better starting point for enzyme activation studies. On the other hand, the α-amylase was not soluble in *E. coli* heterologous expression system, and thus failed to satisfy the desired criteria of an ideal target for FBLD. A catalytically active enzyme that could be readily generated through *E. coli* over-expression system would allow rapid and easy enzyme production for the library screening, crystallisation and possible introduction of mutations for compound tethering.

Chapter 7: Materials and methods

7.1. Chemicals and reagents

The fragment library was assembled as described by Schulz *et al.* [146] and dissolved in DMSO-d$_6$ to a concentration of 200 mM and quality controlled using ^{1}H NMR spectroscopy. The library is a subset of the Maybridge Ro3 diversity library (https://www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__230/DesktopDefault.aspx). The entire library contains 2500 compounds of which 560 compounds were selected for maximum diversity in chemical representation, for improved distribution across a PMI plot in comparison to the Ro3 library and more reliability in terms of solubility.

7.2. Cloning

Genes from *Geobacillus stearothermophilus* ("Termamyl"-like *a-amylase*) and *Trichoderma reesei* (*Trbgl2*; AB003110) were synthesized and optimised to the favored codon usage for *E. coli* by Invitrogen. The *a-amylase* gene fragment was amplified by polymerase chain reaction (PCR) with forward 5'-CCAGGGACCAGCAATGGCAGCACCGTTTAATGGC-3' and reverse 5'-GAGGAGAAGGCGCGTTATTACGGCCATGCAACCAGAC-3' primers (overhangs underlined). The *Trbgl2* gene fragment was amplified by PCR with forward 5'-CCAGGGACCAGCA ATGCTGCCGAAAGATTTTCAGTGG-3' and reverse 5'-GAGGAGAAGGCGCGTTATGCTGCTGCAATCAGTTCATCAAAC-3' primers (overhangs underlined). Gene from *Clostridium cellulovorans* (*CcbglA*; NCBI accession number AY268940) was isolated from the genomic

DNA that was purchased by the DSMZ company. The *CcBglA* gene fragment was amplified by PCR with forward 5'-CCAGGGACCAGCAATGGAAAAGCTAAGATTTCCCA-3' and reverse 5'-GAGGAGAAGGCGCGTTATTACTTATTAGATCTTTCTATAAGCTCC-3' primers (overhangs underlined). Each PCR product of the three genes was cloned into the expression vector of pET-YSBLIC3C using the NEBuilder HiFi DNA Assembly Cloning kit (NEB) which allows rapid assembly of DNA fragments without any ligation step due to complementary overhangs on insert and vector. The vector linearised by PCR with forward 5'-TTGCTGGTCCCTGGAACAGAACTTCC-3' and reverse 5'-CGCGCCTTCTCCTCACATATGGCTAGC-3' primers. The expression vector of pET-YSBLIC3C carries the T7 promoter, a hexahistidine (His$_6$) cleavable extension on the N-terminus and kanamycin resistance gene. All the resulted DNA constructs pET-YSBLIC3C_*a-amylase*, pET-YSBLIC3C_*Trbgl2* and pET-YSBLIC3C_*CcbglA* were sequenced and verified.

7.3. Site-directed mutagenesis

Selected codon in the pET-YSBLIC3C_*Trbgl2* recombinant plasmid previously obtained (Section 8.2) was mutated by PCR in order to obtain the E367Q substitution in the *Tr*Bgl2. The *E367Q TrBgl2* gene fragment was produced by polymerase chain reaction (PCR) with forward 5'-CCTATTTATGTTACCCAAAATGGCACCAG-3' and reverse 5'-GCCATTTTGGGTAACATAAATAGGCGGA-3' primers (overhangs underlined). The PCR product was cloned into the expression vector of pET-YSBLIC3C

using the Gibson Assembly Cloning kit (NEB) which allows rapid assembly of DNA fragments without any ligation step due to complementary overhangs. The resulted DNA construct pET-YSBLIC3C_E367Q*Trbgl2* was sequenced and verified.

7.4. Protein expression

7.4.1. Unlabelled protein

The constructs pET-YSBLIC3C_*Trbgl2* pET-YSBLIC3C_E367Q*Trbgl2,* pET-YSBLIC3C_*CcbglA* and pET-YSBLIC3C_*a-amylase* were transformed into *E. coli* Bl21 (DE3) competent cells which are deficient in Lon protease (cytoplasm) and OmpT protease (outer membrane) and are suitable for expression of genes cloned downstream of a T7 promoter. An overnight culture was used to inoculate 0.5 L of LB medium containing 100 µg/mL Kanamycin and was subsequently incubated at 37°C (for α-amylase) and 30°C (for *Tr*Bgl2 and *Cc*BglA) with shaking until the medium reached an optical density, OD$_{600}$ of 0.8-1.0. The temperature was then reduced to 13°C (for *Tr*Bgl2, E367Q *Tr*Bgl2 and *Cc*BglA) for one h before adding the isopropyl β-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM for *Tr*Bgl2 and E367Q *Tr*Bgl2, and 0.2 mM for *Cc*BglA. The cells (for *Tr*Bgl2, E367Q *Tr*Bgl2 and *Cc*BglA) were pelleted by centrifugation after 16 h following the protocol from Jeng *et al*. [2]

7.4.2. ^{15}N Isotope-Labelled TrBgl2 protein

Production of ^{15}N-labelled E367Q TrBgl2 followed a similar protocol to the unlabelled proteins. An overnight TrBgl2 LB medium preculture from a freshly transformed colony was pelleted by centrifugation at room temperature (4,000 rpm, 10 min) and resuspended in baffled flasks with 0.5 L of M9 medium containing 3 g Na_2HPO_4, 1.5 g KH_2PO_4, 0.25 g NaCl, 0.5 g $^{15}NH_4Cl$, 2 mM $MgSO_4$, 100 μM $CaCl_2$, 0.4% glucose, 5 mL Gibco MEM vitamins solution (SIGMA) and 0.5 mL trace elements solution (50 mM $FeCl_3$, 50 mM $ZnSO_4$, 100 mM $MnCl_2$, 10 mM $CuSO_4$) and 100 μg/mL Kanamycin. Baffled flasks increase oxygenation of the medium during shaking for optimal protein expression. The 0.5 L M9 culture was subsequently incubated at 30°C with shaking until the medium reached an optical density, OD_{600} of 0.8-1.0. The temperature was then reduced to 13°C for one h before adding the isopropyl β-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The cells were pelleted by centrifugation after incubation at 13°C overnight with shaking at 200 rpm.

7.4.3. ^{15}N, ^{13}C, ^2H Isotope-Labelled TrBgl2 protein

A 5 mL LB rich growth medium (**Preculture 1**) inoculated with a freshly transformed colony containing 100 μg/mL Kanamycin and was subsequently incubated at 37°C in an orbital shaker, 200 rpm for 6-8 h. A 5 mL M9/H_2O medium (**Preculture 2**) containing 30 mg Na_2HPO_4, 15 mg KH_2PO_4, 2.5 mg NaCl, 5 mg $^{15}NH_4Cl$, 7.5 mg [^2H,^{13}C] glucose, 100 μM $CaCl_2$, 2 mM $MgSO_4$, 5 mL Gibco MEM vitamins solution (SIGMA) and 0.5 mL trace

elements solution (50 mM $FeCl_3$, 50 mM $ZnSO_4$, 100 mM $MnCl_2$, 10 mM $CuSO_4$) and 100 µg/mL Kanamycin was inoculated using Preculture 1 so that the starting OD_{600} = 0.05 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm overnight. Preculture 2 was pelleted by centrifugation at room temperature (4,000 rpm, 10 min) and resuspended in 50 mL conical baffled flask with 10 mL M9/50% H_2O–50% D_2O medium (**Preculture 3**) so that the starting OD_{600} = 0.05 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm for 6-8 h. Conical baffled flasks increase oxygenation of the medium during shaking for optimal protein expression. Preculture 3 was pelleted by centrifugation at room temperature (4,000 rpm, 10 min) and resuspended in 250 mL conical baffled flask with 50 mL M9/D_2O medium (**Preculture 4**) so that the starting OD_{600} = 0.25 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm overnight. Preculture 4 was pelleted by centrifugation at room temperature (4,000 rpm, 10 min) and resuspended in 1 L conical baffled flask with 250 mL M9/D_2O medium (**Preculture 5**) so that the starting OD_{600} = 0.25 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm until the medium reached an optical density, OD_{600} of 0.7-0.8. The temperature was then reduced to 16°C and isopropyl β-thiogalactopyranoside (IPTG/D_2O) was added to a final concentration of 0.5 mM. The cells were pelleted by centrifugation after incubation at 16°C overnight with shaking at 200 rpm.

7.4.4. ^{15}N, ^{13}C, ^{2}H and IVL methyl groups specifically labelled TrBgl2 protein

A 5 mL LB rich growth medium (**Preculture 1**) inoculated with a freshly transformed colony containing 100μg/mL Kanamycin and was subsequently incubated at 37°C in an orbital shaker, 200 rpm for 6-8 h. A 5 mL M9/H_2O medium (**Preculture 2**) containing 30 mg Na_2HPO_4, 15 mg KH_2PO_4, 2.5 mg NaCl, 5 mg $^{15}NH_4Cl$, 7.5 mg [^{2}H,^{13}C] glucose, 100 μM $CaCl_2$, 2 mM $MgSO_4$, 5 mL Gibco MEM vitamins solution (SIGMA) and 0.5 mL trace elements solution (50 mM $FeCl_3$, 50 mM $ZnSO_4$, 100 mM $MnCl_2$, 10 mM $CuSO_4$) and 100 μg/mL Kanamycin was inoculated using Preculture 1 so that the starting OD_{600} = 0.05 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm overnight. Preculture 2 was pelleted by centrifugation at room temperature (4,000 rpm, 10 min) and resuspended in 50 mL conical baffled flask with 10 mL M9/50% H_20−50% D_2O medium (**Preculture 3**) so that the starting OD_{600} = 0.05 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm for 6-8 h. Conical baffled flasks increase oxygenation of the medium during shaking for optimal protein expression. Preculture 3 was pelleted by centrifugation at room temperature (4,000 rpm, 10 min) and resuspended in 250 mL conical baffled flask with 50 mL M9/D_2O medium (**Preculture 4**) so that the starting OD_{600} = 0.25 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm overnight. Preculture 4 was pelleted by centrifugation at room temperature (4,000 rpm, 10 min) and resuspended in 1 L conical

baffled flask with 250 mL M9/D$_2$O medium (**Preculture 5**) so that the starting OD$_{600}$ = 0.25 and was subsequently incubated at 37°C in an orbital shaker, 200 rpm. When the optical density, OD$_{600}$ = 0.4, 30 mg of Val/Leu precursor, 2-Keto-3-(methyl-d$_3$)-butyric acid- 1,2,3,4-^{13}C$_4$, 3-d sodium salt (SIGMA, Cat. 637858) and 17.5 mg of IIe precursor, 2-Ketobutyric acid-^{13}C$_4$,3,3-d$_2$ sodium salt hydrate (SIGMA, Cat. 607541) was added and was subsequently incubated at 37°C in an orbital shaker, 200 rpm until the medium reached an optical density, OD$_{600}$ of 0.7-0.8. The temperature was then reduced to 16°C and isopropyl β-thiogalactopyranoside (IPTG/D$_2$O) was added to a final concentration of 0.5 mM. The cells were pelleted by centrifugation after incubation at 16°C overnight with shaking at 200 rpm.

7.5. Protein purification

7.5.1. Unlabelled protein

The bacterial cells (for *Tr*Bgl2, E367Q *Tr*Bgl2 and *Cc*BglA) resuspended in a lysis buffer containing 20 mM TRIS-HCl pH 7.5, 400 mM NaCl, 10 mM Imidazole and one EDTA-free protease inhibitor cocktail tablet and then destroyed by using the Cell Disruption System (Benchtop, Constant Systems Ltd). The cell debris was removed by centrifuge at 18,000 rpm for 45 min at 4°C. The supernatant was loaded on a chromatography column containing 5 mL Ni-NTA agarose (QIAGEN, Germany), equilibrated with lysis buffer. The His-tagged protein was subsequently washed with the same buffer and subsequently eluted by a linear gradient of 10-300 mM Imidazole. The purified His-tagged proteins were analysed on 12% SDS-

PAGE and concentrated to a volume of 2 mL using 10 kDa cut-off size AMICON-ultra filters. The samples were loaded onto a pre-packed XK16/600 Sephacryl S200 column using an AKTA start purifier system (GE Healthcare). It was carried out using a buffer containing 50 mM TRIS-HCl pH 8.0, 100 mM NaCl, 3 mM dithiothreitol (DTT). The flow rate was 1 mL/min, and elutions was monitored at 280 nm. The His_6 tag *Cc*BglA removed after incubation with 3C protease at 4°C for 18 h. The protein was then re-run through the Ni-NTA agarose column to retain the His_6 tag fragment. His_6-tag free *Cc*BglA was obtained in the flow through fractions, was pooled and loaded again onto an S200 gel filtration column.

Protein quantification was carried out using the optical density values of the sample at 280nm using extinction coefficient of 1.86 for *Tr*Bgl2 and E367Q *Tr*Bgl2 and 2 M^{-1} cm^{-1} for *Cc*BglA as determined by the PROTPARAM program (EXPASY proteomics server).

7.5.2. Isotope-labelled protein

Purification followed the same protocol as unlabelled *Tr*Bgl2 protein above (Section 7.5.1).

7.6. Estimation of purified protein concentration

A Nanodrop ND-1000 spectrophotometer (Thermo Scientific) was used to estimate the concentration of purified protein samples. The theoretical extinction coefficient of each protein was calculated using the ExPASy

ProtParam tool (https://web.expasy.org/protparam/) and entered into the Nanodrop control software.

7.7. Size Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS)

Size exclusion chromatography with multiple-angle light scattering (SEC-MALS) is used to measure molecular mass and rms radius R_g of macromolecules in solution. Size-exclusion chromatography (SEC) separated the molecules based on their hydrodynamic volume and size while MALS measures the absolute molar mass and size of the molecules using the intensity and the light scattered by the sample at a range of angles. Therefore, combining SEC and MALS in an SEC-MALS experiments allows to obtain more accurate information about the size and shape of macromolecules in solution [113].

