3-Phosphoglycerate dehydrogenase

as target in cancer therapy

Judith Edda Unterlass

Thesis submitted for degree of Doctor of Philosophy

September 2015

Northern Institute for Cancer Research

Faculty of Medical Sciences

Newcastle University

Newcastle University
UK | Malaysia | Singapore

I. Abstract

Cancer cells adapt their metabolism to simultaneously fulfil the requirements of energy production and biomass generation necessary to sustain high proliferation rates. This deregulated energy metabolism and the proteins responsible therefor provide a potential new route of targeting cancer that has not been thoroughly explored. 3-Phosphoglycerate dehydrogenase (PHGDH), which takes 3-phosphoglycerate (3-PG) out of the glycolytic pathway and into serine production, has been reported as potential target in certain breast cancer forms and melanoma. There is no known inhibitor of PHGDH to date to fully validate the target.

Inhibition of PHGDH was explored in breast cancer and melanoma cell lines using siRNA and shRNA interference techniques. Greater knockdown was achieved by siRNA resulting in better growth inhibition than when using shRNA expressing cell lines.

The substrate-binding pocket was investigated with substrate analogues and substrate-containing compounds. NAD^+-binding was found to be stabilised by coordinated binding of substrates. The catalytic subunits of human PHGDH were crystallised and revealed a flexible lid domain that moves in response to substrate binding.

The NAD^+-fragment adenosine 5'-diphosphoribose (ADPR) was shown to be a moderate inhibitor of the enzymatic activity of PHGDH and was used for assay validation. Cofactor analogues with different substituents around the pyridine ring were equally suitable to promote the oxidation of 3-PG.

A fragment screen was performed using differential scanning fluorimetry and hits were subsequently validated by competition isothermal titration calorimetry. To investigate the fragments in crystals of human PHGDH, a truncated form of PHGDH (construct 93) was engineered by limited proteolysis. Soaking of fragments into crystals of 93 confirmed binding of seven fragments. Structure-activity relationship studies were initiated around the confirmed hits.

II. Acknowledgments

III. Contents

x

IV. List of Figures, Tables and Equations

IV.1   List of Figures

IV.2 List of Tables

IV.3   List of Equations

V. List of Abbreviations

aa	Amino acid
ADPR	Adenosine 5'-diphosphoribose
CD	Circular dichroism
CV	Coefficient of variation
cv	Column volume
Da	Dalton
DMF	Dimethylformamide
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
$ES^{+/-}$	Electrospray ionisation positive/ negative mode
EtOAc	Ethyl acetate
FBS	Foetal bovine serum
FP	Fluorescence polarisation
G	Gibbs free energy
H	Enthalpy
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HRMS	High resolution mass spectrometry
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K_a	Affinity constant
K_D	Binding constant
LB	Lysogeny broth
LC-MS	Liquid chromatography-mass spectrometry

LE	Ligand efficiency
M. tuberculosis	Mycobacterium tuberculosis
MOI	Multiplicity of infection
mol eq.	Molar equivalent
mRNA	Messenger RNA
m/z	Mass charge ratio
NAD$^+$	Nicotinamide adenine dinucleotide (oxidised from)
NADH	Nicotinamide adenine dinucleotide (reduced form)
OD$_{600}$	Optical density at λ = 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein databank
3-PG	3-Phosphoglycerate
PHP	3-Phosphohydroxypyruvate
R$_f$	Retention factor
Rmsd	Root mean squared deviation
ROS	Reactive oxygen species
RT	Room temperature
S	Entropy
SAR	Structure-Activity relationship
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
shRNA	Short hairpin RNA
siRNA	Short interfering RNA
SOC	Super optimal broth with catabolic repressor
sPHGDH	Short form of PHGDH (aa 3-314)

TB	Terrific broth	
TCEP	Tris(2-carboxyethyl)phosphine	
TEV	Tobacco Etch Virus	
T_m	Melting temperature	
WT	Wild-type	
λ_{em}	Emission wavelength	
λ_{ex}	Excitation wavelength	

Amino acids

Alanine	A	Ala	Leucine	L	Leu
Arginine	R	Arg	Lysine	K	Lys
Asparagine	N	Asn	Methionine	M	Met
Aspartic acid	D	Asp	Phenylalanine	F	Phe
Cysteine	C	Cys	Proline	P	Pro
Glutamic acid	E	Glu	Serine	S	Ser
Glutamine	Q	Gln	Threonine	T	Thr
Glycine	G	Gly	Tryptophan	W	Trp
Histidine	H	His	Tyrosine	Y	Tyr
Isoleucine	I	Ile	Valine	V	Val

Chapter 1. Introduction

Despite a strong research effort, cancer is still among the leading causes of disease worldwide with an estimate of 14.1 million new cancer cases arising in 2012 [1]. In the UK alone, 331,500 people were diagnosed with cancer in 2011, with breast, lung, prostate and bowel cancer accounting for the largest part of the cancer burden [2]. Taking into account increasing population numbers as well as increased life expectancy, the burden of cancer as an age-related diseases, is predicted to increase to 23.6 million new cases each year by 2030 globally [1].

A major problem for the treatment of cancer is the complexity of the disease as illustrated by the large variety of cancer types and their genomic heterogeneity. Recent genetic analysis of over 1 million tumour samples identified over 2 million coding mutations, over 10,000 fusion mutations and more than 60 million gene expression variants [3]. Against such a multifaceted background, usage of the genetic information available on tumours to design tumour-specific drugs seems a challenging task. In addition, malignant cells develop from normal human cells and, as such, retain a very close similarity to healthy cells, which renders them difficult to target without harming healthy cells. In an attempt to find common features in the multistep process of tumorigenesis, a set of six hallmarks of cancer was presented by Hanahan and Weinberg. These include the ability to resist apoptosis, promote angiogenesis, overcome anti-growth signals, sustain growth signalling, activate tissue invasion and metastasis and limitless proliferation [4]. This set of capabilities acquired by cells during tumour development was recently expanded to include two emerging hallmarks: the power to evade immunological destruction and to adapt their energy metabolism [5]. Deregulated energy metabolism is thereby strongly related to the increased proliferation of cancer cells, which necessitates an adjustment of the cells' energy metabolism so as to be able to fuel the increased demands of cell growth and division.

1.1 Cancer cell metabolism

Metabolism can be simply described as the sum of biochemical processes in a living organism involved in the generation or consumption of energy. The principal

metabolic pathways comprise intermediates, which are essential for energy homeostasis and synthesis of macromolecules, such as carbohydrates, fatty acids and amino acids. The core metabolic pathways can be divided into three major categories: processes for construction or synthesis of macromolecules (anabolism), processes for degradation of molecules to release energy (catabolism) and processes to eliminate toxic waste products (waste disposal) [6].

All cells have to carry out metabolic processes to generate energy so as to be able to execute important functions of cell maintenance, such as creation of concentration gradients, dynamic processes of the cytoskeleton, DNA repair, transcription and translation, protein turnover and vesicular trafficking [7]. Normal cells have complex signalling and sensing pathways to regulate nutrient uptake from the environment, and initiate the necessary metabolic pathways so that, under normal conditions, no uncontrolled nutrient uptake is observed in the absence of stimulation by growth factors [8]. Cancer cells seem to be able to overcome this dependency on growth signals. They are able to constitutively take up and metabolise nutrients, particularly glucose and glutamine, to assist cell survival and drive cell growth [6]. For fast proliferating cells, such as cancer cells, the main metabolic tasks are in balancing the need for sufficient energy production with fulfilling the additional anabolic demands of producing sufficient amounts of building blocks for the development of daughter cells.

1.1.1 Glucose metabolism

A major source of cellular energy and biomass is glucose, which is taken up from the environment through glucose transporter proteins (GLUT). In many cancers certain glucose transporter isoforms are upregulated and this can be mediated through activation of oncogenes such as c-Myc [9].

In the presence of oxygen, most cells metabolise glucose to pyruvate through glycolysis followed by the complete oxidation of pyruvate to carbon dioxide in the tricarboxylic acid (TCA) cycle (Figure 1.1). Through the steps of glycolysis, the six carbon sugar, glucose, is oxidized and split resulting in two molecules of pyruvate. In the subsequent processes of oxidative phosphorylation through the TCA cycle two adenosine 5'-triphosphate (ATP) molecules, one guanosine 5'-triphosphate (GTP) as well as nicotinamide adenine dinucleotide (NADH) and flavine adenine dinucleotide ($FADH_2$), are generated. The subsequent use of NADH and $FADH_2$ in the respiratory

chain in the mitochondria increases the ATP production per glucose molecule up to 36 molecules. Under anaerobic conditions, however, normal cells will switch to an alternative fermentation process and reductively metabolise pyruvate to lactate. This process of anaerobic glycolysis is far less efficient with regards to ATP production than the oxidative phosphorylation in the mitochondria as it only generates 2 molecules of ATP per glucose.
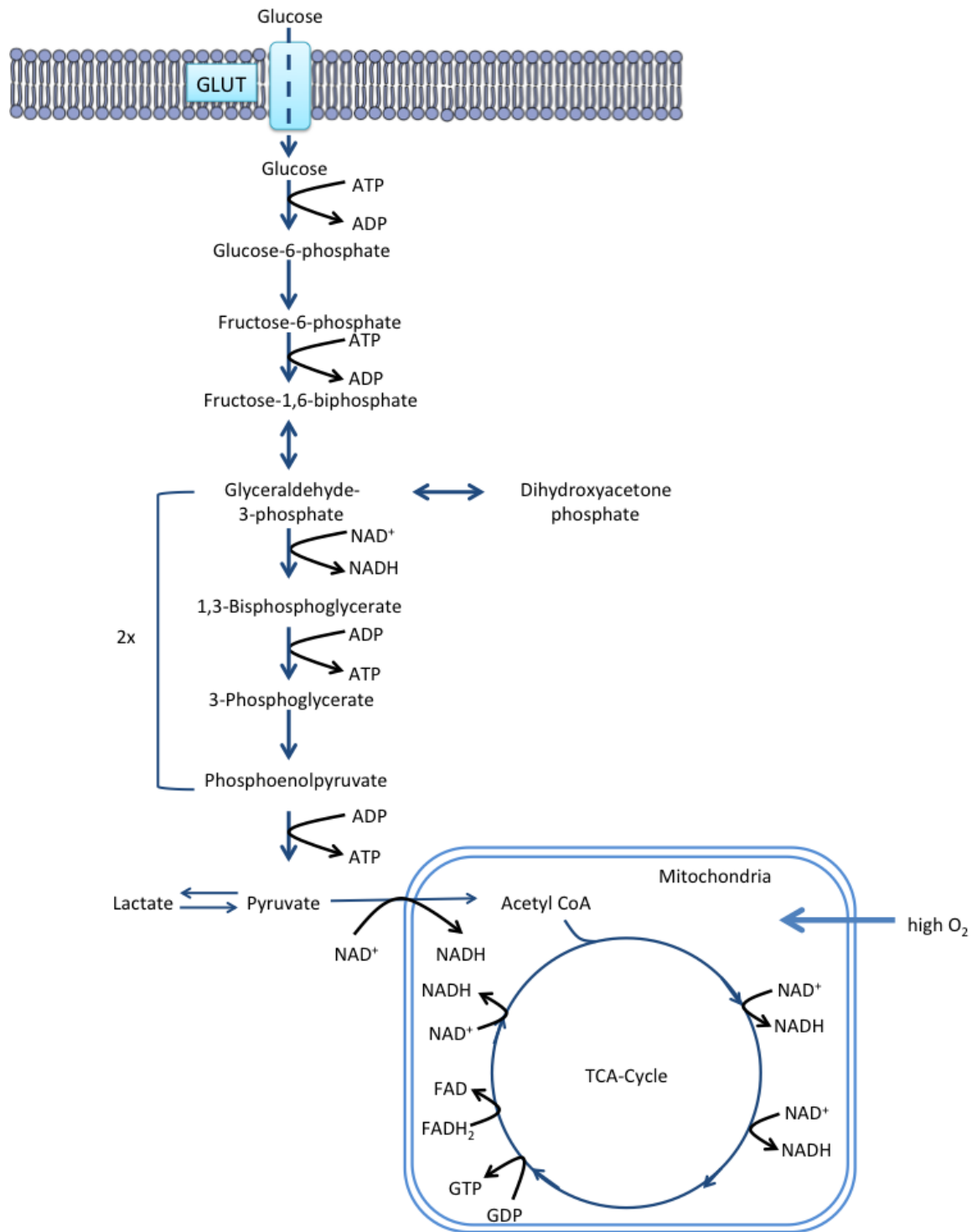
Figure 1.1 - Schematic representation of glycolysis and the subsequent TCA cycle.

Glucose, after being transported into the cell, is cleaved into two three-carbon fragments (glyceraldehyde 3-phosphate and dihydroxyacetone phosphate). In this first part of glycolysis two molecules of ATP are consumed. Glyceraldehyde 3-phosphate is then further metabolised to pyruvate which results in the generation of two molecules of ATP. As the latter steps of the pathway are performed twice per glucose molecule, glycolysis results in an overall generation of two molecules of ATP.

1.1.2 The Warburg effect

In the 1920s, Otto Warburg discovered that even in the presence of saturating amounts of oxygen, cancer cells seem to metabolise glucose mainly through fermentation, resulting in increased lactate production and reduced generation of ATP (Figure 1.2) [10, 11].
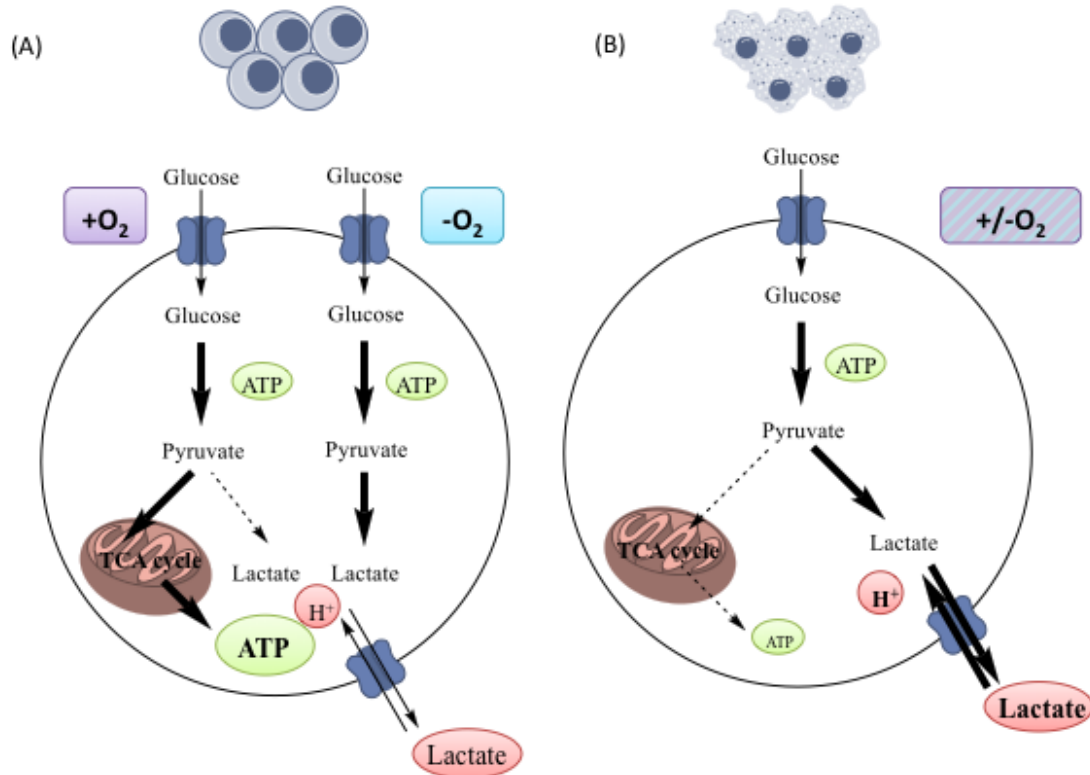
Figure 1.2 – Schematic representation of the Warburg effect

Differentiated, healthy tissue (A) metabolises glucose depending on the cellular oxygen level, so that in the presence of sufficient oxygen large quantities of ATP are generated through the TCA cycle (oxidative phosphorylation). Only if oxygen levels drop, the majority of glucose will be metabolised to lactate which is excreted from the cells to allow further lactate production, thus sustaining a minimum level of ATP production (anaerobic glycolysis). (B) Tumorigenic cells metabolise the majority of glucose to lactate independently of intracellular oxygen levels (Warburg effect or aerobic glycolysis). Picture adapted from Vander Heiden *et al.*, 2009 [8].

This apparently paradoxical phenomenon of aerobic glycolysis, named the "Warburg effect" after its discoverer, has since been demonstrated in many different tumour types [12]. Warburg initially thought that the observed decrease in oxidative phosphorylation was a consequence of defects in mitochondrial respiration resulting in compromised aerobic respiration and subsequent reliance on glycolytic metabolism [11]. However, subsequent evidence emerged to suggest that respiration is generally not impaired in cancer cells [13, 14]. How this metabolic rearrangement

is triggered, and if it is a prerequisite of tumour initiation, has still to be elucidated. However, further investigations have shown the importance of genetic as well as environmental factors in contributing to the Warburg effect. Among the genetic alterations reported are deregulation of transcription factors, such as c-Myc and hypoxia-inducible transcription factor (HIF-1α), that increase the expression of proteins that permit or mediate glycolysis, such as the GLUT1 transporter and the glycolytic enzyme hexokinase, as well as lactate dehydrogenase A (LDH-A) which converts pyruvate to lactate and is therefore a key enzyme in maintaining glycolysis [15]. These genetic alterations are closely related to changes in the tumour microenvironment where certain regions of the tumour are characterised by hypoxia and acidosis. Hypoxia refers to low oxygen levels often found in solid tumours due to irregular vasculature, that in turn results in poor blood flow and consequent upregulation of HIF-1α, which further regulates the expression of genes that mediate various metabolic processes (Figure 1.3). Acidosis can also be regarded as a consequence of hypoxia, since increased aerobic glycolysis increases the amount of intracellular lactate so that more lactate is transported out of the cells. Lactate export occurs through monocarboxylate transporters (MCT1-4). MCTs are symporters that facilitate the transport of lactate together with protons out of the cell and thus acidify the surrounding extracellular tissue (Figure 1.3) [16].
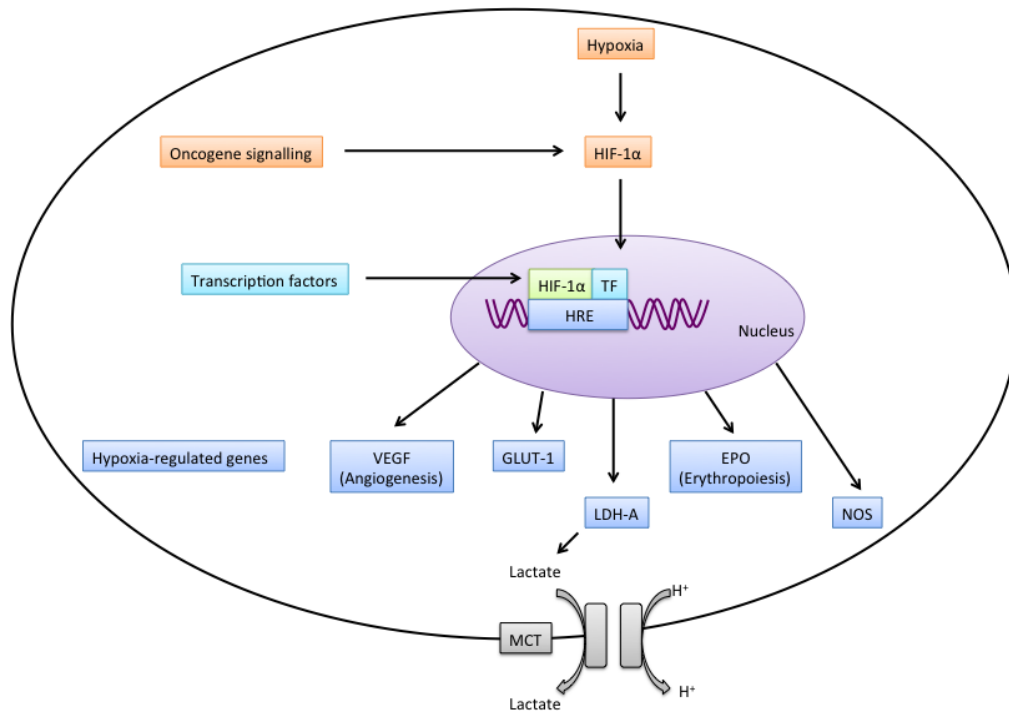
Figure 1.3 – Summary of hypoxia-regulated genes.

Under hypoxic conditions HIF-1α is stabilised and protected from degradation through oncogenic signalling pathways. After translocation into the nucleus, HIF-1α interacts with transcription factors (TF) to initiate transcription of hypoxia-regulated genes (HRE), such as LDH-A. Increased LDH-A activity results in increased lactate production. Lactate is subsequently excreted from the cell along a proton gradient via the MCT symporter. Picture adapted from [17].

However, despite increased understanding of the Warburg effect and its mechanisms of regulation, it is still unclear why cancer cells perform aerobic glycolysis against the background of this metabolic pathway being inferior to oxidative phosphorylation in terms of energy generation. One proposed explanation is that fermentation allows production of ATP at a faster rate (albeit lower efficiency) so that the overall ATP generation rate is higher than with oxidative phosphorylation [18]. However, ATP can also be generated from sources other than glucose, such as fatty acids, and a summary of data from measurements of oxidative and glycolytic ATP production in different cell lines and tissues showed that glycolysis only accounted for an average of 17 % of total ATP production, making this a rather unlikely explanation [14]. Another hypothesis is based on the increase in lactate production, which, through subsequent lactate transport out of the cells results in the aforementioned acidic microenvironment, which causes apoptosis in normal cells while sparing the hypoxia-adapted cancer cells [19]. A third hypothesis proposes that the increased amount of glycolytic intermediates generated through up-regulated glycolysis provides the main

selective advantage underlying the Warburg effect. Sufficient amounts of glycolytic intermediates are necessary to fuel biosynthesis: for example, glucose-6-phosphate, fructose-6-phosphate and glyceraldehyde-3-phosphate all feed into nucleotide synthesis, while 3-phosphoglycerate (3-PG) is a precursor for amino acids and nucleobases [7].

Another consequence of the Warburg effect is that fewer glucose-derived metabolites enter the TCA cycle, which results in cancer cells being more reliant on alternative substrates that can be fed into the TCA cycle, such as glutamine [20].

1.1.3 Glutamine metabolism

Glutamine is the most abundant amino acid in mammals with plasma concentrations of 0.6-0.9 mmol/L and it is used as a major source of nitrogen for the transamination reactions required for the synthesis of amino acids and nucleotides [21, 22]. After entering into the cytosol, glutamine donates nitrogen to purines and pyrimidines via transamination and is thereby converted into glutamate. Glutamate can then donate its remaining amino group to α-keto acids to form non-essential amino acids and α-ketoglutarate, which can be transported into the mitochondria (Figure 1.4) [23]. The conversion of glutamine to α-ketoglutarate can also take place directly in the mitochondria. α-Ketoglutarate then enters into the TCA cycle, thereby mediating the carbon-donor function of glutamine [23]. In addition, glutamine plays an essential role in the suppression of oxidative stress through carbon and nitrogen donation for the synthesis of glutathione, a crucial intracellular antioxidant (Figure 1.4) [24].
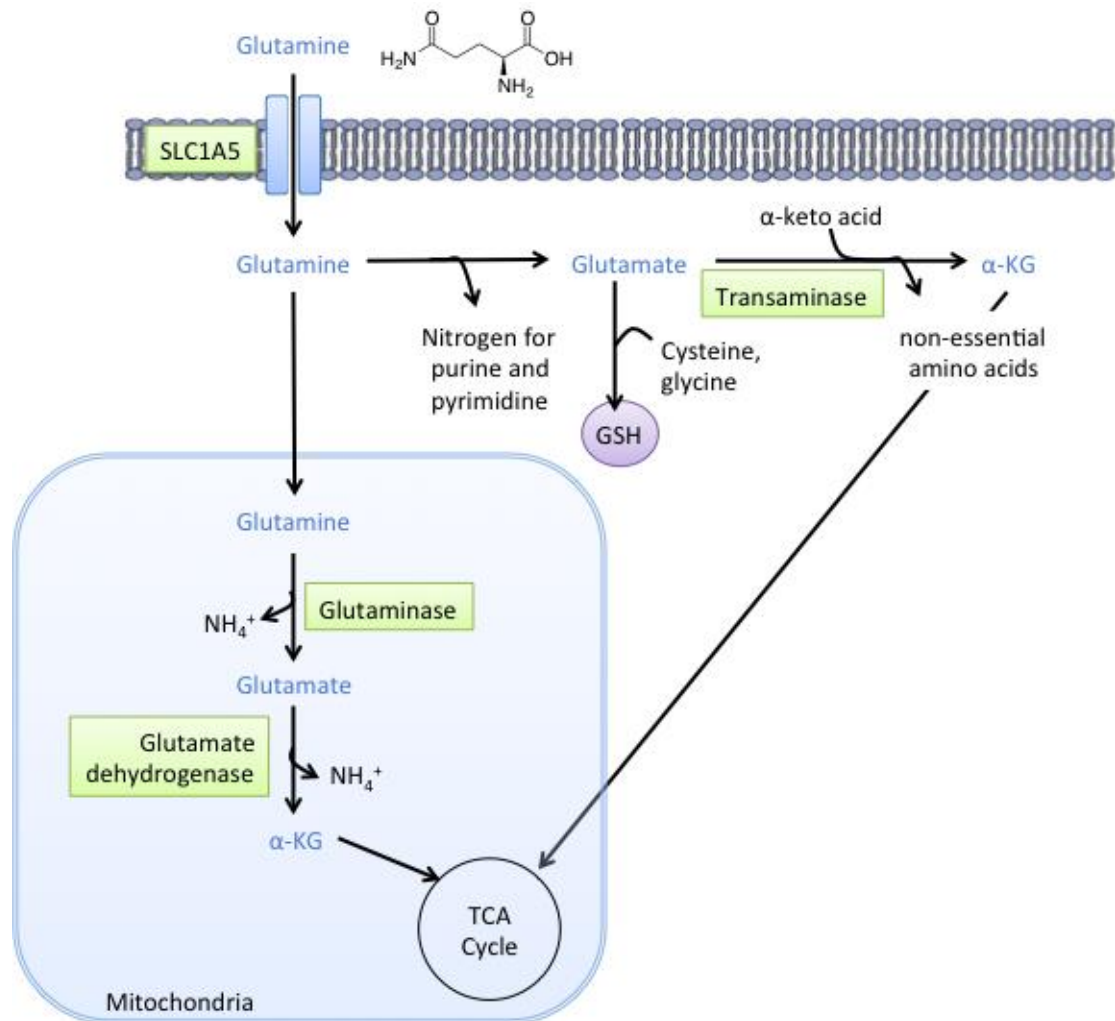
Figure 1.4 – Glutamine metabolism.

Glutamine is imported into the cell via transporters, such as the SLC1A5 and then either directly further metabolised to glutamate and α-ketoglutarate (α-KG) or imported into the mitochondria to be there metabolised to α-KG. Picture adapted from [22].

Recently, additional evidence was presented that underlines the importance of glutamine metabolism in cancer cells. The majority of sporadic clear-cell renal carcinomas (RCC) harbour loss-of-function mutations in the von Hippel-Lindau (*VHL*) tumour suppressor gene resulting in hyperactivation of the transcription factors hypoxia inducible factor 1α and 2α (HIF-1α/ HIF-2α) [25]. Loss of *VHL* was found to be associated with cells performing reductive carboxylation from glutamine to support citrate and lipid synthesis even at normal oxygen levels [26]. The observation that glutamine deprivation with glutaminase inhibitors, inhibiting the enzyme that catalyses the first step in glutamine metabolism, was preferentially cytotoxic to VHL-deficient cells, supports the idea of developing cancer therapeutics to target

glutamine metabolism [25]. However, depletion of glutamine in growing tumours also occurs naturally, due to increased glutamine catabolism [24]. Research undertaken to reveal the signalling pathways activated in response to glutamine deprivation showed that low glutamine levels triggered an increase in reactive oxygen species (ROS) levels, followed by a signalling cascade that leads to the activation of p53 and the induction of pro-survival p53 target genes such as *Cdkn1a* and *Gadd45a* [24]. Furthermore p53-deficient HCT116 cells seemed more sensitive to glutamine deprivation than p53-wild type controls, so that triggering glutamine metabolism in p53-deficient tumours could also be a valuable approach in cancer therapy [24].

1.1.4 Amino acid metabolism

Amino acids play important roles in synthesis of other biomolecules as well as in cell signalling. As such, highly proliferating cells, such as cancer cells, require higher amounts of amino acids and thus have to alter their amino acid management to meet these demands. There are generally four mechanisms by which cells obtain amino acids: *de novo* biosynthesis; dietary uptake of exogenous amino acids from the circulation through transporters; dietary uptake by macropinocytosis; and recycling of amino acids by autophagy [27]. Essential amino acids have to be taken up from the diet as they can not be synthesised *de novo*. In order to regulate how much of each amino acid is currently required by the cell, cellular sensing mechanisms are likely to be involved. These have been shown to exist at least for leucine and arginine [28, 29]. In this context, mammalian target of rapamycin complex 1 (mTORC1), a master regulator of cell growth in cancer cells, has been shown to be activated by amino acids, in particular arginine and leucine, through promotion of the translocation of the complex to the lysosomal surface [30]. The mTORC1 pathway is one of the most frequently mutated signalling pathways in cancer with significant effects on tumour progression, and several mTOR inhibitors are under investigation for therapeutic use in different cancer types [31].

Apart from the implications of L-glutamine metabolism, other amino acids that are not themselves essential can play important roles as intermediary metabolites, including L-serine. L-serine, and in particular its *de novo* synthesis pathway, have been shown to be upregulated in a subset of tumours [32].

1.2 *L*-Serine metabolism

L-serine belongs to the family of non-essential amino acids, as it can be synthesised *de novo.* However, it has been shown that under certain conditions, at least in cell culture, some of the non-essential amino acids become important for optimal cellular growth. In particular, L-serine was required to support the clonogenic survival of HeLa cells at very low densities [33]. L-Serine plays an important role in cell proliferation, as it is the main source of one-carbon units for the synthesis of purine nucleotides and the pyrimidine nucleotide deoxythymidine monophosphate (Figure 1.5) [34]. In this pathway, serine is converted to glycine by serine hydroxymethyl-transferase (SHMT), which transfers one CH_2-unit of serine onto tetrahydrofolate (THF). The thereby-formed methylene-THF is a key intermediate for thymidylate synthesis and an indirect carbon source through other THF cofactors for purine synthesis. The glycine obtained in this SHMT-catalysed reaction is both a carbon and nitrogen source in the purine synthesis, and an important neurotransmitter [35].

In addition to glycine, serine metabolic pathways also provide the precursors for the synthesis of other amino acids such as cysteine and taurine as well as D-serine, a neuromodulator in the brain [36]. Other biomolecules derived from L-serine are sphingosine and phosphatidylserine, which both form part of the phospholipid bi-layers that form cell membranes. Phosphatidylserine plays a particularly important role in the plasma membranes in the brain where it contributes to the activation of signalling pathways [36, 37]. Inborn serine-deficiency disorders, caused by defects in *de novo* L-serine biosynthesis, are associated with major neurological symptoms, further emphasising the importance of L-serine synthesis [38]. Under these conditions the *de novo* synthesis of *L*-serine is insufficient to fulfil the cellular demands, so that L-serine, can be classified as a conditionally essential amino acid [36].
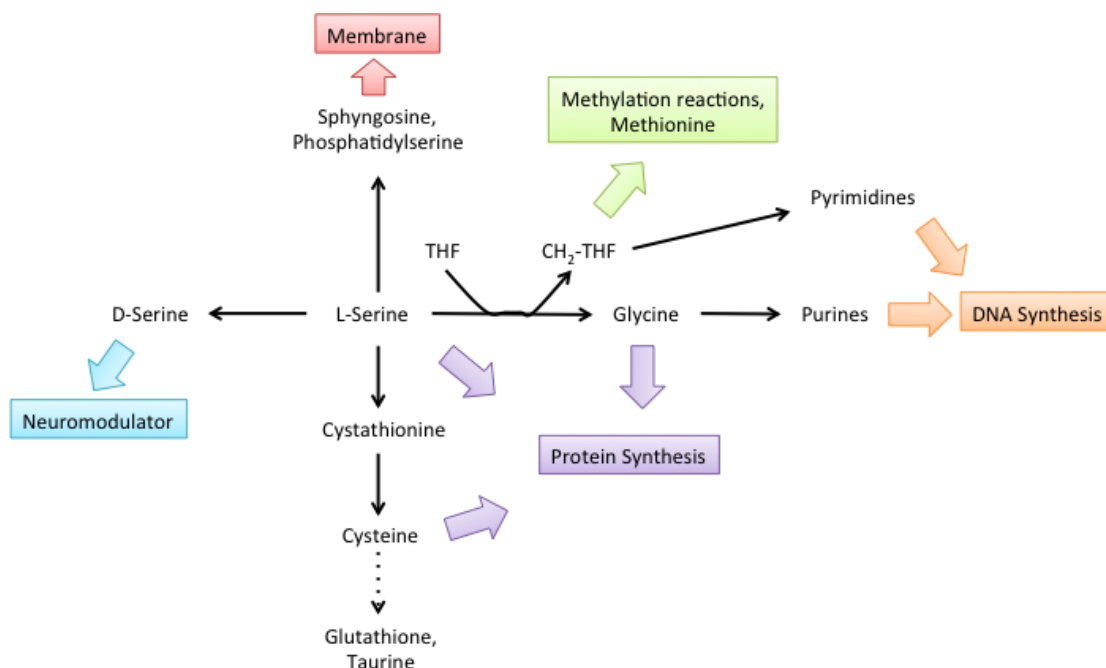
Figure 1.5 – Importance of L-serine in the biosynthesis of other biomolecules.
Schematic representation of how L-serine is linked to the synthesis of various biomolecules.