Sample injection volume of 100 µl CcBglA (2.5 mg/mL) was fractionated on a Superdex S200 10/300 column (GE Healthcare) equilibrated with buffer containing 50 mM HEPES in pH 8.0, 100 mM NaCl, 3 mM dithiothreitol (DTT) at 0.5 mL/min flow rate. The system comprises a UV 280/254 nm detector, a Wyatt HELEOS-II multi-angle light scattering detector and Wyatt Rex refractive index detector for the sample monitoring. Data analysis were performed using ASTRA V software (Wyatt Technology) and MWs were estimated using the Zimm fit method with degree 1. Detector normalization was performed using BSA.

7.8. Enzyme kinetics determination

*p*NPG (4-nitrophenyl-β-D-GlcNAc) and 4-methylumbelliferyl-β-D-glucopyranoside (MUG) substrates were used for kinetic analysis for *Tr*Bgl2 and *Cc*BglA. Both reactions were performed in 160 mM phosphate-citrate buffer in pH 6.0, were initiated by adding 20 µL of the enzyme to give a final concentration of 100 nM and were assayed at 40 and 45°C for *Tr*Bgl2 and *Cc*BglA respectively. Reactions were carried out using 0-12 mM and 0-3 mM *p*NPG for *Tr*Bgl2 and *Cc*BglA respectively, in a total volume of 150 µL per well. The released *p*-nitrophenol was monitored continuously at 400 nm by BMG Labtech POLARstar OPTIMA.

The fluorescent kinetic assay was performed using 31.25-4000 µM MUG in a total volume of 100 µL per well. Fluorescent 4-methylumbelliferyl release was recorded continuously at 355 nm excitation and 460 nm emission for 20 s by BMG Labtech POLARstar OPTIMA. Corrections were made to the detected gradients to adjust for the inner filter effect of MUG and covert the gradients into units of µmol min^{-1} according to a standard curve of MUG. Values quoted throughout are mean and standard deviation of two independent replicates. K_m and k_{cat} values were calculated by fitting the data to the Michaelis- Menten equation.

The same fluorescent kinetic assay was performed for library screening for *Tr*Bgl2 and *Cc*BglA using 500 and 300 µM MUG in a total volume of 150 per well. Compounds dissolved in DMSO were added to each well to yield a final concentration from 1 to 2 mM (2.3% DMSO). The reactions were initiated

by adding 20 µL of the enzyme to give a final concentration of 100 nM. Control experiments to evaluate a possible interference from innate compounds library fluorescence were done by repeating the assay in the absence of the enzyme.

Inhibition of TrBgl2 and CcBglA by isofagomine (IC_{50}) was performed using a final concentration of 100 nM protein and 500 µM MUG with various concentrations of isofagomine (0-100 µM) in a total volume of 100 µL per well. The reactions were performed in 160 mM phosphate-citrate buffer in pH 6.0 and were assayed at 40 and 45°C for TrBgl2 and CcBglA respectively. Reactions were initiated with the addition of enzyme, and reactions progress were monitored continuously at 355 nm excitation and 460 nm emission. All curve fitting for enzyme kinetics and inhibition experiments was performed using GraphPad Prism 5.0.

AC_{50} measurements of fragment activators were determined in a similar manner to the IC_{50} values. Assays were carried out in a 96 well plate (Black Nunc F96 MicroWell Plates) at 40 and 45°C for TrBgl2 and CcBglA respectively in a total volume of 150 µl per well, buffer (160 mM phosphate-citrate pH 6.0, 2.5% DMSO) with 500 µM MUG, as substrate, 100 nM protein and various concentrations of activators (0-8 mM) depending on the solubility limits of each compound. Reactions were initiated by addition of 20 µl of protein to a final concentration of 100 nM, followed by 5 s of shaking. Fluorescent 4-methylumbelliferyl release was recorded continuously at 355 nm excitation and 460 nm emission for 20 s by BMG

Labtech POLARstar OPTIMA. Rates were calculated using GraphPad Prism 5.0 as the slope of the curve of the collected data. The initial slopes were fitted to sigmoidal dose response curves using GraphPad Prism 5.0 and the AC_{50} values extracted. For generating AC_{50} values assays were repeated 3 times. Final AC_{50} values are reported as mean and standard deviation of 3 AC_{50} curve fits.

Apparent K_m and k_{cat} values were calculated from a Michaelis-Menten curve fit to 9-point substrate titrations at a constant concentration of the activator. The covariation experiments used the same range of substrate concentrations across an activator concentration range. The data obtained from these experiments was fit to the nonessential activator equation:

$$v = \frac{V_{max}[S]}{K_M \left[\dfrac{1 + \dfrac{[A]}{K_A}}{1 + \dfrac{\beta[A]}{\alpha K_A}} \right] + [S] \left[\dfrac{1 + \dfrac{[A]}{\alpha K_A}}{1 + \dfrac{\beta[A]}{\alpha K_A}} \right]}$$

[S] is the concentration of the enzyme, [A] the concentration of the activator, α modifies K_m, β modifies V_{max} and K_A is the half max concentration of the activator.

7.9. Crystallization and structure determination

7.9.1. Crystallisation

Crystallization conditions for *Tr*Bgl2 and *Cc*BglA were screened using the sitting-drop vapor diffusion method. The drops were made up of 0.5 µL of protein solution with an equal volume of reservoir solution and were

equilibrated against 54 μL reservoir solution at 20°C. The concentration of the *Tr*Bgl2 and *Cc*BglA protein was 1.2 mg mL^{-1} in a buffer containing 50 mM TRIS pH 8.0, 100 mM NaCl and 3 mM DTT. Initial crystallization screening was performed using commercially available crystallization kits (Table 7.1).

Table 7.1. *Commercially available screens.*

Crystal Screen HT (Hampton Research)
The JCSG+ (Molecular Dimensions)
Peg Ion HT (Hampton Research)
PDB Screen (Jobie's)
MPD screen (Qiagen)
Morpheus (Molecular Dimensions)
Index (Hampton Research)

The crystals of native *Tr*Bgl2 were obtained with 35% (w/v) PEG 3350, 0.2 M NaCl and 0.1 M TRIS pH 8.5 at 293 K. The final crystal size was reached in 3 d. Crystals of *Tr*Bgl2/isofagomine complex were obtained at similar conditions to that of native *Tr*Bgl2 containing saturated isofagomine at 1 mM final concentration. The crystals transferred to cryo-protectant solutions containing mother liquor supplemented with 15% ethyl glycol and cryo-cooled using liquid N$_2$. Diffracting crystals were sent to Diamond Light Source for data collection using a Pilatus 6M-F detector (DecTRIS) at 100 K. An oscillation range of 0.1° degree was chosen and 180° were collected.

7.9.2. Structure determination

Diffraction images for crystals of *apo Tr*Bgl2 and for *Tr*Bgl2/isofagomine complex were indexed and integrated using iMOSFLM [147]. POINTLESS was used to determine the crystallographic space group [148] and structure factor amplitudes were scaled and merged using AIMLESS [149].

The crystals of *apo Tr*Bgl2 and complexed with isofagomine grew in a P1 21 1 space group, and these crystals were almost isomorphous to each other. The *apo* structure was solved by molecular replacement of (MOLREP [150]) using the structure of already determined structure of *apo Tr*Bgl2 (PDB ID 3AHY) as a model. MOLREP was run using default parameters. The model was improved by manual rebuilding and real-space refinement in COOT [151] and maximum-likelihood refinement using Refmac [152].

As the space group of *Tr*Bgl2 liganded with isofagomine was isomorphous to the *apo* crystal, coordinates from the apo structure were used directly for determination of the isofagomine complex structure. R_{free} sets were assigned as they had been for the apo-structure to maintain the uniqueness and integrity of the cross-validation Free R set throughout. Figures were created using CCP4MG [114].

7.10. Thermal Shift Assay (TSA)

Thermal shift assay (TSA) is a technique used to determine protein thermal stability. It is usually used to optimise buffer conditions for crystallographic screening aiming to maximise the protein stability and to assess the effects of ligands on protein targets. TSA exploits an environmentally sensitive dye, SYPRO Orange, which binds non-specifically to hydrophobic surfaces

and water quenches its fluorescence. By gradually heating a protein and dye mixture buried hydrophobic sidechains are revealed and bind the dye as the protein unfolds. When the dye is exposed to this hydrophobic environment the fluorescence intensity of the sample increases by excluding the water molecules and enables an estimation of a protein melting point (T_m). Addition of stabilizing or destabilising ligands the T_m will be altered, this change is known as a thermal shift, or $\Delta T_{m.}$ [153].

Thermal shift screens were performed in 96-well polypropylene PCR plates (Agilent 401334) with a final volume 25 μl/well. Well components were E367Q *Tr*Bgl2 (300 nM), SYPRO orange dye (8X), DMSO (5.2%), substrate, MUG (4 mM) and compound at 4 mM in the same buffer as used for the enzymatic assay (phosphate-citrate, pH 6.0). Melting data was collected using the Agilent Stratagene Mx3005P rtPCR machine ramping from 25°C to 95°C at 30 s per degree. Data was analysed with a combination of the Agilent MxPro software, MTSA [154], JTSA (http://paulsbond.co.uk/jtsa) and the DSF analysis spreadsheet in Excel [153].

7.11. NMR Spectroscopy

NMR experiments were carried out on a Bruker Avance Neo 700 MHz NMR spectrometer equipped with a 5 mm triple-resonance nitrogen-cooled cryoprobe. All NMR data were collected at 298 K.

7.11.1. STD NMR

The Saturation Transfer Difference (STD) NMR experiment relies on the fast exchange between the bound and the free ligand state (dissociation

constant, K_D, ranging from 10^{-8} mol L^{-1} to 10^{-3} mol L^{-1}). In STD-NMR experiment radiofrequency pulses are applied by irradiating at a region of the spectrum that contains only resonances of the protein (on-resonance spectrum); this magnetization is transferred to binding ligands via spin diffusion through Nuclear Overhauser effect (NOE) during the timescale of the experiment leading to partial saturation of ligand resonances (Figure 7.1). The binding of a ligand can be detected as a reduction in signal intensity in the 1D spectrum recorded for the ligand subtracting the *on-resonance* spectrum from the *off-resonance* spectrum which recorded without protein saturation [155].
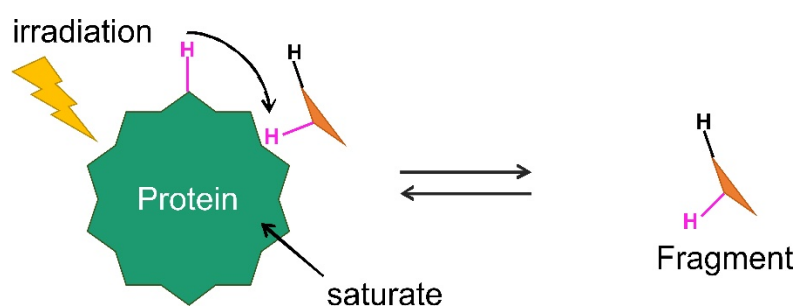
Figure 7.1. *Schematic illustration of the STD experiment.*

The samples contained 20 μM of the enzyme, fragments at a concentration of 500 μM per compound and 100 μM DSS in 50 mM phosphate buffer containing 25 mM NaCl and 5% D_2O at pH 6.0 with or without 500 μM MUG substrate depending on the experiment performed. The experiments were recorded by using a Bruker standard pulse sequence (stddiffesgp.3 [156]). The on-resonance frequency was set to 0.25 ppm and off-resonance irradiation was applied at 40 ppm. STD NMR spectra were collected using standard pulse sequences with a total of 16 scans (NS), 5 s delay between

the scans (d1) and 1.5 s acquisition time (AQ). Data were processed using Topspin 3.2 software.

7.11.2. WaterLOGSY

WaterLOGSY is based on the NOESY experiment involving transfer of magnetization from bulk water molecules during the mixing time to the bound ligand via an intermolecular NOE and spin diffusion (Figure 7.2.). The observed WaterLOGSY signs are opposite for signals of free (positive NOE) and protein-bound ligands (negative NOE) due to different tumbling that they experience. This allows easily discrimination of binders and non-binders [122] [123].

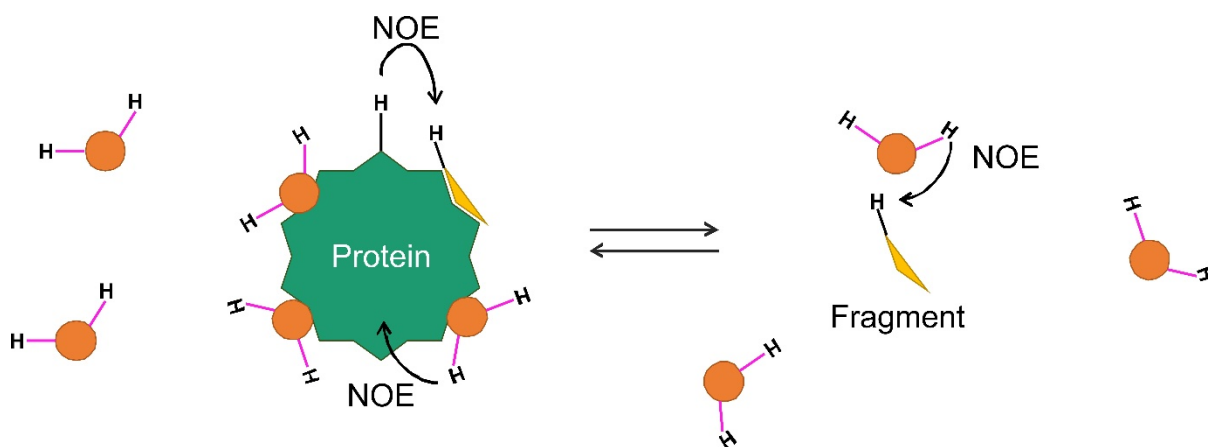**Figure 7.2.** *Schematic illustration of the WaterLOGSY experiment.*

The sample contained 20 µM of the inactive mutant E367Q *Tr*Bgl2, 500 µM of compound **30**, 500 µM of MUG substrate and 100 µM DSS in 50 mM phosphate buffer containing 25 mM NaCl and 5% D_2O at pH 6.0. WaterLOGSY spectra were collected by using a Bruker standard pulse sequence (ephogsygpno.2 [123]) with a total of 32 scans (NS), 3 s delay

between the scans (d1), 1.5 s acquisition time (AQ) and 1.5 s mixing time (D8). Data were processed using Topspin 3.2 software.

7.11.3. TROSY NMR

Small molecules in solution relaxes slowly resulting in a long transverse relaxation time (*T*2) (large *T*2 values) which is translated into narrow line widths (Δv) after Fourier transformation (FT). As molecular masses increase (masses greater than 25 kDa), spectra become crowded due to large numbers of resonances and lines becomes broad due to fast transverse relaxation (smaller *T*2 values). Although, the overlap of NMR signals can be overcome by a proper choice of isotope-labelling, there is still the limitation caused by parts of the signals that relax quickly. In a standard experiment all the parts (parts that relax fast and slow) get mixed together, so everything appears to relax at an averaged fairly fast rate. Transverse Relaxation Optimized Spectroscopy (TROSY) carefully keeps these components separate, and just detects the slow relaxation part which improves spectra resolution and sensitivity of large molecules [124]. TROSY experiment transfers magnetisation via $^1J_{HN}$ coupling from amide protons to the attached nitrogen and then the magnetisation is transferred back to the proton for detection. The result is a correlation of a protein's amide proton and nitrogen resonances which can be considered as a structural ''fingerprint'', generating one peak for each backbone amide, with the exception of proline, and some additional peaks for amino acid sidechains that contain NH groups where buffer exchange is slow enough for detection.

The schematic diagram below (Figure 7.3) indicate the detected nuclei and the flow of magnetisation between them during the experiment.
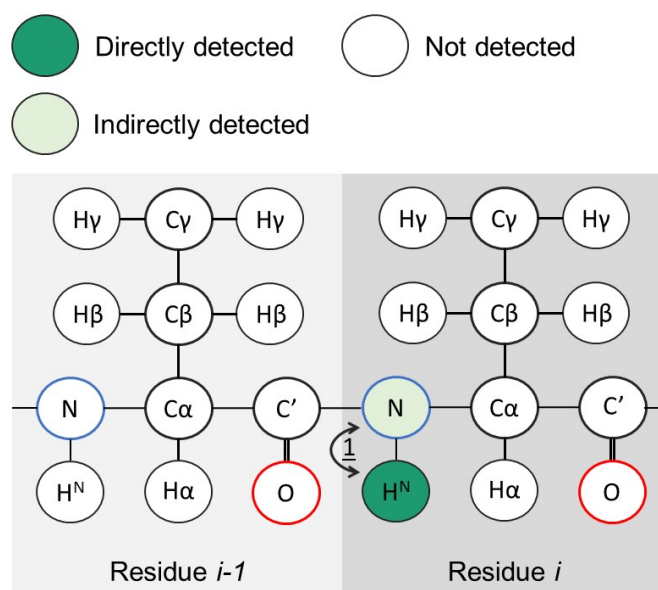

Figure 7.3. *Schematic diagram of 2D ^{1}H ^{15}N TROSY experiment. <u>Magnetisation: $H^N \rightarrow N(t1) \rightarrow H^N(t2)$.</u>*

Ligand binding will induce changes in chemical shifts of the protein's ^{1}H and ^{15}N nuclei. The largest changes will be observed for the ^{1}H ^{15}N that correspond to amide backbone groups of the protein that are closest to the ligand binding site.

Each sample was prepared with 120 µM of the wild type *Tr*Bgl2 or inactive mutant E367Q *Tr*Bgl2, 8 mM/16 mM of compound **27**/compound **30**, 5 mM of MUG substrate and 100 µM DSS in 50 mM TRIS buffer containing 100 mM NaCl and 10% D_2O at pH 8.0. DMSO concentration was usually not above 3.8%. TROSY spectra of the enzyme with DMSO in the buffer were collected as controls. ^{1}H ^{15}N TROSY spectra were collected by using a Bruker standard pulse sequence (trosyetf3gpsi [157, 158]) with 128 scans (NS),

1.5 s delay between the scans (d1), 35 ppm spectra width (SW), 256-time domain data size (TD) and 117 ppm irradiation carrier frequency offset (O1P).

7.11.4. Constant time (CT)^{1}H ^{13}C HSQC

^{1}H ^{13}C HSQC is the carbon equivalent of the ^{1}H ^{15}N HSQC and shows all H-C correlations. ^{1}H ^{13}C HSQC experiment transfers magnetisation via ^{1}J$_{HN}$ coupling from ^{1}H to ^{13}C and then back to the proton for detection. The constant time (CT) version of ^{1}H ^{13}C HSQC is used to circumvents the splitting of signals due to homonuclear ^{13}C ^{13}C J-coupling constant allowing to get higher resolution in the carbon dimension in cases that relaxation is not a major issue (despite the size of *Tr*Bgl2 here). The constant time refers to the evolution period between the two INEPT[1] steps which is kept constant during the experiment [159].
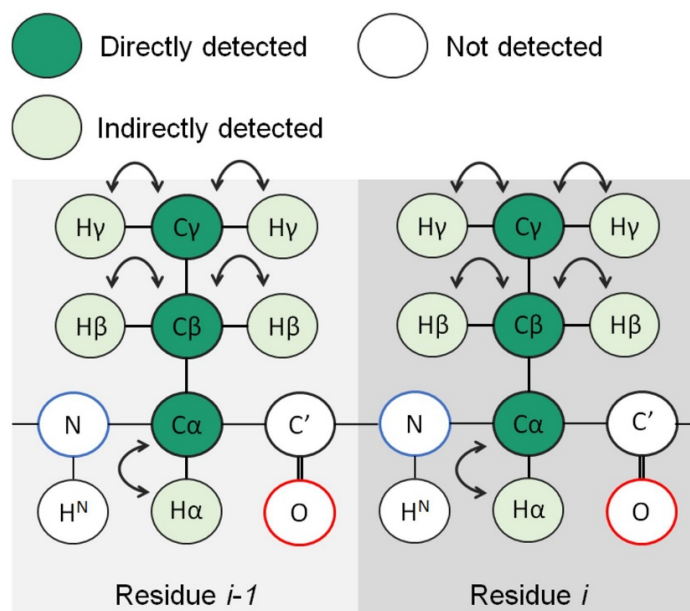

Figure 7.4. *Schematic diagram of 2D ^{1}H ^{13}C HSQC experiment.* <u>*Magnetisation:*</u>

<u>*H^C →C(t1) →H^C(t2).*</u>

Each sample was prepared with 250 μM of the wild type TrBgl2, 8 mM of compound **27**, in 50 mM TRIS buffer containing 100 mM NaCl and 10% D_2O at pH 8.0. DMSO concentration was usually not above 3.8%. ^1H ^{13}C HSQC spectra of the enzyme with DMSO in the buffer were collected as controls. ^1H ^{13}C HSQC spectra were recorded by using a Bruker standard pulse sequence (hsqcctetgpsp [159]) with 8 scans (NS), 1.5 s delay between the scans (d1), 30 ppm spectra width (SW), 256-time domain data size (TD) and 15 ppm irradiation carrier frequency offset (O1P).

[1] Insensitive nuclei enhanced by polarization transfer (INEPT) is a signal enhancement process which increases the amount of bulk magnetisation of the insensitive nucleus. The enhancement occurs through transfer of nuclear spin polarization from proton to less sensitive nucleus (e.g. ^{13}C, ^{15}N) with the appropriate ^1H-X coupling.

7.12. Protein assignment

7.12.1. Sample preparation

Each sample was prepared with 250 μM of the ^2H, ^{13}C, ^{15}N labelled TrBgl2 for backbone assignments and methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$), V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N, ^{13}C, ^2H] sample of TrBgl2 for methyl groups assignment, 100 μM sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as reference in 50 mM TRIS-HCl containing 100 mM NaCl and 10% D_2O at pH 8.0.

7.12.2. Data collection

The majority of the NMR spectra presented in this section were collected by Dr Alex Heyam at the University of York and Dr Marta Carneiro from the ZoBio B.V. at University of Leiden. Two spectrometers were used, a 700 MHz Bruker Avance Neo and an 800 MHz Bruker Avance II, both controlled by TopSpin 3 (Bruker). Spectra were recorded at 298 K.

Table 7.2. *List of NMR experiments recorded.*

Experiment	Sample labelling
HNCO	^{2}H, ^{13}C, ^{15}N
HN(CA)CO	^{2}H, ^{13}C, ^{15}N
HNCA	^{2}H, ^{13}C, ^{15}N
HN(CO)CA	^{2}H, ^{13}C, ^{15}N
HN(CO)CACB	^{2}H, ^{13}C, ^{15}N
HNCACB	^{2}H, ^{13}C, ^{15}N
HN(CA)CB	^{2}H, ^{13}C, ^{15}N
^{15}N-resolved NOESY	^{2}H, ^{13}C, ^{15}N
^{13}C-resolved NOESY	^{2}H, ^{13}C, ^{15}N
HMCM(CG)CBCA	{I(δ1 only), L(^{13}CH$_3$,^{12}CD$_3$), V(^{13}CH$_3$,^{12}CD$_3$)} U-[^{15}N, ^{13}C, ^{2}H]
CH$_3$-CH$_3$ NOESY	{I(δ1 only), L(^{13}CH$_3$,^{12}CD$_3$), V(^{13}CH$_3$,^{12}CD$_3$)} U-[^{15}N, ^{13}C, ^{2}H]

The spectra collected are described in Appendix A.

7.12.3. Software and data analysis

3D spectra were collected using non-uniform sampling (NUS) for all the backbone and side-chain assignment experiments. The advantage of NUS is that the evolution period (t1) changes in a semi-random way requiring significantly fewer t1 increments to achieve the same level of resolution as

in a normal 2D experiment. Backbone and sidechain assignments were automated using FLYA [133]. Analysis and assignment were aided by several scripts written by Eiso AB (ZoBio B.V.).

7.13. Molecular Docking

HADDOCK (High Ambiguity Driven DOCKing) Web server [139, 140] used 1000 protein structures for initial rigid body docking solutions. The best 200 solutions based on their intermolecular energy were used for a semi-flexible refinement. After initial docking iterations, a final explicit solvent refinement was performed on all 200 models which then clustered based on root mean square deviation criteria. Analysis of HADDOCK scores which is a weighted sum of van der Waals, electrostatic, desolvation and restraint violation energies, was performed to select the best docking solutions. HADDOCK calculations were driven by the available chemical shift perturbations (CSPs) recorded after binding of compound **27** to *Tr*Bgl2. Three-dimensional (3D) conformations of the ligand (**27**) were generated using the SMILES as input.

Appendix A: NMR experiments used for the assignment

Here, it is provided clear descriptions for the variety of 3D NMR experiments used for the protein assignment during this thesis.

Directly detected — Magnetised, not detected

Indirectly detected — Not detected
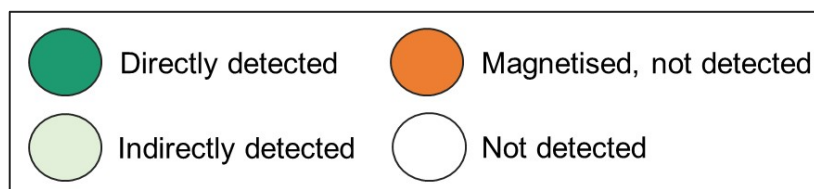
HNCO [160-162]

HNCO is the simplest and most sensitive backbone experiment. It correlates the H^N-N pair of one residue (*i*) with the carbonyl C' resonance of the preceding residue (*i-1*). The magnetisation is transferred from H^N to N via $^1J_{HN}$ coupling using an INEPT pulse sequence and then from N to C' via $^1J_{NC'}$ coupling. This results in a 3D spectrum containing a single peak for each C'-N connection correlating the resonances of each of the three nuclei. The experiment was recorded by Dr Alex Heyam (University of York) using the pulse sequence named trhncogp3d [163].
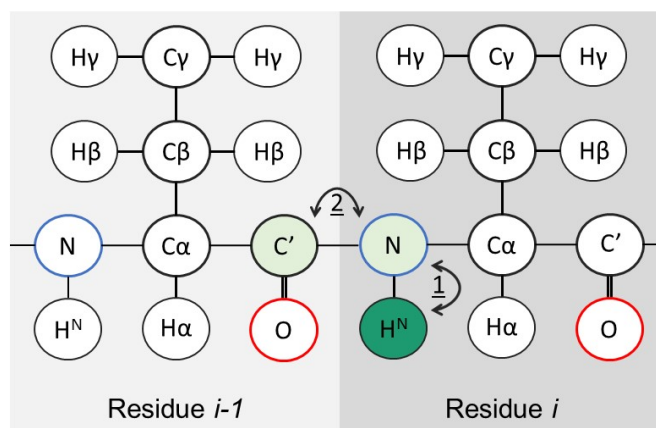
Figure A.1. *Schematic diagram of 3D HNCO experiment. <u>Magnetisation: $H^N \to N$ $\to C'(t1) \to N(t2) \to H^N(t3)$.</u>*

HN(CA)CO [164]

HN(CA)CO is useful in combination with the HNCO. Similar to HNCO J-couplings is used to correlate the H^N-N pair and the C' resonance via the Cα without evolving the chemical shift. This results in a 3D spectrum containing a strong peak for the C' of the residue *i*, and a weak one for the C' of the residue *i-1*. The experiment was recorded by Dr Alex Heyam (University of York) using the pulse sequence named trhncacogp2h3d [165].
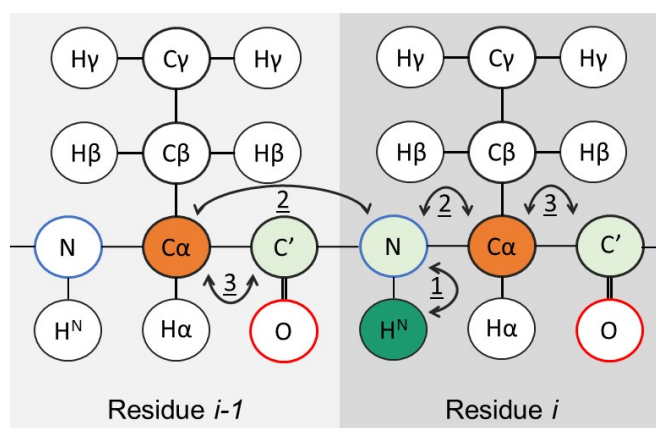

Figure A.2. *Schematic diagram of 3D HN(CA)CO experiment. <u>Magnetisation: H $\to N \to CA \to C'(t1) \to CA \to N(t2) \to H(t3).$</u>*

HNCA [160]

HNCA correlates the H^N-N pair with the Cα resonance of the same and previous residue. The magnetisation is transferred from H^N to N and then via N- Cα- J- coupling to the Cα and then back again to ^{15}N and ^1H. This results in a 3D spectrum containing the Cα peak of the residue *i* and may the Cα peak of the previous residue *i-1*, but much weaker. The experiment was recorded by Dr Alex Heyam using the pulse sequence named trhncagp3d [163].