1.3 *L*-Serine biosynthesis

There are four potential sources of L-serine: (i) uptake through the diet, (ii) *de novo* synthesis from the glycolytic metabolite 3-phosphoglycerate (3-PG), (iii) interconversion with glycine and (iv) degradation of proteins and phospholipids [36]. The contribution of each pathway to L-serine homeostasis in humans remains to be fully elucidated. A study on humans after 24 hours of fasting revealed protein breakdown as a minor L-serine source, contributing 20 % of total L-serine. In this study, the major contribution was from L-serine synthesis, with 27 % and 53 % derived from glycine and 3-PG respectively [39]. L-Serine homeostasis is likely to vary between the different organs as well as depending on the development stage [36]. As the transfer of L-serine across the blood-brain barrier is very low, most of the L-serine in the brain has to be derived from synthesis rather than through dietary uptake [40]. Investigations of L-serine biosynthesis in rat kidneys showed a prevalence of L-serine synthesis from 3-PG over the conversion of glycine [41]. In contrast, the majority of L-serine in foetal lamb liver was found to derive from glycine with the flux through SHMT accounting for over 50 % of L-serine production [42].

1.3.1 Glycine degradation and SHMT

SHMT catalyses the reversible conversion of L-serine to glycine through transfer of a CH_2-group from L-serine onto THF to form 5,10-methylene THF (Figure 1.6). In humans two SHMT genes are known (*SHMT1* and *SHMT2*) encoding for the cytoplasmic (*c*SHMT) and mitochondrial (*m*SHMT) isozymes. Mutagenesis studies in Chinese hamster ovary cells showed that cells lacking *m*SHMT were unable to synthesize glycine indicating that *c*SHMT functions could not compensate for the loss of *m*SHMT [43]. Thus *m*SHMT accounts for the generation of the majority of one-carbon groups resulting in the formation of formate that can then be transported into the cytosol to be used in nucleotide synthesis [44]. *c*SHMT on the other hand functions as a metabolic switch in the cytosol to regulate the supply of one-carbon units for folate synthesis and methylation of homocysteine to methionine. Through the subsequent adenylation of methionine, S-adenosylmethionine (SAM) is generated. SAM is a cofactor for the methylation of many biomolecules, such as DNA and RNA [45].

Glycine degradation is also catalysed by glycine decarboxylase (GLDC), which in humans is part of a multienzyme complex that couples the decarboxylation of glycine with the synthesis of 5,10-methylene THF [46]. 5,10-methylene THF can then in turn function as a substrate of SHMT for the synthesis of L-serine.
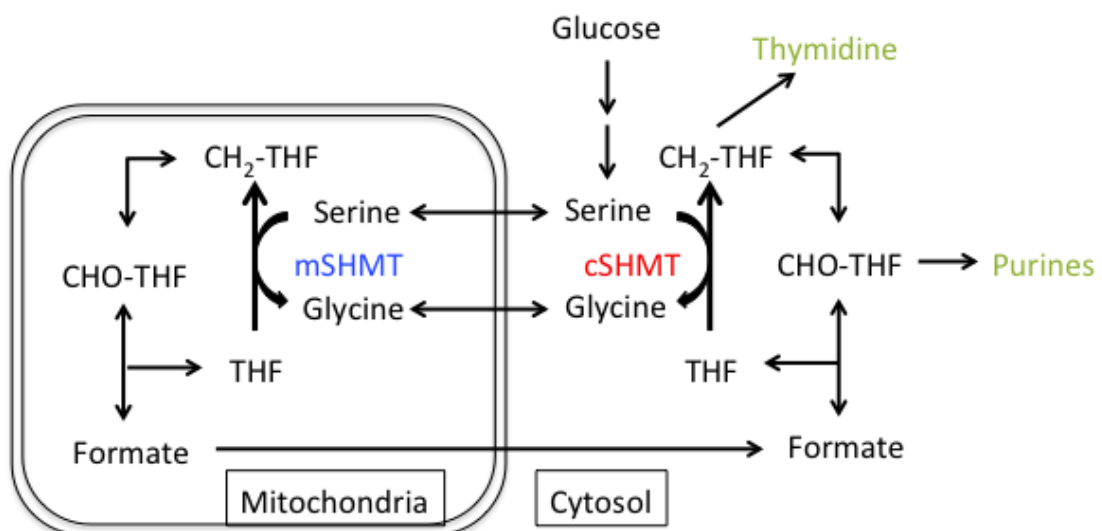
Figure 1.6 – SHMT in the regulation of serine, glycine and folate metabolism. Picture adapted from Fu et *al.* [44].

1.3.2 The phosphorylated L-serine biosynthesis pathway and PHGDH

For *de novo* serine biosynthesis two pathways were originally proposed: the 'phosphorylated pathway', which uses the glycolytic intermediate 3-PG as a substrate, and the 'non-phosphorylated pathway', which involves D-glycerate and hydroxypyruvate. However, the non-phosphorylated pathway was later shown to preferentially participate in serine catabolism, in which L-serine is used for the synthesis of pyruvate to fuel gluconeogenesis. This does not occur in the phosphorylated pathway as the final dephosphorylation step is irreversible [36, 47]. Since the role of the non-phosphorylated pathway is primarily in L-serine degradation, the phosphorylated pathway has now been recognised as the primary route of L-serine synthesis [48, 49].

In the phosphorylated pathway, 3-phosphoglycerate dehydrogenase (PHGDH) catalyses the first step, at which glycolytic flux is diverted into the serine synthesis pathway (Figure 1.7). In this step 3-PG is oxidized to 3-phosphohydroxypyruvate (PHP) using NAD^+ as a cofactor. NAD^+ is concomitantly reduced to NADH as part of the reaction [50]. PHP is then metabolised to 3-phosphoserine by the enzyme 3-phosphoserine aminotransferase (PSAT). For this reaction glutamate is used as a nitrogen donor, releasing α-ketoglutarate as a by-product. The final conversion of phosphoserine to serine is catalysed by 3-phosphoserine phosphatase (PSPH) [51]. The dephosphorylation of phosphoserine is irreversible, and explains the propagation of the reaction cascade to serine production despite the fact that the two preceding reaction steps in the serine synthesis cascade are reversible. The enzymes of the phosphorylated L-serine synthesis pathway are found in many organs and tissues with particular prevalence in kidney, testis, spleen and brain. Lower activity of these enzymes is seen in heart and skeletal muscle [32]. Studies on the enzymatic activity of PHGDH and PSPH in rat liver show an influence of the protein content in the diet, with both enzymes exhibiting low activity in a high protein diet, as well as hormonal factors, such as glucagon and corticosteroids [32].
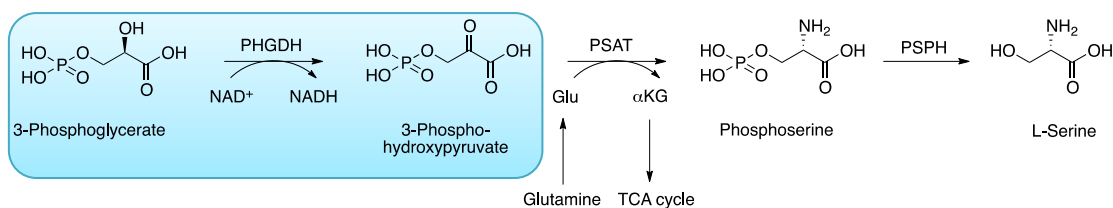
Figure 1.7 – Phosphorylated *L*-serine synthesis pathway

1.3.3 Clinical relevance of L-serine

The importance of L-serine in the cell is highlighted by the effect of modulating serine levels on disease development. An example are the serine deficiency disorders, a group of neurometabolic disorders that arise due to defects in L-serine biosynthesis [52]. Although serine deficiency disorders can be caused by defects in all three enzymes involved in L-serine synthesis, most patients diagnosed with this disease to date have a defect in PHGDH [52]. So far this defect is known to be caused by one of two identified mutations in *PHGDH*: a V490M or V425M substitution [53]. Patients were reported to suffer from congenital microcephaly, severe psychomotor retardation and intractable seizures, which can be partly treated through oral administration of high doses of L-serine and glycine [40].

There is increasing evidence for the involvement of enzymes necessary for L-serine synthesis or metabolism in tumorigenesis. For example, GLDC has been identified as a metabolic oncogene, with overexpression of GLDC in primary cells causing cell transformation [46]. On the other hand an increased supply of glycine through elevated *m*SHMT levels and activity have been reported in colon carcinoma and neuroblastoma [35, 54]. Interestingly high levels of *m*SHMT correlated strongly with high PHGDH levels in human neuroblastoma patient samples indicating that *m*SHMT could be one of the major utilizer of the increasing amounts of *L*-serine synthesised *via* the phosphorylated pathway [54]. Elevated levels of PSAT1 and PHGDH have also been reported in different tumour types, such as PSAT1 and PHGDH upregulation in hepatocellular carcinoma [55] and colon carcinoma [35], respectively. In this context, both enzymes have been proposed as potential targets in cancer therapy with the work presented herein focusing on PHGDH as a target.

1.4 PHGDH

1.4.1 Structure

3-Phosphoglycerate dehydrogenase (PHGDH) is ubiquitously found in all organisms although existing in at least three different basic structural forms, referred to as type I, II and III (Figure 1.8) [56]. These forms do not appear to be strictly organism specific as certain bacteria and lower eukaryotes, such as yeast and Neurospora, have a structurally similar PHGDH to the *Escherichia coli* enzyme, while mammalian PHGDH shows structural homology with the enzyme from the pathogenic bacterium *Mycobacterium tuberculosis* [56].
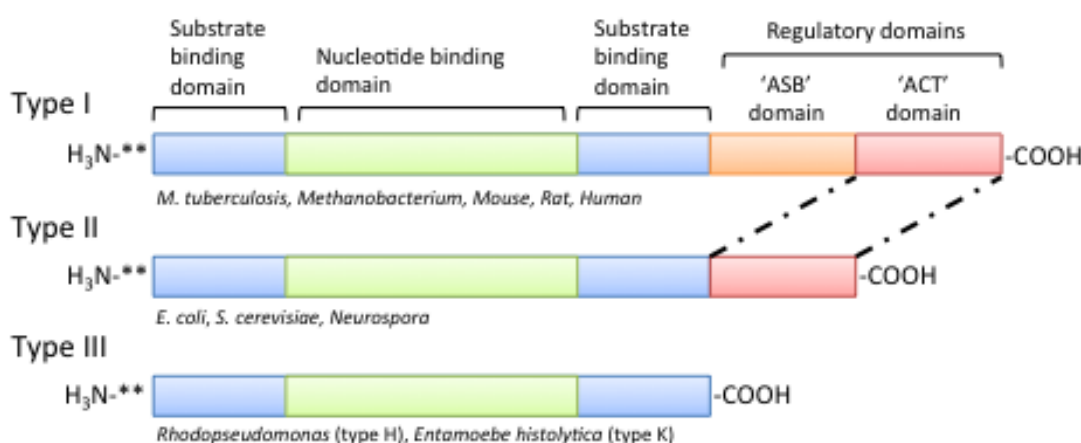
Figure 1.8 – Basic domain structure of PHGDH (adapted from Grant, 2012).

The figure shows the basic domain structure found within the three enzyme types of PHGDH coloured by domain. Additional amino acids at the N-terminus are not explicitly shown as variations in length and composition of this part of the protein depend on the species (indicated by **). Two forms of the type III enzyme exist depending on whether lysine (type K) or histidine (type H) is present at the active site[57].

All three types of PHGDH contain two common domains, the substrate-binding domain and the nucleotide-binding domain, both of which are necessary for substrate binding and catalysis [57]. The substrate-binding domain starts at the amino terminus of the polypeptide chain, and is then interrupted by the entire nucleotide-binding domain before being completed. Type III enzymes are composed of only these two domains, however, a variant of type III exists where the catalytic histidine residue is replaced by a lysine (type K). The group of type II enzymes contain an additional regulatory ACT domain at the C-terminus, which was reported to function as a binding site for serine in feedback inhibited enzymes, such as *E. coli* PHGDH. This

16

ACT domain was first discovered in *E. coli* and is composed of approximately 60-70 amino acids forming a βαββαβ structure [58]. Later, the ACT domain was recognized as a recurring motif that is mainly found in metabolic enzymes and transcriptional regulators involved in amino acid metabolism [59]. The ACT domain is also found in mammalian type I PHGDH, however the mammalian enzyme is no longer sensitive to serine due to mutations of key serine-binding residues [57]. Type I enzymes also contain an additional structural domain between the ACT domain and the substrate-binding domain, the allosteric substrate binding (ASB) domain. The first ASB domain was observed in the structure of *M. tuberculosis* PHGDH and was shown to comprise 150 amino acid residues that form an αβααββ motif [60]. In the type I enzymes from rat and *M. tuberculosis*, the ASB domain was found to constitute another serine binding site, being able to bind the substrate and induce substrate inhibition [61, 62]. Rat and human PHGDH share a high sequence homology of 94 % (Figure 1.9) [63]. Two mRNA transcripts of human PHGDH have been identified with the larger 2.1 kb transcript being the dominant transcript with high expression levels in prostate, testis, ovary, brain, liver, kidney, and pancreas. The 710 bp transcript, although also detected at low levels in most tissues where the larger transcript was expressed, is the main transcript in heart and the exclusive one in skeletal muscle [63].

```
SERA_HUMAN    1    MAFANLRKVLISDSLDPCCRKILQDGGLQVVEKQNLSKEELIAELQDCEGLIVRSATKVT    60
SERA_RAT      1    MAFANLRKILISDSLDPCCRKILQDGGLQVVEKQNLSKEELIAELQDCEGLIVRSATKVT    60
SERA_MYCTU    1    ---MSLPVVLIADKLAPSTVAALGDQ-VEVRWVDGPDRDKLLAAVPEADALLVRSATTVD    56
                     .*  :**:*.* *.    * *  ::*    :  .:::*:* : :.:.*:*****.*

SERA_HUMAN   61    ADVINAAEKLQVVGRAGTGVDNVDLEAATRKGILVMNTPNGNSLSAAELTCGMIMCLARQ   120
SERA_RAT     61    ADVINAAEKLQVVGRAGTGVDNVDLEAATRKGVLVMNTPNGNSLSAAELTCGMLMCLARQ   120
SERA_MYCTU   57    AEVLAAAPKLKIVARAGVGLDNVDVDAATARGVLVVNAPTSNIHSAAEHALALLLAASRQ   116
                   *:*: ** **:::*.***.*:****:;*** :*:*:*:*:.*  **** : .:::. :**

SERA_HUMAN  121    IPQATASMKDGKWERKKFMGTELNGKTLGILGLGRIGREVATRMQSFGMKTIGYDPIISP   180
SERA_RAT    121    IPQATASMKDGKWDRKKFMGTELNGKTLGILGLGRIGREVAARMQAFGMKTVGYDPIISP   180
SERA_MYCTU  117    IPAADASLREHTWKRSSFSGTEIFGKTVGVVGLGRIGQLVAQRIAAFGAYVVAYDPYVSP   176
                   ** * **::: .*.*..* ***: ***:*;:*****: ** *: :** .:.*** :**

SERA_HUMAN  181    EVSASFGVQQLPLEEIWPLCDFITVHTPLLPSTTGLLNDNTFAQCKKGVRVVNCARGGIV   240
SERA_RAT    181    EVAASFGVQQLPLEEIWPLCDFITVHTPLLPSTTGLLNDSTFAQCKKGVRVVNCARGGIV   240
SERA_MYCTU  177    ARAAQLGIELLSLDDLLARADFISVHLPKTPETAGLIDKEALAKTKPGVIIVNAARGGLV   236
                   :*.:*:: * *::: .***:** *  *.*:**::..::*: * ** :**.****:*

SERA_HUMAN  241    DEGALLRALQSGQCAGAALDVFTEEPPRDRALVDHENVISCPHLGASTKEAQSRCGEEIA   300
SERA_RAT    241    DEGALLRALQSGQCAGAALDVFTEEPPRDRALVDHENVISCPHLGASTKEAQSRCGEEIA   300
SERA_MYCTU  237    DEAALADAITGGHVRAAGLDVFATEPCTDSPLFELAQVVVTPHLGASTAEAQDRAGTDVA   296
                   **.** *: .*: .*.****:** **  *.: :*: ******* ***.*.* :*

SERA_HUMAN  301    VQFVDMVKGKSLTGVVNAQALTSAFSPHTKPWIGLAEALGTLMRAWAGSPKGTIQVITQG   360
SERA_RAT    301    VQFVDMVKGKSLTGVVNAQALTSAFSPHTKPWIGLAEALGTLMHAWAGSPKGTIQVVTQG   360
SERA_MYCTU  297    ESVRLALAGEFVPDAVNVG--GGVVNEEVAPWLDLVRKLGVLAGVLSDELPVSLSVQVRG   354
                    ..   : *:: : .**.  ....  .. **:*.. **.*  . :.    ::.* .:*

SERA_HUMAN  361    TSLKNAGNCLSPAVIVGLLKEASKQADVNLVNAKLLVKEAGLNVTTSHSPAAPGEQGFGE   420
SERA_RAT    361    TSLKNAGTCLSPAVIVGLLREASKQADVNLVNAKLLVKEAGLNVTTSHSPGVPGEQGIGE   420
SERA_MYCTU  355    ELAAEEVEVLRLSALRGLFSAV-IEDAVTFVNAPALAAERGVTAEICKASESPNHRSVV-   412
                   :      *   :.: **:  . : *.:*** *. * **:.. .::   *  .:.:

SERA_HUMAN  421    CLLAVALAGAPYQAVGLVQGTTPVLQGLNGAVFRPEVPLRR-----------DLPLLLFR   469
SERA_RAT    421    CLLTVALAGAPYQAVGLVQGTTPMLQMLNGAVFRPEVPLRR-----------GQPLLLFR   469
SERA_MYCTU  413    ---DVRAVGA-----------DGSVVTVSGTLYGPQLSQKIVQINGRHFDLRAQGINLII   458
                      * .**          : :.*::: *::  :        : *:

SERA_HUMAN  470    TQTSDPAMLPTMIGLLAEAGVRLLSYQTSLVSDGETWHVM-GIS---------SLLPSLE   519
SERA_RAT    470    AQPSDPVMLPTMIGLLAEAGVQLLSYQTSKVSDGDTWHVM-GLS---------SLLPSLD   519
SERA_MYCTU  459    HYVDRPGALGKIGTLLGTAGVNIQAAQLSEDAEGPGATILLRLDQDVPDDVRTAIAAAVD   518
                    . *  * .: **. ***.: : * *  ::*   :: :.      :: :::

SERA_HUMAN  520    AWKQHVTEAFQFHF                                                533
SERA_RAT    520    AWKQHVSEAFQFCF                                                533
SERA_MYCTU  519    AYKLEVVDLS----                                                528
                   *:* .* :
```

Figure 1.9 – Sequence alignment of human, rat and *M. tuberculosis* PHGDH.
Sequence alignment was performed in UniProt [64] and amino acids involved in nucleotide-binding are highlighted in green.

To date there has not been a published analysis of the structure of human PHGDH, although the structure of a truncated form (aa 3-314, sPHGDH) with malate and NAD$^+$ bound to the active site has been solved (PDB 2G76) (Figure 1.10). This truncated form lacks the regulatory ACT and ASB domains, thus comprising only the substrate and nucleotide-binding sites, and is present as a homodimer in the crystal asymmetric unit.

Figure 1.10 - Crystal structure of the truncated version of human sPHGDH (aa 3-314) and interactions of malate with sPHGDH

(A) Ribbon diagram of human PHGDH with two molecules present in the asymmetric unit (coloured by chain). The α-helices are shown as helical bands and the β-sheets are displayed as arrows with the tip pointing towards the C-terminus. For both chains the bound malate and NAD$^+$ molecules are shown by atom type with the carbons in green. (B) Amino acids of PHGDH involved in malate-binding. Black dashed lines represent H-bonds and salt bridges.

Crystallisation of human sPHGDH was performed in 0.1 M MMT buffer, which is composed of a 1:2:2 mixture of DL-malic acid: MES monohydrate: Tris base. As a result of the high *DL*-malate concentration in the buffer, binding of *D*-malate to the active site of the protein could be observed. In the crystal, malate, a substrate analogue of 3-PG, was shown to be favourably aligned for the catalytic reaction, including the formation of a salt bridge between the carboxylic group of malate and Arg235 of the enzyme (Figure 1.10 B) [57].

Despite the availability of the structure of the catalytic subunits of human PHGDH, the structure of full-length enzyme has not been solved to date. In fact, no atomic structure of any mammalian PHGDH is known, although purification of rat PHGDH indicated that the protein formed a tetramer on size exclusion chromatography [62]. *M. tuberculosis* PHGDH is thus the closest homologue for which the full-length structure has been solved. PHGDH from *M. tuberculosis* was shown to crystallise as a tetramer with the catalytic domains adopting the same orientation in all four subunits, whereas the regulatory domains adopt differing conformations (Figure 1.11) [60]. The catalytic subunit of PHGDH could not adopt the tetrameric form, most probably because it lacks the regulatory domains.
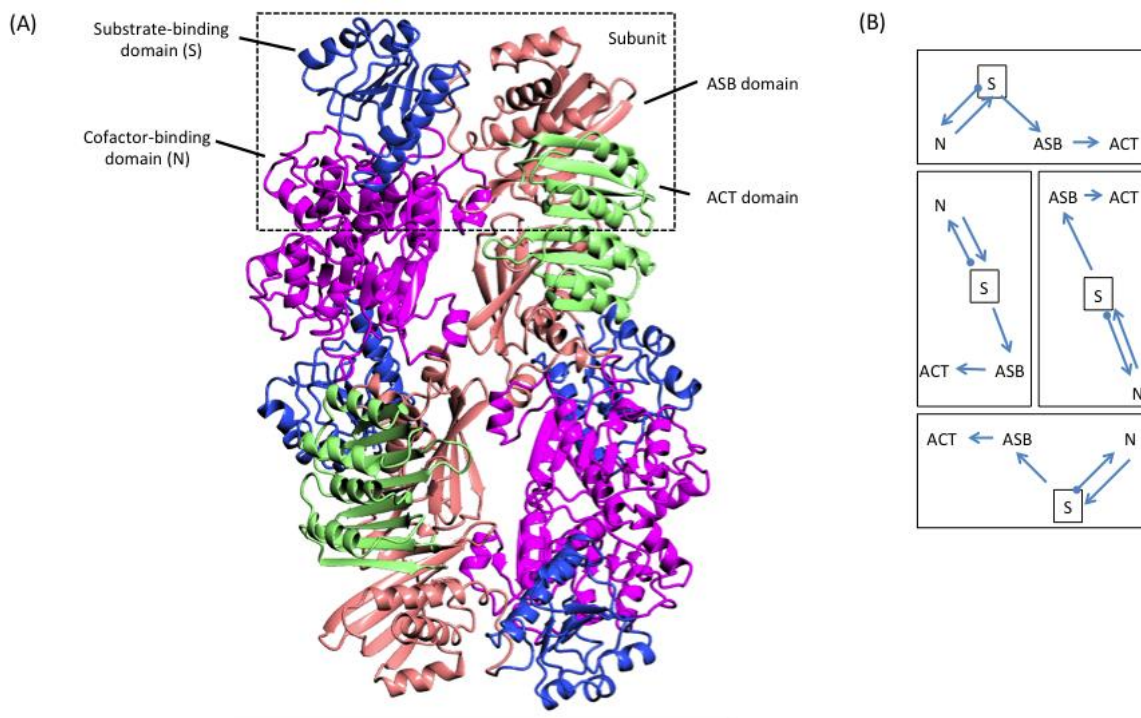
Figure 1.11 – Structure of active *M. tuberculosis* PHGDH tetramer

(A) The structure of *M. tuberculosis* PHGDH (PDB 1YGY) is coloured according to its different domains with the cofactor-binding domains in magenta, the substrate-binding domains in blue, the ASB domain in red and the ACT domain in green. The black square highlights one of the four subunits, which form the active enzyme. (B) Scheme of the arrangement of the different domains in each subunit. The arrows indicate the direction of the amino acid sequence starting with the N-terminus from the substrate-binding domain (Pictures adapted from Grant *et al.* [57]).

1.4.2 Catalytic mechanism

Under physiological conditions, the PHGDH-catalysed reaction proceeds in the direction of the oxidation of 3-PG because the final, irreversible step in L-serine biosynthesis draws the reaction in the direction of PHP synthesis. However, *in vitro* PHGDH activity has mainly been assessed by following reduction of PHP, since that reaction is readily promoted by exploiting the high affinity of PHGDH for NADH (K_m of human PHGDH for NADH is 20 µM) [57].

As the PHGDH-catalysed reaction uses both a substrate and a cofactor, different orders of binding of both ligands to the enzyme are possible. Although E. coli and M. tuberculosis PHGDH function both function *via* an ordered Bi-Bi mechanism, their respective orders of ligand binding are not consistent, with the substrate binding first to *M. tuberculosis* PHGDH, but the co-factor binding first to that of *E. coli* [57]. The catalytic mechanism for human PHGDH remains to be elucidated.

E. coli PHGDH, which belongs to the class of type II enzymes, can catalyse α-ketoglutarate (α-KG) reduction to *D*-2-hydroxyglutaric acid (D-2HG), whereas rat and *M. tuberculosis* PHGDH were reported not to catalyse the reaction with α-KG as substrate [57]. This led to the wider hypothesis that mammalian PHGDHs in general might not be able to use α-KG as a substrate. However, more recent literature reports that human PHGDH makes a significant contribution to the production of D-2HG in cells through reduction of α-KG (Figure 1.12). Fan and colleagues showed that knockdown of PHGDH reduced cellular D-2HG levels by 50 % in the breast cancer cell line MDA-MB-468 and overexpression of PHGDH in MDA-MB-231 cells yielded a two-fold increase in intracellular D-2HG levels [65]. In addition, oxaloacetate was shown to function as a substrate for recombinant purified human PHGDH yielding D-malate, which is in agreement with the reported co-crystal structure of the catalytic subunit of PHGDH with D-malate bound to its active site (PDB 2G76) [65].
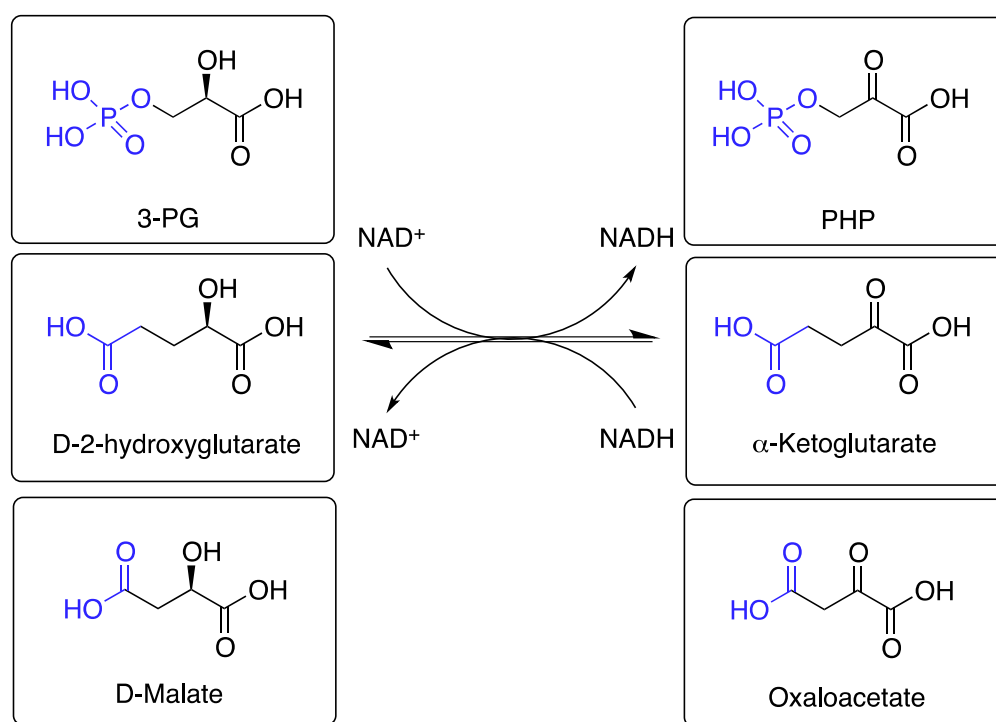
Figure 1.12 – Reactions catalysed by human PHGDH

Among the chemical structures of substrates and products, common sub-structures are coloured in black and differences are highlighted in blue.

1.4.3 PHGDH and cancer

With increasing evidence of altered metabolic pathways in cancer cells, the importance of understanding how cancer cells adapt their metabolic processes to drive the macromolecular biosynthesis necessary for proliferation and rapid cell growth becomes apparent. Early research pointed towards the relevance of the serine synthesis pathway in cancer, with work from Snell and colleagues reporting significantly enhanced PHGDH activity in rat hepatomas [66, 67]. PHGDH activity was increased in slow- and fast-growing hepatomas, and was therefore concluded to be relevant for tumour initiation and progression, although the increase in activity in fast-growing hepatomas of 50-75-fold was most striking [66]. Subsequent investigation of PHGDH activity in human colon carcinoma revealed a 10-fold increase compared to normal human colon mucosa, and, although only a small cohort of four patients was investigated, the importance of PHGDH activity in human neoplasia was shown [35]. The work described so far compared the enzymatic activity of PHGDH in healthy and tumorous tissue without investigation of any genetic alterations. It was not until 20 years later that a pooled analysis of somatic copy number alterations revealed *PHGDH*, the gene encoding for PHGDH, as a frequently amplified gene across 3131 cancer samples [68]. The *PHGDH* gene is located on chromosome 1p12, a region that displays repeated copy number gain in 16 % of the cancers assessed [69]. Among the cancers examined for *PHGDH* gain, the most common tumour type was melanoma, with 40 % frequency in *PHGDH* amplification [69]. Moreover, analysis of 106 human tumour samples by correlating protein levels with mRNA expression revealed that high PHGDH expression was associated with certain breast cancer subsets such as triple-negative breast cancer and basal subtypes [69]. Previous studies had already shown a correlation between high PHGDH mRNA levels and poor five-year survival in ER-negative basal breast cancers [70, 71]. This was confirmed by a study showing that 70 % of ER-negative breast tumours had increased levels of PHGDH mRNA and protein [72]. Examination of a panel of breast cancer cell lines revealed that PHGDH protein levels were also elevated in two ER-negative cell lines lacking the *PHGDH* copy number gain [72]. This indicates that, in addition to genomic amplifications of *PHGDH*, other mechanisms must exist to increase PHGDH protein levels.

Consistent with these genetic data, it was found that shRNA mediated knockdown of PHGDH in melanoma cell lines with *PHGDH* copy number gain resulted in decreased proliferation compared to the non-amplified cell lines [69]. Similar experiments using RNAi-mediated suppression of PHGDH in different cell lines also found decreased cell proliferation only in cells with elevated PHGDH expression, both with and without *PHGDH* amplification [72]. Importantly, PHGDH suppression was also shown to affect the growth of established mammary tumours in mice. When mice bearing MDA-MB-468 tumours with either a doxycycline-inducible PHGDH or off-target (GFP) shRNA were treated with doxycycline there was a remarkable and significant suppression of tumour growth in the PHGDH knockdown but not the off-target control [72]. Another study on cervical adenocarcinoma samples revealed high PHGDH expression in this cancer type [73]. Consistent with the earlier knockdown experiments, PHGDH knockdown in HeLa cells significantly inhibited cell proliferation and halted tumour progression *in vivo* and was additionally shown to potentiate the effects of cisplatin in both cells and xenografts [73]. Another strong indication of the importance of PHGDH in tumorigenesis was provided by Possemato and colleagues who showed that ectopic expression of PHGDH in non-tumorigenic MCF-10A cells induced phenotypic alterations. Among these alterations were an abnormal large nuclear morphology, failure to orient in a uniform fashion adjacent to the basal acinar membrane and enhanced proliferation [72].

Interestingly, PHGDH suppression with shRNA resulted in reduced proliferation of cells containing the PHGDH amplification even when growing in media with additional serine or cell-permeable methyl-serine-ester [72]. This finding suggests that the PHGDH-mediated serine synthesis pathway promotes cell proliferation independently of serine synthesis. Recently, human PHGDH was shown to catalyse the NADH-dependent reduction of α-ketoglutarate (αKG) to 2-hydroxyglutarate (2-HG) [65]. The oncometabolite 2-HG is a competitive inhibitor of α-KG-requiring enzymes and, as such, blocks the activity of for example DNA and histone demethylases [74]. Previously, increased 2-HG production was associated with the expression of mutant isocitrate dehydrogenase (IDH), however it has now been detected in human breast cancers with wild-type IDH [75]. Another proposed mechanism for the reduction in growth of certain tumours after PHGDH knockdown is through increased redox stress as PHGDH catalyses the oxidation of 3-PG in

combination with reduction of NAD$^+$ to NADH, thus affecting the NAD$^+$/NADH ratio and the cellular redox status [76]. In glioma cells, PHGDH was also found to have a crucial non-metabolic role through stabilising FOXM1, which is an oncogenic transcription regulator of for example cell cycle proteins, at the protein level [77].