Figure A.3. *Schematic diagram of 3D HNCA experiment.* <u>*Magnetisation: H →N →CA(t1) →N(t2) →H(t3).*</u>

HN(CO)CA [160]

HN(CO)CA correlates the H^N- N pair of the residue *i* with the Cα resonance of the residue *i-1*. The magnetisation is transferred from H^N to N of the residue *i* and then to C' of the residue *i-1*. From C' of the residue *i-1* the magnetisation is passed to Cα of the same residue and then back again via C' to ^{15}N and ^1H of the residue *i* for detection. This results in a 3D spectrum

containing the Cα peak of the residue *i-1*. The experiment was recorded by Dr Alex Heyam using the pulse sequence named trhncocagp2h3d [166].
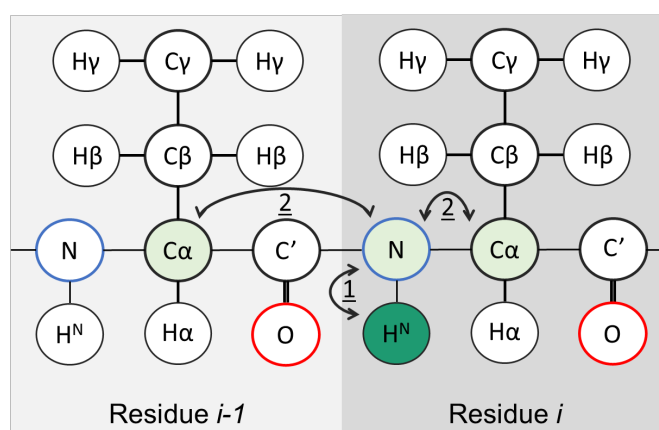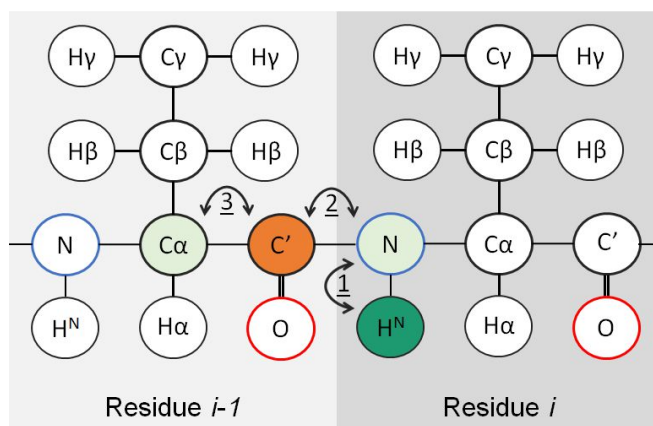

Figure A.4. *Schematic diagram of 3D HNCA experiment. Magnetisation: H →N →C' →CA(t1) →C' →N (t2) → H (t3)*

HN(CO)CACB [160]

HN(CO)CACB contains useful information for making sequential matches and inferring residue type correlating Cα, Cβ, N and HN resonances via J-coupling. Magnetisation is transferred from Hα and Hβ to Cα and Cβ and then it is passed via the C', without detection, and on to the N and then HN for detection. This results in a 3D spectrum containing two peaks for each HN group for the Cα and Cβ of the *i-1* residue. The experiment was recorded by Dr Alex Heyam using the pulse sequence named trhncocacbgp2h3d [166].
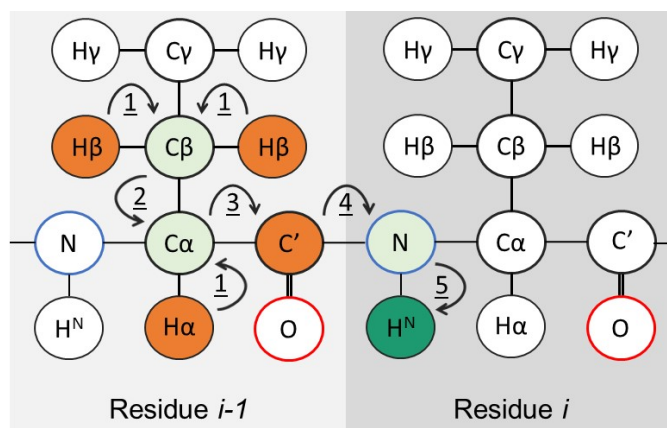
Figure A.5. *Schematic diagram of 3D HN(CO)CACB experiment. Magnetisation:* *HA+HB →CA+CB(t1) →C' →N(t2) →H(t3).*

HNCACB [167]

HNCACB correlates Cα, Cβ, N and H resonances via J-coupling. Magnetisation is transferred from Hα and Hβ to Cα and Cβ and from there it is passed directly to the N and then H^N for detection. This results in a 3D spectrum containing the Cα and Cβ shifts of the *i* residue and may the Cα and Cβ shifts of the previous *i-1* residue, but much weaker. The experiment was recorded by Dr Alex Heyam using the pulse sequence named trhncacbgp2h3d [166].
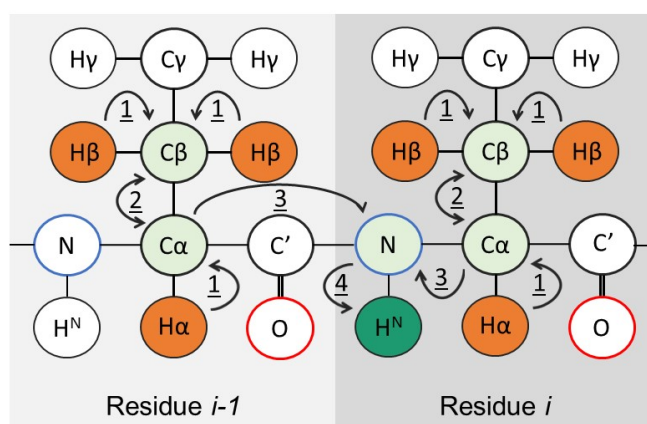
Figure A.6. *Schematic diagram of 3D HNCACB experiment. Magnetisation:* *HA+HB →CA+CB(t1) →N(t2) →H(t3).*

HN(CA)CB [167]

HN(CA)CB correlates the H^N-N pair with the Cβ resonance of the same residue. The magnetisation is transferred from H^N to N of the residue *i* and then via N- Cα- J-coupling to the Cα of the same residue. From Cα of the residue *i* the magnetisation is passed to Cβ of the same residue *i* and then back again via Cα to ^{15}N and ^1H for detection. This results in a 3D spectrum containing the Cβ peak of the residue *i* and may the Cβ peak of the previous residue *i-1*, but much weaker. This experiment was recorded by Dr Marta Carneiro at ZoBio (The Netherlands).
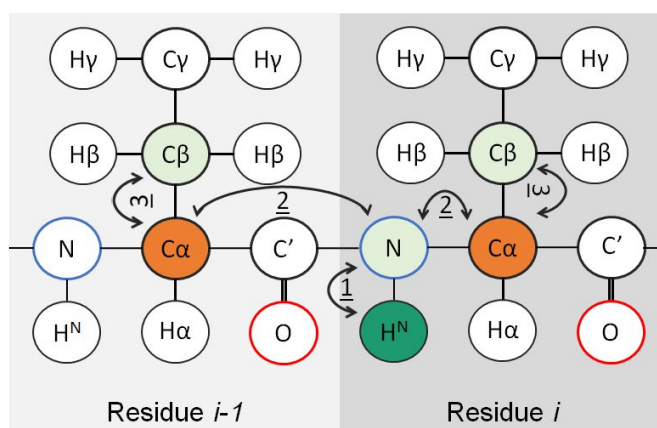
Figure A.7. *Schematic diagram of 3D HN(CA)CB experiment. Magnetisation: H →N →CA →CB(t1) →CA →N(t2) →H(t3).*

^{15}N-resolved NOESY [168, 169]

3D ^1H ^{15}N NOESY correlates all ^1H spins within NOE distance of the amide proton. ^1H nuclei mixes spins within approximately 5 Å distance allow NOE transfer between them. J-coupling between H^N and N is preceded before detection at the H^N proton. The experiment was recorded by Dr Alex Heyam

(University of York) using the pulse sequence named noesyhsqcgpsm3d.2 [170].
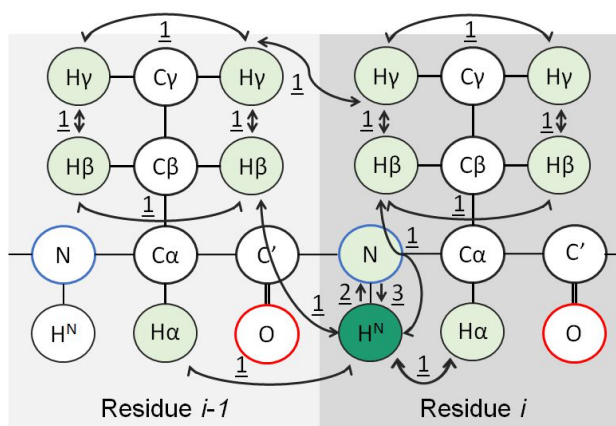
Figure A.8. *Schematic diagram of 3D ^{1}H- ^{15}N NOESY experiment.* *Magnetisation:* *^{1}H-NOE (t1) →N (t2) →H^N (t3).*

^{13}C-resolved NOESY [168, 169]

3D ^{1}H ^{13}C NOESY correlates all ^{1}H spins within NOE distance of the H^C proton. ^{1}H nuclei mixes spins within approximately 5 Å distance allow NOE transfer between them. J-coupling between H and C is preceded before detection at the H^C proton. This experiment was recorded by Dr Marta Carneiro at ZoBio (The Netherlands).
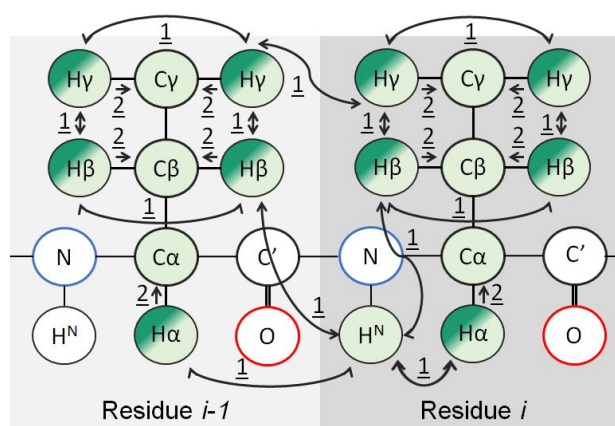
Figure A.9. *Schematic diagram of 3D ^{1}H- ^{13}C NOESY experiment.* *Magnetisation:* *^{1}H-NOE (t1) →C (t2) →H^C (t3).*

HMCM(CG)CBCA [126, 135]

HMCM(CG)CBCA methyl detected experiment relays magnetisation in an out-and-black manner from methyl carbon and proton spins of Ile (I), Leu (L) and Val (V) residues to all aliphatic carbons (HMCM(CG)CBCA) and from methyl groups to carbonyls and back (Val-HMCM(CBCA)CO and Ile, Leu-HMCM(CGCBCA)CO) in which the shifts of each carbon are collected in separate experiments. This experiment was recorded by Dr Marta Carneiro at ZoBio (The Netherlands).
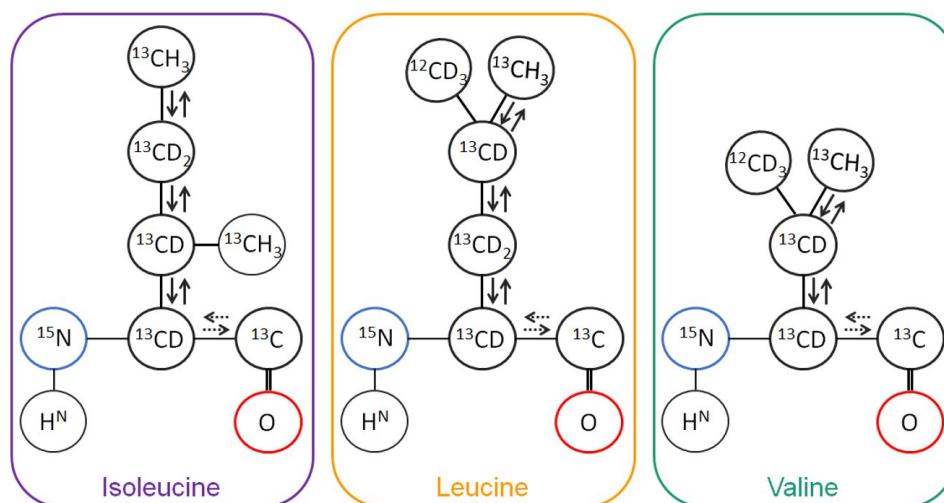
Figure A.10. *Schematic diagram of the magnetization along the carbon skeletons of I, L, V residues. In dashed lines is shown the final transfer from $^{13}C_a$ to ^{13}CO.*

CH$_3$-CH$_3$ NOESY [135, 171]

Methyl-methyl NOESY experiment correlates all methyl proton spins within NOE distance. ^{1}H nuclei mixes spins within approximately 5 Å distance allow NOE transfer between them. It is important to note that there is not intra-residue NOE cross-peaks because only one of two methyl groups of each V

and L residue is protonated. This experiment was recorded by Dr Marta Carneiro at ZoBio (The Netherlands).
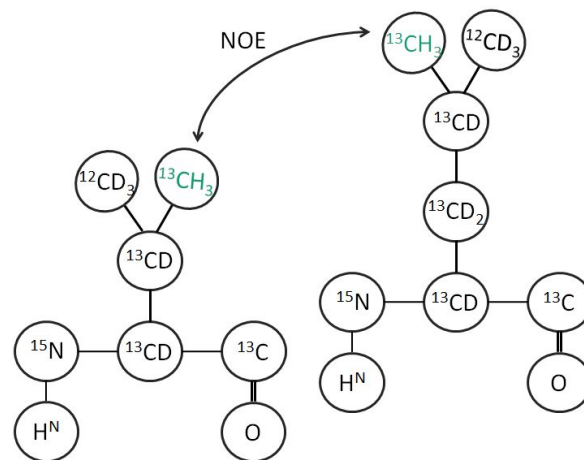
Figure A.11. *Magnetization transfer for the methyl- methyl NOESY experiment.*

Appendix B: Structural overlay of bound and unbound
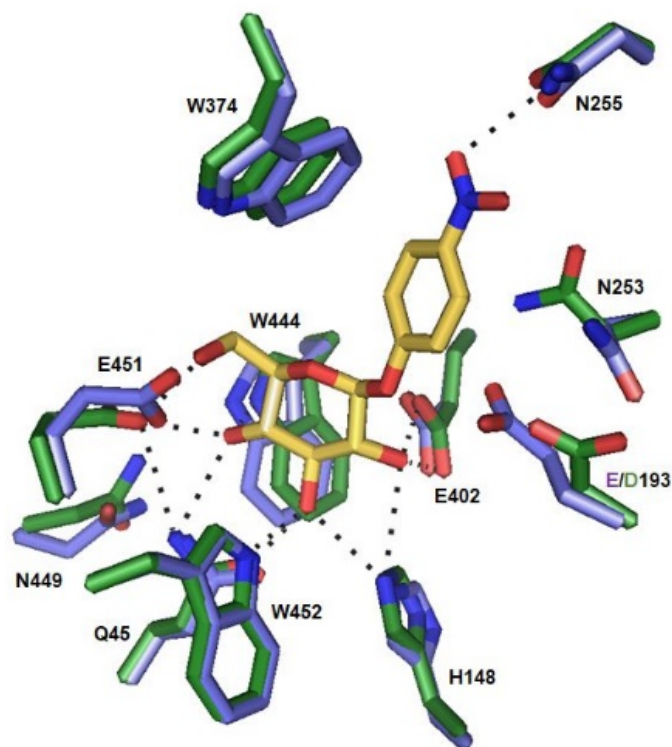
NkBgl *active site*

Figure B.1. Structural overlay of bound and unbound NkBgl active site is shown as a stick diagram. NkBgl-E193D bound to pNPG (PDB ID 3AI0) is shown in green and the apo NkBgl (PDB ID 3AHZ) is shown in purple. pNPG is shown as yellow stick diagram. Oxygen atoms are shown in red and nitrogen in blue. Hydrogen bonds are shown as black dashed lines.

Appendix C: CSPs of compound 30 mapped on TrBgl2
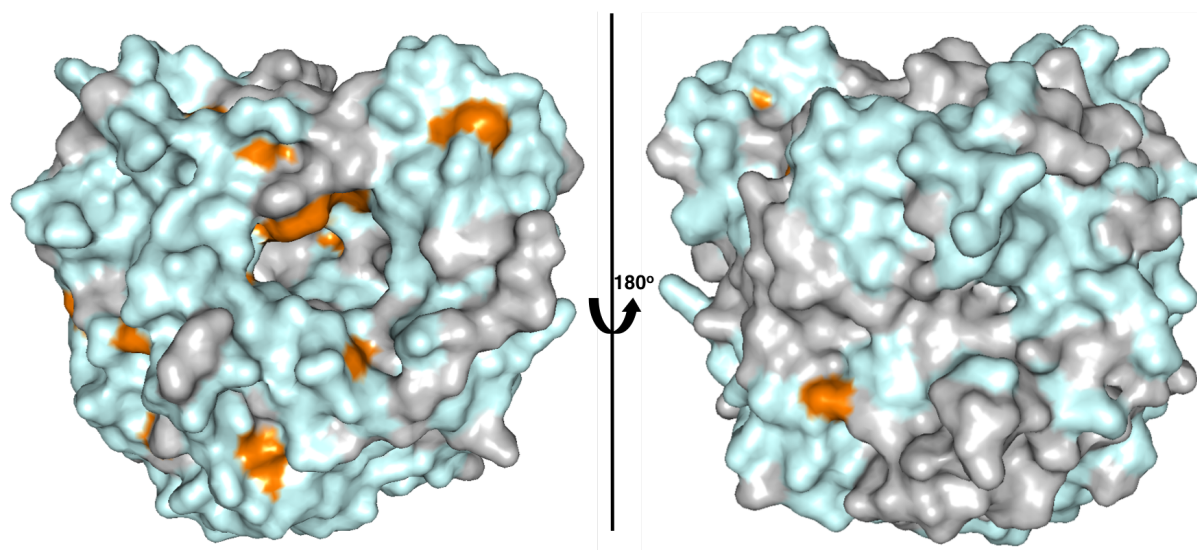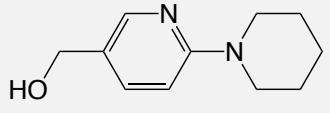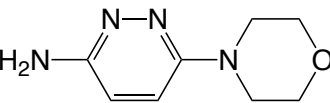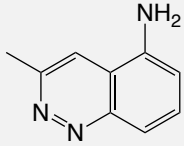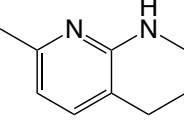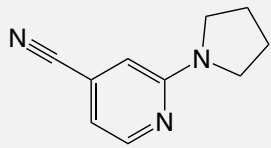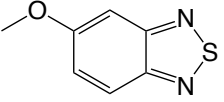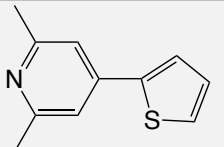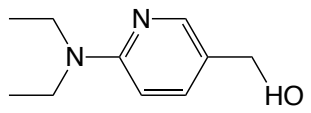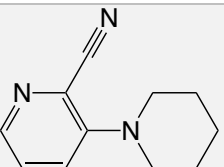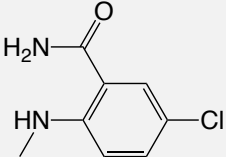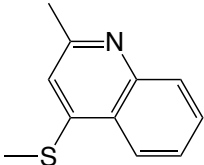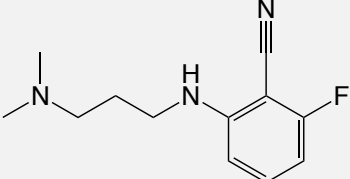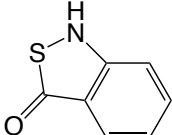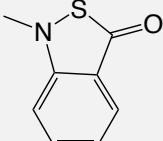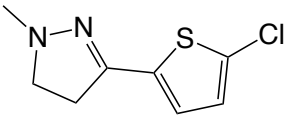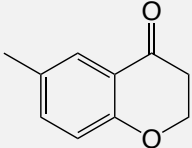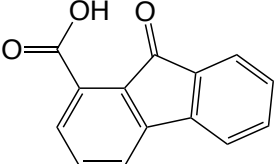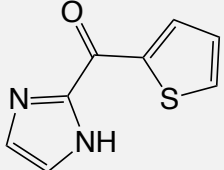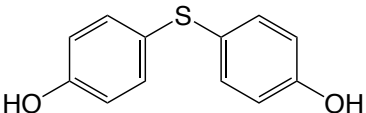
Figure C.1. *Amide chemical shift perturbations caused by the presence of **30** mapped to the protein surface of the crystal structure of TrBgl2. Backbone CSPs are coloured orange. Unassigned residues are coloured grey. The figure was created in PyMOL.*
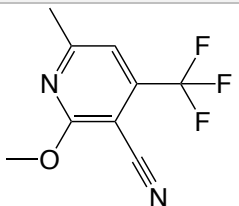
Appendix D: Chemical structure of compound 31
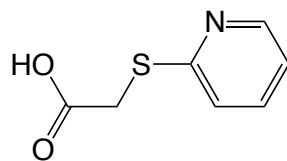
Appendix E: List of fragments caused more than 50 % inhibition after library screening for TrBgl2

Chemical structure	% inhibition [1]
	50
	58
	72
	50
	67
	58
	53
	50
	52

Chemical structure	% inhibition [1]
	87
	78
	65
	80
	75
	52
	57
	50
	80
	70

Chemical structure	% inhibition [1]
	91

[1] Inhibition obtained at 1 mM fragment concentration
(substrate conc. = 500 µM)

Appendix F: Trbgl2 gene sequence and transcript in pET-YSBL3C vector

F.1. Trbgl2 gene sequence in pET-YSBL3C vector

ATGCTGCCGAAAGATTTTCAGTGGGGTTTTGCAACC

GCAGCATATCAGATTGAAGGTGCAGTTGATCAGGAT

GGTCGTGGTCCGAGCATTTGGGATACCTTTTGTGCA

CAGCCTGGTAAAATTGCAGATGGTAGCAGCGGTGTT

ACCGCATGTGATAGCTATAATCGTACCGCAGAAGAT

ATT GCACTGCTGAAAAGCCTGGGTGCAAAAGCTAT

CGTTTTAGCATTAGCTGGTCACGCATTATTCCGGAA

GGTGGTCGCGGTGATGCCGTTAATCAGGCAGGTATT

GATCACTATGTGAAATTCGTTGATGATCTGCTGGAT

GCAGGTATTACCCCGTTTATTACCCTGTTTCATTGGG

ATCTGCCGGAAGGTCTGCATCAGCGTTATGGTGGTC

TGCTGAATCGCACCGAATTTCCGCTGGATTTTGAAA

ATTATGCCCGTGTTATGTTTCGTGCCCTGCCTAAAGT

TCGTAATTGG ATTACCTTTAATGAACCGCTGTGTAGC

GCAATTCCTGGTTATGGTAGCGGCACCTTTGCACCG

GGTCGTCAGAGCACCAGCGAACCGTGGACCGTTGGT

CATAATATTCTG GTTGCACATGGTCGTGCAGTTAAA

GCATATCGTGATGATTTCAAACCGGCAAGCGGTGAT

GGTCAGATTGGTATTGTTCTGAATGGCGATTTTACCT

ATCCGTGGGATGCAGCAGATCCTGCAGATAAAGAAG

CAGCAGAACGTCGTCTGGAATTTTTCACCGCATGGTT
TGCAGATCCGATTTATCTGGGTGATTATCCGGCAAG
CATGCGTAAACAGCTGGGTGATCGTCTGCCGACCTT
TACACCGGAAGAACGTGCACTGGTGCATGGTAGCAA
TGATTTTTATGGCATGAACCACTATACCAGCAACTAT
ATTCGTCATCGTAGCAGTCCGGCAAGTGCAGATGAT
ACCGTTGGTAATGTTGATGTGCTGTTTACCAACAAA
CAGGGCAATTGTATTGGTCCGGAAACACAGAGCCCG
TGGCTGCGTCCGTGTGCAGCAGGTTT
TCGTGATTTTCTGGTTTGGATTAGCAAACGCTATGGT
TATCCGCCTATTTATGTTACCGAAAATGGCACCAGTA
TTAAAGGTGAAAGCGATCTGCCCAAAGAAAAGATTCT
GGAAGATGATTTTCGCGTGAAGTACTACAATGAATAT
ATCCGTGCAATGGTTACCGCAGTGGAACTGGATGGT
GTTAATGTGAAAGGTTATTTTGCATGGTCCCTGATG
GATAATTTTGAATGGGCAGATGGTTATGTGACCCGT
TTTGGTGTTACCTATGTGGATTATGAAAACGGCCAG
AAACGCTTTCCGAAAAAAGCGCAAAAAGCCTGAAA
CCGCTGTTTGATGAACTGATTGCAGCAGCATAA