1.5 Targeting cancer cell metabolism

As outlined previously, metabolic pathways could represent attractive therapeutic targets in cancer therapy due to the fact that cancer cell metabolism is altered in certain aspects compared to the metabolism of normal cells by virtue of the demands of highly proliferating malignant cells (Figure 1.13). Nevertheless, this strategy generates the challenges of achieving a therapeutic window and developing non-invasive methods for patient stratification, as not all tumours rely on the same metabolic pathways.
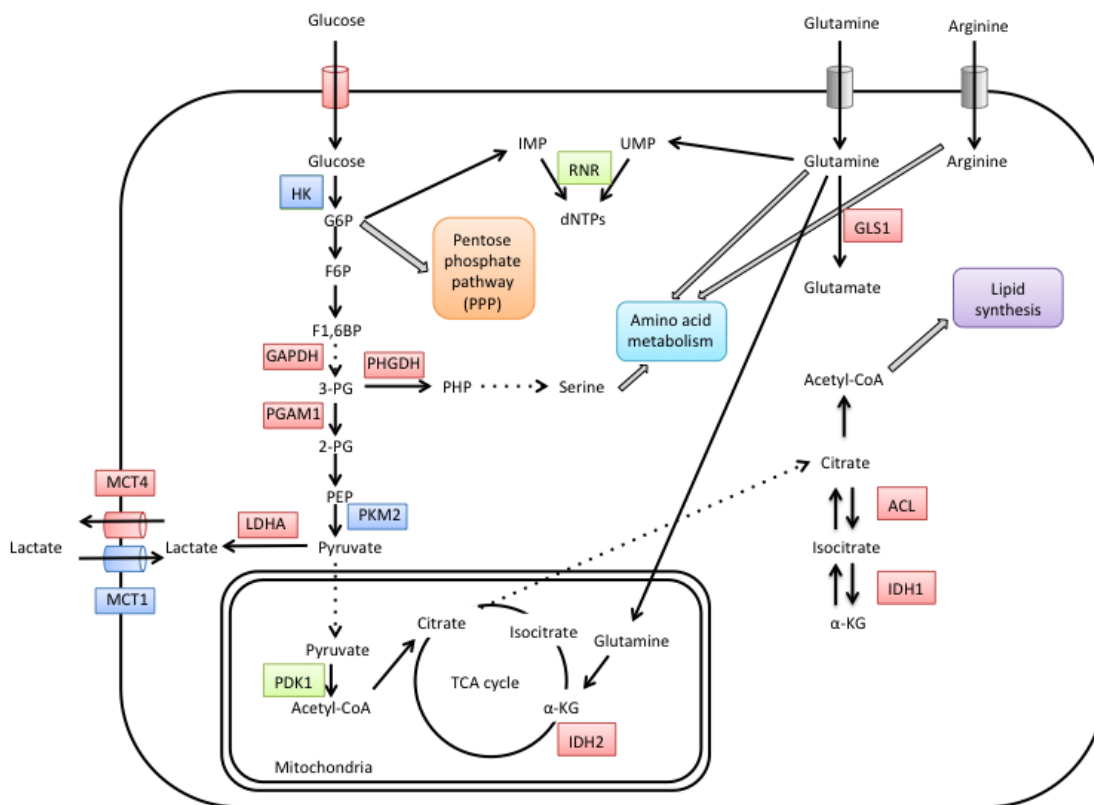
Figure 1.13 - Core metabolic pathways and potential metabolic targets in cancer

Metabolic enzymes which are under preclinical investigation are highlighted in red. Targets in early and more advanced clinical trials are highlighted in blue and green, respectively. (Figure adapted from [78, 79]).

In the late 1940's successful targeting of cancer cell metabolism was already achieved using the antifolate drug aminopterin to induce remission in children with acute lymphoblastic leukaemia [80]. Today antifolate drugs, such as the dihydrofolate reductase (DHFR) inhibitor methotrexate, are effectively used in the treatment of various cancers [81]. Anti-folate drugs were a starting point for the development of other metabolite analogues as anticancer agents that interfere with nucleotide synthesis, such as 5-fluorouracil, hydroxyurea or gemcitabine. In addition, there are a variety of drugs for the treatment of other metabolic diseases already on the market. Such drugs can now be reassessed for their therapeutic effect in cancer patients (Table 1.1). For example, the hexokinase II inhibitor metformin, which is on the market for the treatment of diabetes II, is now being assessed in several clinical trials for cancer therapy [82].

Compound	Target	Tumour / cancer cell type	Clinical stages
Glycolysis			
2-Deoxyglucose	Glucose transporter, Hexokinase 2	Prostate cancer	Phase I (terminated)
Silybin/silibinin	Glucose transporter	Prostate cancer	Phase I/II
Gossypol/AT-101	Lactate dehydrogenase A	Multiple cancers	Phase I/II
Nucleic acid synthesis			
Methotrexate	Folate metabolism (DHFR)	Multiple cancers	Approved agent
5-Fluorouracil	Thymidine synthesis	Multiple cancers	Approved agent
Hydroxyurea	Deoxynucleotide synthesis	Multiple cancers	Approved agent
Gemcitabine	Nucleotide incorporation	Multiple cancers	Approved agent
Amino acid metabolism			
L-Asparaginase	Asparagine	Acute lymphoblastic leukaemia	Approved agent
Arginine deaminase	Arginine	Multiple cancers	Phase II
NAD metabolism			
FK866/APO866	Nicotinamide phospho-ribosyltransferase	Cutaneous T-cell lymphoma, B-cell chronic lymphocytic leukemia, melanoma	Phase II

Table 1.1 - Therapeutic compounds that target metabolic enzymes with potential for use in cancer therapy (adapted from [21, 81]).

Based on the increased understanding of the Warburg effect and the importance of glucose metabolism for cancer cells, silybin, an inhibitor of glucose transporters was developed and is currently in clinical studies [83]. 2-deoxyglucose, by contrast, was established as a molecule that is able to enter cells *via* the glucose transporters but cannot be metabolised beyond 2-deoxyglucose-6-phosphate and therefore

accumulates, thus inhibiting glucose catabolism [21]. Despite concerns that the toxicity of 2-deoxyglucose (especially brain toxicity) might be a limiting issue in the use of this drug, a completed phase I clinical study reported a dose of 45 mg/kg as recommended for follow-up phase II trials [84].

Although clinical trials have significantly added to the understanding of cancer metabolism, not all compounds showed clinical benefit and often treatment outcomes were most efficacious in combination with radiation and/or chemotherapy [79]. However, increased benefit of anti-metabolic drugs is expected with a better understanding of patient stratification for this class of chemotherapeutics.

1.6 Targeting strategies for NAD$^+$-requiring enzymes

Although the multitude of NAD$^+$-requiring enzymes mirrors their importance in biochemical processes, it also represents a challenge when it comes to selective targeting of specific NAD$^+$-requiring enzymes. In particular, the idea of targeting the NAD$^+$-binding pocket seems an irrational approach as protein structural motifs that bind NAD$^+$/NADH are often very similar among related enzymes. In the case of dehydrogenases, a frequently-observed structural motif is the Rossmann fold, a βαβαβ subunit which provides a basic frame for binding nucleotides in general and coenzymes in particular [85]. With this strong structural homology among coenzyme binding pockets of NAD$^+$ requiring enzymes, it has been assumed that selectively targeting these enzymes with NAD$^+$-competitive therapeutics would be challenging and therefore rather lead to general toxicity [86]. Contradicting this hypothesis, the number of NAD$^+$ or nicotinamide analogues, which have been found to target certain NAD+-utilising enzymes with high specificity, is increasing.

1.6.1 Analogues of NAD$^+$

An early example of selective inhibition of an NAD$^+$-requiring enzyme was reported for horse liver alcohol dehydrogenase in 1993 [87]. It was found that of two isosteric analogues of NAD$^+$, one a C-nucleotide of nicotinamide (C-NAD) and the other picolinamide (C-PAD), C-NAD was an extremely potent inhibitor of alcohol dehydrogenase (K_i = 1.1 nM). In this study, C-PAD was found to be 20000 fold less potent (K_i = 20 μM) (Figure 1.14) [87]. Crystallographic analysis revealed, that in the C-NAD complex, the nitrogen of the pyridine ring replaced the alcohol as the fourth

coordination ligand to the active site zinc atom whereas such a zinc-nitrogen interaction is not observed in the complex with C-PAD [88].
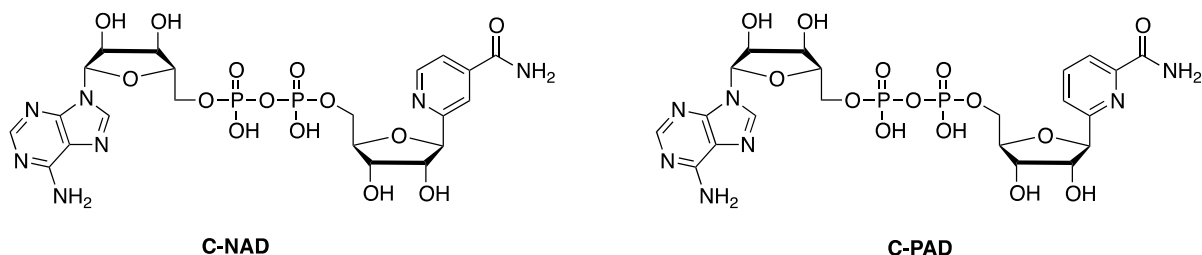
Figure 1.14 - Isosteric analogues of NAD: C-NAD and C-PAD.

Another example of selective inhibition of an NAD$^+$-requiring enzyme has been described for inosine monophosphate dehydrogenase (IMPDH), a key enzyme in the *de novo* GTP synthesis pathway [89]. IMPDH is selectively inhibited by the oncolytic C-nucleoside tiazofurin, which is metabolically activated by phosphorylation to give the 5'-monophosphate followed by conversion to the NAD analogue thiazole-4-carboxamide adenine dinucleotide (Figure 1.15) [89]. Tiazofurin was tested in phase I and II clinical studies and found to induce complete haematological remission in patients with acute nonlymphocytic leukaemia or chronic myelogenous leukemia in blast crisis [90]. In addition, another inhibitor of IMPDH has been identified: mycophenolic acid. Mycophenolic acid inhibits IMPDH with a K$_i$ of 7 nM (for the type 2 isoform of IMPDH) by mimicking the interaction of the nicotinamide mononucleotide moiety of NAD with the enzyme [86]. To date mycophenolic acid appears to be specific for IMPDH as no other NAD$^+$-requiring enzymes have been reported to be targets and the compound is used clinically as an immunosuppressant [86].
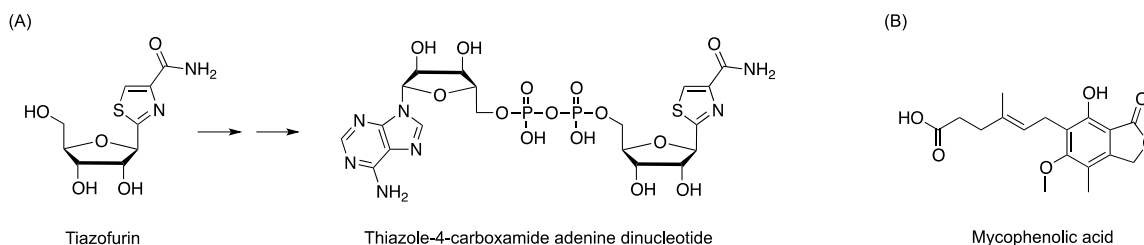
Figure 1.15 - Structures of IMPDH inhibitors, Tiazofurin and Mycophenolic acid.
Tiazofurin is metabolised intracellularly to thiazole-4-carboxamide adenine dinucleotide (TAD).

1.6.2 Nicotinamide analogues

The feasibility of targeting NAD^+-requiring enzymes was also demonstrated by poly(ADP-ribose) polymerase (PARP) inhibitors, which are based on the nicotinamide moiety of NAD^+. PARP-1 and 2 were shown to have a significant role in the recognition and repair of DNA damage, thus fulfilling a fundamental role in the maintenance of genomic stability. PARP inhibitors, originally designed to chemo- and radiosensitise by inhibiting repair of the DNA damage inflicted, have subsequently been shown to have remarkable activity in tumours with defective homologous recombination DNA repair.

Based on the structure of nicotinamide, a naturally-occurring PARP inhibitor, the first generation of PARP inhibitors were benzamides and 3-substituted benzamides, which are structural analogues of nicotinamide with the nitrogen in the aromatic ring being substituted by a carbon (Figure 1.16). Although these inhibitors had a somewhat higher potency, being 10-20 fold more potent than nicotinamide, they still lacked specificity and potency [91].

Nicotinamide   3-Aminobenzamide
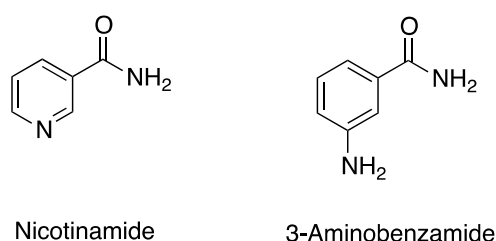
Figure 1.16 - First generation of PARP inhibitors.

A second generation of PARP inhibitors was developed in the 1990s when Banasik and colleagues used an "analogue by catalogue" approach to screen a panel of 170 compounds for their inhibitory effect against PARP-1. They found more potent inhibitors (4-amino-1,8-naphthalimide, 6(5*H*)- and 2-nitro-6(5*H*)-phenanthridinone and 1,5-dihydroxyisoquinoline) with IC_{50} values approximately two orders of magnitude lower than 3-aminobenzamide (Figure 1.17) [92]. Other groups have used structure activity relationships (SAR) to drive their drug discovery programmes. For example, Suto *et al.* performed further work on the dihydroisoquinolinone core leading to the compound PD128763, which has an IC_{50} value 50-fold lower than 3-aminobenzamide [93]. In parallel, two other structural classes, imidazole and benzoxazole carboxamides, were identified as potent PARP-1 inhibitors (Figure 1.17)

[94]. In these compounds the imidazole nitrogen functions as an intramolecular acceptor of the amide hydrogen giving rise to PARP inhibitors with IC_{50} values in the nM range [91].
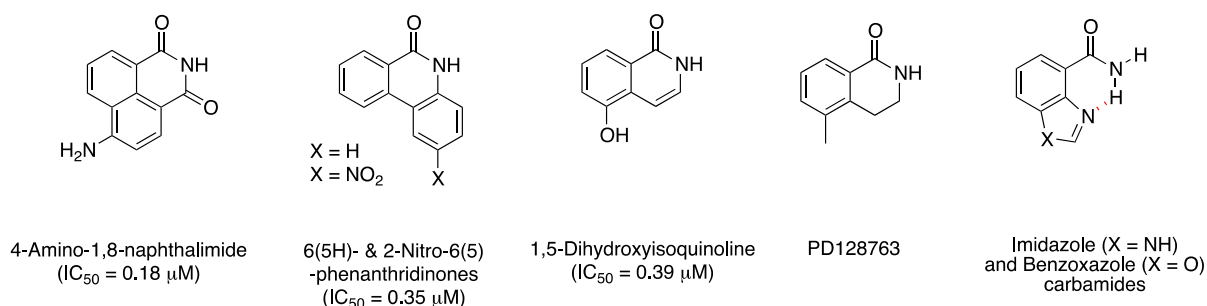
Figure 1.17 - Second generation of PARP inhibitors.

Built on these findings, further PARP-1/2 inhibitors have been developed using both SAR and structure-based drug design, and several are currently undergoing clinical trials. Among the most advanced compounds in clinical studies are AG014699 (Rucaparib), Veliparib and Olaparib, which was licenced for the treatment of ovarian cancer in 2015 (Figure 1.18) [95].
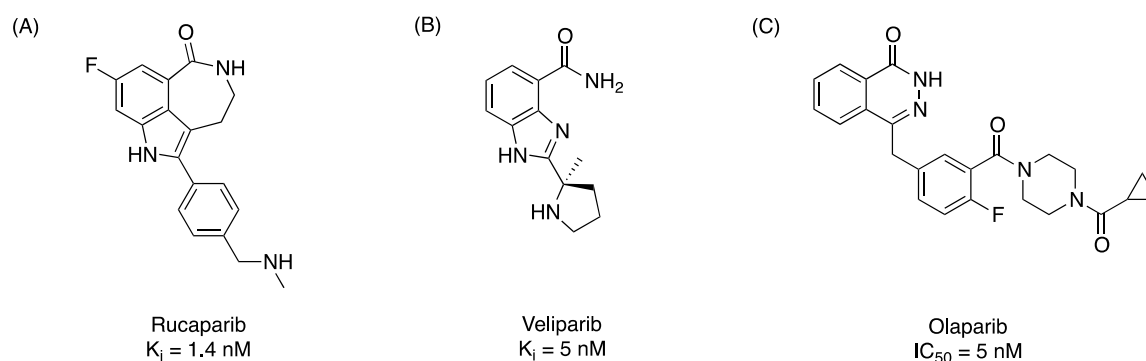
Figure 1.18 - Potent PARP inhibitors in clinical trials.

1.6.3 Bisubstrate analogue inhibitors

One strategy to increase potency and specificity in enzyme inhibition is to target more than one subpocket of the enzyme using so-called bisubstrate or multisubstrate analogue inhibitors. A high degree of specificity is expected from these compounds as they combine features of two substrates (e.g. substrate and coenzyme) required by the target enzyme and should therefore not target enzymes that only use one of these substrates [96]. In addition, bisubstrate analogue inhibitors should also be

30

favourable from a thermodynamic point of view; as such a compound has the theoretical potential to have a binding free energy equal to the sum of the binding energies of both substrates [96], with a possible further contribution from reduced entropy. To date, this concept has mainly been used to design structural probes to elucidate enzymatic mechanism as in the case of isocitrate dehydrogenase. In this approach, bisubstrate inhibitors for isocitrate dehydrogenase were obtained by covalently linking 2-oxoglutarate with NAD$^+$ and NADP$^+$. These were shown to be able to distinguish between the two isoforms of pig heart mitochondrial isocitrate dehydrogenase where one isoform is known to use NAD$^+$ as coenzyme and the other NADP$^+$ (Figure 1.19) [97].
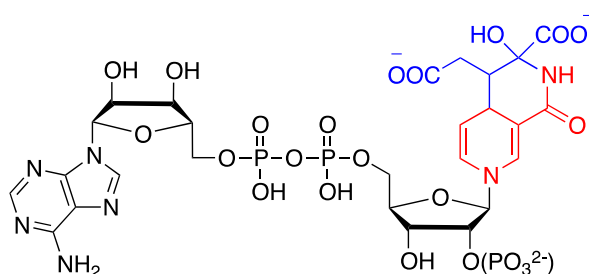
NAD(P)-oxoglutarate adduct

Figure 1.19 - Structure of bisubstrate inhibitor for isocitrate dehydrogenase.
The nicotinamide portion is represented in red and the oxoglutarate moiety in blue.

The concept of bisubstrate inhibitors has also led to potent therapeutic compounds, such as Mupirocin, a femtomolar inhibitor of bacterial isoleucyl-tRNA synthetase widely used as a topical antibiotic [98].

1.7 Aims and Objectives

Increasing evidence for particular differences in the metabolism of cancer cells compared to normal cells are reported, indicating the importance of metabolic targets in cancer. However, to date antimetabolites are mostly targeted at the nucleic acid metabolism. A potential novel area for therapeutic targeting is represented by upregulated anaerobic glycolysis in cancer cells. In this context, PHGDH, which utilises the glycolytic metabolite 3-PG, has been proposed as a potential target in cancer therapy. PHGDH, a member of the NAD^+-requiring dehydrogenase family, might be difficult to specifically target due to similarities within its protein family. However, other protein targets have been shown to overcome specificity issues as seen with inhibitors of the NAD^+-consuming PARPs and also dehydrogenase inhibitors. In the particular case of dehydrogenases, specific inhibition was shown to improve using bi-substrate mimics.

The main aims of the investigations reported here were:

1. To further explore PHGDH as a target in different cancer types and validate the target through knockdown experiments.

2. To design potential PHGDH inhibitors based on a bisubstrate approach through investigation of the substrate-binding site.

3. To investigate the cofactor-binding site of PHGDH with NAD^+ analogues.

4. To develop site-directed mutants of PHGDH to reduce binding affinity for the cofactor.

5. To screen for PHGDH inhibitors based on a fragment library and develop a FBDD cascade for PHGDH.

Chapter 2. Materials and Methods

2.1 Chemicals and Reagents

All chemicals listed in the biological materials and methods section were of analytical grade and obtained from Sigma Aldrich, St. Louis, MO, USA, unless otherwise stated. Chemicals used for compound synthesis are described in chapter 9.

2.2 Purification buffers

Extraction buffer I   50 mM NaH_2PO_4, pH 8, 10 mM Imidazole, 300 mM NaCl, 2 mM DTT or 0.5 mM TCEP, 0.01 mg/mL DNase, 0.05 mg/mL RNase, 0.25 mg/mL Lysozyme, 5 mM $MgCl_2$, 1 EDTA-free complete^{TM} protease inhibitor tablet (Roche, Basel, Switzerland)/ 50 mL

Extraction buffer II   50 mM NaH_2PO_4, pH 8, 25 mM Imidazole, 150 mM NaCl, 10 % (v/v) glycerol, 0.01 mg/mL DNase, 0.05 mg/mL RNase, 0.25 mg/mL Lysozyme, 5 mM $MgCl_2$, 1 EDTA-free complete^{TM} protease inhibitor tablet (Roche, Basel, Switzerland)/ 50 mL

Binding buffer I   50 mM NaH_2PO_4, pH 8, 10 mM Imidazole, 300 mM NaCl, 2 mM DTT or 0.5 mM TCEP

Binding buffer II   50 mM NaH_2PO_4, pH 8, 25 mM Imidazole, 150 mM NaCl, 10 % (v/v) glycerol

Elution buffer I   50 mM NaH_2PO_4, pH 8, 250 mM Imidazole, 300 mM NaCl, 2 mM DTT or 0.5 mM TCEP

Elution buffer II   50 mM NaH_2PO_4, pH 8, 800 mM Imidazole, 150 mM NaCl, 10 % (v/v) glycerol

SEC buffer   25 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM DTT or 0.5 mM TCEP

Low salt buffer   20 mM Tris, pH 8, 40 mM NaCl, 2 mM DTT or 0.5 mM TCEP

High salt buffer   20 mM Tris, pH 8, 40 mM NaCl, 2 mM DTT or 0.5 mM TCEP

2.3 Transformation of competent cells

Transformation, which is the process of introducing foreign DNA into a host cell, was performed using the two chemically competent *E. coli* strains DH5α (Life

Technologies, Carlsbad, CA, USA) and Rosetta (DE3) (Merck, Darmstadt, Germany). The DH5α strain was used for the production of high quality DNA for sequencing and Rosetta (DE3) was used as an expression strain for large-scale protein production.

For the transformation, 1 µL purified plasmid DNA (approximately 20 ng/µL) was added to 20-50 µL competent cells and incubated on ice for 5 – 30 minutes. Cells underwent heat shock for 30 seconds at 42 °C, followed by incubation on ice for 15-30 min. After addition of 80 – 180 µL of prewarmed LB or SOC media (Sigma Aldrich, St Louis, MO, USA) the cells were incubated at 37 °C for 1 h while shaking at 250 rpm. 25 – 80 µL of cells were plated onto LB agar plates containing 50 µg/mL kanamycin. Plates were incubated for 18 h at 37 °C and single colonies used to inoculate aliquots of 5 mL LB medium containing 50 µg/mL kanamycin. These cultures were used to prepare glycerol stocks (15 % (v/v) glycerol) as well as to prepare vector DNA stocks (for DH5α only).

2.4 DNA sequencing

Plasmid DNA was extracted from the 5 mL overnight culture using a Spin Miniprep Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. The integrity of all DNA constructs was verified by DNA sequencing. DNA sequencing was carried out by Source Bioscience (Nottingham, UK) in forward and reverse directions using universal T7 primers.

2.5 Protein expression

2.5.1 PHGDH, sPHGDH and site-directed mutants thereof

The pNIC28-Bsa4 plasmids containing the human PHGDH cDNA encoding for the full-length enzyme (aa 1-533) as well as a truncated version (sPHGDH, aa 3-314) were kindly donated by Wyatt Yue (SGC Oxford, Oxford, UK). Both plasmids are fused with an N-terminal His$_6$-tag and a Tobacco Etch Virus (TEV) protease recognition site (Figure 2.1). For large-scale protein production, the proteins were expressed in *E. coli* Rosetta (DE3) grown in TB medium (Sigma Aldrich, St. Louis, MO, USA) containing 50 µg/mL kanamycin.

Figure 2.1, pNIC28-Bsa4 vector map and amino acid sequence of PHGDH.

(A) pNIC28-Bsa4 vector map highlighting the T7 promoter and terminator regions as well as different restriction sites. The plasmid contains a kanamycin resistance gene (kanR) and a SacB gene allowing for antibiotic selection and negative selection on 5 % sucrose, respectively. Picture adapted from vector information sheet provided by SGC Oxford [99]. (B) Amino acid sequence of PHGDH with His$_6$-tag and TEV cleavage site highlighted in red and blue, respectively. Amino acid sequence of sPHGDH is highlighted in bold.

Glycerol stocks were used to inoculate 5 mL aliquots of TB medium containing 50 µg/mL kanamycin. Starter cultures were grown overnight at 37 °C and 1 mL of the

overnight culture used to inoculate 1 L TB medium containing 50 µg/mL kanamycin. 1 L cultures were grown at 37 °C with shaking (200 rpm) until an optical density at λ = 600 nm (OD_{600}) of 0.8 was reached. Then, the temperature was lowered to 25 °C and cells further incubated for one hour to allow for the temperature to equilibrate. Protein expression was induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cells were grown overnight before harvesting by centrifugation at 4000 rpm, 4 °C for 15 minutes (Beckman Avanti J-26, JLA 8.100 rotor). Cell pellets were stored at -80 °C until further use.

2.5.2 TEV protease

E. coli Rosetta (DE3) cells containing the pRK793 plasmid with cDNA encoding for TEV protease fused with an N-terminal His_6-tag were kindly donated by Vincent Rao (Newcastle University, Newcastle-upon-Tyne, UK). For large-scale protein production, a starter culture was prepared in 5 mL LB medium containing 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. Cells were grown at 37 °C overnight and used to inoculate 1 L LB medium containing 100 µg/mL ampicillin and 30 µg/mL chloramphenicol. Large-scale bacterial growth was performed at 37 °C with shaking (200 rpm) until an OD_{600} of > 0.5 was reached. The temperature was lowered to 25 °C and cells further incubated for 30 minutes to allow for the temperature to equilibrate. Protein expression was induced with IPTG at a final concentration of 1 mM, and cells incubated for another 16 hours. Cells were harvested by centrifugation at 4000 rpm, 4 °C for 15 minutes (Beckman Avanti J-26, JLA 8.100 rotor), and cell pellets stored at -80 °C until further use.

2.6 Protein purification

2.6.1 General procedure for PHGDH, sPHGDH and site-directed mutants thereof, and construct 93

Cell pellets were resuspended in extraction buffer I and the cells lysed by sonication on ice for 3.5 min (intervals of 20 sec sonication and 40 sec rest). The lysate was cleared by centrifugation (20000 rpm, 4 °C, 1 h, Beckman Avanti J-26, JA 25.50 rotor) and the supernatant filtered through 0.45 µm serum acrodiscs (Sartorius AG, Göttingen, Germany) before applying to a HisTrap Ni-Sepharose column (GE Healthcare, Little Chalfont, UK) equilibrated in binding buffer I. Protein was eluted

with a linear 10 column volume (cv) gradient up to a maximum concentration of imidazole of 250 mM (elution buffer II). To remove the His$_6$-tag from the protein, His$_6$-tagged TEV protease was added at a mass ratio of 1:25 protease: protein and the sample incubated at 4 °C overnight. The sample was applied to a gravity flow column containing Ni-Sepharose beads (GE Healthcare, Little Chalfont, UK) and the protein samples collected from the flow-through, leaving uncleaved protein and tagged TEV protease bound to the column. The protein was concentrated using Vivaspin concentrators with 10 – 30 kDa molecular weight cut-off (VivaProducts, Littleton, MA, USA). The concentrated protein fraction was further purified by size exclusion chromatography on a Superdex^{TM} 75 26/60 gel filtration column (GE Healthcare, Little Chalfont, UK) in SEC buffer. The protein containing fractions were combined, concentrated and buffer exchanged into low salt buffer using a PD-10 column (GE Healthcare, Little Chalfont, UK). The protein was further purified by anion exchange chromatography using a linear gradient up to 1 M NaCl (high salt buffer). The fractions were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and those containing the protein of interest were pooled and concentrated to about 15 mg/mL. Protein concentration was determined using the extinction coefficients and calculated molecular mass as determined by ProtParam [100].

2.6.2 General procedure for TEV protease

Cell pellets were resuspended in extraction buffer II and the cells lysed by sonication on ice for 3.5 min (intervals of 20 sec sonication and 40 sec rest). The lysate was cleared by centrifugation at 20000 rpm, 4 °C for 1 h (Beckman Avanti J-26, JA 25.50 rotor) and the supernatant filtered through 0.45 μm serum acrodiscs (Sartorius AG, Göttingen, Germany) before applying to a HisTrap Ni-Sepharose column (GE Healthcare, Little Chalfont, UK) equilibrated in binding buffer II. Protein was eluted with a linear 20 cv gradient up to a maximum concentration of imidazole of 800 mM using binding and elution buffer II. Fractions containing TEV protease were determined according to the UV trace (280 nm) and SDS-PAGE. DTT was added to TEV protease-containing fractions to a final concentration of 1 mM. Fractions were pooled and the concentration determined using a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA) in combination with the specific extinction coefficient,

31970 $M^{-1}cm^{-1}$, and the molecular weight, 28617 Da, as calculated by ProtParam [100]. The TEV protease was stored in aliquots at -80 °C.

2.7 SDS-PAGE

SDS-PAGE is a technique to separate proteins according to their molecular weight by loading the proteins onto a gel and apply an electric field. SDS is an anionic detergent, which linearizes the proteins and gives them a negative charge. The linearized proteins move through the gel along the electric field according to their size with smaller proteins moving faster towards the anode compared to larger proteins.

Protein samples were mixed in a 3:1 ratio with RunBlue LDS Sample buffer (Expedeon, San Diego, CA, USA) and denatured through heating at 95 °C for 5 minutes. Samples are loaded onto SDS Precast gels (Expedeon, San Diego, CA, USA) with at least one well being loaded with PageRuler prestained protein ladder (Thermo Fisher Scientific, Waltham, MA, USA) to allow for approximate localisation of the protein bands by weight. Gels were run in RunBlue RAPID Running buffer (Expedeon, San Diego, CA, USA) at 150 V for 60 minutes. Protein bands were visualised by staining the gel with the Coomassie-based gel stain InstantBlue (Expedeon, San Diego, CA, USA).

2.8 In vitro PHGDH activity assay

The enzymatic activity of PHGDH or its catalytic domain (sPHGDH) was measured in the direction of 3-PG oxidation to PHP by coupling with a resazurin reduction reaction to allow fluorescence detection (Figure 2.2). The red fluorescent product, resorufin can be detected by monitoring emission at 580 nm upon excitation at 540 nm.
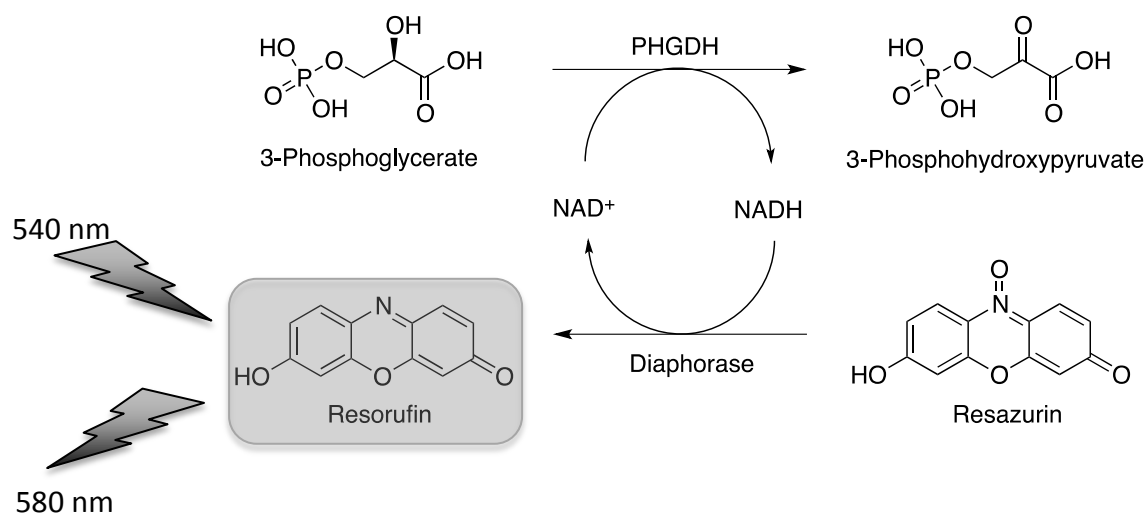
Figure 2.2 - Scheme of the reaction catalysed by PHGDH.

PHGDH converts 3-PG to PHP with concomitant reduction of the cofactor NAD+ to NADH. The reaction is coupled to a second enzymatic reaction in which diaphorase catalyses the conversion of resazurin to resorufin and concomitantly regenerates NAD+ by oxidation of NADH. The red fluorescent dye, resorufin, is measured by monitoring the emission at 580 nm upon excitation at 540 nm.

The assay mixture contained 18 mM BIS TRIS propane, pH 7.5, 1.7 mM EDTA, pH 7.5, 10 µM Triton X-100, 3.3 mM glutathione, 140 µM 3-PG (Santa Cruz Biotechnology, Dallas, Texas, USA), 25 µM NAD$^+$ and 75 nM purified human PHGDH or sPHGDH. The assay was performed in black 96-well plates (Sterilin Ltd, Gwent, UK) in 150 µL total volume of which 25 µL was the NAD$^+$ recycling mixture containing the diaphorase and resazurin (Amplite Fluorimetric assay buffer, Stratech Scientific Ltd., Suffolk, UK). The fluorescence was measured on an Omega Platereader (BMG Labtech, Ortenberg, Germany) over 40 – 50 minutes. Analysis was performed in GraphPad Prism by determining the initial velocity from the slope of increase in fluorescence over time. The initial velocity was corrected for baseline increase in fluorescence in the absence of protein.