F.2. TrBgl2 transcript in pET-YSBL3C vector

MGSSHHHHHHSSGLEVLFQGPA M L P K D F Q W G F A T A A Y Q I E G A V

D Q D G R G P S I W D T F C A Q P G K I A D G S S G V T A C D S Y N R T

A E D I A L L K S L G A K S Y R F S I S W S R I I P E G G R G D A V N Q A

G I D H Y V K F V D D L L D A G I T P F I T L F H W D L P E G L H Q R Y G

G L L N R T E F P L D F E N Y A R V M F R A L P K V R N W I T F N E P L C

S A I P G Y G S G T F A P G R Q S T S E P W T V G H N I L V A H G R A V K

A Y R D D F K P A S G D G Q I G I V L N G D F T Y P W D A A D P A D K E

A A E R R L E F F T A W F A D P I Y L G D Y P A S M R K Q L G D R L P T F

T P E E R A L V H G S N D F Y G M N H Y T S N Y I R H R S S P A S A D D T

V G N V D V L F T N K Q G N C I G P E T Q S P W L R P C A A G F R D F L V

W I S K R Y G Y P P I Y V T E N G T S I K G E S D L P K E K I L E D D F R V

K Y Y N E Y I R A M V T A V E L D G V N V K G Y F A W S L M D N F E W A

D G Y V T R F G V T Y V D Y E N G Q K R F P K K S A K S L K P L F D E L I

A A A

Appendix G: E367Q Trbgl2 gene sequence and transcript in pET-YSBL3C vector

G.1. E367Q Trbgl2 gene sequence in pET-YSBL3C vector

ATGCTGCCGAAAGATTTTCAGTGGGGGTTTTGCAACC

GCAGCATATCAGATTGAAGGTGCAGTTGATCAGGAT

GGTCGTGGTCCGAGCATTTGGGATACCTTTTGTGCA

CAGCCTGGTAAAATTGCAGATGGTAGCAGCGGTGTT

ACCGCATGTGATAGCTATAATCGTACCGCAGAAGAT

ATT GCACTGCTGAAAAGCCTGGGTGCAAAAAGCTAT

CGTTTTAGCATTAGCTGGTCACGCATTATTCCGGAA

GGTGGTCGCGGTGATGCCGTTAATCAGGCAGGTATT

GATCACTATGTGAAATTCGTTGATGATCTGCTGGAT

GCAGGTATTACCCCGTTTATTACCCTGTTTCATTGGG

ATCTGCCGGAAGGTCTGCATCAGCGTTATGGTGGTC

TGCTGAATCGCACCGAATTTCCGCTGGATTTTGAAA

ATTATGCCCGTGTTATGTTTCGTGCCCTGCCTAAAGT

TCGTAATTGG ATTACCTTTAATGAACCGCTGTGTAGC

GCAATTCCTGGTTATGGTAGCGGCACCTTTGCACCG

GGTCGTCAGAGCACCAGCGAACCGTGGACCGTTGGT

CATAATATTCTG GTTGCACATGGTCGTGCAGTTAAA

GCATATCGTGATGATTTCAAACCGGCAAGCGGTGAT

GGTCAGATTGGTATTGTTCTGAATGGCGATTTTACCT

216

ATCCGTGGGATGCAGCAGATCCTGCAGATAAAGAAG

CAGCAGAACGTCGTCTGGAATTTTTCACCGCATGGTT

TGCAGATCCGATTTATCTGGGTGATTATCCGGCAAG

CATGCGTAAACAGCTGGGTGATCGTCTGCCGACCTT

TACACCGGAAGAACGTGCACTGGTGCATGGTAGCAA

TGATTTTTATGGCATGAACCACTATACCAGCAACTAT

ATTCGTCATCGTAGCAGTCCGGCAAGTGCAGATGAT

ACCGTTGGTAATGTTGATGTGCTGTTTACCAACAAA

CAGGGCAATTGTATTGGTCCGGAAACACAGAGCCCG

TGGCTGCGTCCGTGTGCAGCAGGTTTTCGTGATTTT

CTGGTTTGGATTAGCAAACGCTATGGTTATCCGCCT

ATTTATGTTACC**CAA**AATGGCACCAGTATTAAAGGT

GAAAGCGATCTGCCCAAGAAAGATTCTGGAAGAT

GATTTTCGCGTGAAGTACTACAATGAATATATCCGTG

CAATGGTTACCGCAGTGGAACTGGATGGTGTTAATG

TGAAAGGTTATTTTGCATGGTCCCTGATGGATAATTT

TGAATGGGCAGATGGTTATGTGACCCGTTTTGGTGT

TACCTATGTGGATTATGAAAACGGCCAGAAACGCTT

TCCGAAAAAAGCGCAAAAAGCCTGAAACCGCTGTT

TGATGAACTGATTGCAGCAGCATAA

G.2. E367Q TrBgl2 transcript in pET-YSBL3C vector

MGSSHHHHHHSSGLEVLFQGPA M L P K D F Q W G F A T A A Y Q I E G A V

D Q D G R G P S I W D T F C A Q P G K I A D G S S G V T A C D S Y N R T

A E D I A L L K S L G A K S Y R F S I S W S R I I P E G G R G D A V N Q A

G I D H Y V K F V D D L L D A G I T P F I T L F H W D L P E G L H Q R Y G

G L L N R T E F P L D F E N Y A R V M F R A L P K V R N W I T F N E P L C

S A I P G Y G S G T F A P G R Q S T S E P W T V G H N I L V A H G R 3A V

K A Y R D D F K P A S G D G Q I G I V L N G D F T Y P W D A A D P A D K

E A A E R R L E F F T A W F A D P I Y L G D Y P A S M R K Q L G D R L P T

F T P E E R A L V H G S N D F Y G M N H Y T S N Y I R H R S S P A S A D D

T V G N V D V L F T N K Q G N C I G P E T Q S P W L R P C A A G F R D F L

V W I S K R Y G Y P P I Y V T **Q** N G T S I K G E S D L P K E K I L E D D F

R V K Y Y N E Y I R A M V T A V E L D G V N V K G Y F A W S L M D N F E

W A D G Y V T R F G V T Y V D Y E N G Q K R F P K K S A K S L K P L F D E

L I A A A

Appendix H: Ccbgla gene sequence and transcript in pET-YSBL3C vector

H.1. Ccbgla gene sequence in pET-YSBL3C vector

A T G G A A A A G C T A AG ATT T C C C A A A G A T T T T A T T TT T G G

A A C A G C C A CT G C A G CA T A T C A A A T T G A A G G A G C T T A C

A A A G A A G A T G A G A A G G T G A A T C T A T T T G G G A T A G G

T T T A G T C A T A T A C C A G G A A A T G T A G C T A A A A T G C A T A

A T G G T G A T A T T G C T T G T G A T C A C T A T C A T A G A T A T A A

A G A A G A T G T T C A G C T A T T A A A A A G C C T T G G A A T T A A

A A G T T A T A G G T T T T C A A T T G C T T G G C C T A G A A T T T T C

C C A A A A G G T T T T G G C G A G A T A A A C C A G A A G G G A A T T

C A G T T CT A T A G G G A T T T A A T T G A T G A A C T A A T T A A A A

A T G A T A T A G A A C C A G C T A T A A C A A T T T A TC A T T G G G A

T C T T C C A C A A A A G C T T C A G G A T A T T G G A G G G T G G G C

A A A T C C G C A A G T T G C TG A T T A C T A T G T T G A T T A T G C A

A A C T T A T T A T T C A G A G A G T T C G G A G A T A G A G T A A A A

A C A T G G A T A A C T C A T A A T G A G C C A T G G G T T G C A T C A

T A T C T T G G C T A T G C T T T A G G A G T T C A T G C T CC A G G A A

TT A A A G A T A T G A A A A T G G C A T T G T T A G C T G C A C A T A A

C A T A T T A T T A T C G C A C T T T A A G G C A G T T A A A G C T T A T

A G A G A A T T A G A A C A A G A T G G G C A A A T A G G T A T A A C A

T T A A A T C T T T C A A C C TG T T A T T C A A A T T C A G C T G A T G

A A G A A G A T A T T G C T G C A G C C C A T A G A A G T G A T G G A T

GGAACAACAGATGGTTTTTAGATGCTGCATTAAAAGG

AACTTATCCTGAGGATATGATAAAAATCTTTAGCGAT

ACAAATATTATGCCTGAACTACCTAAAGAGTTATTTA

CTGAGGTATTTGAAACTTCTGATTTTTTAGGAATAAA

TTATTATACGACAAGTTGTAAAGAATAACTCTGAA

GCTTTTATCGGTGCTGAAAGTGTAGCAATGGATAAT

CCTAAAACAGAAATGGGTTGGGAGATATATCCGCAA

GGGCTTTATGATTTGCTAACGAGGATACACAGGGATT

ATGGGAACATAGATTTATACATAACAGAAACGGTG

CAGCTTTTAATGATATGGTTAATAGAGACGGTAAAG

TTGAAGATGAAAATAGATTAGATTATTTATACACTCA

TTTTGCTGCTGCATTAAGTGCTATAGAAGCGGGAGT

ACCTTTAAAGGGATATTATATTTGGTCTTTCATGGAT

AATTTTGAGTGGGCTGAAGGATATGAAAAAGATTTG

GAATAGTACATGTAAACTATAAAACTCAGGAGAGAA

CAATAAAGAAGAGTGCTTATTGGTATAAGGAGCTTAT

AGAAAGATCTAATAAGTAA

H.2. CcBglA transcript in pET-YSBL3C vector

MGSSHHHHHHSSGLEVLFQGPA M E K L R F P K D F I F G T A T A A Y Q I E
G A Y K E D E K G E S I W D R F S H I P G N V A K M H N G D I A C D H Y
H R Y K E D V Q L L K S L G I K S Y R F S I A W P R I F P K G F G E I N Q K
G I Q F Y R D L I D E L I K N D I E P A I T I Y H W D L P Q K L Q D I G G W
A N P Q V A D Y Y V D Y A N L L F R E F G D R V K T W I T H N E P W V A
S Y L G Y A L G V H A P G I K D M K M A L L A A H N I L L S H F K A V K A
Y R E L E Q D G Q I G I T L N L S T C Y S N S A D E E D I A A A H R S D G
W N N R W F L D A A L K G T Y P E D M I K I F S D T N I M P E L P K E L F
T E V F E T S D F L G I N Y Y T R Q V V K N N S E A F I G A E S V A M D N
P K T E M G W E I Y P Q G L Y D L L T R I H R D Y G N I D L Y I T E N G A
A F N D M V N R D G K V E D E N R L D Y L Y T H F A A A L S A I E A G V P
L K G Y Y I W S F M D N F E W A E G Y E K R F G I V H V N Y K T Q E R T I
K K S A Y W Y K E L I E R S N K

Appendix I: α-amylase gene sequence and transcript in pET-YSBL3C vector

I.1. α-amylase gene sequence in pET-YSBL3C vector

ATGCCAGGGACCAGCAGCAGCACCGTTTAATGGCAC

CATGATGCAGTATTTTGAATGGTATCTGCCGGATGA

TGGCACCCTGTGGACCAAAGTTGCAAATGAAGCAAA

TAATCTGAGCAGCCTGGGTATTACCGCACTGTGGCTG

CCTCCGGCATATAAAGGCACCAGCCGTAGTGATGTT

GGTTATGGTGTTTATGATCTGTATGACCTGGGCGAAT

TTAATCAGAAAGGCACCGTTCGTACCAAATATGGCAC

CAAAGCACAGTATCTGCAGGCAATTCAGGCAGCACA

TGCAGCAGGTATGCAGGTTTATGCAGATGTTGTGTT

TGATCATAAAGGTGGTGCAGATGGCACCGAATGGGT

TGATGCAGTTGAAGTTAATCCGAGCGATCGCAATCAA

GAAATTAGCGGCACCTATCAGATTCAGGCATGGACC

AAATTTGATTTTCCGGGTCGTGGTAATACCTACAGC

AGCTTTAAATGGCGTTGGTATCATTTTGATGGTGTGG

ATTGGGATGAAAGCCGTAAACTGAGCCGTATTTACAA

ATTTCGTGGTAAAGCATGGGATTGGGAAGTGGATACC

GAATTTGGCAATTATGATTATCTGATGTATGCCGATCT

GGATATGGATCATCCGGAAGTTGTTACCGAACTGAA

AAATTGGGGTAAATGGTATGTGAACACCACCAACAT

TGATGGTTTTCGTCTGGATGCAGTGAAACACATCAA

ATTTAGCTTTTTTCCGGATTGGCTGAGCTATGTTCGT
AGCCAGACCGGTAAACCGCTGTTTACAGTTGGTGAA
TATTGGAGCTATGATATCAACAAACTGCACAACTACA
TCACCAAAACCGATGGTACAATGAGCCTGTTTGATGC
ACCGCTGCATAACAAATTCTATACCGCAAGCAAAAGC
GGTGGTGCATTTGATATGCGTACCCTGATGACCAAT
ACACTGATGAAAGATCAGCCGACCCTGGCAGTTACC
TTTGTGGATAATCATGATACCGAACCGGGTCAGGCA
CTGCAGAGCTGGGTTGATCCGTGGTTCAAACCGCTG
GCATATGCATTTATTCTGACCCGTCAAGAAGGTTATC
CTTGCGTTTTTTATGGCGATTATTATGGCATTCCGCA
GTATAACATTCCGAGCCTGAAAAGCAAAATTGATCC
GCTGCTGATTGCACGTCGTGATT ATGCCTATGGCAC
CCAGCACGATTATCTGGATCATAGCGATATTATTGG
TTGGACCCG TGAAGGTGGTACAGAAAACCGGGTA
GCGGTCTGGCAGCACTGATTACCGATGGTCCGGGTG
GTAGCAAATGGATGTATGTTGGTAA ACAGCATGCCG
GTAAAGTGTTTATGACCTGACCGGTAATCGTAGCG
ATACCGTTACCATTAATAGTGATGGTTGGGGTGAATT C
AAAGTGAATGGTGGTAGCGTTAGCGTTTGGGTTCCG
CGTAAAACCACCGTTAGCACCATTGCCCGTCCGATT
ACCACCCGTCCGTGGACCGGTGAATTTGTTCGTTGG
ACCGAACCGCGTCTGGTTGCATGGCCGTAATAACGC
GCCTTCTCCTCTAA

I.2. α-amylase transcript in pET-YSBL3C vector

MGSSHHHHHHSSGLEVLFQGPA M A A P F N G T M M Q Y F E W Y L P D D G

T L W T K V A N E A N N L S S L G IT A L W L P P A Y K G T S R S D V G Y

G V Y D L Y D L G E F N Q K G T V R T K Y G T K A Q Y L Q A I Q A A H A

A G M Q V Y A D V V F D H K G G A D G T E W V D A V E V N P S D R N Q

E I S G T Y Q I Q A W T K F D F P G R G N T Y S S F K W R W Y H F D G V

D W D E S R K L S R I Y K F R G K A W D W E V D T E F G N Y D Y L M Y A

D L D   M D H P E V V T E L K N W G K W Y V N T T N I D G F R L D A V K

H I K F S F F P D W L S Y V R S Q T G K P L F T V G E Y W S Y D I N K L H

N Y I T K T D G T M S L F D A P L H N K F Y T A S K S G G A F D M R T L M

T N T L M K D Q P T L A V T F V D N H D T E P G Q A L Q S W V D P W F K

P L A Y A F I L T R Q E G Y P C V F Y G D Y Y G I P Q Y N I P S L K S K I D P

L L I A R R D Y A Y G T Q H D Y L D H S D I I G W T R E G G T E K P G S G

L A A L I T D G P G G S K W M Y V G K Q H A G K V F Y D L T G N R S D T

VT I N S D G W G E F K V N G G S V S V W V P R K T T V S T I A R P I T T

R P W T G E F V R W T E P R L V A W P

References

[1] Koshland DE. Stereochemistry and the mechanism of enzymatic reactions. Biol Rev. 1953;28:416-36.

[2] Jeng WY, Wang NC, Lin MH, Lin CT, Liaw YC, Chang WJ. Structural and functional analysis of three beta-glucosidases from bacterium Clostridium cellulovorans, fungus Trichoderma reesei and termite Neotermes koshunensis. J Struct Biol. 2011;173:46-56.

[3] Bollag G, Hirth P, Tsai J, Zhang J, Ibrahim PN, Cho H. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. Nature. 2010;467:596-9.

[4] Souers AJ, Leverson JD, Boghaert ER, Ackler SL, Catron ND, Chen J. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nature Medicine. 2013;19:202-8.

[5] Roberts AW, Davids MS, Pagel JM, Kahl BS, Puvvada SD, Gerecitano JF. Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. N Engl J Med. 2016;374:311-22.

[6] Roberts AW, Huang D. Targeting BCL2 With BH3 Mimetics: Basic Science and Clinical Application of Venetoclax in Chronic Lymphocytic Leukemia and Related B Cell Malignancies. Clin Pharmacol Ther. 2017;101:89-98.

[7] Wylie AA, Schoepfer J, Jahnke W, Cowan-Jacob SW, Loo A, Furet P. The allosteric inhibitor ABL001 enables dual targeting of BCR–ABL1. Nature. 2017;543:733-7.

[8] Cowan-Jacob SW, Jahnke W, Knapp S. Novel approaches for targeting kinases: allosteric inhibition, allosteric activation and pseudokinases. Future Med Chem. 2014;6:541-61.

[9] Jia M, Dahlman-Wright K, Gustafsson JA. Estrogen receptor alpha and beta in health and disease. Best Pract Res Clin Endocrinol Metab. 2015;29:557-68.

[10] Cooper B, Handin RI, Young LH, Alexander RW. Agonist regulation of the human platelet alpha-adrenergic receptor. Nature. 1978;274:703-6.

[11] Kuwahara H, Nishizaki M, Kanazawa H. Nuclear localization signal and phosphorylation of Serine350 specify intracellular localization of DRAK2. J Biochem. 2008;143:349-58.