2.9 Differential scanning fluorimetry

Differential scanning fluorimetry is used to monitor protein unfolding with increasing temperature and in the presence of a fluorescent dye, e.g. Sypro Orange (Figure 2.3).

The assay allows for the quantification of protein stability by derivation of the protein's melting temperature (T_m), which is the temperature at which half of the protein is unfolded. In the presence of a bound ligand, the protein stability and

therefore the T_m is increased due to the contribution of free energy of binding from the ligand.
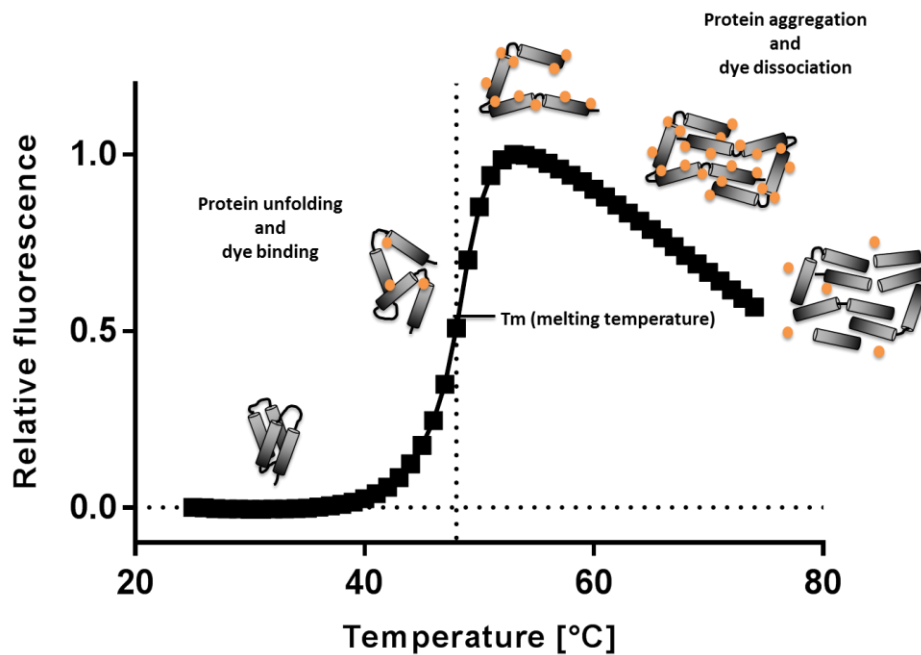
Figure 2.3 - Typical fluorescence emission curve for thermally induced protein unfolding in the presence of a fluorescent dye.

Protein and dye are mixed and excited at 490 nm and emission detected at 600 nm when using SYPRO orange as fluorescent dye. As the temperature is increased the protein unfolds and exposes hydrophobic patches to which the dye binds resulting in increased fluorescence emission. After all the protein has unfolded, the observed fluorescence decreases due to protein aggregation and dissociation of the dye from the protein. (Picture adapted from [101].)

Unless otherwise stated the assay mixture contained 25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 10 x Sypro Orange (Thermo Fisher Scientific, Waltham, MA, USA), 2 % DMSO and 15 µM protein. The final assay volume was 15 µL (384-well plate) or 20 µL (96-well plate). The heat denaturation curves were recorded using an RT-PCR instrument (ViiA7, Applied Biosystems, Warrington, UK). The temperature was increased by 3 °C/min starting at 25 °C up to 95 °C and the fluorescence of Sypro Orange measured with excitation and emission wavelengths of 470 and 570 nm, respectively. Analysis was performed using the Boltzmann equation in GraphPad Prism.

2.10 Isothermal titration calorimetry

Isothermal titration calorimetry measures enthalpy changes arising from interactions between molecules, e.g. protein-ligand or protein-protein binding, through

comparison of the temperature in a reaction cell to that of a reference cell. In addition to changes in enthalpy (ΔH) other thermodynamic parameters of the binding reaction can be obtained, such as the equilibrium constant (K_a), changes in entropy (ΔS) and the stoichiometry of the interaction (n).

For the determination of the thermodynamic parameters the calorimetric data is fitted by nonlinear least squares curve fitting using the Wiseman isotherm (Equation 2.1) [102].

$$\frac{dQ}{d[X]_t} = \Delta H^\circ V_0 \left[\frac{1}{2} + \frac{1 - X_R - r}{2\sqrt{(1 + X_R + r)^2 - 4X_R}} \right] with\ r = \frac{1}{K_a[M]_t}$$

Equation 2.1 - Wiseman isotherm for ligand – receptor binding at 1:1 stoichiometry.

Abbreviations: X = Ligand, M = Receptor, $[M]_t$ = total receptor concentration, $dQ/d[X]_t$ = change in heat/ moles of ligand added, ΔH = change in enthalpy, V_0 = effective volume of calorimetric cell, X_R = $[X]_t/[M]_t$ = ratio of ligand to receptor concentration, K_a = binding constant.

With the Wiseman isotherm, changes in enthalpy (ΔH), the binding constant and the stoichiometry of the reaction can be determined. To obtain the enthropic contribution to the binding, the experimental data is combined with the Gibbs-Helmholtz equation (Equation 2.2).

$$\Delta G = -RTlnK_a = \Delta H - T\Delta S$$

Equation 2.2 – Gibbs-Helmholtz equation.

Abbreviations: G = Gibbs free energy, T = temperature, R = ideal gas constant, K_a = binding constant, H = enthalpy, S = entropy

The relation of the measured equilibrium binding constant K_a and the total receptor concentration, which also defines the shape of the binding isotherm, is summarised in the Wiseman c-parameter (Equation 2.3).

$$c = n[M]_t K_a$$

Equation 2.3 – Equation for the Wiseman c-parameter.

Abbreviations: n = number of binding sites per receptor, $[M]_t$ = total receptor concentration, K_a = binding constant.

ITC experiments were performed at 25 °C in 25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP with the protein in the sample cell and the ligand in the syringe using a

MicroCal iTC200 (GE Healthcare, Little Chalfont, UK). In general, 1 x 0.5 µL, followed by 17 x 2 µL injections of ligand into the protein with 120 seconds spacing between the injections were conducted. Data was corrected for the heat of dilution by subtracting ligand: buffer titration heats. All thermodynamic parameters of the reaction were determined following peak integration and best-fit of data to the one-site binding model using the ORIGIN software, version 7.0.

2.11 Fluorescence polarisation assay

In a fluorescence polarisation experiment, the amount of polarised light emitted from a fluorophore (e.g. NADH), which itself is excited with polarised light, is measured. The degree of polarisation of the fluorophore is related to its rotation such that for a small fluorophore, which is rapidly rotating most of the emitted light is not polarised. When the rotation of the fluorophore is hindered through binding to a larger and therefore more slowly tumbling molecule such as a protein, the amount of polarised light is increased. When an inhibitor binds to the same binding site on the protein as the fluorophore, the fluorophore is displaced resulting in a decrease in the amount of polarised light.

Fluorescence polarisation assays were carried out in black 96 well microplates (Sterilin Ltd, Gwent, UK) with a final assay volume of 60 µL per well. The assay mixture contained 40 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP or 2 mM DTT, 1 µM NADH and 2 µM PHGDH or sPHGDH. Compounds were screened at 0.5 and 1 mM in 5 % (v/v) DMSO by adding the compounds at a ratio of 1:3 to the assay mixture from appropriate stock solutions in 15 % (v/v) DMSO. Samples were incubated for 30 min at RT and fluorescence polarisation was measured using a PHERAstar FS plate reader (BMG Labtech, Ortenberg, Germany) with an FP filter module (λ_{ex} = 340 nm and λ_{em} = 440 nm). % Inhibition was calculated using NADH only in FP buffer as positive control (100 % inhibition) and NADH plus protein in FP buffer as negative control (0 % inhibition) (Equation 2.4).

$$\% \ inhibition = \left(\frac{FP_{c-} - FP_{sample}}{FP_{c-} - FP_{c+}} \right) \times 100$$

Equation 2.4 – Calculation of % protein inhibition.

FP stands for fluorescence polarisation signal of positive (c+) and negative (c-) control.

The robustness of the assay was evaluated by determining the Z'-factor which is defined as the ratio of the separation band to the signal dynamic range of the assay (Equation 2.5) [103].

$$Z' = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|}$$

Equation 2.5 – Calculation of Z'-factor.

Standard deviation (σ) of positive (c+) and negative (c-) control, mean values (μ) from positive (c+) and negative (c-) control.

2.12 Crystallography

Initial sitting drop vapour diffusion crystallisation trials were set up against commercially available sparse matrix screens: Index (Hampton Research, Aliso Viejo, CA, USA), JCSG+, Structure, Proplex and Morpheus (all from Molecular Dimensions, Newmarket, Suffolk, UK) in 96-well MRC crystallisation plates (Molecular Dimensions, Newmarket, Suffolk, UK). Plates were set up using a Mosquito LCP liquid handler (TTP Labtech, Melbourn, UK) with two drops of 300 nL protein (15 mg/mL) mixed with 300 or 600 nL precipitant and a shared reservoir solution of 70 µL precipitant. Crystal trials were stored at 4 or 20 °C.

Deep-well blocks of optimisation screens were prepared using a Biomek (Beckman Coulter, Brea, CA, USA) and drops pipetted as described above. The precipitant was optimised by varying the buffer pH as well as the concentrations of the different chemicals in the condition (salt, PEG).

2.13 Data collection and analysis

Crystals, for which datasets were collected and presented in this work, were grown in a variety of conditions (Table 2.1).

Name	Protein construct	Ligand	Growth condition
sPHGDH 01	aa 3-314	NAD^+, α-ketoglutarate	0.1 M BIS-TRIS, pH 6.5, 0.2 M NH_4Ac, 27 % (w/v) PEG3350
sPHGDH 02	aa 3-314	NAD^+, *L*-tartrate	0.1 M TRIS, pH 8.5, 0.2 M $MgCl_2$ hexahydrate, 20 % (w/v) PEG3350
sPHGDH 03	aa 3-314	ADPR	0.1 M BIS-TRIS, pH 6, 0.3 M NH_4Ac, 23 % (w/v) PEG3350
sPHGDH 04	aa 3-314	TAD	0.1 M BIS-TRIS, pH 6, 21 % (w/v) PEG3350
sPHGDH 05	aa 3-314	AMP	0.1 M Sodium HEPES/ MOPS buffer, pH 7.5, 0.12 M monosaccharides, 30 % (w/v) PEG550-MME and PEG20K
sPHGDH 06	aa 3-314	1:1000 chymotrypsin	0.1 M HEPES, pH 7.5, 0.02 M $MgCl_2$, hexahydrate, 22 % (w/v) Poly(acrylic acid sodium salt) 5,100
sPHGDH 07	aa 3-314	1:1000 chymotrypsin, fragment E	0.2 M Sodium formate, 20 % (w/v) PEG3350
93-10	aa 93-298	NAD^+	0.1 M TRIS, pH 8, 0.2 M NH_4Cl, 20 % PEG6000
93	aa 93-298	apo-protein for soaking experiments	0.1 M PCTP buffer, pH 7, 22-28 % PEG1500

Table 2.1 – Crystal growth conditions

Crystals were harvested using CryoLoops (Hampton Research, Aliso Viejo, CA, USA) matched in size to the crystal and cryo-protected in reservoir solution supplemented with 20 % (w/v) PEG400 or ethylene glycol before being flash cooled in liquid nitrogen. For crystals grown in the presence of ligand or used for ligand soaking the cryo-condition additionally contained the ligand at the concentration stated in the relevant section.

Data were processed using xia2, an automated combination of XDS [104], XSCALE [104] and programs from the CCP4 suite [105] and used for molecular replacement with PHASER [106] as implemented within CCP4i2 suite [105], using the published structure of human PHGDH (PDB 2G76) as search model. The structure was refined using cycles of manual correction in COOT [107] followed by refinement with Refmac5 [108] within CCP4i2 [105]. Models were validated using the validation available in COOT [107], e.g. checking for geometry and Ramachandran outliers, and MolProbity [109]. Figures were produced using CCP4mg [110].

2.14 Statistics

Statistical analyses were performed in GraphPad Prism (version 6) using the appropriate analysis parameters. $* = < 0.1$, $** = < 0.01$, $*** = < 0.001$, $**** = < 0.0001$ and ns = not significant.

Chapter 3. Target validation

3.1 Introduction

The drug discovery pipeline is a complex process with huge costs estimated to amount to be over $ 1 billion for the development of a launched drug [111]. The development process of a drug can be broken down into the four stages of target selection, hit discovery, candidate identification and subsequent clinical trials with the candidate compound (Figure 3.1) [112].
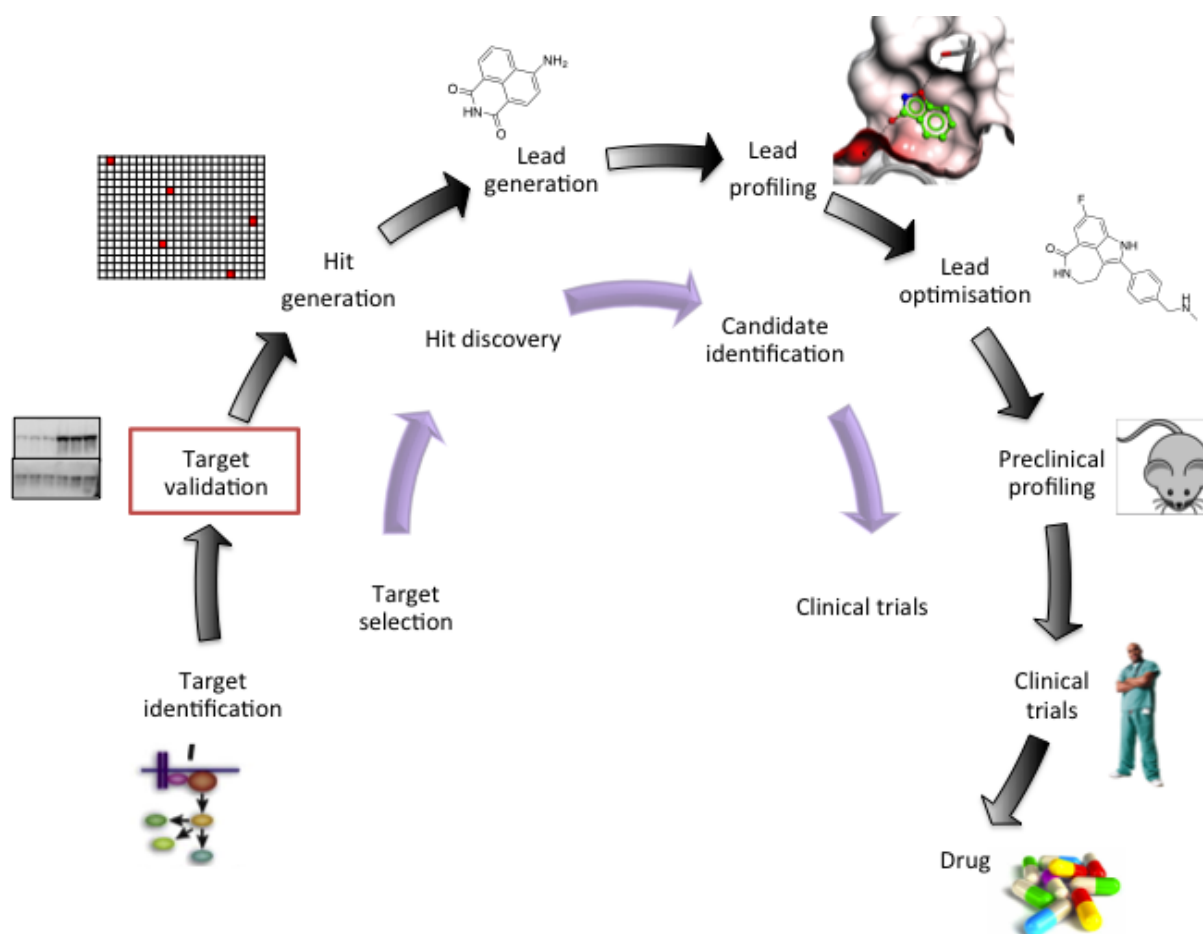
Figure 3.1 – Drug discovery pipeline. Picture adapted from [112].

Historically, phenotypic screenings have been performed, in which compounds are assessed in cellular models to identify compounds that modify the behaviour of the cell in the desired way, followed by identification of the molecular target. However, these days target-based drug discovery is more frequently undertaken. A target-based drug discovery program may be initiated following the identification of a novel molecular target. Most often this target is upregulated or

activated, driving the disease process, and downregulation or inhibition of the target causes a phenotypic change indicating amelioration of the disease. Target validation is a critical milestone in the drug discovery cascade. Its importance is emphasised by the main reasons for candidate drug failure in clinical trials, which are insufficient safety, or the drug simply does not work; both are often a consequence of poor target validation early on [113]. Difficulties in reproducing published target validation experiments are an additional well-known obstacle. A recent analysis by Bayer® on their in-house target identification and validation projects showed that only in 35 % of cases could literature findings be fully or partially repeated [114].

Previous studies have shown the overexpression of PHGDH in certain melanoma and breast cancer cell lines as well as the dependence on PHGDH expression for cell growth in these cell lines and xenograft models [69, 72]. A follow up study on breast cancer as a target reproduced growth inhibition effects in *in vitro* models following PHGDH knockdown, but failed to be translated into *in vivo* efficacy [115]. As a possible explanation for the differences seen in the *in vivo* studies, the stage of the tumour was proposed as an important characteristic, such that PHGDH inhibition might only be a beneficial treatment at early stages of tumour development [115]. However, recent literature further supports a role for PHGDH in cancer, including one in cervical cancer with positive in vivo data, but also the potential of using PHGDH expression levels as prognostic marker [73, 116].

To a large extent, literature data points towards PHGDH as being a putative oncogene associated with progression in several tumour types. Its role as a therapeutic target is supported by several cell-based studies, but to date the *in vivo* data are somewhat contradictory.

3.2 Aims of the project

The aims of these investigations were: Firstly, to determine PHGDH expression in a panel of cell lines including as far as possible cell lines that had been used in the published literature studies [72, 115]. PHGDH expression levels in the various cell lines were also analysed in regards to their p53 status, as PHGDH has been reported to be a target of p53 regulation [117]. In addition, expression of other proteins, i.e. c-Myc and PARP, was investigated to see if a correlation with PHGDH expression could be identified. Secondly, PHGDH knockdown experiments using siRNA or stable, transduced shRNA cell lines were performed to investigate if the reported target validation could be repeated by looking at the effect of PHGDH knockdown on cell growth.

3.3 Materials and Methods

All chemicals listed in the biological materials and methods section were of analytical grade and obtained from Sigma Aldrich, St. Louis, MO, USA, unless otherwise stated.

3.3.1 Cell culture

All tissue culture plasticware was obtained from Corning (Corning, NY, USA). Cell culture reagents were purchased from Invitrogen Life Technologies (Carlsbad, CA, USA) unless otherwise stated. Cell culture work was performed in class II microbiological safety cabinets (BioMAT2, Medical Air Technology, Manchester, UK) using aseptic techniques.

Breast cancer cell lines MDA-MB-231, MDA-MB-468 and Hs578T as well as the highly transfectable HEK293T cells were obtained from the American Type Culture Collection (Manassas, VA, USA). The three breast cancer cell lines are triple-negative (ER-, PR- and HER2-). MDA-MB-468 has amplified *PHGDH* whereas Hs578T and MDA-MB-231 have no *PHGDH* amplification [72, 118]. Melanoma cell lines SK-MEL-28, containing a *PHGDH* amplification [119] and its BRAF-mutant clone were provided by Dr Penny Lovat (Newcastle University, Newcastle-upon-Tyne, UK). All cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10 % fetal bovine serum (FBS). Except for HEK293T cells, cell media was further supplemented with penicillin (100 U/mL) and streptomycin (1.0 mg/mL). Cell lines were maintained at exponential growth at 37 °C in a humidified atmosphere of 5 % CO_2 and 95 % air. All cells were confirmed to be mycoplasma free by regular testing (MycoAlert mycoplasma detection kit, Lonza, Basel, Switzerland). Flasks of exponentially growing cells of a variety of cancer cell lines for protein expression analysis were generously donated by researchers in the NICR.

3.3.2 Preparation of cell lysates for SDS-PAGE and Western blotting

After removal of media, the cells were washed with ice-cold PBS and covered with a thin layer of Trypsin/EDTA (0.05 % trypsin and 0.02 % EDTA in PBS). Cells were left until they detached and then resuspended in media. The cell suspension was transferred into Eppendorf tubes and centrifuged at 500 x g for 5 minutes at 4 °C. The supernatant was discarded and the cell pellet resuspended in an approximately equivalent volume of Phosphosafe extraction reagent (EMD

Millipore, Billerica, MA, USA). The cells were extracted for 10 min at RT, then further broken up by sonication for 8 seconds at 6 mHz using a Soniprep 150 Plus (MSE, London, UK). Cell debris was pelleted by centrifugation at 16000 x g for 5 minutes at 4 °C. The supernatant was transferred into a fresh tube and stored at -20 °C until further analysis by SDS-PAGE and Western blotting. The protein content in supernatants was measured using a Pierce BCA Protein Assay (Thermo Fisher Scientific, Waltham, MA, USA).

3.3.3 SDS-PAGE

SDS-PAGE was performed according to section 2.7 with the difference that protein lysates were loaded onto Criterion TGX Precast 7.5 % gels (BIO-RAD, Hercules, CA, USA) and the gel run at 100 V for 90 minutes.

3.3.4 Western blotting

After separation by SDS-PAGE proteins are transferred onto a membrane, which allows for detection of specific proteins through binding of specific antibodies. Those specific antibodies are themselves recognised by secondary antibodies conjugated to horseradish peroxidase (HRP). HRP catalyses the oxidation of luminol to 3-aminophthalate resulting in the emission of light and this signal can be further amplified and detected.

Protein transfer onto Hybond^{TM}-C Extra nitrocellulose membrane (Amersham Bioscience, GE Healthcare, Little Chalfont, UK) was performed at 100 V for 60 minutes in TRIS-Glycine transfer buffer (Thermo Fisher Scientific, Waltham, MA, USA). At this point the membrane was stained for total protein content using Ponceau S stain (see section 3.3.5) After Ponceau S staining, the membrane was blocked in 5 % milk or 5 % BSA in TRIS-buffered saline with 0.1 % Tween-20 (TBS-T) followed by incubating with primary antibody (Table 3.1) in 5 % milk or 1 % BSA in TBS-T at 4 °C overnight. Following incubation with primary antibodies, the membrane was washed twice in TBS-T before incubation with HRP-conjugated secondary antibody (Table 3.1) in 5 % milk or 1 % BSA in TBS-T for one hour at RT. Unbound secondary antibody was removed by washing the membrane twice in TBS-T. Proteins were detected by enhanced chemiluminescence using Amersham ECL Western blotting detection reagent

(GE Healthcare, Little Chalfont, UK) and visualised and analysed in a FujiFilm LAS-3000 (Fujifilm, Minato, Tokyo, Japan).

Protein	Molecular weight of detected proteins (kDa)	Antibody Type	Clone	Isotype	Supplier	Cat No.	Dilution
Primary Antibodies							
PHGDH	57 kDa	Polyclonal	-		Sigma Aldrich	HPA021241	1:1000
c-Myc	57 kDa (predicted: 49 kDa)	Monoclonal	Y69	IgG	Abcam	Ab32072	1:500
PARP	114 kDa PARP-1 holoenzyme 85 kDa apoptosis-related fragment 50/62/74 kDa necrosis-related fragments	Monoclonal	C2-10	IgG1	Trevigen	4338-MC-50	1:2000
GAPDH (loading control)	37 kDa	Polyclonal	-	IgG	Santa Cruz Biotechnology Inc.	Sc-25778	1:1000
Secondary Antibodies							
Goat Anti-Mouse IgG, HRP conjugate	-	Polyclonal	-	IgG	Dako	P 0447	1:1000
Goat Anti-Rabbit IgG, HRP conjugate	-	Polyclonal	-	IgG	Dako	P 0448	1:2000

Table 3.1 – Primary and secondary antibodies used in this study.

3.3.5 Ponceau S staining

Ponceau S staining solution was used to reversibly visualise the total amount of protein on the Western blot membrane.

After protein transfer, the membrane was stained for 5 minutes with 0.1 % (w/v) Ponceau S in 5 % acetic acid. After washing the membrane twice for 5 minutes with distilled water to remove any non-specific background staining, a picture of the membrane was taken. The total amount of protein per lane was quantified by integrating the stained area using the FujiFilm LAS-3000 (Fujifilm, Minato, Tokyo, Japan). Ponceau S stain was removed by washing the membrane twice for 5 minutes in TBS-T. Subsequently, the membrane could be used for blocking and immunological detection of desired proteins.

3.3.6 Transient siRNA knockdown

Small interfering RNA (siRNA) was used for transient silencing of protein coding genes by binding to target mRNA, which was subsequently targeted for cleavage and degradation.

Transient siRNA-mediated knockdown of PHGDH was achieved in cells using 5 nM ON-TARGET plus PHGDH siRNA (ON-TARGETplus, 009518, Dharmacon, Lafayette, CO, USA) by either using individual siRNAs or a pool of four individual siRNAs (Table 3.2), with each siRNA at 1.25 nM. Scrambled (sc) siRNA (as a pool of four different sc siRNAs) was used in parallel as negative control (Table 3.2). siRNA was transfected into 2×10^5 cells/well of a 6-well plate using lipofectamine RNAiMax reagent (Invitrogen, Life Technologies, Thermo Fisher Scientific, Waltham, Massachusetts, USA) in OptiMEM medium (Life Technologies, Thermo Fisher Scientific, Waltham, Massachusetts, USA). Cells were allowed to grow for 24 – 240 hours in antibiotic-free medium before being analysed or used for further experiments.

siRNA	Target sequence	Company details
PHGDH 05	GGAAAUUGCUGUUCAGUUC	Dharmacon, Lafayette,CO, USA
PHGDH 06	CGACAGGCUUGCUGAAUGA	Dharmacon, Lafayette,CO, USA
PHGDH 07	GACCCUUGCUGCCGGAAAGA	Dharmacon, Lafayette,CO, USA
PHGDH 08	UGAACUUGGUGAACGCUAA	Dharmacon, Lafayette,CO, USA
sc siRNA (pool of 4)	UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUUACAUGUUUUCCUA	Dharmacon, Lafayette,CO, USA

Table 3.2 - Sequences of siRNAs targeting PHGDH and control scrambled siRNA (sc siRNA).

3.3.7 Development of inducible shRNA-containing cell lines

Short hairpin RNA (shRNA) is used for long-term knockdown of the target mRNA by endogenous expression of shRNA. As the shRNA sequence is integrated into the host genome and therefore continuously synthesised, the knockdown of the target protein is expected to last longer compared to siRNA treatment of cells. To be able to integrate the shRNA into the host genome, the sequence is incorporated into a lentiviral vector, which is then transduced into the host. Once integrated into the host genome, the shRNA is transcribed by RNA polymerases to yield an RNA sequence that is incorporated into the RNA-induced silencing complex (RISC) and directs the RISC to the target mRNA sequence to allow for its degradation.

In the following experiments, the pTRIPZ inducible shRNA vector (Figure 3.2) was used. In the pTRIPZ vector, the gene for the shRNA is under the control of TRE promoter, which also drives the expression of the TurboRFP reporter and thus allows observation of expression from the TRE promoter and, ergo, expression of the shRNA DNA through detection of red fluorescence.
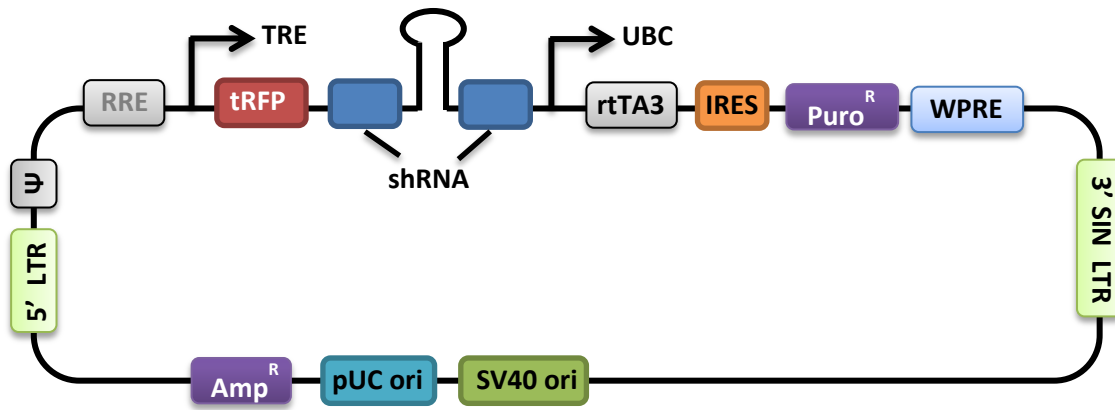
Vector element	Utility
TRE	Tetracycline-inducible promoter
tRFP	TurboRFP reporter for visual tracking of transduction and shRNA expression
shRNA	MicorRNA-adapted shRNA (based on miR-30) for gene knockdown
UBC	Human ubiquitin C promoter for constitutive expression of rtTA3 and puromycin resistance genes
rtTA3	Reverse tetracycline-transactivator 3 for tetracycline-dependent induction of the TRE promoter
Puro^R	Puromycin resistance permits antibiotic selective pressure and propagation of stable integrants
IRES	Internal ribosomal entry site allows expression of rtTA3 and puromycin resistance genes in a single transcript
5' LTR	5' long terminal repeat
3' SIN LTR	3' self-inactivating long terminal repeat for increased lentivirus safety
Ψ	Psi packaging sequence allows viral genome packaging using lentiviral packaging systems
RRE	Rev response element enhances titer by increasing packaging efficiency of full-length viral genomes
WPRE	Woodchuck hepatitis posttranscriptional regulatory element enhances transgene expression in the target cells
Amp^R	Ampicillin resistance
pUC ori	Origin of replication for propagation in *E. coli*
SV40 ori	Origin for replication in mammalian cells

Figure 3.2 - Schematic vector map of the pTRIPZ inducible lentiviral shRNA vector including description of components

(Picture adapted from suppliers' technical manual:
http://dharmacon.gelifesciences.com/shrna/tripz-lentiviral-shrna-controls/?Parent=12884902171
#ProductStrip-2)

3.3.8 Plasmid preparation

To prepare the plasmid DNA of the pTRIPZ expression vector, 5 mL Luria Bertani (LB) medium containing 100 µg/mL ampicillin were inoculated with the respective glycerol stocks of the pTRIPZ-inducible shRNA *E. coli* clones (Table 3.3) and grown at 37 °C for 8 hours. The starter inoculum was then diluted 1:500 into a larger culture volume of 200 mL LB medium supplemented with 100 µg/mL ampicillin and grown at 37 °C for 18 hours. Bacterial cells were harvested by centrifugation at 6000 x g for 15 minutes at 4 °C. The plasmid DNA was purified from the cell pellet using the EndoFree Plasmid Maxi Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. The final DNA content was measured using a Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) and aliquots stored at -20 °C until further use.

Abbreviation	Name	Mature antisense sequence
B03	V2THS_91260	TAGAAGTGGAACTGGAAGG
B04	V2THS_91258	TGCTGTACTACAGGGTCAG
B05	V3THS_327719	CATTCAGCAAGCCTGTCGT
B08	V3THS_327721	CTGACACCAGTGAAGTCTG
Non-targeting control (NTC)	-	CTTACTCTCGCCCAAGCGAGAG

Table 3.3 - Mature antisense sequences of shRNA produced from TRIPZ inducible shRNA clones.

shRNA clones were from Dharmacon, Lafayette, CO, USA. The NTC was kindly provided by Dr Natalia Martinez-Soria, Newcastle University, UK.

3.3.9 Virus particle preparation

For virus particle preparation, 2 x 10^6 HEK293T cells were seeded into a 75 cm^2 tissue culture flask and left to attach and grow. After 24 hours the medium was removed from the cells and replaced with fresh medium supplemented with packaging mixture. The packaging mixture was prepared twenty minutes prior to usage and contained 2.25 µg pMD2.G envelope plasmid (plasmid # 12259, Addgene, Cambridge, MA, USA), 6.75 µg psPA2X packaging plasmid (plasmid #12260, Addgene, Cambridge, MA, USA), 9 µg pTRIPZ expression plasmid for PHGDH (Dharmacon, Lafayette, CO, USA) or pTRIPZ non-silencing control (Dharmacon, Lafayette, CO, USA), 54 µL LT1 transfection reagent (MirusBio, Madison, WI, USA)

and 1 mL OptiMEM medium. Viral supernatants were collected after 72 hours, centrifuged at 1000 x g for 5 minutes at RT and filtered through a 0.45 µm filter to remove any cells. The supernatant was then ultracentrifuged at 119 944 x g for 2 hours at 4 °C. The supernatant was decanted, the virus pellet resuspended in 1 mL basal medium, and 100 µL aliquots were stored at -80 °C.

3.3.10 Assessment of puromycin sensitivity

The pTRIPZ vectors transduced into mammalian cells contain a puromycin resistance gene. Administration of the antibiotic will therefore allow selection for cells that have been successfully transduced with the pTRIPZ vector, hence generating stable transfected cell lines. In order to avoid killing any transfected cells at the beginning of the selection process, the lowest amount of puromycin required to kill all non-transfected cells had to be established.

5×10^4 cells were seeded per well of a 24-well plate. After 24 hours medium was exchanged for fresh medium containing varying concentrations of puromycin (0, 0.1, 0.25, 0.5, 1, 2, 3, 5 µg/mL). Following 72 hours incubation, cells were fixed by addition of 0.1 mL ice-cold Carnoy's fixative (3 parts methanol and 1 part acetic acid) and left at 4 °C for 1 hour. Plates were then washed several times with water and left to dry. Cells were stained with 0.4 mL 0.4 % SRB solution (0.4 % Sulforhodamine B dissolved in 1 % acetic acid) for 30 minutes. Unbound dye was washed off with 1 % acetic acid and plates left to dry. SRB was solubilised in 0.4 mL 10 mM Tris (pH 10.5) and absorbance of the coloured solution measured at 570 nm.

3.3.11 Transduction of target cells

The day before transduction 2×10^5 cells/well were seeded into a 6-well plate. The next day medium was removed from the cells and the cells were treated with different amounts (10 and 50 µL) of virus stocks (target sequences B03 and B04, and non-silencing control) in serum-free RPMI 1640 medium in a total volume of 1 mL. After 8 hours of incubation at 37 °C, 1 mL of full RPMI 1640 medium (containing 10 % FBS, penicillin and streptomycin) was added to the cells. At 48 hours post-transduction, the medium was removed from the cells and replaced with 2 mL full medium containing 1 µg/mL puromycin. Every 2-3 days, medium was replaced with freshly prepared selective medium (medium containing 1 µg/mL puromycin) to continue selection for transduced cells.

3.3.12 TRIzol RNA extraction

TRIzol, a monophasic solution containing phenol and guanidine isothiocyanate, is used to lyse cells and dissolve cell components without compromising the integrity of the RNA, thus allowing the extraction of RNA from cell samples.

Cells were removed from the plastic surface using trypsin-EDTA (0.05 % trypsin and 0.02 % EDTA in PBS) and centrifuged for 5 min at 500 x g at 4 °C. The supernatant was removed and the cell pellet mixed vigorously with 500 µL TRIzol. After incubating for 5 min at RT, 100 µL chloroform was added and the mixture incubated for a further 5 min. To separate the aqueous and organic layer the sample was spun at 12000 x g for 15 min at 4 °C. The aqueous layer was transferred into a fresh tube and mixed with 250 µL isopropanol. The samples were incubated at RT for 10 min to allow for complete RNA precipitation before centrifugation at 12000 x g for 10 min at 4 °C. The supernatant was removed from the RNA pellet and the pellet washed by adding 1 mL 75 % ethanol. The sample was spun for 5 min at 7500 x g at 4 °C and the supernatant removed and discarded. The remaining pellet was air dried and then resuspended in 20 µL diethylpyrocarbonate-treated (DEPC) water and stored at -20 °C until required.

3.3.13 Two step quantitative reverse transcription polymerase chain reaction (qRT-PCR)

Step one: Reverse Transcription Reaction

In order to be able to analyse the RNA content in cell samples via qRT-PCR, the RNA must first be transcribed into complementary DNA by the Moloney murine leukemia virus reverse transcriptase (M-MLV-RT).

For the reverse transcription reaction, 1-2 µg RNA in DEPC water was incubated at 65 °C for 5 min to disrupt the RNA secondary structure, followed by 2 min incubation at 37 °C. The RNA was mixed with M-MLV RT buffer (Promega, Fitchburg, WI, USA), M-MLV reverse transcriptase (Promega, Fitchburg, WI, USA) at a final concentration of 3 units/µL, dNTP mix (Promega, Fitchburg, WI, USA) at a final concentration of 0.4 mM and Oligo(dT)15 primer (Promega, Fitchburg, WI, USA) at a final concentration of 2.5 µM. The mixture was incubated at 37 °C for one hour, followed by 5 min incubation at 95 °C to inactivate the reverse transcriptase.

Quantitative, real-time PCR allows for the detection and quantification of cDNA levels as the reaction progresses. As the cDNA is amplified a fluorescent probe (*e.g.* SYBR green) present in the reaction mix, binds to the synthesised double-stranded DNA and emits fluorescence.

Real-time qPCR was performed with 0.5-0.75 µg cDNA in a final volume of 10 µL using PHGDH, GAPDH or β_2-microglobulin (β2m) -specific primers (Table 3.4). A serial dilution standard of a control cDNA sample (treated with scrambled siRNA) was used to determine the linear regression. RT-qPCR was performed with 7900 Real-Time PCR system (Applied Biosystems, Waltham, MA, USA) following the cycling profile below (Figure 3.3). Reaction was run in the presence of Platinum® SYBR® Green qPCR SuperMix-UDG (Thermo Fisher Scientific, Waltham, MA, USA), and analysed through correction for GAPDH transcription level.

Gene	Orientation	Abbreviation	Primer Sequence (5'-3')
PHGDH	Forward	RT1F	CACGACAGGCTTGCTGAATGA
	Reverse	RT1R	CTTCCGTAAACACGTCCAGTG
GAPDH	Forward	fGAPDH	CGACCACTTTGTCAAGCTCA
	Reverse	rGAPDH	GGGTCTTACTCCTTGGAGGC
β2m	Forward	Pb2mF	GAGGCTATCCAGCGTACTCCA
	Reverse	Pb2mR	CGGCAGGCATACTCATCTTTT

Table 3.4 - Primers for real-time qPCR
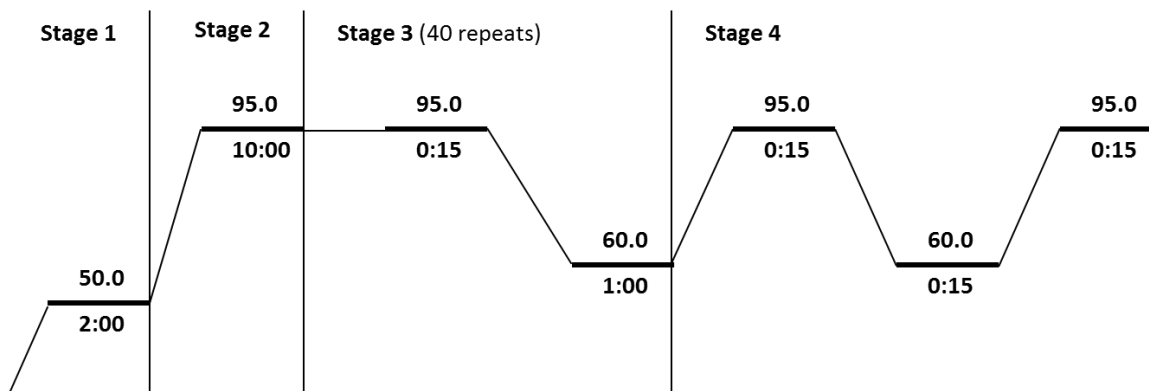
Figure 3.3 - Thermal cycling profile of real-time qPCR

3.3.14 Clonogenic survival assays

Clonogenic survival assays were used to determine the ability of cells to proliferate and form colonies. Cells are seeded at very low density and left to undergo > 5 division cycles so that colonies comprising > 30 cells are obtained and can be observed by eye.

Method 1: siRNA treated cells

Cells were treated with siRNA as described in section 3.3.6 and incubated for 48 hours. Cells were then harvested through addition of trypsin-EDTA (0.05 % trypsin and 0.02 % EDTA in PBS) and seeded at specific densities in 2 mL final volume of full medium into 6-well plates. Cells were left to replicate for 10-16 days until colony formation was observed. Following colony formation, medium was removed from the cells and the cells were fixed by addition of 1 mL Carnoy's fixative (methanol: acetic acid, 3:1, v:v), and incubation for 10 minutes at RT. Carnoy's fixative was removed from the cells and the plates left to dry. Cells were stained with 1 % crystal violet solution in 10 % aqueous ethanol for 20 minutes. Crystal violet was removed, residual stain washed away with water and the plates allowed to dry. Colonies were then counted by eye and the cloning efficiency calculated (Equation 3.1).

$$Cloning\ efficiency = \frac{Number\ of\ colonies\ counted}{Number\ of\ cells\ seeded} \times 100\ \%$$

Equation 3.1 – Calculation of cloning efficiency.

Method 2: shRNA treated cells

Cells were treated with medium with and without 1 µg/mL doxycycline for 72 hours; medium was replaced with fresh medium with and without doxycycline every 24 hours. Cells were harvested using trypsin-EDTA (0.05 % trypsin and 0.02 % EDTA in PBS) and seeded at specific densities in 2 mL final volume of medium with or without doxycycline. Cells were left to replicate for 10-16 days until colony formation was observed and subsequent visualisation and analysis were performed as described for method 1.

3.3.15 Growth inhibition assay

Growth inhibition assays were performed to investigate cellular proliferation over a fixed period of time, typically 7-10 days. For the quantification of cells, cells were

stained with a coloured dye, such as sulforhodamine B (SRB) for staining of protein or 4',6'-diamidino-2-phenylindole (DAPI) for staining of DNA.

Exponentially growing cells were cultured for three days in media with or without 1 µg/mL doxycycline by replacing media with fresh media (with or without doxycycline) every 24 hours. Cells were then counted and a fixed number of cells seeded into wells of a 96-well plate in media with or without 1 µg/mL doxycycline. Cells were allowed to grow for up to seven days with one plate being fixed every 24 hours. Cells were fixed by adding 25 µL of Carnoy's fixative and incubated for 10 min at RT. Carnoy's fixative was removed and remaining traces washed away with water. Plates were dried and stored at 4 °C until required. To measure the amount of cells per well, cell protein content was determined by staining with 0.4 % SRB in 1 % acetic acid which bound under these mild acidic conditions to basic amino acid residues. The stain was removed after 30 minutes and plates washed four times by submersion in 1 % acetic acid solution. After the plates had dried, the SRB dye was solubilised in 10 mM Tris, pH 10.5 (100 µL/well) and absorbance measured at 570 nm on a PHERAstar FS platereader (BMG Labtech, Ortenberg, Germany).

3.3.16 Live-cell imaging

Cells were treated with siRNA as described in section 3.3.6 and incubated for 48 hours. Cells were then harvested through addition of trypsin-EDTA and seeded at specific densities in 2 mL final volume of full medium into 6-well plates. Cells were imaged using the IncuCyte Zoom HD system (Essen bioscience, Ann Arbor, MI, USA) by capturing pictures of 16 separate regions per well every 6 hours. During the live-cell image, cultures were permanently kept in a 37 °C incubator.

3.4 Results

Julia Reber, a summer project student working under my supervision, performed part of the work with the melanoma cell lines (section 3.4.3).

3.4.1 PHGDH expression varies in a panel of cancer cell lines

To confirm differences in PHGDH expression reported for melanoma and breast cancer cell lines [69, 72] and to investigate PHGDH expression beyond the reported cell lines, a panel of cancer cell lines was assessed by SDS-PAGE and Western blotting Due to the variety of cancer cell lines being assessed it was found that levels of so-called house-keeping proteins, such as GAPDH, vinculin or Histone H1 varied between the cell lines (Figure 3.4). To overcome this, each sample was normalised to the total protein content of the sample as visualised and determined by PonceauS staining (Figure 3.4).
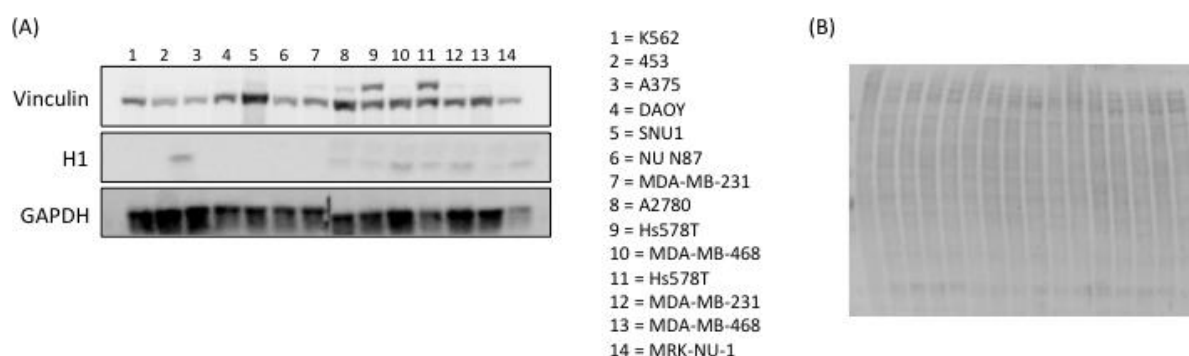
Figure 3.4 - Comparison of expression levels of different "house-keeping" proteins with total protein content per sample
Cells were harvested and 40 ng of protein lysate loaded per well and analysed by SDS-PAGE and Western blotting. (A) Representative blots of GAPDH, histone H1 and vinculin expression measured by chemiluminescence using the Fuji. (B) Total protein content on Western blot visualised by Ponceau S staining.

Due to the large number of cell lines being assessed multiple experiments had to be performed, therefore protein expression levels in K562 cell lysates were used as an internal normalisation standard. In this context, the measurement of inter- as well as intra-assay variability was important to correct for the reliability of the measured protein expression. The intra-assay coefficient of variation (CV) was found to be reasonably low at 6.0 % for the PHGDH/Ponceau S ratio. As expected, the inter-assay CV was higher with a PHGDH/Ponceau S ratio at 28.9 % after normalisation to the respective standard sample, but found to be in a reasonable range for Western blot experiments.

The observed PHGDH expression levels varied from almost no expression to 6-fold higher expression compared to K562 cells (Figure 3.5). For some cell lines, a large difference in expression levels was observed in cell lysates harvested from different passages of the cell culture. Analysis of PHGDH protein expression according to the cancer types the cell lines were derived from showed that breast cancer and melanoma were upon the cancer types with the highest PHGDH expression levels (Figure 3.5 C). However, no cancer type was associated with significantly higher overall PHGDH expression compared to the other cancer type.
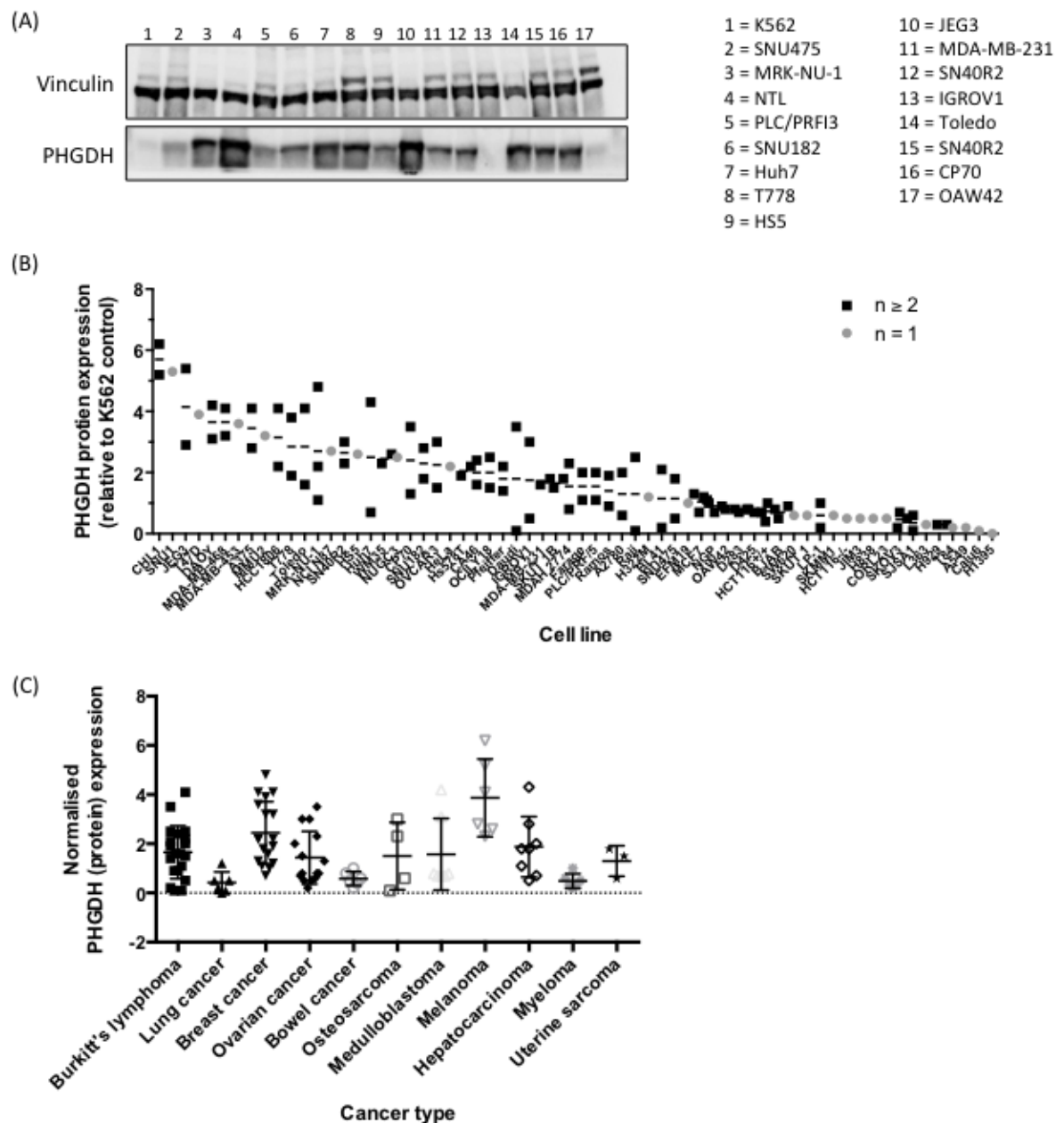
Figure 3.5 - PHGDH expression in a panel of cancer cell lines

Cells were harvested, lysed and the proteins separated by gel electrophoresis. Proteins were detected using Western blotting and quantified by chemiluminescence using the Fuji. Bands were normalised to total protein content visualised by PonceauS staining and quantified in relation to protein expression in K562 as the chosen standard cell line. (A) One representative blot showing PHGDH expression in 16 different cell lines. (B) PHGDH expression in all cell lines assessed relative to PHGDH expression in K562 control samples. Protein expression levels were assessed in one sample (grey circles) or if available in two or more samples (black squares). (C) PHGDH levels, normalised to PHGDH protein level in K562 cells, were group according to their tissue of origin.

In melanoma cell lines, wt p53 has been shown to regulate PHGDH mRNA and protein levels where activation of p53 through Nutlin-3 treatment resulted in decreased PHGDH mRNA and protein levels. With this regulation mechanism in

mind we investigated the PHGDH protein expression levels in context of the p53 status in the various cell lines. Overall the p53 mutant cell lines showed higher PHGDH protein levels (p = 0.023) compared to cell lines with wt p53 (Figure 3.6 A). As the investigated cell line panel spans cancers of different origins, the p53 status was also taken into account in the individual cancer types, but no significant difference in PHGDH expression according to the p53 status could be detected (Figure 3.6 B).
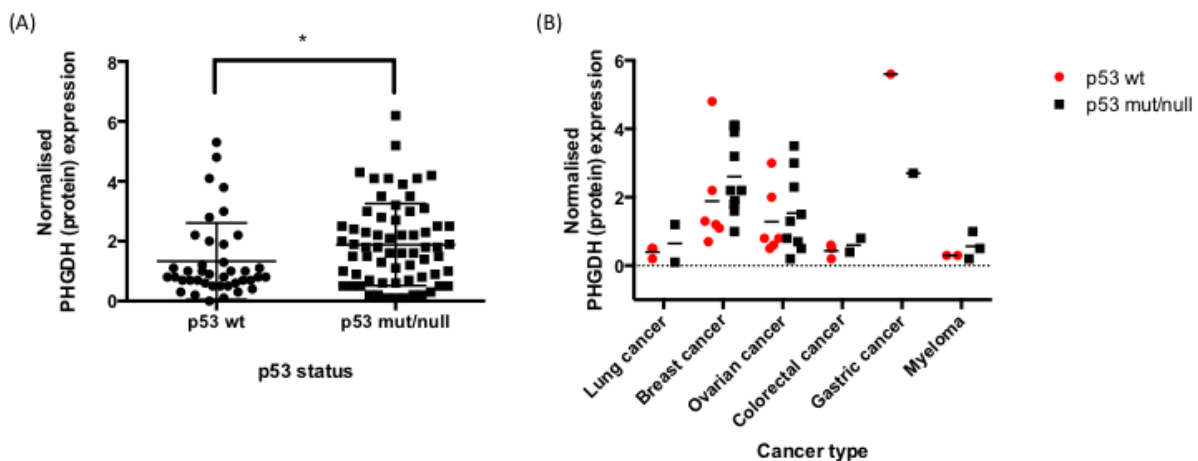
Figure 3.6 – Analysis of PHGDH protein expression in context of the cell lines' p53 status

PHGDH protein levels were determined as described before and normalised to PHGDH expression in K562 cells as reference. p53 status of the different cell lines was determined from IARC TP53 database, version R17, November 2013 [120]. For the following cell lines the p53 status was determined from the cited literature: D425 [121], D283 [122], T778 [123], Hs578T [124], OCI-LY18 [125], Pfeiffer [126]. Cell lines for which the p53 status was not known are excluded from the analysis. (A) PHGDH expression grouped according to the cell lines' p53 status. Data was analysed in GraphPad Prism using a paired t.test. p = 0.023. (B) Separation of PHGDH expression according to cancer type and p53 status. Only cancer types with at least one cell line per p53 status were included.

3.4.2 PHGDH expression correlates negatively with c-Myc

Alongside PHGDH expression levels, expression levels of c-Myc were investigated in the reported cell lines. As outlined earlier c-Myc is implicated in the regulation of many metabolic pathways, such as glutamine, glucose and nucleotide metabolism, and was recently shown to positively regulate mRNA expression of proteins involved in serine synthesis, such as PHGDH [127, 128]. As the band for c-Myc runs at 57 kDa, the same molecular weight as PHGDH, lysates were split and run on two gels to allow for detection of both proteins (Figure 3.7 A). A variety of c-Myc expression levels were detected within the panel of cell lines with very high levels in the

medulloblastoma cell lines D425 and D283 and in the ovarian cancer cell line MDAH-2774. D425 contain a c-Myc gene amplification and re-arrangement of one allele was found in D283, both factors reported to increase c-Myc expression substantially in those cell lines [129]. In order to investigate if the expression levels of c-Myc and PHGDH were correlated, a Spearman correlation analysis was performed using the expression levels of cell lines that were assessed at least twice. Intriguingly, a negative correlation (Spearman r = -0.4716, p = 0.0014) was determined where high c-Myc levels were observed in cells with low PHGDH expression (Figure 3.7 C). Despite the relatively high correlation coefficient, analysis of the correlation blot by eye rather indicates an unordered distribution of both protein levels (Figure 3.7 D).
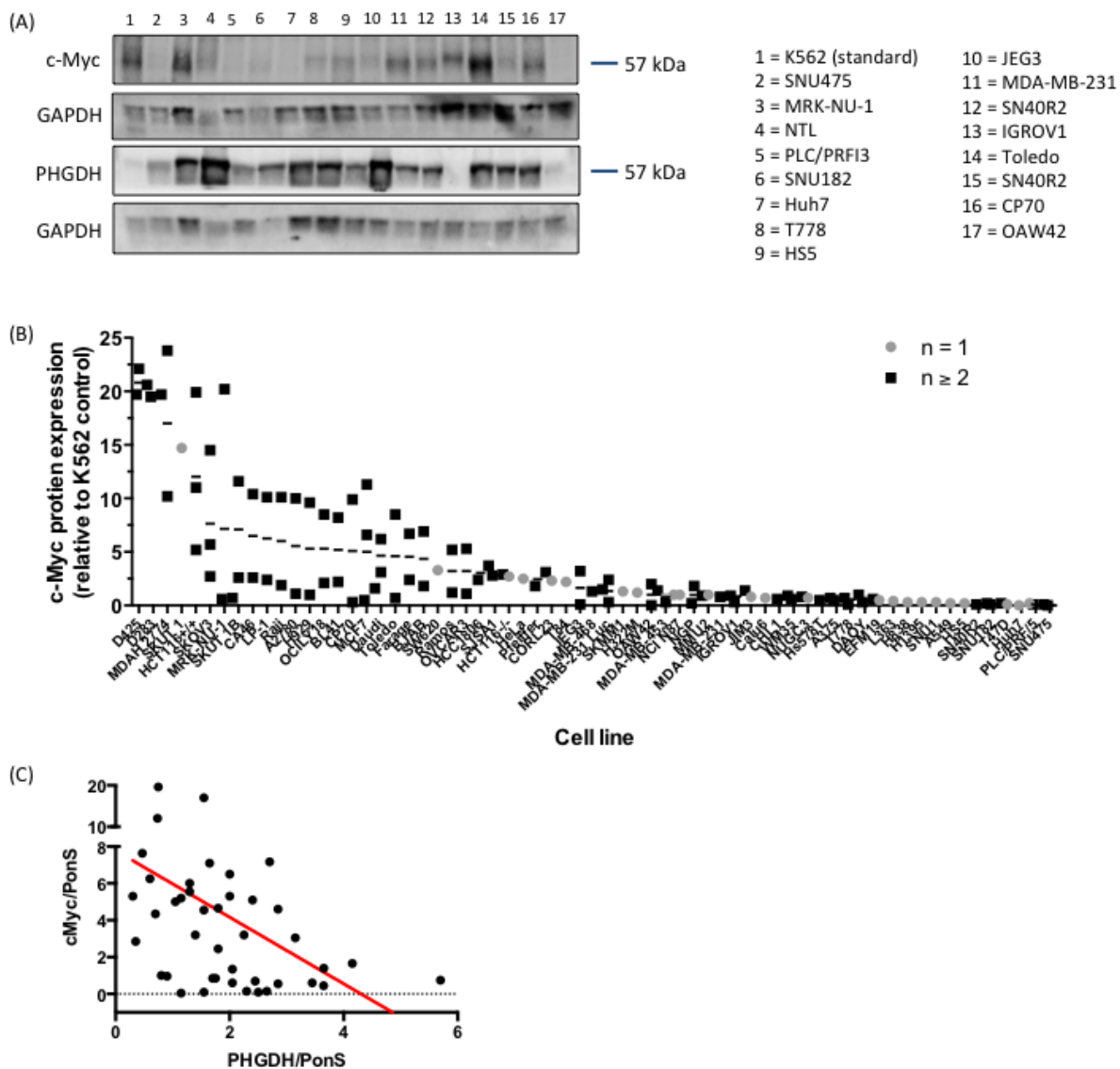
Figure 3.7 – c-Myc expression in a panel of cancer cell lines

Cells were harvested, lysed and proteins separated by gel electrophoresis. Proteins were detected using Western blotting and quantitated by chemiluminescence using the Fuji. Bands were normalised to total protein content visualised by PonceauS staining and quantified in relation to protein expression in K562 as the chosen standard cell line. (A) One representative blot showing c-Myc and PHGDH expression in 16 different cell lines. (B) c-Myc expression in all cell lines assessed relative to c-Myc expression in K562 control sample. Protein expression levels were assessed in one sample (grey circles) or if available in two or more samples (black squares). (C) Spearman correlation analysis of c-Myc and PHGDH expression levels in 43 different cancer cell lines was performed in GraphPad Prism. The red line indicates the correlation line (r = -0.4716).

In addition to c-Myc and PHGDH, protein levels of poly(ADP-ribose) polymerase 1 (PARP1) were also investigated. PARP1 catalyses the transfer of the ADP-ribosyl moiety of NAD^+ onto various target proteins, a post-translational modification called PARylation. PARP1 plays important roles in different cellular processes, such as

DNA repair, genomic stability, cell death signalling and cellular metabolism [130]. In the context of cellular metabolism, PARP1 activation by DNA damage depletes NAD^+ levels, this affecting the glycolytic pathway, which uses NAD^+ as a key cofactor [131]. Since PHGDH is a NAD^+-requiring enzyme linked to glycolysis, a possible correlation of PARP and PHGDH expression levels was investigated.

Among the cell lines investigated, a range of PARP1 expression levels was observed (Figure 3.8 A and B). However, there was no significant correlation (p = 0.0805) seen between PHGDH and PARP1 protein levels in these cell lines (Figure 3.8 C).
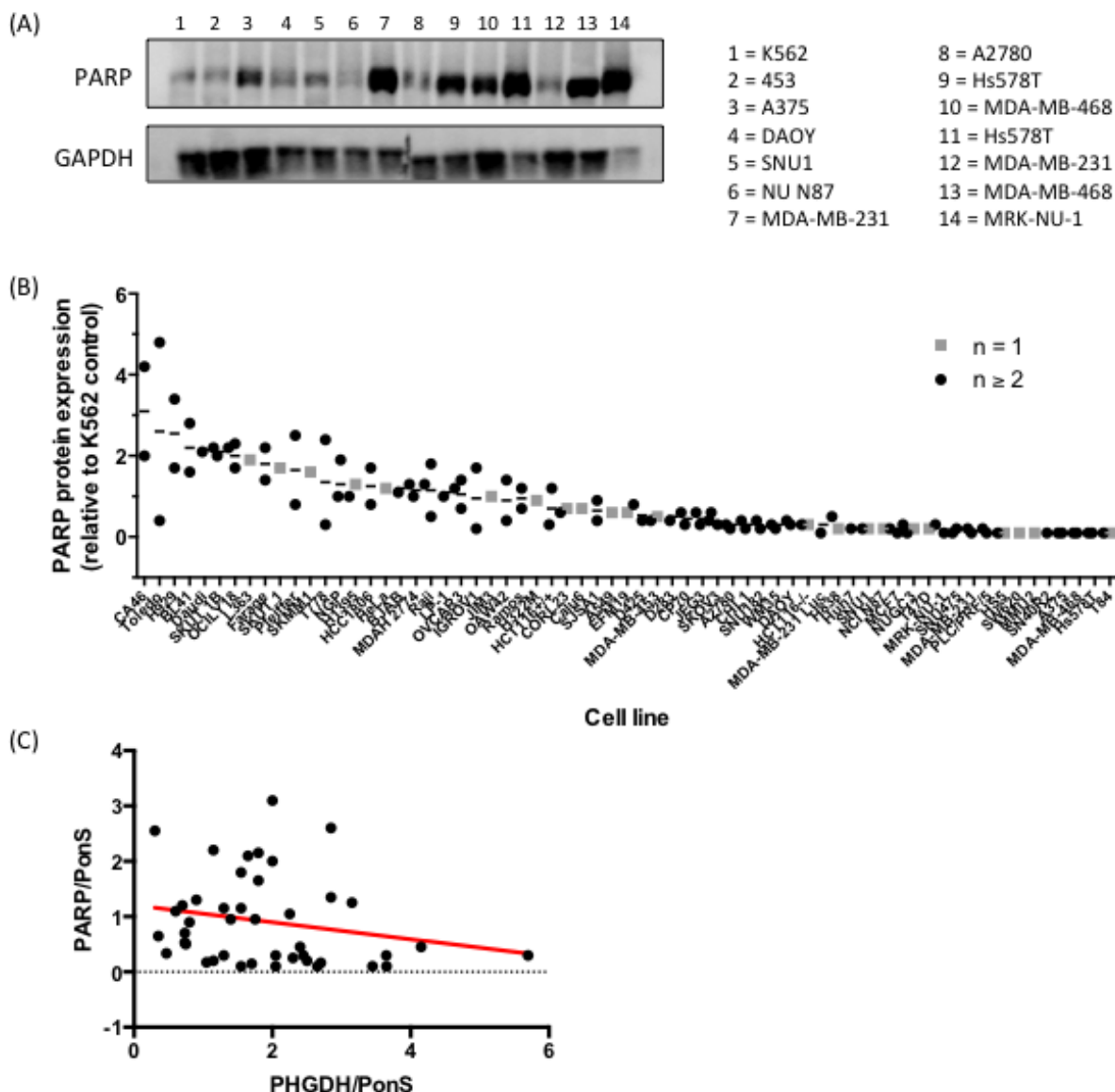
Figure 3.8 – PARP1 expression in a panel of cancer cell lines

Cells were harvested, lysed and proteins separated by gel electrophoresis. Proteins were detected using Western blotting and quantitated by chemiluminescence. Bands were normalised to total protein content visualised by PonceauS staining, and quantified in relation to protein expression in K562 cells. (A) One representative blot showing PARP expression in 12 different cell lines. (B) PARP expression in all cell lines assessed relative to PARP expression in K562 control sample. Protein expression levels were assessed in one sample (grey circles) or if available in two or more samples (black squares). (C) Spearman correlation analysis of PARP and PHGDH expression levels in 43 different cancer cell lines was performed in GraphPad Prism. The red line indicates the correlation line (r = -0.2693).

3.4.3 siRNA-mediated knockdown of PHGDH

In order to validate PHGDH as a potential target, siRNA-mediated knockdown was performed in breast cancer cell lines. Based on the literature, three breast cancer cell lines were chosen: MDA-MB-231 (non-amplified *PHGDH*), MDA-MB-468 (amplified

PHGDH) and Hs578T (non-amplified *PHGDH,* but with high expression) [72]. The effects on cell growth after *PHGDH* knockdown were expected in the PHGDH amplified or highly expressing cell lines.