[12] Gresset A, Hicks SN, Harden TK, Sondek J. Mechanism of Phosphorylation-induced Activation of Phospholipase C-γ Isozymes. 2010.

[13] Canagarajah B. Activation Mechanism of the MAP Kinase ERK2 by Dual Phosphorylation. Cell. 1997;90:859-69.

[14] Cardenas ML, Cornish-Bowden A. Characteristics necessary for an interconvertible enzyme cascade to generate a highly sensitive response to an effector. Biochem J. 1989;257:339-45.

[15] Zorn JA, Wells JA. Turning enzymes ON with small molecules. Nat Chem Biol. 2010;6:179-88.

[16] Castellano S, Spannhoff A, Milite C, Dal Piaz F, Cheng D, Tosco A. Identification of small-molecule enhancers of arginine methylation catalyzed by coactivator-associated arginine methyltransferase 1. J Med Chem. 2012;55:9875-90.

[17] Kukday SS, Manandhar SP, Ludley MC, Burriss ME, Alper BJ, Schmidt WK. Cell-Permeable, Small-Molecule Activators of the Insulin-Degrading Enzyme. J Biomol Screen. 2012;17:1348-61.

[18] Kashani-Amin E, Larijani B, Ebrahim-Habibi A. Neohesperidin dihydrochalcone: presentation of a small molecule activator of mammalian alpha-amylase as an allosteric effector. FEBS Lett. 2013;587:652-8.

[19] Smith HQ, Smith TJ. Identification of a Novel Activator of Mammalian Glutamate Dehydrogenase. Biochemistry. 2016;55:6568-76.

[20] Meng H, McClendon CL, Dai Z, Li K, Zhang X, He S. Discovery of Novel 15-Lipoxygenase Activators To Shift the Human Arachidonic Acid Metabolic Network toward Inflammation Resolution. J Med Chem. 2016;59:4202-9.

[21] Mike LA, Dutter BF, Stauff DL, Moore JL, Vitko NP, Aranmolate O. Activation of heme biosynthesis by a small molecule that is toxic to fermenting Staphylococcus aureus. Proc Natl Acad Sci U S A. 2013;110:8206-11.

[22] Surdel MC, Horvath DJ, Jr., Lojek LJ, Fullen AR, Simpson J, Dutter BF. Antibacterial photosensitization through activation of coproporphyrinogen oxidase. Proc Natl Acad Sci U S A. 2017;114:E6652-e9.

[23] Geldenhuys WJ, Bergeron SA, Mullins JE, Aljammal R, Gaasch BL, Chen WC. High-content screen using zebrafish (Danio rerio) embryos identifies a novel kinase activator and inhibitor. Bioorg Med Chem Lett. 2017;27:2029-37.

[24] Hwang S, Mruk K, Rahighi S, Raub AG, Chen CH, Dorn LE. Correcting glucose-6-phosphate dehydrogenase deficiency with a small-molecule activator. Nat Commun. 2018;9:4045.

[25] Cardenas ML, Cornish-Bowden A, Ureta T. Evolution and regulatory role of the hexokinases. Biochim Biophys Acta. 1998;1401:242-64.

[26] Shiota C, Coffey J, Grimsby J, Grippo JF, Magnuson MA. Nuclear import of hepatic glucokinase depends upon glucokinase regulatory protein, whereas export is due to a nuclear export signal sequence in glucokinase. J Biol Chem. 1999;274:37125-30.

[27] Matschinsky FM, Magnuson MA, Zelent D, Jetton TL, Doliba N, Han Y. The network of glucokinase-expressing cells in glucose homeostasis and the potential of glucokinase activators for diabetes therapy. Diabetes. 2006;55:1-12.

[28] Matschinsky FM. Banting Lecture 1995. A lesson in metabolic regulation inspired by the glucokinase glucose sensor paradigm. Diabetes. 1996;45:223-41.

[29] Grimsby J, Sarabu R, Corbett WL, Haynes NE, Bizzarro FT, Coffey JW. Allosteric activators of glucokinase: potential role in diabetes therapy. Science. 2003;301:370-3.

[30] Kamata K, Mitsuya M, Nishimura T, Eiki J, Nagata Y. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. Structure. 2004;12:429-38.

[31] Hariharan N, Farrelly D, Hagan D, Hillyer D, Arbeeny C, Sabrah T. Expression of human hepatic glucokinase in transgenic mice liver results in

decreased glucose levels and reduced body weight. Diabetes. 1997;46:11-6.

[32] Matschinsky FM. Assessing the potential of glucokinase activators in diabetes therapy. Nat Rev Drug Discov. 2009;8:399-416.

[33] Filipski KJ, Pfefferkorn JA. A patent review of glucokinase activators and disruptors of the glucokinase--glucokinase regulatory protein interaction: 2011-2014. Expert Opin Ther Pat. 2014;24:875-91.

[34] Nakamura A, Terauchi Y. Present status of clinical deployment of glucokinase activators. J Diabetes Investig. 2015;6:124-32.

[35] Brocklehurst KJ, Payne VA, Davies RA, Carroll D, Vertigan HL, Wightman HJ. Stimulation of Hepatocyte Glucose Metabolism by Novel Small Molecule Glucokinase Activators. Diabetes. 2004;3:535-41.

[36] Efanov AM, Barrett DG, Brenner MB, Briggs SL, Delaunois A, Durbin JD. A novel glucokinase activator modulates pancreatic islet and hepatocyte function. Endocrinology. 2005;146:3696-701.

[37] Coope GJ, Atkinson AM, Allott C, McKerrecher D, Johnstone C, Pike KG. Predictive blood glucose lowering efficacy by Glucokinase activators in high fat fed female Zucker rats. Br J Pharmacol. 2006;149:328-35.

[38] Futamura M, Hosaka H, Kadotani A, Shimazaki H, Sasaki K, Ohyama S. An allosteric activator of glucokinase impairs the interaction of glucokinase and glucokinase regulatory protein and regulates glucose metabolism. J Biol Chem. 2006;281:37668-74.

[39] Fyfe MC, White JR, Taylor A, Chatfield R, Wargent E, Printz RL. Glucokinase activator PSN-GK1 displays enhanced antihyperglycaemic and insulinotropic actions. Diabetologia. 2007;50:1277-87.

[40] Ohyama S, Takano H, Iino T, Nishimura T, Zhou YP, Langdon RB. A small-molecule glucokinase activator lowers blood glucose in the sulfonylurea-desensitized rat. Eur J Pharmacol. 2010;640:250-6.

[41] Nakamura A, Shimazaki H, Ohyama S, Eiki J, Terauchi Y. Effect of long-term treatment with a small-molecule glucokinase activator on glucose metabolism, lipid profiles and hepatic function. J Diabetes Investig. 2011;2:276-9.

[42] Futamura M, Yao J, Li X, Bergeron R, Tran JL, Zycband E. Chronic treatment with a glucokinase activator delays the onset of hyperglycaemia and preserves beta cell mass in the Zucker diabetic fatty rat. Diabetologia. 2012;55:1071-80.

[43] Du X, Hinklin RJ, Xiong Y, Dransfield P, Park J, Kohn TJ. C5-Alkyl-2-methylurea-Substituted Pyridines as a New Class of Glucokinase Activators. ACS Med Chem Lett. 2014;5:1284-9.

[44] Sharma R, Litchfield J, Bergman A, Atkinson K, Kazierad D, Gustavson SM. Comparison of the circulating metabolite profile of PF-04991532, a hepatoselective glucokinase activator, across preclinical species and humans: potential implications in metabolites in safety testing assessment. Drug Metab Dispos. 2015;43:190-8.

[45] Lu M, Li P, Bandyopadhyay G, Lagakos W, Dewolf WE, Jr., Alford T. Characterization of a novel glucokinase activator in rat and mouse models. PLoS One. 2014;2:e88431.

[46] Park K, Lee BM, Hyun KH, Han T, Lee DH, Choi HH. Design and Synthesis of Acetylenyl Benzamide Derivatives as Novel Glucokinase Activators for the Treatment of T2DM. ACS Med Chem Lett. 2015;6:296-301.

[47] Dransfield PJ, Pattaropong V, Lai S, Fu Z, Kohn TJ, Du X. Novel Series of Potent Glucokinase Activators Leading to the Discovery of AM-2394. ACS Med Chem Lett. 2016;7:714-8.

[48] Paczal A, Balint B, Weber C, Szabo ZB, Ondi L, Theret I. Structure-Activity Relationship of Azaindole-Based Glucokinase Activators. J Med Chem. 2016;59:687-706.

[49] Bonadonna RC, Heise T, Arbet-Engels C, Kapitza C, Avogaro A, Grimsby J. Piragliatin (RO4389620), a novel glucokinase activator, lowers plasma glucose both in the postabsorptive state and after a glucose challenge in patients with type 2 diabetes mellitus: a mechanistic study. J Clin Endocrinol Metab. 2010;95:5028-36.

[50] Xu H, Sheng L, Chen W, Yuan F, Yang M, Li H. Safety, tolerability, pharmacokinetics, and pharmacodynamics of novel glucokinase activator HMS5552: results from a first-in-human single ascending dose study. Drug Des Devel Ther. 2016;10:1619-26.

[51] Pal M. Recent advances in glucokinase activators for the treatment of type 2 diabetes. Drug Discov Today. 2009;14:784-92.

[52] Bourbonais FJ, Chen J, Huang C, Zhang Y, Pfefferkorn JA, Landro JA. Modulation of glucokinase by glucose, small-molecule activator and glucokinase regulatory protein: steady-state kinetic and cell-based analysis. Biochem J. 2012;441:881-7.

[53] Bowler JM, Hervert KL, Kearley ML, Miller BG. Small-Molecule Allosteric Activation of Human Glucokinase in the Absence of Glucose. ACS Med Chem Lett. 2013;4.

[54] Luo J, Manning BD, Cantley LC. Targeting the PI3K-Akt pathway in human cancer: rationale and promise. Cancer Cell. 2003;4:257-62.

[55] Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nat Rev Mol Cell Biol. 2006;7:85-96.

[56] Frodin M, Antal TL, Dummler BA, Jensen CJ, Deak M, Gammeltoft S. A phosphoserine/threonine-binding pocket in AGC kinases and PDK1 mediates activation by hydrophobic motif phosphorylation. Embo j. 2002;21:5396-407.

[57] Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell. 2002;109:275-82.

[58] Biondi RM. Phosphoinositide-dependent protein kinase 1, a sensor of protein conformation. Trends Biochem Sci. 2004;29:136-42.

[59] Biondi RM, Cheung PC, Casamayor A, Deak M, Currie RA, Alessi DR. Identification of a pocket in the PDK1 kinase domain that interacts with PIF and the C-terminal residues of PKA. Embo j. 2000;19:979-88.

[60] Biondi RM, Komander D, Thomas CC, Lizcano JM, Deak M, Alessi DR. High resolution crystal structure of the human PDK1 catalytic domain

defines the regulatory phosphopeptide docking site. Embo j. 2002;21:4219-28.

[61] Engel M, Hindie V, Lopez-Garcia LA, Stroba A, Schaeffer F, Adrian I. Allosteric activation of the protein kinase PDK1 with low molecular weight compounds. Embo j. 2006;25:5469-80.

[62] Stroba A, Schaeffer F, Hindie V, Lopez-Garcia L, Adrian I, Frohner W. 3,5-Diphenylpent-2-enoic acids as allosteric activators of the protein kinase PDK1: structure-activity relationships and thermodynamic characterization of binding as paradigms for PIF-binding pocket-targeting compounds. J Med Chem. 2009;52:4683-93.

[63] Hindie V, Stroba A, Zhang H, Lopez-Garcia LA, Idrissova L, Zeuzem S. Structure and allosteric effects of low-molecular-weight activators on the protein kinase PDK1. Nat Chem Biol. 2009;5:758-64.

[64] Wei L, Gao X, Warne R, Hao X, Bussiere D, Gu XJ. Design and synthesis of benzoazepin-2-one analogs as allosteric binders targeting the PIF pocket of PDK1. Bioorg Med Chem Lett. 2010;20:3897-902.

[65] Blander G, Guarente L. The Sir2 family of protein deacetylases. Annu Rev Biochem. 2004;73:417-35.

[66] Sanders BD, Jackson B, Marmorstein R. Structural basis for sirtuin function: what we know and what we don't. Biochim Biophys Acta. 2010;1804:1604-16.

[67] Howitz KT, Bitterman KJ, Cohen HY, Lamming DW, Lavu S, Wood JG. Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. Nature. 2003;425:191-6.

[68] Milne JC, Lambert PD, Schenk S, Carney DP, Smith JJ, Gagne DJ, et al. Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. Nature. 2007;450:712-6.

[69] Kaeberlein M, McDonagh T, Heltweg B, Hixon J, Westman EA, Caldwell SD, et al. Substrate-specific activation of sirtuins by resveratrol. J Biol Chem. 2005;280:17038-45.

[70] Pacholec M, Bleasdale JE, Chrunyk B, Cunningham D, Flynn D, Garofalo RS. SRT1720, SRT2183, SRT1460, and resveratrol are not direct activators of SIRT1. J Biol Chem. 2010;285:8340-51.

[71] Dai H, Kustigian L, Carney D, Case A, Considine T, Hubbard BP. SIRT1 activation by small molecules: kinetic and biophysical evidence for direct interaction of enzyme and activator. J Biol Chem. 2010;285:32695-703.

[72] Hubbard BP, Gomes AP, Dai H, Li J, Case AW, Considine T. Evidence for a Common Mechanism of SIRT1 Regulation by Allosteric Activators. 2013.

[73] Darby JF, Landstrom J, Roth C, He Y, Davies GJ, Hubbard RE. Discovery of selective small-molecule activators of a bacterial glycoside hydrolase. Angew Chem Int Ed Engl. 2014;53:13419-23.

[74] Yuan H, Marmorstein R. Structural basis for sirtuin activity and inhibition. J Biol Chem. 2012;287:42428-35.

[75] Gertz M, Nguyen GT, Fischer F, Suenkel B, Schlicker C, Franzel B. A molecular mechanism for direct sirtuin activation by resveratrol. PLoS One. 2012;7:e49761.

[76] Schlicker C, Boanca G, Lakshminarasimhan M, Steegborn C. Structure-based development of novel sirtuin inhibitors. Aging (Albany NY). 2011;3:852-72.