Optimisation of the protocol for siRNA-mediated knockdown of PHGDH was performed in Hs578T cells. Hs578T cells were treated with four different siRNAs (siRNA 5-8) or scrambled siRNA (sc siRNA) at a final concentration of 10 nM for 72 hours. PHGDH mRNA and protein expression levels were determined every 24 hours by RT-PCR and Western blotting, respectively. Although PHGDH mRNA levels were successfully decreased by more than 90 % after 24 hours of siRNA treatment, a decrease in protein levels was only seen after 48 hours, indicating a protein half-life between 24 and 48 hours (Figure 3.9). After 72 hours of siRNA treatment, 30 % of protein remained (Figure 3.9 C).
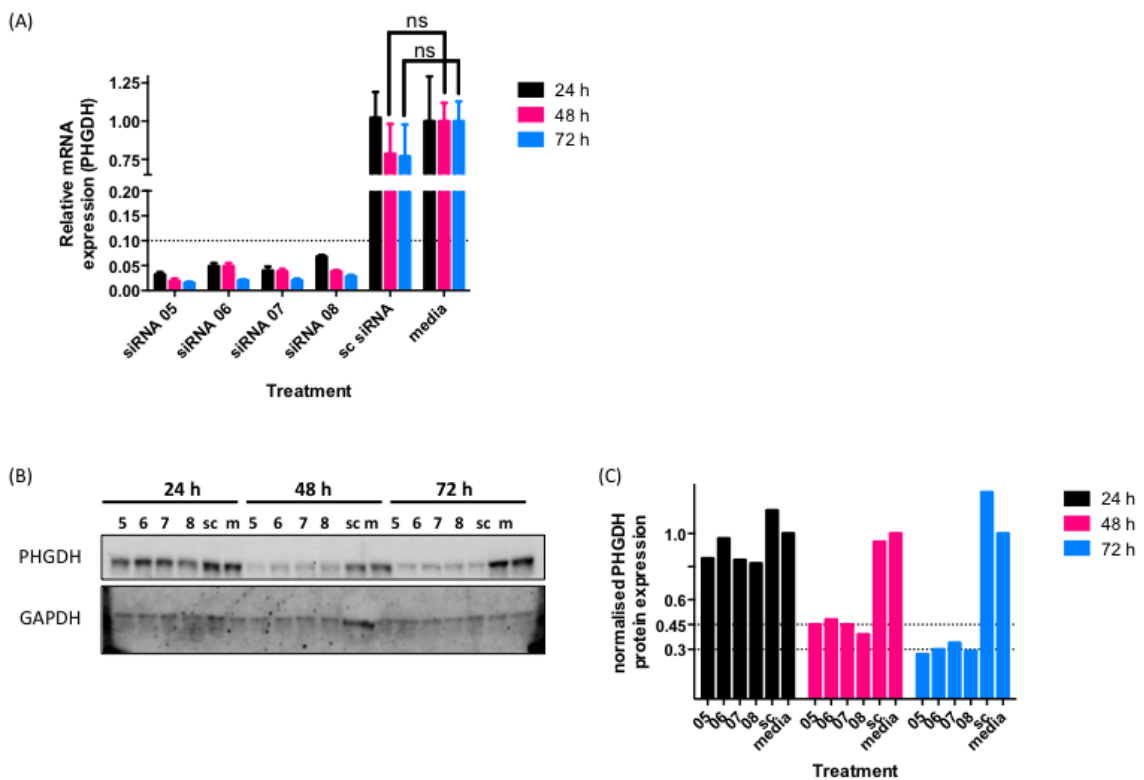
Figure 3.9 – siRNA-mediated PHGDH knockdown in Hs578T cells

Cells were seeded into wells of a 6-well plate at a density of 2×10^5 cells/well and treated with 5 nM on-target or scrambled siRNA or medium only control. Cells were harvested and lysed or RNA isolated after 24, 48 and 72 hours. Proteins were separated by gel electrophoresis and detected using Western blotting before visualisation by chemiluminescence. (A) Relative mRNA expression levels as detected by RT-PCR using GAPDH as reference gene and normalised to respective media control. (B) Western blots showing PHGDH and GAPDH protein levels. (C) Quantification of PHGDH protein levels detected by Western blotting and normalised to total protein content of the sample and to the media control.

Further initial experiments, different concentrations of siRNA were investigated to see if the maximum possible knockdown of PHGDH was achieved with 10 nM siRNA, as well as to investigate if a lower siRNA concentration would result in the same level of knockdown and thereby minimise possible off-target effects of the siRNA. In addition to individual on-target siRNAs, a pool of all four siRNAs was used as they were of similar potency and this allowed the use of smaller amounts of each individual siRNA, thus further minimising potential off-target effects. As 24 hours had been shown not to be a sufficient time of knockdown to see a reduction in protein levels, only 48 and 72 hours were investigated. The experiment confirmed that 5 nM of individual siRNA as well as the pool of four individual siRNAs was sufficient to induce the same level of knockdown as seen with 10 or 20 nM (Figure 3.10). Although PHGDH protein expression was already reduced by 50 % after 48 hours of siRNA treatment, a further reduction was seen after 72 hours (Figure 3.10).
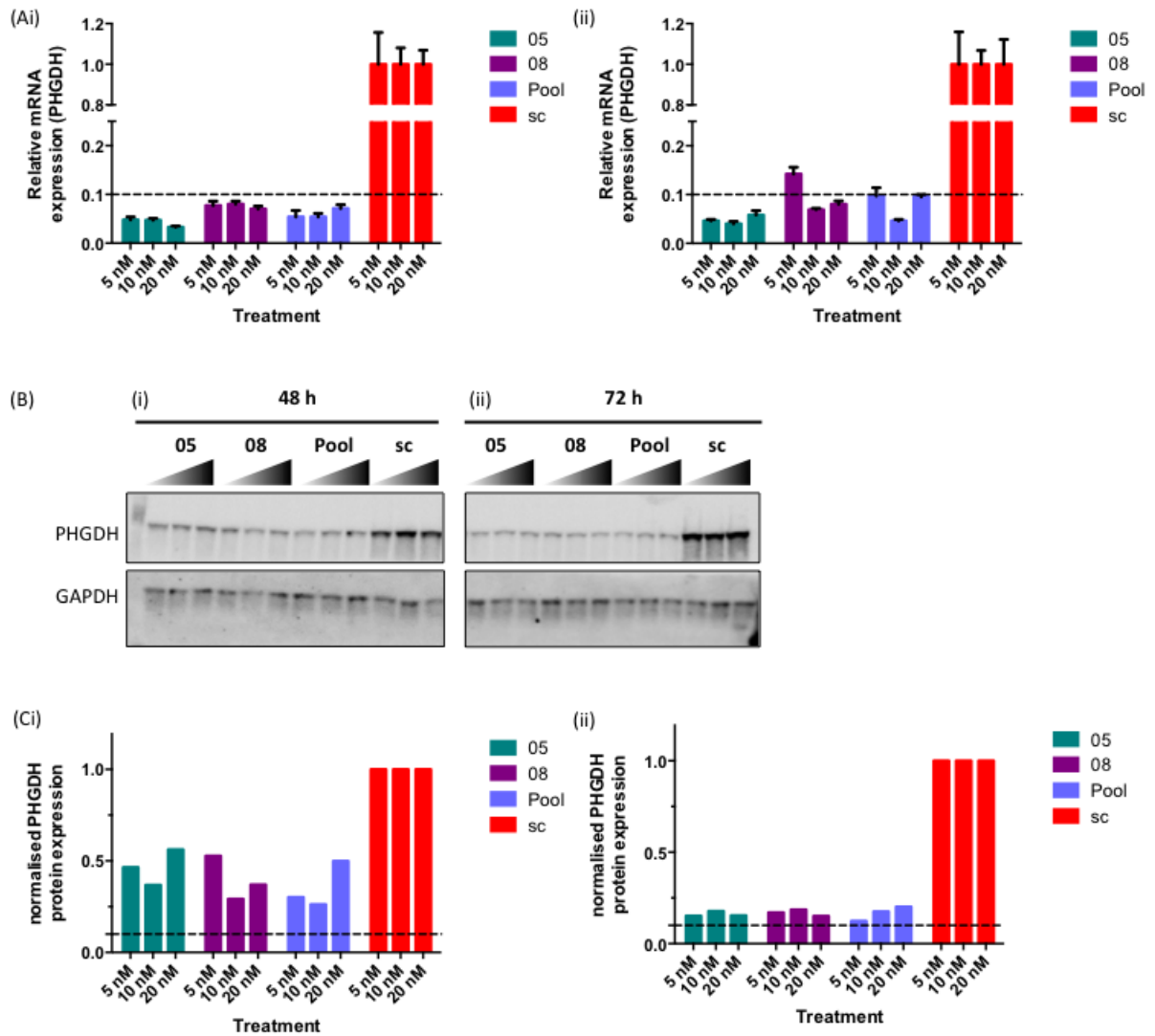
Figure 3.10 – siRNA-mediated knockdown of PHGDH in Hs578T with different concentrations of siRNA.

Cells were seeded into wells of a 6-well plate at a density of 2×10^5 cells/well and treated with 5, 10 or 20 nM individual siRNA05, siRNA08, a pool of four individual siRNAs or scrambled siRNA. Cells were harvested and lysed or RNA isolated after 48 (i) and 72 (ii) hours. Proteins were separated by gel electrophoresis and detected using Western blotting before visualisation by chemiluminescence. (A) Relative mRNA expression levels as detected by RT-PCR using GAPDH as reference gene and normalised to respective scrambled siRNA control. (B) Western blots showing PHGDH and GAPDH protein levels. (C) Quantification of PHGDH protein levels detected by Western blotting and normalisation to total protein content of the sample and to the media control.

The experimental conditions established (5 nM pool siRNA, ≥ 72 h) were then used to assess the effect of PHGDH knockdown in all three breast cancer cell lines. mRNA levels were significantly reduced in all three cell lines over 120 hours resulting in PHGDH mRNA levels below 10 % (Figure 3.11 A). The reduction in mRNA levels was directly translated into PHGDH protein levels, which were also significantly

reduced in Hs578T and MDA-MB-468 cells over 120 hours, whereas the reduction in PHGDH levels in MDA-MB-231 was only about 50 %, however, this result is likely to be also due to the low PHGDH protein content in this cell line, which was difficult to detect by Western blotting (Figure 3.11 B and C).
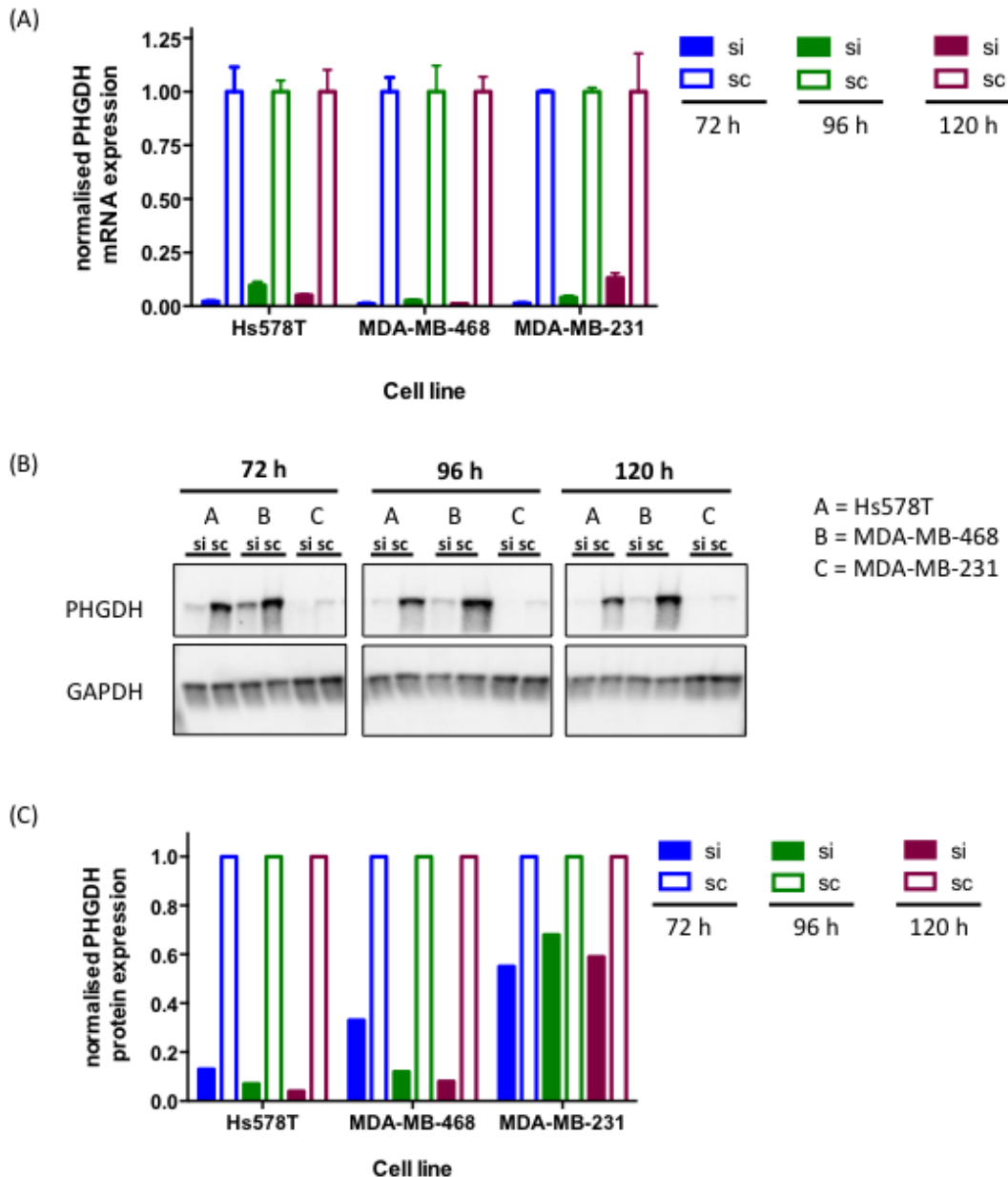
Figure 3.11 – siRNA-mediated knockdown of PHGDH in breast cancer cell lines

Cells were seeded into wells of a 6-well plate at a density of 2×10^5 cells/well and treated with 5 nM siRNA (pool of four individual siRNAs) or scrambled siRNA. Cells were harvested and lysed or RNA isolated after 72, 96 or 120 hours. Proteins were separated by gel electrophoresis and detected using Western blotting before visualisation by chemiluminescence. (A) Relative mRNA expression levels as detected by RT-PCR using GAPDH as reference gene and normalised to respective scrambled siRNA control. (B) Western blots showing PHGDH and GAPDH protein levels. (C) Quantification of PHGDH protein levels detected by Western blotting and normalisation to total protein content of the sample and to the scrambled siRNA control.

76

To determine whether PHGDH knockdown had an effect on cell survival, the three breast cancer cell lines were treated with 5 nM siRNA for 48 hours and assessed for colony formation. In general, scrambled and mock controls had only minor impact on cell survival, with the exception of mock treatment of MDA-MB-468 (Figure 3.12 B). Comparison of the siRNA-treated cells with the respective scrambled controls showed that as expected, cell growth of MDA-MB-231 was not impaired and the growth of MDA-MB-468 cells was significantly reduced by about 47 % (Figure 3.12 A). In contrast to literature findings, the growth of Hs578T cells was not affected by PHGDH knockdown, indicating that a *PHGDH* amplification has to be present so that knockdown induces a growth inhibiting effect.
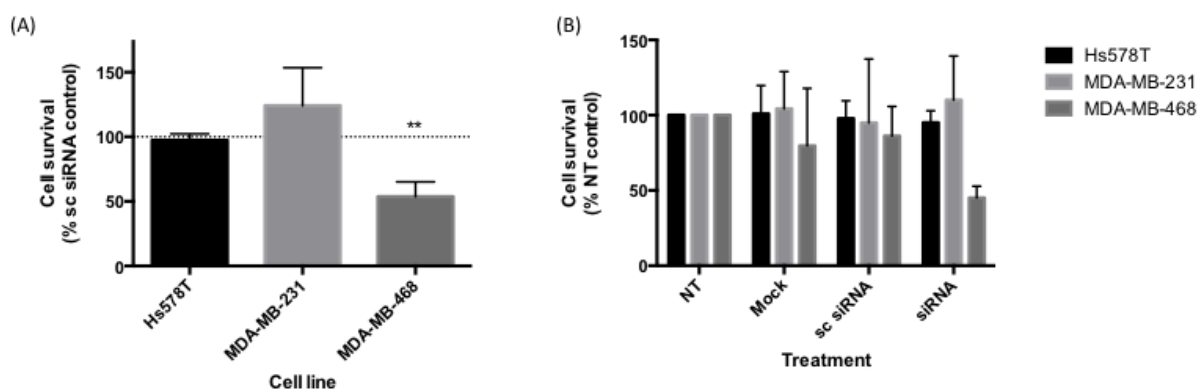
Figure 3.12 – Sensitivity of breast cancer cell lines to PHGDH knockdown

Hs578T, MDA-MB-231 and MDA-MB-468 cells were seeded at 2×10^5 cells/well in a 6-well plate and treated with 5 nM on-target or scrambled siRNA, lipofectamine (mock control) or medium only (NT control). After 48 hours, cells were reseeded in fresh medium at low density and grown for 14 days to allow colony formation. Colonies (> 30 cells) were fixed using Carnoy's fixative, stained with 0.4 % crystal violet and counted. The plating efficiency was calculated and normalised to (A) scrambled control or (B) NT control. Graphs represent averages and SD of three independent experiments with three intra-assay replicates. Significant reduction of cell growth compared to control sample was determined by paired t-test; ** $p = 0.0022$.

In addition, live cell imaging was used to measure cell proliferation in the three breast cancer cell lines over 7 days following 48 h of siRNA knockdown. In agreement with colony formation assays, knockdown of PHGDH in MDA-MB-468 cells had an effect on cell growth compared to not-treated (NT) cells or control treatment (sc siRNA or mock) (Figure 3.13B). MDA-MB-231 were overall unaffected in their growth by PHGDH knockdown whereas data for Hs578T cells was inconclusive as sc siRNA, and to a larger extent, mock treatment impaired cell growth (Figure 3.13 A and C). However, maximum confluency achieved after 7 days was low for all cell lines, so

that either seeding out more cells or observing the cells for longer would have obtained a better read-out.
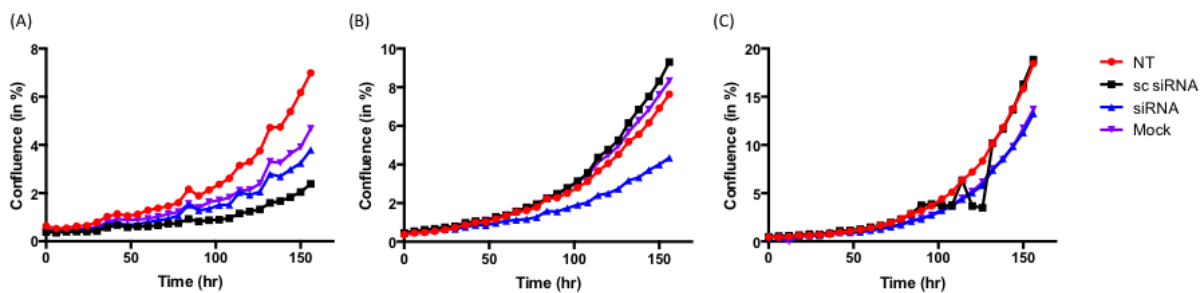
Figure 3.13 – Analysis of cell confluence using live cell imaging.

Hs578T (A), MDA-MB-468 (B) and MDA-MB-231 (C) cells were seeded at 2×10^5 cells/well in a 6-well plate and treated with 5 nM on-target or scrambled siRNA. After 48 h cells were reseeded and lower density and grown at 37 °C over 7 days and imaged every 6 hours. Cell confluence analysis was performed using the IncuCyte analysis software. Data shows one representative experiment.

PHGDH knockdown had also been shown to impair cell growth in melanoma cell lines with *PHGDH* copy number gain. Therefore the melanoma cell lines SK-Mel28, which contains a *PHGDH* amplification, and a BRAF-mutant cell line with low PHGDH protein expression were investigated. Cell lines were treated with 5 nM siRNA for 48 hours and reseeded to allow analysis of subconfluent cell samples for longer times after initial siRNA treatment. PHGDH protein levels were still significantly reduced after 168 h and 240 h in SK-Mel28 and BRaf-mutant cells, respectively (Figure 3.14 A and B). This was in agreement with reduced mRNA levels with the exception of the 120 h time-point in BRaf-mutant cells, which might be due to an experimental error in the RT-PCR as later timepoints showed significantly reduced mRNA levels in this cell line (Figure 3.14 C).
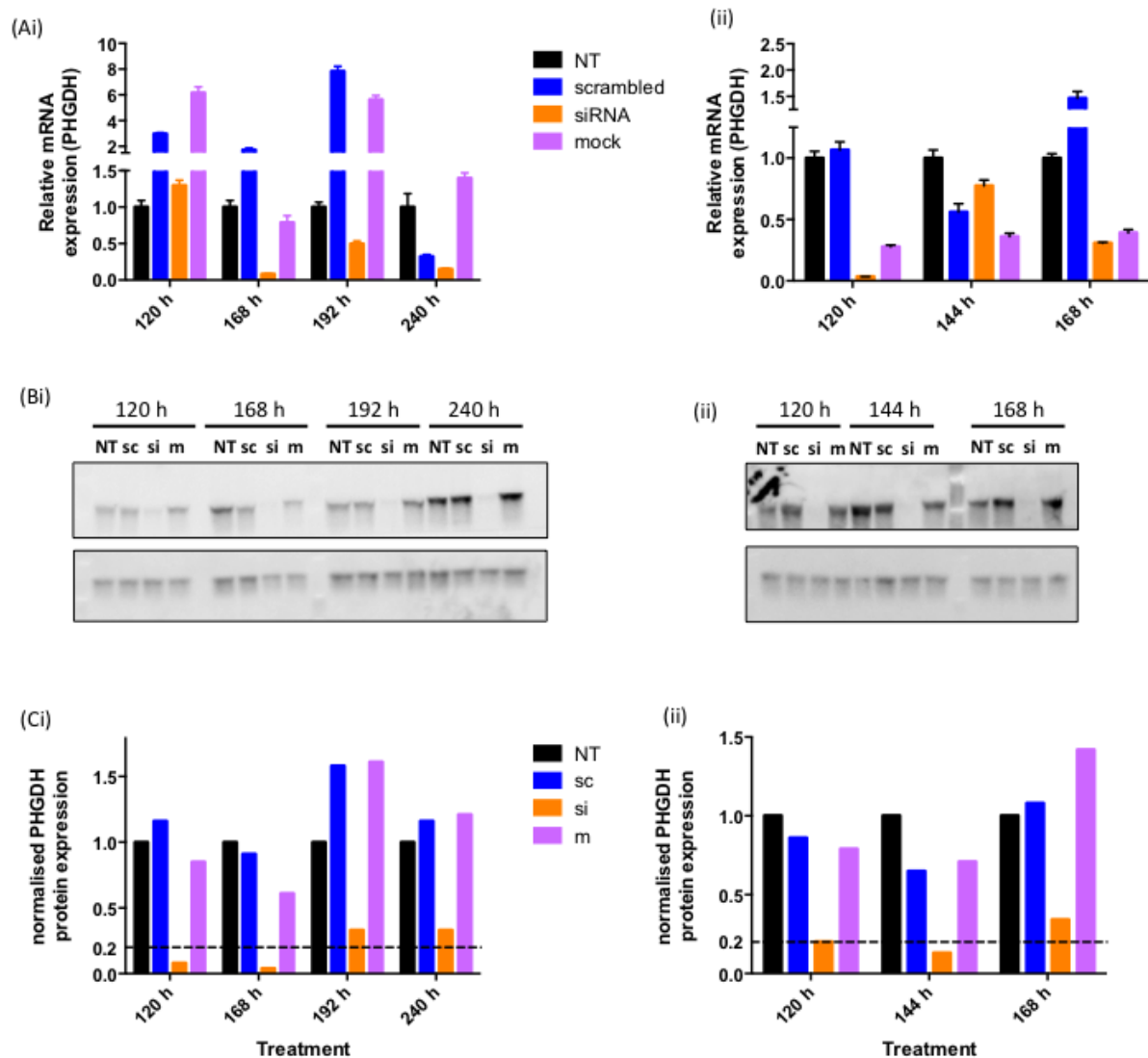
Figure 3.14 - siRNA-mediated knockdown of PHGDH in SK-Mel28 and BRaf-mutant cell line

Cells were seeded into wells of a 6-well plate at a density of 2×10^5 cells/well and treated with 5 nM of a pool of four on-target siRNAs, scrambled siRNA or lipofectamine only control (mock, m) or not-treated (NT). Cells were harvested after 48 hours, reseeded at lower density and grown for different times. At indicated timepoints cells were lysed or RNA isolated. Proteins were separated by gel electrophoresis and detected using Western blotting before visualisation by chemiluminescence. (A) Relative mRNA expression levels as detected by RT-PCR using (i) GAPDH or (ii) β_2-microglobulin as reference gene and normalised to respective NT control. (B) Western blots showing PHGDH and GAPDH protein levels in (i) BRaf-mutant cells and (ii) SK-Mel28. (C) Quantification of PHGDH protein levels detected by Western blotting and normalisation to total protein content of the sample and to the NT control.

Both cell lines were investigated in clonogenic survival assays following PHGDH knockdown. Although cell proliferation was slightly reduced in both cell lines compared to medium only control, the reduction was not significant when compared to cells treated with scrambled siRNA (Figure 3.15).
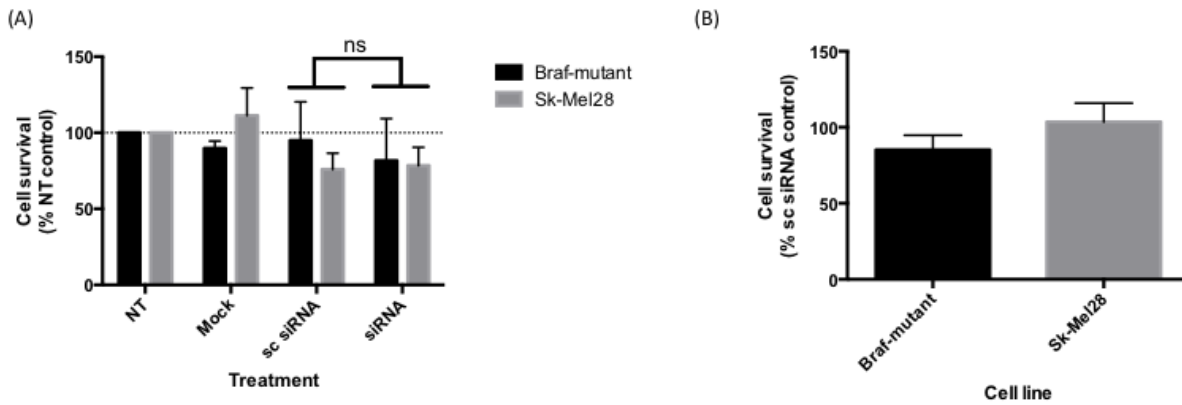
Figure 3.15 – Sensitivity of melanoma cell lines to PHGDH knockdown

BRaf-mutant and SK-Mel28 cells were seeded at $2x10^5$ cells/well into a 6-well plate and treated with 5 nM on-target or scrambled siRNA, lipofectamine (mock control) or medium only (NT control). After 48 hours cells were reseeded in fresh medium at low density and grown for 14 days for colony formation. Colonies (> 30 cells) were fixed using Carnoy's fixative, stained with 0.4 % crystal violet and counted. The plating efficiency was calculated and normalised to (A) NT control (B) scrambled control. Graphs represent averages and SD of three independent experiments with three intra-assay replicates.

3.4.4 Development of stable cell lines with tetracycline-inducible shRNA expression

The reduction in cell growth in MDA-MB-468 cells after siRNA-mediated knockdown, though significant, was less prominent than reported in the literature when using shRNA, whereas no effect at all was seen in Hs578T [72]. siRNA is typically used for short-term knockdown of target molecules, whereas shRNA-mediated knockdown can be performed for longer times [132]. With PHGDH being a relatively stable protein, this difference in the duration of knockdown was suspected to have an influence on the growth inhibition studies. Although siRNA and shRNA both direct mRNA cleavage and degradation by binding to their complementary mRNA strand, the delivery mechanism of the double-stranded RNA into the cells is different. siRNA is brought into the cells with the help of a delivery system, such as lipofectamine, whereas the DNA encoding the shRNA is integrated into the cell genome with the help of lentviral vectors.

E. coli cells containing the pTRIPZ expression vector with the DNA sequence of the respective shRNA were amplified and the plasmid DNA extracted. The obtained plasmid DNA, together with the packaging psPA2X vector and the pMD2.G envelope plasmid, were used to transfect HEK293 cells from which the final virus particles could be extracted. The breast cancer cell lines Hs578T, MDA-MB-468 and MDA-

MB-231 were transduced with 10 and 50 µL of two on-target virus stocks or the non-target control virus. After 48 hours, puromycin selection of the transduced cells was started using medium containing 1 µg/mL puromycin, as this was the lowest dose of puromycin found to kill 100 % of non-transduced cells (Figure 3.16).
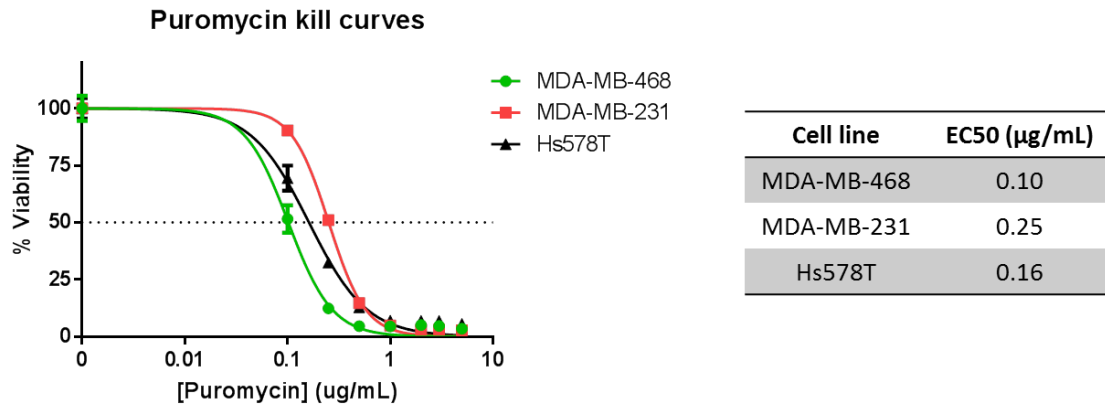
Figure 3.16 - Puromycin kill curves for different breast cancer cell lines.

Untransduced MDA-MB-468, MDA-MB-231 and Hs578T cells were treated with increasing amounts of puromycin (0-5 µg/mL) and left under puromycin selection for 72 hours. 1 µg/mL was found to be a sufficient amount of puromycin to kill all non-transduced cells within 72 hours in all three cell lines.

In order to test if the cell lines had been efficiently transduced, each newly generated shRNA-expressing cell line was seeded into 6-well plates in medium with or without 0.5 or 1 µg/mL doxycycline for 72 hours. Doxycycline binds to the tetracycline-inducible TRE promoter and thus activates the expression of red-fluorescent protein and shRNA, so that efficiency of transduction can be observed using a fluorescent microscope. All three cell lines transduced with both amounts of virus stock (10 and 50 µL) showing efficient transduction (Figure 3.17).
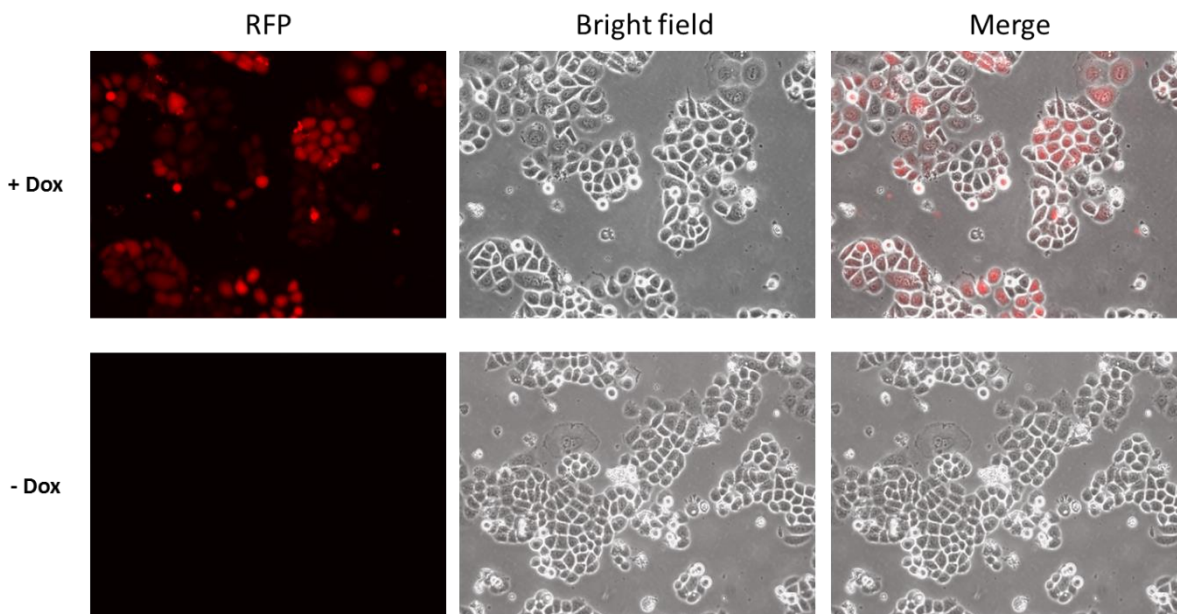
Figure 3.17 - RFP expression in transduced MDA-MB-468 following exposure to doxycycline.

MDA-MB-468 cells were transduced with 10 µL shB03 virus and left to grow for 24 hours. Addition of doxycycline (1 µg/mL) to the medium resulted in RFP and shRNA expression with RFP expression being measured after 72 hours.

All shRNA-containing cell lines were assessed for the effectiveness of PHGDH knockdown following 72 hours of doxycycline treatment at 0, 0.5 or 1 µg/mL through analysis of mRNA and protein levels. Analysis of the mRNA levels showed overall good reduction of PHGDH mRNA levels after doxycycline treatment in all three breast cancer cell lines transduced with on-target shRNA. In Hs578T and MDA-MB-231 the knockdown in the on-target cell lines (B03 and B04) was of similar level independent of the amount of virus (10 or 50 µL) used for transduction, whereas in the respective cell lines of MDA-MB-468, the cells transduced with higher amounts of virus had better knockdown results at least after treatment with 0.5 µg/mL doxycycline (Figure 3.18). Analysis of the protein levels in the different cell lines showed that, as expected, the NTC-shRNA had no effect on PHGDH levels, whereas cell lines containing the plasmid for the two on-target shRNAs (B03 and B04) showed a reduction in PHGDH protein levels after doxycycline treatment (Figure 3.19). However, PHGDH protein levels were only reduced about 50-60 % with no significant increase in the cell lines transduced with more virus particles (50 µL). Overall, in cells treated with 1 µg/mL doxycycline the PHGDH levels were slightly lower compared to 0.5 µg/mL doxycycline treatment.
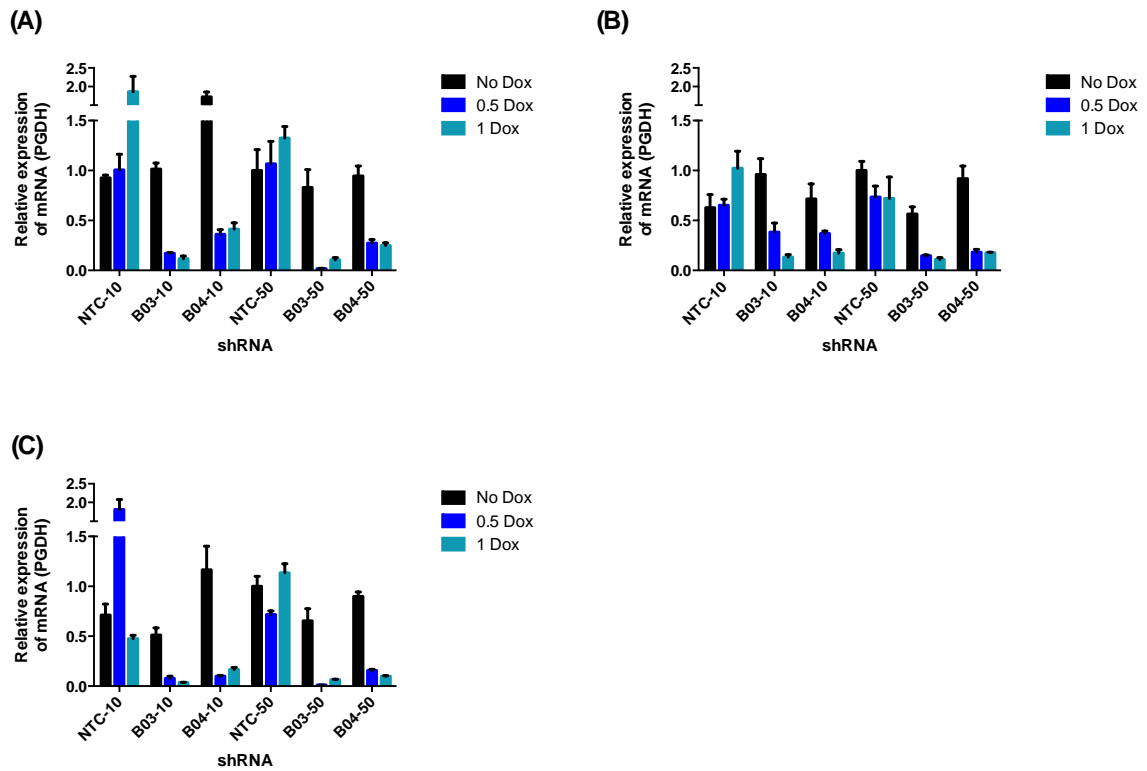
Figure 3.18 – mRNA levels in shRNA-expressing breast cancer cell lines after doxycycline treatment.

Cells were seeded into wells of a 6-well plate at a density of 2×10^5 cells/well and treated with 0 (No Dox), 0.5 (0.5 Dox) or 1 (1 Dox) µg/mL doxycycline for 72 hours with fresh medium applied every 24 hours. Cells were harvested, lysed and mRNA levels determined by RT-PCR in (A) Hs578T, (B) MDA-MB-468 and (C) MDA-MB-231. mRNA levels of PHGDH were normalised to mRNA levels of β2m and expressed relative to the No Dox treatment of the respective NTC-50 cell line.
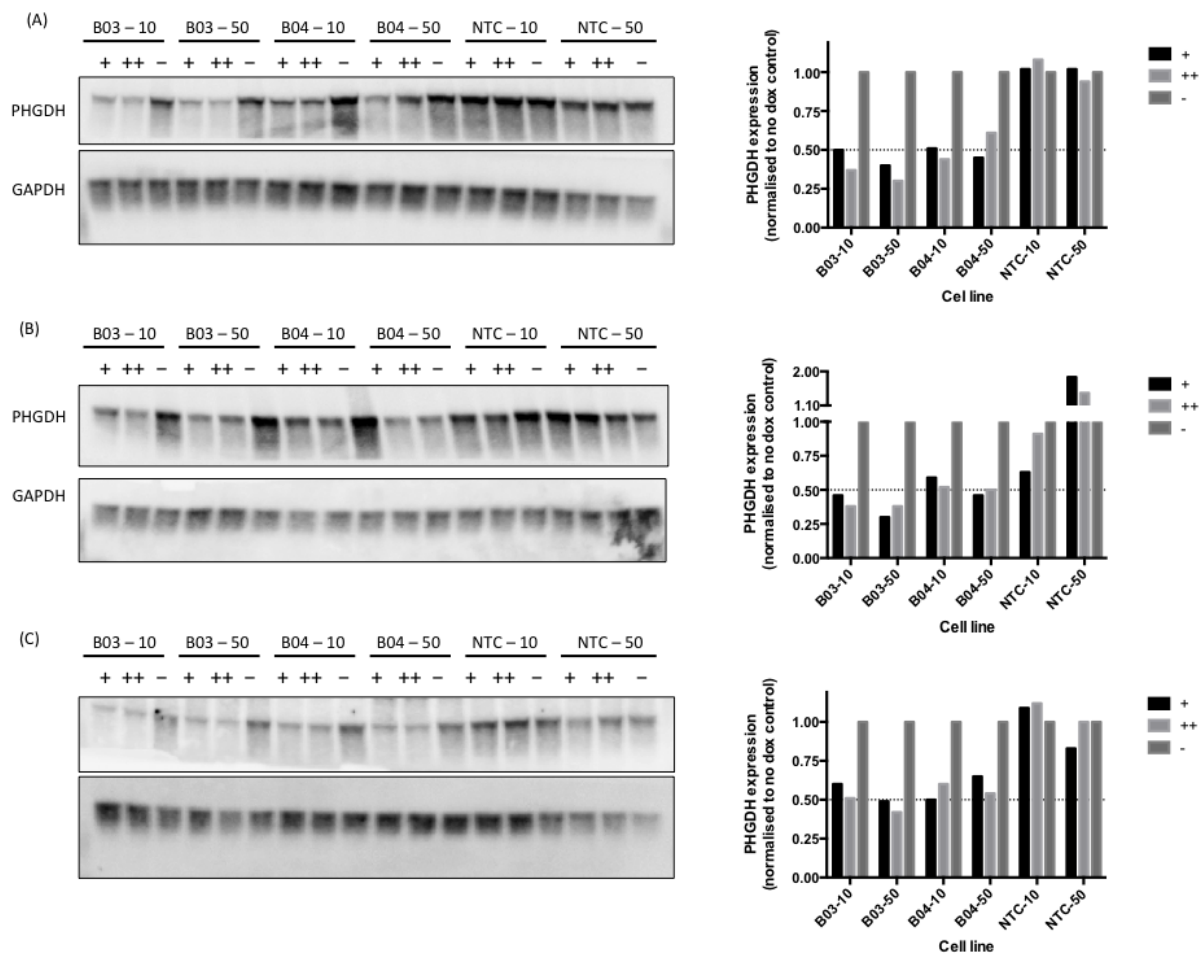
Figure 3.19 – PHGDH protein levels in shRNA-expressing breast cancer cell lines after doxycycline treatment.

Cells were seeded into wells of a 6-well plate at a density of 2×10^5 cells/well and treated with 0 (-), 0.5 (+) or 1 (++) µg/mL doxycycline for 72 hours with fresh medium applied every 24 hours. Cells were harvested, lysed and proteins were separated by gel electrophoresis and detected using Western blotting before visualisation by chemiluminescence. Western blots showing PHGDH and GAPDH protein levels as well as quantification of protein levels in (A) Hs578T, (B) MDA-MB-468 and (C) MDA-MB-231.

Since transduction with 50 µL virus did not significantly improve the knockdown of PHGDH compared to 10 µL virus, cells transduced with the lower amount of virus were taken forward in order to minimise the likelihood of integration of multiple lentiviral vector copies per cell.

The effect of PHGDH knockdown was assessed in the transduced cell lines in colony formation assays. Cells were maintained in medium with (1 µg/mL) or without doxycycline for 72 hours with fresh medium applied every 24 hours. Cells were then seeded at low density in medium with or without doxycycline and left to proliferate for 14 days. Comparison of the doxycycline-treated cell lines compared to the non-

treated counterparts showed that treatment with doxycycline increased cell proliferation independently of which shRNA was expressed.
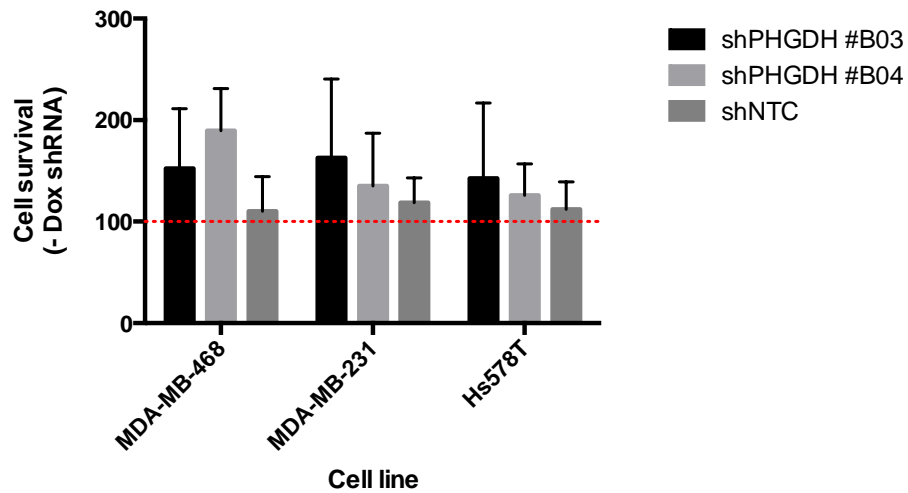
Figure 3.20 - Sensitivity of breast cancer cell lines to PHGDH knockdown assessed by colony formation.

Hs578T, MDA-MB-231 and MDA-MB-468 cells stably transduced with on-target (B03 and B04) or control (NTC) shRNA were split into two flasks with one being treated with 1 µg/mL doxycycline. After 72 hours of doxycycline treatment, the cells were reseeded in fresh medium (+/- doxycycline) at low density and grown for 14 days for colony formation. Colonies (> 30 cells) were fixed using Carnoy's fixative, stained with 0.4 % crystal violet and counted. The plating efficiency was calculated and normalised to the respective non-treated cell sample. Data represent mean and error of two independent experiments with three intra-assay replicates.

In addition to colony formation assays, growth inhibition assays were performed to investigate the effect of PHGDH knockdown on cell proliferation over a shorter period of time (7 days) with a daily read-out. No effect was seen on cell growth in any of the transduced cell lines after doxycycline treatment (Figure 3.21).
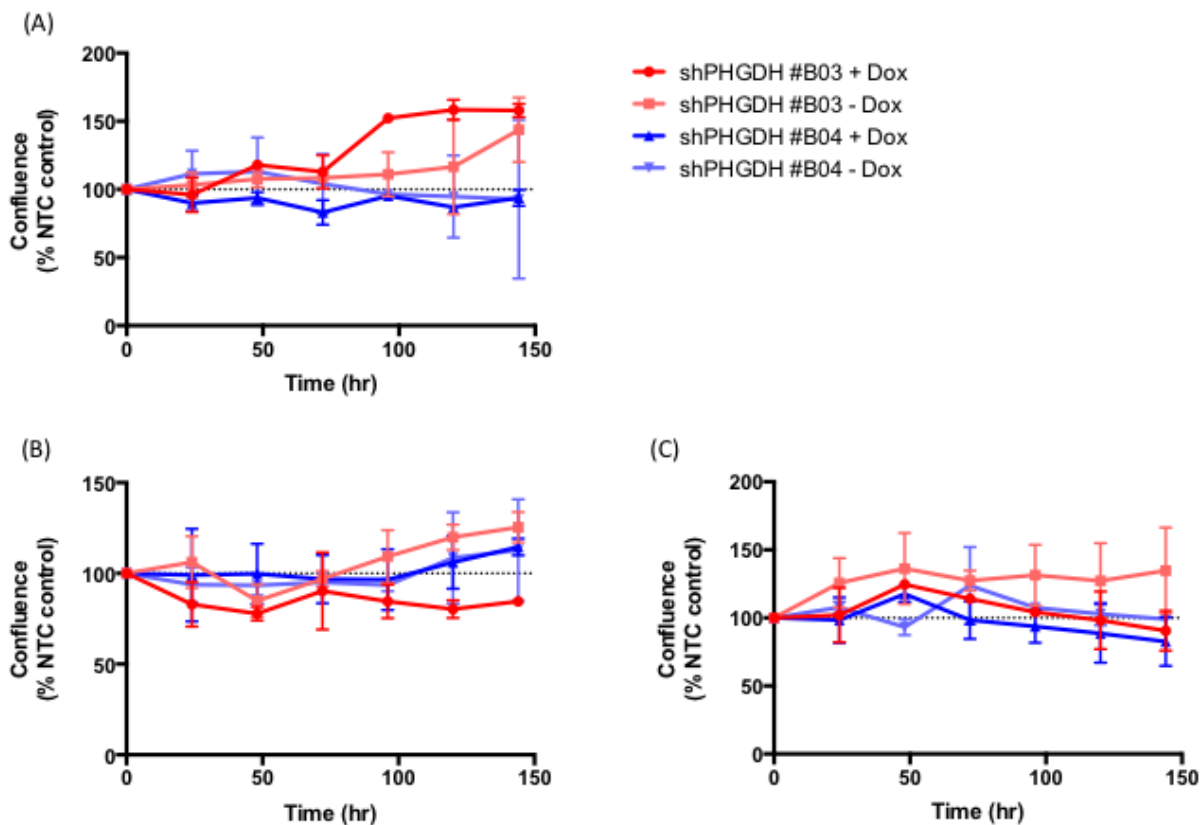
Figure 3.21 - Sensitivity of breast cancer cell lines to PHGDH knockdown assessed by SRB growth assay.

(A) MDA-MB-468, (B) Hs578T and (C) MDA-MB-231 cells stably transduced with on-target (B03 and B04) or control (NTC) shRNA were split into two flasks with one being treated with 1 µg/mL doxycycline. After 72 hours of doxycycline treatment, the cells were reseeded in fresh medium (+/-doxycycline) into wells of a 96-well plate with 1 x 10^4 cells (MDA-MB-231) or 2.5 x 10^4 cells/well (MDA-MB-468, Hs578T) at low density and grown for 7 days. Every 24 hours, one plate per cell line was fixed using Carnoy's fixative, stained with 0.4 % SRB before being solubilised and SRB absorbance measured. Data represent mean and error of two independent experiments with six intra-assay replicates.

3.5 Discussion

PHGDH has been validated by previous studies as a potential target in breast cancer and melanoma cell lines [69, 72]. Thorough target validation is a critical step in the drug discovery process and disagreement with literature findings is not uncommon. Therefore the work described in this chapter focused on repeating the published target validation as well as exploring PHGDH expression levels in a wider set of cancer cell lines.

PHGDH expression was investigated in a panel of 65 cancer cell lines, in order to extend the set of potential tumours that would benefit from PHGDH inhibitors beyond breast cancer and melanoma. The different cell lines investigated included cancers from about 15 different organs, which made it difficult to find a housekeeping protein to use for normalisation of protein levels in all cell lines. Variations in protein and mRNA levels of the preferentially used housekeeping genes, such as GAPDH or β-actin, have been reported with very high differences seen between different organs as well as between tumorigenic and healthy tissues [133, 134]. Therefore, to exclude introduction of variations resulting from changes in housekeeping proteins, the protein levels were normalised to the total amount of protein per lane detected by Ponceau S staining and quantified by densitometry. PHGDH protein levels were compared to PHGDH expression in K562 as reference cell line. Among the cell lines with 3.5-6-fold higher PHGDH expression compared to K562, there were, as expected, breast cancer (MDA-MB-468, MDA-MB-453 and TD47) and melanoma (CHL-1, A375) cell lines, but also a gastric cancer cell line (SNU1), a medulloblastoma cell line (DAOY) and the choriocarcinoma cell line JEG-3, which is the only one reported to have a *PHGDH* amplification [3]. In contrast to published data, T47D cells were shown to express high amounts of PHGDH [115]. However, with only one sample analysed for this cell line together with an inter-assay CV of about 29 % additional repeats would be necessary to set T47D in the right context. To date, no implication of PHGDH in gastric or uterine cancers has been reported or investigated. However, recent studies observed elevated PHGDH expression levels in astrocytic tumours with PHGDH inhibition resulting in reduced proliferation in glioma cells [77]. Hence, further investigations into other cancer types might expand the potential usage of PHGDH inhibitors.

Interestingly, PHGDH expression was correlated with c-Myc expression. c-Myc has been shown to positively regulate enzymes of the serine synthesis pathway through their mRNA expression under nutrient deprivation conditions [127]. Surprisingly, in this study c-Myc and PHGDH protein levels correlated negatively with a Spearman correlation coefficient of -0.47. However, the cell lines analysed were cultured in complete medium and therefore it would be interesting to see if different protein expression levels could be detected under nutrient deprivation. In addition, some of the cell lines reported herein have only been tested once, so that more repeats would be necessary to strengthen the findings.

Besides c-Myc, PARP1 protein levels were also explored with PARP1 being among the major NAD^+-consuming enzymes and thus regulating cellular NAD^+ levels, which in turn has implications for glycolytic enzymes using NAD^+ as a major cofactor. In the cell lines investigated, a variety of PARP1 expression levels was detected, which, however, did not correlate with PHGDH expression levels.

To validate the reported effects on cell proliferation following PHGDH knockdown, siRNA-mediated knockdown of PHGDH in three breast cancer cell lines was investigated. MDA-MB-231 which have very low PHGDH expression, Hs578T that do not contain a *PHGDH* amplification, but were found to have high PHGDH protein expression, and MDA-MB-468, which have a genomic amplification of *PHGDH* [72]. Optimisation, performed in Hs578T cells showed that 5 nM of a pool of four individual siRNAs resulted in about 90 % reduction in mRNA levels and ≥ 80 % reduction in protein levels. Similar effects on mRNA levels after PHGDH knockdown were achieved in MDA-MB-231 and MDA-MB-468 cells resulting in good reduction of protein levels in MDA-MB-468 cells. To determine the effect on cell growth, clonogenic survival after siRNA treatment was investigated. The survival rate was significantly reduced to about 50 % in MDA-MB-468 cells, which is in agreement with reported data of 57 % growth inhibition after seven days using clonogenic survival assays [115]. However, the original study from Possemato and colleagues reported a much more prominent reduction in cell growth of MDA-MB-468 cells of 80-90 % [72]. Interestingly, knockdown of PHGDH in Hs578T cells did not affect cell growth, which is in contrast to literature data reporting a similar reduction in cell growth as seen with *PHGDH* amplified MDA-MB-468 [72]. Live cell imaging experiments, the second technique used to assess cell growth, similar cell growth was observed in cells with

and without PHGDH knockdown in MDA-MB-231 cells, whereas knockdown impaired cell growth in MDA-MB-468 cells. In Hs578T cells, the mock control showed a stronger effect on cell growth compared to siRNA treated cells indicating off-target effects. Live cell imaging analysis is based on the quality of the pictures of the cells, and, although being able to distinguish between cells and plastic surface, the software only analyses cell confluence and therefore does not distinguish between larger cell numbers and enlarged cells, which might be a drawback when using this technique.

Likewise, successful PHGDH knockdown resulting in a reduction of at least 80 % at the protein level had no significant effect on cell growth in the melanoma cell line SK-Mel28 and a BRaf-mutant cell line, despite SK-Mel28 being reported as being dependent on PHGDH expression [69]. Experiments represented in the literature had been performed with stably transduced shRNA-expressing cell lines, which allow more precise manipulation and longer knockdown of the target gene, and might be a reason for differences seen compared to this study where transient siRNA-mediated knockdown was performed. On the other hand, despite siRNA-mediated knockdown being of a transient nature, analysis of PHGDH levels over 240 h in BRaf-mutant cells showed that siRNA-mediated knockdown was able to reduce PHGDH levels over this extended period.

For a better comparison with literature experiments, stably transduced Hs578T, MDA-MB-468 and MDA-MB-231 cell lines were produced. It was chosen to take the whole transduced and puromycin-selected cell populations forward, compared to picking individual clones, in order to account for cell heterogeneity. By keeping the transduced cell lines all the time in puromycin-containing medium, a constant selective pressure was imposed on the cells, so that cells lacking the transduced vector, because of unsuccessful transduction or cell leakage, should not survive. Transduction was successfully performed in all three breast cancer cell lines as seen through expression of the red-fluorescent protein (RFP) and reduced PHGDH mRNA levels. However, PHGDH protein levels after 3 days of doxycycline treatment of transduced cell lines compared to protein levels after 3 days in siRNA-treated cells showed that the knockdown was less effective in the stably transduced cell lines with about 50 % reduction in PHGDH compared to 80-90 % when using siRNA. Insufficient knockdown may be the reason why no effect on cell growth, either

determined by colony formation or in a growth inhibition assay based on the measurement of cellular protein content (SRB assay), was detected in all three breast cancer cell lines. Despite stable knockdown of PHGDH in cells transduced with two different on-target shPHGDH constructs, Chen and colleagues reported differential effects on growth inhibition in the same cell line depending on which shPHGDH construct was expressed as determined by monitoring cellular ATP content [115]. However, one has to bear in mind that measurement of the ATP content is an indirect read-out of cell proliferation as through ATP it measures active cell metabolism and, despite being frequently used, has been shown to be less reliable for cell enumeration compared to SRB assay or colony formation assays [135].

This study has extended the range of cell lines revealing various protein expression levels of PHGDH and showing that cancer types other than breast cancer and melanoma are worth further investigation and might be targetable with PHGDH inhibitors. Attempts to reproduce literature evidence for dependence of cell proliferation on PHGDH were partly successful as a significant effect was seen after siRNA knockdown in the breast cancer cell line MDA-MB-468 and no effect in the control MDA-MB-231 cells. However, reported growth inhibition after PHGDH knockdown in Hs578T cells or the melanoma cell line SK-Mel28 could not be reproduced. From the siRNA and shRNA-mediated knockdown experiments, it was found that the level of knockdown of PHGDH was crucial to affect cell growth and that 50 % PHGDH reduction at protein level was insufficient to affect cell growth. Furthermore, the performed experiments showed that the assay format used to assess cellular growth was critical.

Chapter 4. Investigating the substrate-binding pocket

4.1 Introduction

Among the most remarkable characteristics of enzymes is their specificity in regards to the substrate used as well as the reaction catalysed. The degree of specificity can vary from absolute specificity, where the enzyme will only catalyse a reaction for one specific substrate, which is the expected case for PHGDH, to relatively broad specificity where the enzyme catalyses a reaction on a group of substrates. For example, lipases hydrolyse ester bonds in a variety of triglycerides [136]. To inhibit an enzyme through interaction with the substrate-binding site, it is necessary to understand which structural features of the enzyme define its substrate specificity and also how to avoid off-target effects on other proteins. It is especially critical to consider potential effects on other members of the same protein family as a high degree of similarity in sequence and/or tertiary structure is often observed within a protein family.

Dehydrogenases belong to the family of oxidoreductases, which accounts for about 23 % of the enzymes currently targeted with FDA-approved drugs [137]. Dehydrogenases catalyse redox reactions in which the substrate is oxidised through the transfer of a hydride ion (H^-) onto a cofactor such as $NAD(P)^+$ or FAD (Flavin adenine dinucleotide). For this transfer reaction to take place, the nicotinamide moiety of the cofactor and the substrate have to be in close proximity (**Figure 4.1**). There are many NAD^+-requiring dehydrogenases and the binding site for NAD^+ is very similar among them [85]. Specificity rather derives from the substrate-binding site, making this site more interesting for the design of specific inhibitors. The substrate-binding site of dehydrogenases is often a hydrophilic pocket preferentially accommodating molecules that are charged at physiological pH, such as acids. Mimicry of an acid functional group is, however, not ideal for a drug due to limited passive diffusion of charged species across biological membranes as well as undesirable metabolic effects and toxicity in *in vivo* studies [138].
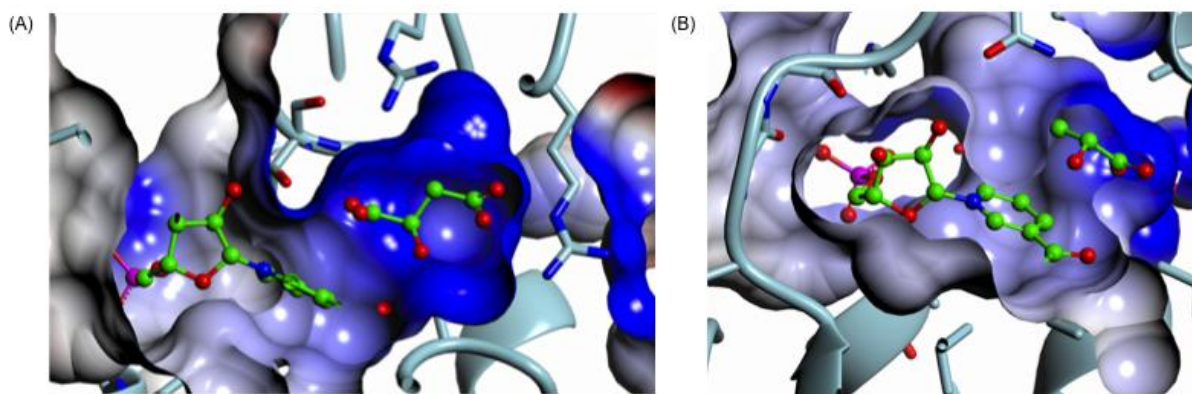
Figure 4.1 – Substrate-binding sites of human NAD$^+$-requiring dehydrogenases.

Substrate binding site of (A) PHGDH with *D*-malate bound to the substrate-binding site (PDB 2G76) and (B) lactate dehydrogenase with *D*-lactate bound (PDB 4QO8). The protein surface around the substrate-binding site is coloured by electrostatic potential: red is negatively charged and blue is positively charged. Ligands are coloured by atom type with green = carbon, red = oxygen and blue = nitrogen.

Despite these drawbacks, carboxylic acid functionalities are found in a large number (>450) of approved drugs [138]. In addition, for another dehydrogenase, lactate dehydrogenase A (LDHA), recent examples of inhibitors of the human enzyme all contained an ionisable, acidic functional group, which was found to be important to sustain the inhibitory activity [139]. To overcome problems of limited trans-cellular absorption, ester prodrugs, which mask the acidic function, have been developed [140, 141]. Esterification improves the lipophilicity of the compound and allows better transfer across the cell membrane and is often reversible in the cell by hydrolysis of the ester to the active acidic acid. However, development of ester prodrugs with a suitable level of ester stability and desired pharmacokinetic parameters can be difficult [141].

4.2 Aims and Objectives

The overall aim of this chapter was to investigate the substrate-binding pocket of PHGDH and to identify compounds that could potentially target this site. As a starting point, fragments containing the substrate analogue malate were investigated, as a crystal structure of a truncated form of human PHGDH (sPHGDH) with *D*-malate bound to the substrate-binding site was known (PDB 2G76). In addition, a wider range of substrate analogues was probed as potential binding partners for PHGDH. A further objective was the crystallisation of sPHGDH in the presence of substrate analogues or potential substrate-binding site-directed inhibitors.

4.3 Materials and Methods

All chemicals listed in the materials and methods section were of analytical grade and obtained from Sigma Aldrich, St. Louis, MO, USA, unless otherwise stated.

4.3.1 Buffers

All buffers were as described in section 2.2 apart from the following:

Affinity binding buffer 50 mM NaH$_2$PO$_4$, pH 8, 300 mM NaCl, 10 mM Imidazole, 2 mM DTT

Affinity wash buffer 50 mM NaH$_2$PO$_4$, pH 8, 300 mM NaCl, 50 mM Imidazole, 2 mM DTT

Affinity elution buffer 50 mM NaH$_2$PO$_4$, pH 8, 300 mM NaCl, 250 mM Imidazole, 2 mM DTT

4.3.2 Differential scanning fluorimetry

Differential scanning fluorimetry assays were run as described in section 2.9. The assay contained 25 mM HEPES, pH 7.5 with 1 x Sypro Orange, 1 µM PHGDH and 200 µM ligand. All ligands were prepared fresh prior to the assay by dissolving in 25 mM HEPES, pH 7.5 and the pH of the solution was adjusted to about 7.5. Concentrations of stocks of cofactors NAD$^+$ and NADH in 25 mM HEPES, pH 7.5, were determined using the Beer-Lambert law (Equation 4.1) by measuring the absorbance at 260 and 340 nm respectively and using the specific extinction coefficients of 18 000 M^{-1}cm^{-1} and 6 220 M^{-1}cm^{-1} for NAD$^+$ and NADH respectively.

$$A = \varepsilon \times c \times d$$

Equation 4.1 – Beer-Lambert law

A = absorbance, ε = extinction coefficient (M^{-1} cm^{-1}), c = concentration of solution (M), d = length of solution the light passes through (cm) (generally 1 cm cuvette).

4.3.3 Fluorescence polarisation assay

Fluorescence polarisation assays were carried out as described in section 2.11. Briefly, the assay mixture contained 40 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP or 2 mM DTT, 1 µM NADH and 2 µM PHGDH or sPHGDH. Adenosine 5'-diphosphoribose (ADPR) was dissolved in 40 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP or 2 mM DTT and screened at a concentration range from 0.5 and 1 mM. The reaction mixture with ADPR was incubated for 30 min at RT and fluorescence

polarisation was measured using a PHERAstar FS platereader (BMG Labtech, Ortenberg, Germany) with an FP filter module (λ_{ex} = 340 nm and λ_{em} = 440 nm).

4.3.4 Protein purification

Full length PHGDH (aa 1-533) and its truncated version sPHGDH (aa 3-314) were expressed in *E. coli* Rosetta(DE3) cells and purified according to section 2.6.1 except that for initial purification the affinity purification step was performed by gravity-flow chromatography using Ni-NTA agarose beads (QIAGEN, Limburg, Netherlands). The gravity-flow column was equilibrated with affinity binding buffer prior to application of clarified protein lysate. Non-specifically bound protein was removed by washing the column with 10 column volumes (cv) of affinity binding buffer, followed by 10 cv of affinity wash buffer. The desired protein was eluted with 5 cv affinity elution buffer. Fractions were analysed by SDS-PAGE, concentrated using centrifuge concentrators with 10 kDa or 30 kDa molecular weight cut-off membranes for sPHGDH and PHGDH respectively. Purification was continued by size exclusion and anion exchange chromatography as described in section 2.6.1. Both proteins were concentrated to 10 – 15 mg/mL and either used directly or flash frozen in liquid nitrogen and stored at -80 °C.

4.3.5 Crystallography

Initial screening of sPHGDH against sparse matrix screens and crystal optimisation followed the protocols described in section 2.12. For co-crystallisation, sPHGDH was mixed with ligands and incubated at 4 °C overnight before setting up sitting-drop vapour diffusion crystal trials. Crystals for which datasets were collected were analysed as described in section 2.13.

4.4 Results

4.4.1 Investigation of malate-containing inhibitors

The co-crystal structure of sPHGDH with NAD$^+$ was available in the public domain (PDB 2G76) and showed NAD$^+$ as well as *D*-malate bound to the active site with *D*-malate sitting in the substrate-binding site (Figure 4.2 A). Incorporation of *D*-malate into the sPHGDH-NAD$^+$ crystals was an unintended consequence of the presence of 10 mM malate in the crystallisation buffer, which contained 0.1 M MMT buffer. MMT buffer is a mixture of *DL*-malate, MES and TRIS (1:2:2 ratio). *D*-Malate is a structural analogue of 3-phosphoglycerate (Figure 4.2 B) and is presumed to bind in a similar manner to the substrate-binding site as the bona fide substrate.
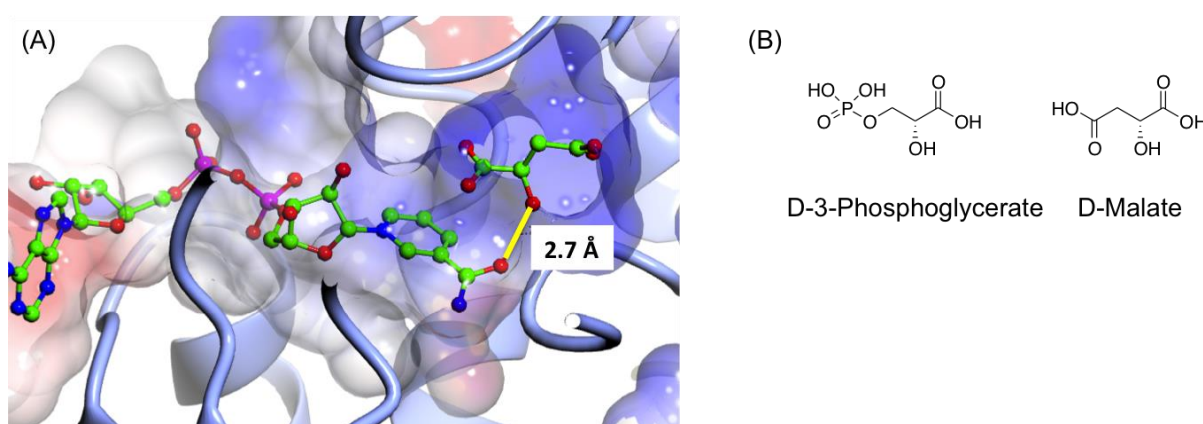
Figure 4.2 - Detail of the active site of sPHGDH (PDB 2G76) with bound NAD$^+$ and *D*-malate.