[77] You W, Rotili D, Li TM, Kambach C, Meleshin M, Schutkowski M. Structural Basis of Sirtuin 6 Activation by Synthetic Small Molecules. Angew Chem Int Ed Engl. 2017;56:1007-11.

[78] Huang Z, Zhao J, Deng W, Chen Y, Shang J, Song K. Identification of a cellularly active SIRT6 allosteric activator. Nat Chem Biol. 2018;14:1118-26.

[79] Sebastian JS. Reversible activators of enzymes. J Chem Edu. 1987;64:1031-2.

[80] Darby JF, Atobe M, Firth JD, Bond P, Davies GJ, O'Brien P. Increase of enzyme activity through specific covalent modification with fragments. 2017.

[81] DeSantis G, Berglund P, Stabile MR, Gold M, Jones JB. Site-directed mutagenesis combined with chemical modification as a strategy for altering the specificity of the S1 and S1' pockets of subtilisin Bacillus lentus. Biochemistry. 1998;37:5968-73.

[82] DeSantis G, Jones JB. Chemical modification of enzymes for enhanced functionality. Curr Opin Biotechnol. 1999;10:324-30.

[83] Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM. Site-directed ligand discovery. Proc Natl Acad Sci U S A. 2000;97:9367-72.

[84] Sadowsky JD, Burlingame MA, Wolan DW, McClendon CL, Jacobson MP, Wells JA. Turning a protein kinase on or off from a single allosteric site via disulfide trapping. Proc Natl Acad Sci U S A. 2011;108:6056-61.

[85] Brethauer S, Studer MH. Biochemical Conversion Processes of Lignocellulosic Biomass to Fuels and Chemicals – A Review. International Journal for Chemistry. 2015;10:572-581.

[86] Service RF. Cellulosic ethanol. Biofuel researchers prepare to reap a new harvest. Science. 2007;5818:1488-91.

[87] Sticklen MB. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. Nat Rev Genet. 2008;9:433-43.

[88] Singhania RR, Patel AK, Sukumaran RK, Larroche C, Pandey A. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. Bioresour Technol. 2013;127:500-7.

[89] Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev. 2002;66:506-77.

[90] Zhang YH, Lynd LR. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. Biotechnol Bioeng. 2004;88:797-824.

[91] Bommarius AS, Katona A, Cheben SE, Patel AS, Ragauskas AJ, Knudson K. Cellulase kinetics as a function of cellulose pretreatment. Metab Eng. 2008;10:370-81.

[92] Resa P, Buckin V. Ultrasonic analysis of kinetic mechanism of hydrolysis of cellobiose by beta-glucosidase. Anal Biochem. 2011;415:1-11.

[93] Du F, Wolger E, Wallace L, Liu A, Kaper T, Kelemen B. Determination of product inhibition of CBH1, CBH2, and EG1 using a novel cellulase activity assay. Appl Biochem Biotechnol. 2010;161:313-7.

[94] Berlin A, Maximenko V, Gilkes N, Saddler J. Optimization of enzyme complexes for lignocellulose hydrolysis. Biotechnol Bioeng. 2007;97:287-96.

[95] Chen M, Zhao J, Xia L. Enzymatic hydrolysis of maize straw polysaccharides for the production of reducing sugars. Carbohydrate polymers 2008. 2008;71:411-5.

[96] Beguin P. Molecular biology of cellulose degradation. Annu Rev Microbiol. 1990;44:219-48.

[97] Bhatia Y, Mishra S, Bisaria VS. Microbial beta-glucosidases: cloning, properties, and applications. Crit Rev Biotechnol. 2002;22:375-407.

[98] Brzobohaty B, Moore I, Kristoffersen P, Bako L, Campos N, Schell J. Release of active cytokinin by a beta-glucosidase localized to the maize root meristem. Science. 1993;262:1051-4.

[99] Leah R, Kigel J, Svendsen I, Mundy J. Biochemical and molecular characterization of a barley seed beta-glucosidase. J Biol Chem. 1995;270:15789-97.

[100] Zagrobelny M, Bak S, Moller BL. Cyanogenesis in plants and arthropods. Phytochemistry. 2008;69:1457-68.

[101] Street SGWaIP. β-Glucosidases: Mechanism and Inhibition. 1989:597-607.

[102] Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. Curr Opin Struct Biol. 1997;7:637-44.

[103] Schmid G, Wandrey C. Purification and partial characterization of a cellodextrin glucohydrolase (beta-glucosidase) from Trichoderma reesei strain QM 9414. Biotechnol Bioeng. 1987;30:571-85.

[104] Gao J, Qian Y, Wang Y, Qu Y, Zhong Y. Production of the versatile cellulase for cellulose bioconversion and cellulase inducer synthesis by genetic improvement of Trichoderma reesei. Biotechnol Biofuels. 2017;10:272.

[105] Tiwari P, Misra BN, Sangwan NS. β-Glucosidases from the Fungus Trichoderma: An Efficient Cellulase Machinery in Biotechnological Applications. BioMed Res Int. 2013.

[106] Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37:D233-8.

[107] Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J. 1991;280:309-16.

[108] Wang Q, Trimbur D, Graham R, Warren RA, Withers SG. Identification of the acid/base catalyst in Agrobacterium faecalis beta-glucosidase by kinetic analysis of mutants. Biochemistry. 1995;34:14554-62.

[109] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42:W320-4.

[110] Seshadri S, Akiyama T, Opassiri R, Kuaprasert B, Cairns JK. Structural and enzymatic characterization of Os3BGlu6, a rice beta-glucosidase hydrolyzing hydrophobic glycosides and (1->3)- and (1->2)-linked disaccharides. Plant Physiol. 2009;151:47-58.

[111] Gloster TM, Davies GJ. Glycosidase inhibition: assessing mimicry of the transition state. Org Biomol Chem. 2010;8:305-20.

[112] Roberts SM, Davies GJ. The crystallization and structural analysis of cellulases (and other glycoside hydrolases): strategies and tactics. Methods Enzymol. 2012;510:141-68.

[113] Tarazona I, Chisvert A, Salvador A. Determination of benzophenone-3 and its main metabolites in human serum by dispersive liquid-liquid microextraction followed by liquid chromatography tandem mass spectrometry. Talanta. 2013;116:388-95.

[114] McNicholas S, Potterton E, Wilson KS, Noble ME. Presenting your structures: the CCP4mg molecular-graphics software. Acta Crystallogr D Biol Crystallogr. 2011;67:386-94.

[115] Dong W, Jespersen T, Bols M, Skrydstrup T, Sierks MR. Evaluation of isofagomine and its derivatives as potent glycosidase inhibitors. Biochemistry. 1996;35:2788-95.

[116] Morera S, Vigouroux A, Stubbs KA. A potential fortuitous binding of inhibitors of an inverting family GH9 beta-glycosidase derived from isofagomine. Org Biomol Chem. 2011;9:5945-7.

[117] Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of fragments on drug discovery. Nature Reviews Drug Discovery. 2016;15:605.

[118] Marion D. An Introduction to Biological NMR Spectroscopy. Mol Cell Proteomics. 2013;11:3006-25.

[119] van Nuland NA, Dobson CM, Regan L. Characterization of folding the four-helix bundle protein Rop by real-time NMR. Protein Eng Des Sel. 2008;21:165-70.

[120] Bieri M, Kwan AH, Mobli M, King GF, Mackay JP, Gooley PR. Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. Febs j. 2011;278:704-15.

[121] LaPlante SR, Carson R, Gillard J, Aubry N, Coulombe R, Bordeleau S. Compound aggregation in drug discovery: implementing a practical NMR assay for medicinal chemists. J Med Chem. 2013;56:5142-50.

[122] Dalvit C, Pevarello P, Tato M, Veronesi M, Vulpetti A, Sundstrom M. Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. J Biomol NMR. 2000;18:65-8.

[123] Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B. WaterLOGSY as a method for primary NMR screening: practical aspects and range of applicability. J Biomol NMR. 2001;21:349-59.

[124] Fernandez C, Wider G. TROSY in NMR studies of the structure and function of large biological macromolecules. Curr Opin Struct Biol. 2003;13:570-80.

[125] Sugiki T, Kobayashi N, Fujiwara T. Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists. Comput Struct Biotechnol J. 2017;15:328-39.

[126] Tugarinov VT, Kay LE. Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. J Am Chem Soc.2003;45:13868-78.

[127] Ruschak AM, Kay LE. Methyl groups as probes of supra-molecular structure, dynamics and function. J Biomol NMR. 2010;46:75-87.

[128] Tugarinov V, Hwang PM, Kay LE. Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. Annu Rev Biochem. 2004;73:107-46.

[129] Sinha K, Jen-Jacobson L, Rule GS. Divide and conquer is always best: sensitivity of methyl correlation experiments. J Biomol NMR. 2013;56:331-5.

[130] LeMaster DM, Richards FM. NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteriation. Biochemistry. 1988;27:142-50.

[131] Sattler M, Fesik SW. Use of deuterium labeling in NMR: overcoming a sizeable problem. Structure. 1996;4:1245-9.

[132] Sattler M. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progress in Nuclear Magnetic Resonance Spectroscopy. 1999;34:93–158.

[133] Schmidt E, Guntert P. A new algorithm for reliable and general NMR resonance assignment. J Am Chem Soc. 2012;134:12817-29.

[134] Shen Y, Bax A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J Biomol NMR. 2013;56:227-41.

[135] Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007;445:618-22.

[136] Maurer T. NMR studies of protein-ligand interactions. Methods Mol Biol. 2005;305:197-214.

[137] Ziarek JJ, Peterson FC, Lytle BL, Volkman BF. Binding site identification and structure determination of protein-ligand complexes by NMR a semiautomated approach. Methods Enzymol. 2011;493:241-75.

[138] Williamson MP. Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc. 2013;73:1-16.

[139] de Vries SJ, van Dijk M, Bonvin AM. The HADDOCK web server for data-driven biomolecular docking. Nat Protoc. 2010;5:883-97.

[140] van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016;428:720-5.

[141] Kay LE, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. J Am Chem Soc. 1992;26:10663-65.

[142] Davis AL, Keeler J, Laue ED, Moskau D. Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients. Journal of Magnetic Resonance. 1992;98:207-216.

[143] Shah DM, A BE, Diercks T, Hass MA, van Nuland NA, Siegal G. Rapid protein-ligand costructures from sparse NOE data. J Med Chem. 2012;55:10786-90.

[144] Lee HL, Chang CK, Jeng WY, Wang AH, Liang PH. Mutations in the substrate entrance region of beta-glucosidase from Trichoderma reesei improve enzyme activity and thermostability. Protein Eng Des Sel. 2012;25:733-40.

[145] Offen WA, Viksoe-Nielsen A, Borchert TV, Wilson KS, Davies GJ. Three-dimensional structure of a variant `Termamyl-like' Geobacillus stearothermophilus alpha-amylase at 1.9 A resolution. Acta Crystallogr F Struct Biol Commun. 2015;71:66-70.

[146] Schulz MN, Landstrom J, Bright K, Hubbard RE. Design of a fragment library that maximally represents available chemical space. J Comput Aided Mol Des. 2011;25:611-20.

[147] Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr. 2011;67:271-81.

[148] Evans P. Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr. 2006;62:72-82.

[149] Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013;69:1204-14.

[150] Vagin A, Teplyakov A, IUCr. MOLREP: an Automated Program for Molecular Replacement. Journal of Applied Crystallography. 1997;30:1022-5.

[151] Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66:486-501.

[152] Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr. 2011;67:355-67.

[153] Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc. 2007;2:2212-21.

[154] Schulz MN, Landstrom J, Hubbard RE. MTSA--a Matlab program to fit thermal shift data. Anal Biochem. 2013;433:43-7.

[155] Viegas A, Manso J, Corvo MC, Marques MM, Cabrita EJ. Binding of ibuprofen, ketorolac, and diclofenac to COX-1 and COX-2 studied by saturation transfer difference NMR. J Med Chem. 2011;54:8555-62.

[156] Mayer M, Meyer B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. Angew Chem Int Ed Engl. 1999;38:1784-8.

[157] Czisch M, Boelens R. Sensitivity enhancement in the TROSY experiment. J Magn Reson. 1998;134:158-60.

[158] Pervushin KV, Wider G, Wuthrich K. Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY. J Biomol NMR. 1998;12:345-8.

[159] Vuister GW, Bax A, Vuister GW, Laboratory of Chemical Physics NIoDaDaKD, National Institutes of Health, Bethesda, Maryland 20892, USA, Bax A, Laboratory of Chemical Physics NIoDaDaKD, National Institutes of Health, Bethesda, Maryland 20892, USA. Resolution enhancement and spectral editing of uniformly. Journal of Magnetic Resonance. 1992;98:428.

[160] Grzesiek S, Bax A. Improved 3D Triple-Resonance NMR Techniques Applied to a 31-kDa Protein. 1992.

[161] Muhandiram DR, Kay LE. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. Journal of Magnetic Resonance. 1994;103:203-16.

[162] Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR Spectroscopy of isotopically enriched proteins. 1990. J Magn Reson. 2011;213:423-41.

[163] Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc Natl Acad Sci U S A. 1998;95:13585-90.

[164] Clubb RT. A Constant-Time Three-Dimensional Triplr-Resonance Pulse Scheme to Correlate Intraresidue HN, 15N, and 13C Chemical Shifts in ^{15}N-^{13}C-Labeled Proteins. Journal of magnetic resonance. 1992;97:213-7.

[165] Salzmann M, Wider G, Pervushin K, Wuthrich K. Improved sensitivity and coherence selection for [15N,1H]-TROSY elements in triple resonance experiments. J Biomol NMR. 1999;15:181-4.

[166] Eletsky A, Kienhofer A, Pervushin K. TROSY NMR with partially deuterated proteins. J Biomol NMR. 2001;20:177-80.

[167] Grzesiek S, Bax A. Correlating Backbone Amide and Side-Chain Resonances in Larger Proteins By Multiple Relayed Triple Resonance NMR. J Am Chem Soc. 1992;16:6291-93.

[168] Zuiderweg ER, Fesik SW. Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. Biochemistry. 1989;28:2387-91.

[169] Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry. 1989;28:6150-6.

[170] Pascal SM, Singer AU, Gish G, Yamazaki T, Shoelson SE, Pawson T. Nuclear magnetic resonance structure of an SH2 domain of phospholipase C-gamma 1 complexed with a high affinity binding peptide. Cell. 1994;77:461-72.

[171] Wen J, Zhou P, Wu J. Efficient acquisition of high-resolution 4-D diagonal-suppressed methyl-methyl NOESY for large proteins. J Magn Reson. 2012;218:128-32.