Ribbon diagram of the active site of one molecule of sPHGDH shown in blue with bound NAD$^+$ and *D*-malate coloured by atom type as in **Figure 4.1**. The distance between the amide oxygen of NAD$^+$ and the hydroxyl group of *D*-malate was measured as 2.7 Å and is indicated with a yellow line. The substrate surface is coloured by electrostatic potential: red is negatively charged and blue is positively charged.

With *D*-malate known to bind to PHGDH, the first set of potential inhibitors to be investigated comprised malate-containing fragments. To generate a bisubstrate inhibitor, a malate moiety (mimicking *D*-3-Phosphoglycerate) should ideally be linked to an aromatic group (mimicking the nicotinamide moiety of NAD$^+$). A search among commercially available compounds identified a small library of seven malate-containing fragments (Figure 4.3). Only one of these fragments, 1,2,3-butanetricarboxylic acid (**7**), did not contain an aromatic group, but rather an aliphatic side-chain with a carboxylic group that could potentially be used for further derivatisation. For the remaining compounds (**1**-**6**), the linker between malate and the

aromatic ring varied from zero to three atoms. The distance between the hydroxyl group of malate and the oxygen of the amide group in nicotinamide was measured to be 2.7 Å. The malate-containing fragments with different linker lengths therefore provide the possibility to experimentally determine the best length (Figure 4.2).

2-[2-(4-Hydroxyphenyl)-2-oxoethyl]succinic acid	Phenylsuccinic acid	3-Indoleacetyl-DL-aspartic acid	DL-Benzylsuccinic acid
(1)	**(2)**	**(3)**	**(4)**

Carbobenzyloxy-DL-aspartic acid	Diethyl 2-[(4-Aminobenzyl)amino] succinate	1,2,4-Butanetricarboxylic acid
(5)	**(6)**	**(7)**
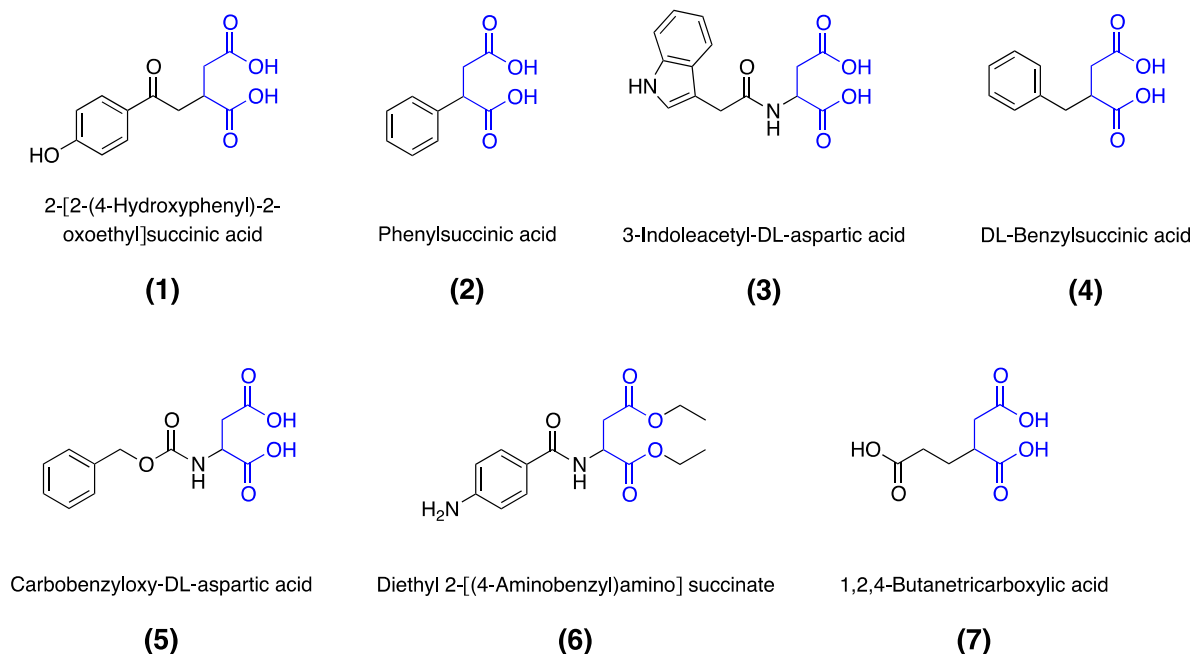
Figure 4.3 - Structures of investigated malate-containing fragments.

The malate moiety of each fragment is highlighted in blue.

These malate-containing compounds were screened in a fluorescence polarisation assay with NADH as fluorophore. The principle of the fluorescence polarisation assay exploits the finite lifetime of a fluorophore's excited state. When excited with polarized light, the short rotational correlation time of a rapidly tumbling fluorophore allows it to adopt a random orientation during the lifetime of the excited state, and therefore emit depolarized light. Upon binding to a larger molecule, e.g. a protein, the fluorophore will tumble more slowly so that light it emits remains. If an inhibitor binds to the binding site of the fluorophore, a decrease in polarised light can be measured as the fluorophore is released once again into solution (Figure 4.4). NADH has a fluorescence emission peak at 440 nm after excitation at 340 nm and has been successfully used as fluorophore in fluorescence polarisation experiments [142].
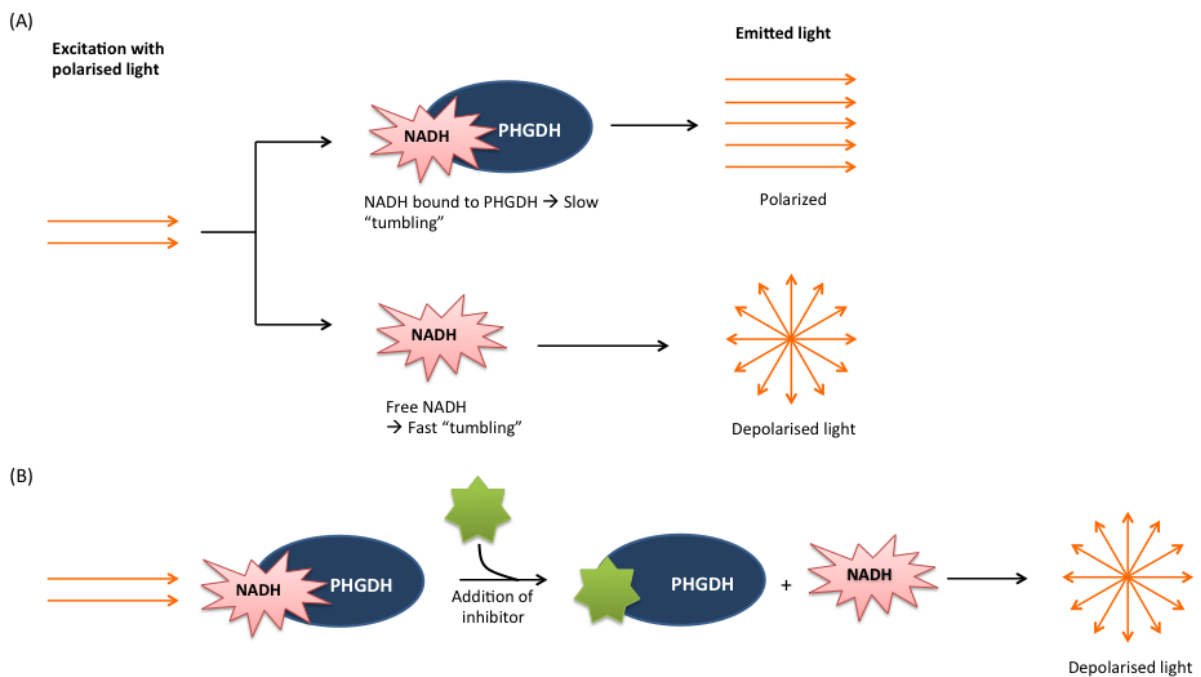
Figure 4.4 – Principle of fluorescence polarisation assay.

(A) After excitation of a fluorophore (e.g. NADH) in buffer solution, no polarised light can be detected as the small fluorophore is tumbling fast. It is not until the rotational correlation time of the fluorophore is increased by binding to a larger molecule, *e.g.* protein, that polarised light can be measured. (B) Addition of an inhibitor to the fluorophore-protein mixture replaces the fluorophore at its binding site on the protein, so that an increase in depolarised light is detected.

A fluorescence polarisation (FP) assay protocol was established using 1 µM NADH in 40 mM HEPES, pH 7.5 with 0.05 % BSA added to account for nonspecific protein binding. The direct binding curves of sPHGDH or PHGDH into 1 µM NADH gave a K_D for NADH of 2.0 ± 0.3 µM and 2.4 ± 0.5 µM for PHGDH and sPHGDH, respectively (Figure 4.5 A). Titration of NADH into buffer showed that up to 1 µM NADH the fluorescence polarisation was constant (Figure 4.5 B).
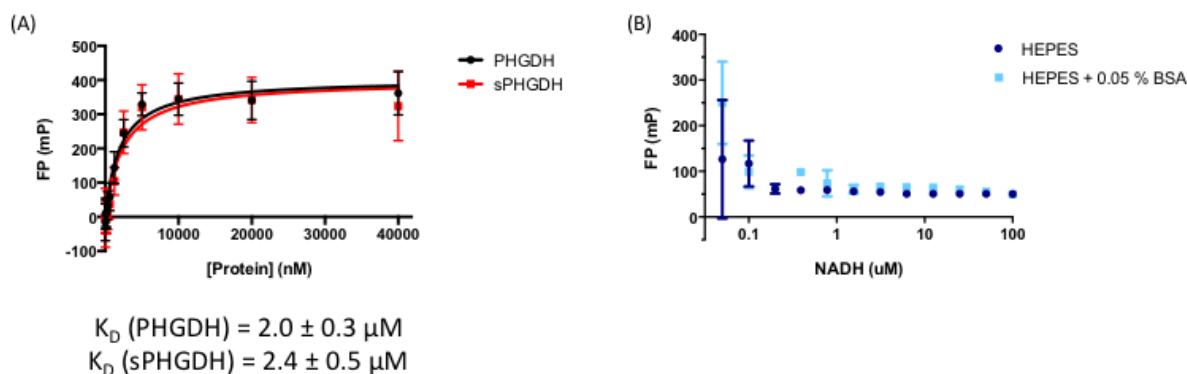
Figure 4.5 – Protein and fluorophore titration curves for FP Assay

FP ($\lambda_{ex}/\lambda_{em}$= 340/440 nm) was measured with increasing concentrations (20-40 000 nM) of sPHGDH or PHGDH (A) in the presence of 1 µM NADH. (B) Titration of increasing concentrations of NADH (0.05-100 µM) into buffer only (+/- BSA).

To validate the assay the cofactor analogue ADPR was used. ADPR is structurally identical to NAD$^+$ but lacks the nicotinamide moiety (Figure 4.6 C). With the addition of increasing concentrations of ADPR, a decrease in polarised light was observed due to NADH displacement by ADPR. The data were transformed into % displacement of NADH on PHGDH and sPHGDH, and IC$_{50}$ values of 282 ± 1 µM and 137 ± 1 µM were obtained, respectively. These are the concentrations of ADPR at which 50 % of NADH was displaced from the respective protein. However, although a displacement of NADH was observed with increasing concentrations of ADPR, a general increase in fluorescence was also seen indicating interference of ADPR with the fluorescence of NADH and therefore rendering the calculated IC$_{50}$ values less reliable. Moreover, the Z-score, which is the number of standard deviations from the mean, was found to vary for the different runs of the experiments, ranging from -0.3 to 0.9. In addition, NAD$^+$ was also investigated for its potential to displace NADH at the binding pocket and was found to be a much weaker binder compared to ADPR, being only able to displace about 35 % NADH at the maximum NAD$^+$ concentration of 1 mM (Figure 4.6 B).
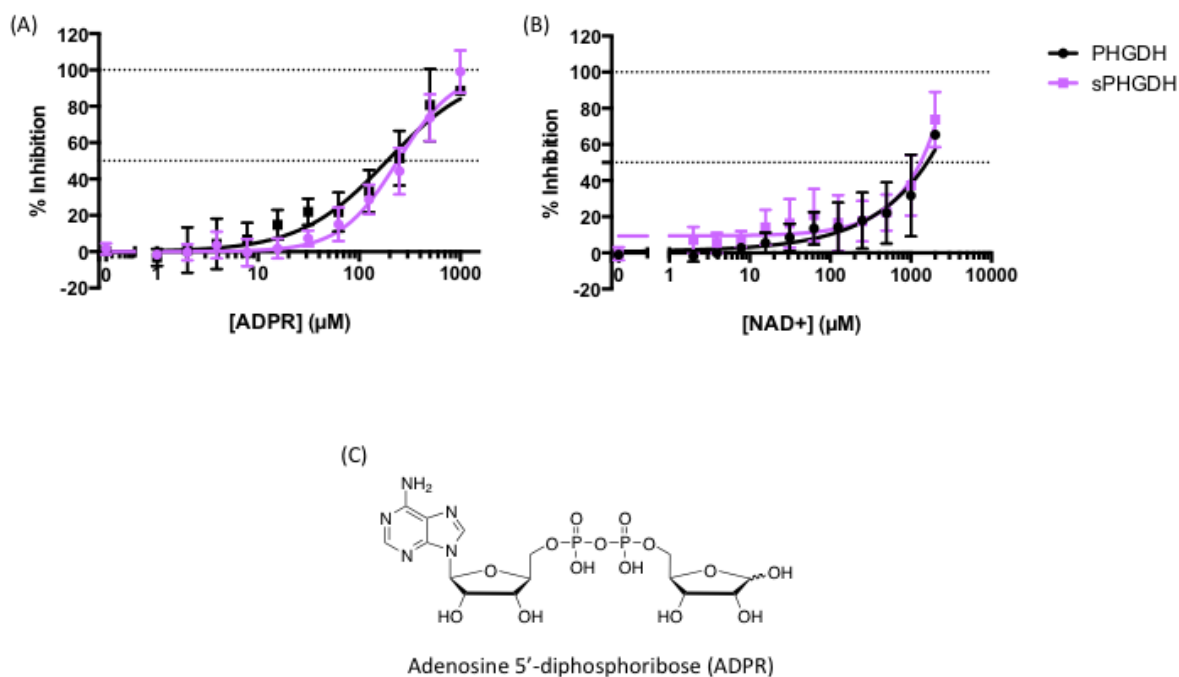
Figure 4.6 – Displacement of NADH by ADPR (A) and NAD$^+$ (B) in fluorescence polarisation assay.

Fluorescence polarisation of NADH was measured in the presence of increasing concentrations of ADPR (A) or NAD$^+$ (B). For each compound two independent measurements were performed with two replicates per measurement. Data were fit by non-linear regression using GraphPad Prism. (C) Chemical structure of ADPR.

The malate-containing fragment library (Figure 4.3) was tested in the fluorescence polarisation assay at 0.5 mM and 1 mM compound concentrations. Two fragments decreased fluorescence polarisation remarkably, namely 3-indole-*DL*-aspartic acid (**3**) and diethyl 2-[(4-aminobenzyl) amino] succinate (**6**). 3-Indole-*DL*-aspartic acid at 1 mM exhibited the strongest response with 55 % and 40 % displacement of NADH from PHGDH and sPHGDH respectively, as well as showing a concentration dependent response when comparing displacement of NADH from PHGDH and sPHGDH at 0.5 mM and 1 mM compound concentration.
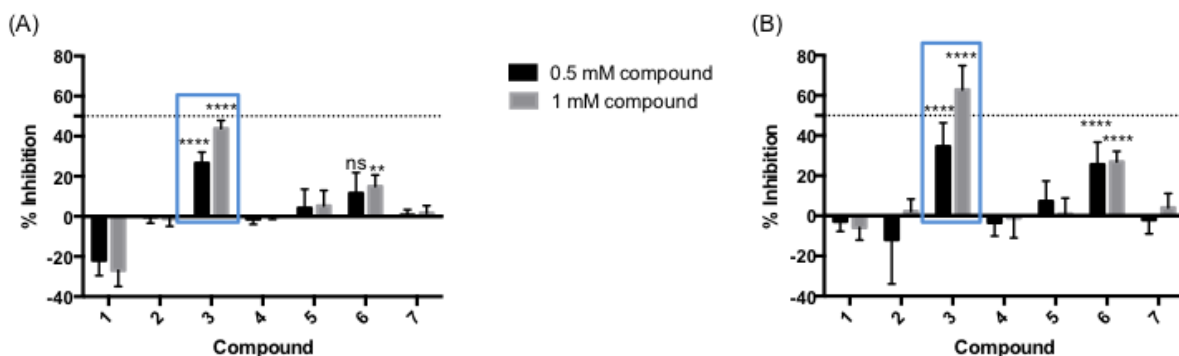
Figure 4.7 – Inhibition of sPHGDH (A) and PHGDH (B) with malate-containing fragments.

All malate-containing fragments (for structures see Figure 4.3) were tested in two independent fluorescence polarisation experiments in duplicate. % Inhibition of PHGDH and sPHGDH through the addition of fragments was calculated by using a sample containing only NADH as 100 % displacement control and a sample containing NADH and protein as 0 % displacement reference. Significant inhibition of sPHGDH or PHGDH was determined in GraphPad Prism using two-way ANOVA with ns = not significant, ** = p = 0.008, **** = p < 0.0001.

4.4.2 Synthesis and evaluation of 3-Indole-DL-aspartic acid analogues

Of the malate-containing fragments investigated, 3-indole-*DL*-aspartic acid showed the most promising effect in the FP assay. As the purchased 3-Indole-*DL*-aspartic acid was a mixture of the respective enantiomers, it was important to investigate whether one of the enantiomers was a more active component than the other. Therefore synthesis of the separate enantiomers was attempted.

Initially, the reaction of 3-indoleacetyl chloride with the aspartic acid was investigated following described procedures (Scheme 4.1) [143, 144]. Although 3-indoleacetyl chloride was synthesised, the subsequent coupling reaction did not work, when using either aqueous conditions in the presence of NaOH (Schotten-Baumann conditions), or pyridine as solvent with catalytic properties.
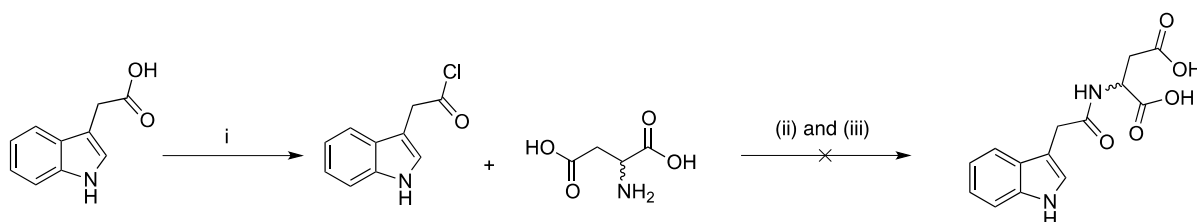
Scheme 4.1 – *Reagents and conditions:* (i) PCl$_5$, Et$_2$O, 0 °C to RT, 1 h, 68 %; (ii) NaOH, Et$_2$O, 0 °C to RT, 2 h; (iii) pyridine/ THF, DMAP, 0 °C to RT, 8 h.

As an alternative synthetic route, the active nitrophenylester of indoleacetic acid was prepared. Although not such a good leaving group as chloride, active esters are expected to react cleanly with amines under mild conditions [145, 146]. Under these conditions 3-indole-*D*-aspartic acid (**9**) and 3-indole-*L*-aspartic acid (**10**) were obtained in high yields of 51 % and 95 %, respectively (Table 4.1).

In addition to generating the resolved enantiomers of the hit compound, synthesis of a small library of analogues was undertaken. Because 3-indole-*DL*-aspartic acid had been shown to bind to PHGDH and sPHGDH in only a single biochemical assay format, only a small library of analogues was prepared. With the synthetic route in place using 3-indoleacetic acid as starting material, other heterocycles than indole, namely isoxazole and benzothiophene with an acetic acid functional group in the respective 3-positions were used as starting materials. While the synthesis of the active nitrophenol ester gave good yields (46-91 %), the following amide coupling did not work very well for other heterocycles than indole (yields 8–15 %) and would have required further optimisation.
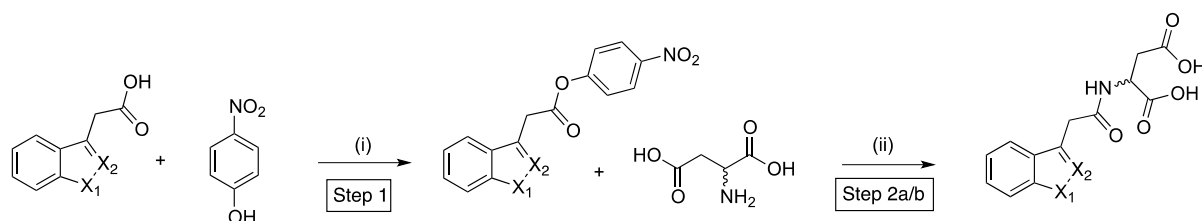
	Step 1	Step 2a	Step 2b
X_1 = NH, X_2 = C	55 %	51 % (**10**)	95 % (**9**)
X_1 = O, X_2 = N	46 %	8 % (**13**)	11 % (**12**)
X_1 = S, X_2 = C	91 %	13 % (**16**)	15 % (**15**)

Table 4.1 – Summary of yields for the two step synthesis of enantiopure indole-3-aspartic acid and analogues.

Reagents and conditions: (i) DCC, EtOAc, 0 °C to RT, 90 min; (ii) *L*-aspartic acid (step 2a) or *D*-aspartic acid (step 2b), tetramethylguanidine, aq. MeOH, RT, 48 h. The number of the synthesised compounds is given in brackets.

Comparison of the enantiopure 3-indole-aspartic acids showed a significantly higher displacement of NADH from sPHGDH with 3-indole-*L*-aspartic acid (**10**) compared to the *D*-enantiomer (**9**) at both concentrations of compound investigated. This effect was less statistically significant for the displacement of NADH from PHGDH at 0.5

mM compound concentration alone and was not significant at 1 mM (Figure 4.8). Interestingly, all heterocyclic analogues investigated were less potent than the compounds containing the indole moiety, but still exhibited a dose-dependent NADH displacement from sPHGDH or PHGDH when comparing the displacement of NADH at 0.5 and 1 mM compound concentration.
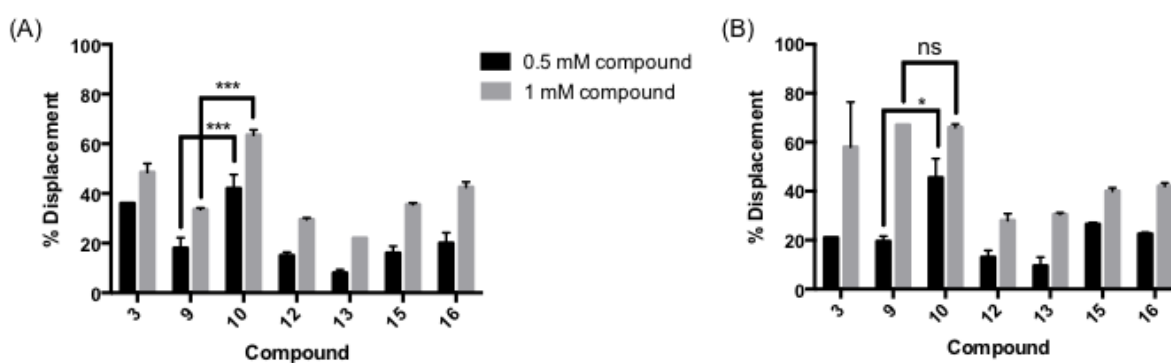
Figure 4.8 - Inhibition of sPHGDH (A) and PHGDH (B) with 3-indole-*DL*-aspartic acid and analogues thereof.

3-Indole-*DL*-aspartic acid and its analogues were tested in a fluorescence polarisation experiment in duplicate. % Inhibition of PHGDH and sPHGDH through the addition of fragments is calculated by using a sample containing only NADH as 100 % inhibition control and a sample containing NADH and protein as 0 % inhibition reference. Statistic analysis was performed in GraphPad Prism using two-way ANOVA; * = p = 0.02, *** = p = 0.0002, **** = p < 0.0001, ns = non significant.

4.4.3 Investigating additional substrate analogues

In addition to malate, other substrate analogues were investigated that were similar in size to 3-phosphoglycerate (3-PG) as well as containing the 2-hydroxypropanoic acid moiety (Figure 4.9 A). Unfortunately, the other physiological binding partner of PHGDH, 3-phosphohydroxypyruvate (PHP) was not commercially available, so that only structural analogues thereof could be investigated (Figure 4.9 B). Binding of these analogues was assessed using a differential scanning fluorimetry assay. This technique monitors protein unfolding through the use of a fluorescent dye that specifically binds to unfolded protein. Based on the fluorescence signal a so-called melting temperature (T_m) of the protein can be defined, which is the temperature at which half of the protein is unfolded. DSF also allows investigating ligand binding to the protein as a bound ligand contributes to the total Gibb's free energy of protein unfolding and thus results in an increased T_m (see also section 2.9) [101, 147].
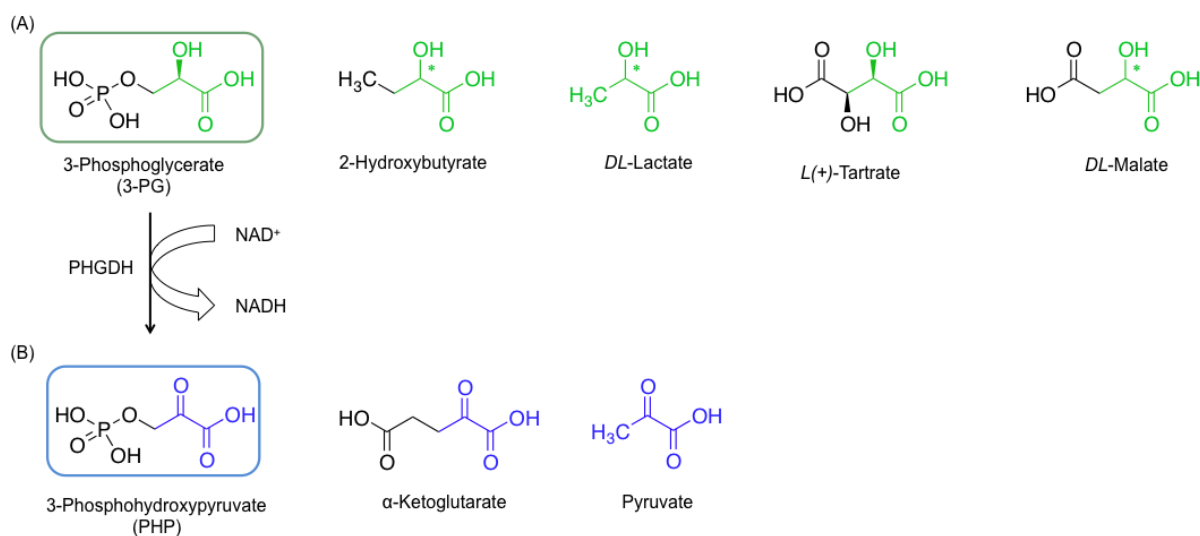
Figure 4.9 – Structures of the substrate 3-phosphoglycerate (A) and the product 3-phosphohydroxypyruvate (B) and their analogues.

For 3-phosphoglycerate and its analogues, the common 2-hydroxypropanoic acid moiety is highlighted in green. For 3-phosphohydroxypyruvate and its analogues, the 2-oxopropanoic acid moieties are coloured in blue.

Of the substrate and analogues tested in the DSF assay, only 3-phosphoglycerate (3-PG) and α-ketoglutarate induced a shift in the melting temperature of PHGDH greater than or equal to 0.5 °C, and only the natural substrate 3-PG resulted in a statistically significant increase in the T_m of PHGDH of 2.5 ± 0.6 °C (Figure 4.10).
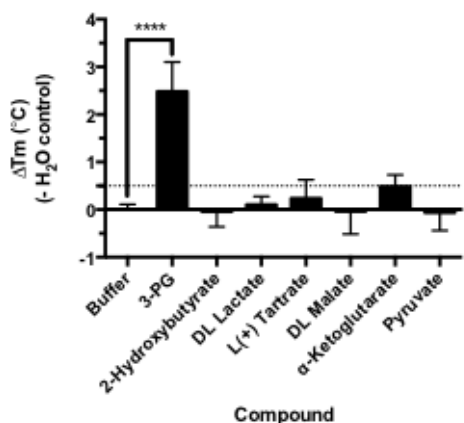
Compound (0.2 mM)	ΔTm (- Buffer control) (°C)
3-PG	2.5 ± 0.6
2-Hydroxybutyrate	0.0 ± 0.3
DL Lactate	0.1 ± 0.2
L(+) Tartrate	0.2 ± 0.4
DL Malate	0.0 ± 0.4
α-Ketoglutarate	0.5 ± 0.2
Pyruvate	-0.1 ± 0.3

Figure 4.10 – Change in T_m of PHGDH upon addition of substrate and analogues.

Thermal denaturation of 1 µM PHGDH in the presence of 0.2 mM substrate analogues. Graphs represent mean and standard deviation of three independent experiments with two replicates per experiment. Statistical analysis was performed in GraphPad Prism (two way ANOVA). **** = $p <$ 0.0001.

Due to the close proximity of the substrate and cofactor-binding sites, the presence of cofactor might be necessary for optimal substrate binding. Testing of NAD^+ in combination with substrate and analogues showed that a significant increase in binding was seen with 3-PG as well as *DL*-malate (Figure 4.11). As expected, NAD^+ alone had only a weak effect on the stability of PHGDH with an increase in T_m of 0.9 ± 0.7 °C. The combined addition of NAD^+ and 3-PG resulted in an increase in T_m of 4.5 ± 0.7 °C compared to the buffer control. This might be due to a synergistic effect of the two compounds rather than an additive effect, which would have only resulted in an increase of 3.4 ± 0.9 °C. A synergistic effect was also observed for the combination of NAD^+ and *DL*-malate, leading to an increase in the T_m of PHGDH of 2.2 ± 0.7 °C, whereas *DL*-malate alone did not affect the T_m of PHGDH at all indicating undetectable binding of *DL*-malate in the absence of NAD^+.

Figure 4.11 – Change in T_m of PHGDH through the addition of substrate (analogues) and NAD⁺.

Thermal denaturation of 1 μM PHGDH in the presence of 0.2 mM NAD⁺ alone and in combination with 0.2 mM substrate analogues, was measured. Graphs represent mean and standard deviation of three independent experiments with two replicates per experiment. Statistical analysis was performed in GraphPad Prism (two way ANOVA). Blue lines indicate statistically significant differences of NAD⁺ compared to NAD⁺ and substrate whereas red lines show statistical significant variations between substrate alone and the respective substrate plus NAD⁺ combination. ** = $p = 0.0026$, **** = $p < 0.0001$.

A similar experiment was performed using the substrate analogues in combination with the cofactor NADH. NADH has a much higher affinity for PHGDH than NAD⁺ resulting in a T_m shift of 9.3 ± 0.5 °C. All substrate and analogues investigated failed to stabilise PHGDH further in the presence of NADH (Figure 4.12). The observed destabilising effect of the substrate analogues was not significant and the decrease in T_m compared to NADH alone ranged from 0.4-1.1 °C.
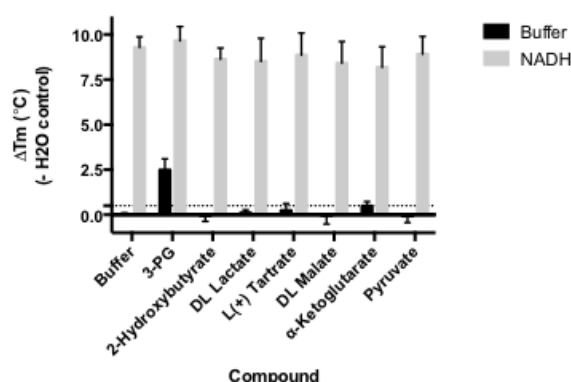
Figure 4.12 – Change in T_m of PHGDH through the addition of substrate (analogues) and NADH.

Thermal denaturation of 1 μM PHGDH in the presence of 0.2 mM NADH alone and in combination with 0.2 mM substrate analogues was measured. Graphs represent mean and standard deviation of three independent experiments with two replicates per experiment. Statistical analysis was performed in GraphPad Prism (two way ANOVA).

4.4.4 Crystallisation of sPHGDH

Crystallisation conditions for PHGDH that would allow for co-crystallisation of the protein with ligands of interest such as the substrate analogues were sought in order to improve molecular understanding of substrate analogue binding.

As the structure of a truncated form of PHGDH (sPHGDH, aa 3-314) co-crystallised with D-malate and NAD$^+$ was available in the public domain (PDB 2G76), this construct was chosen for crystallisation (Figure 4.13).
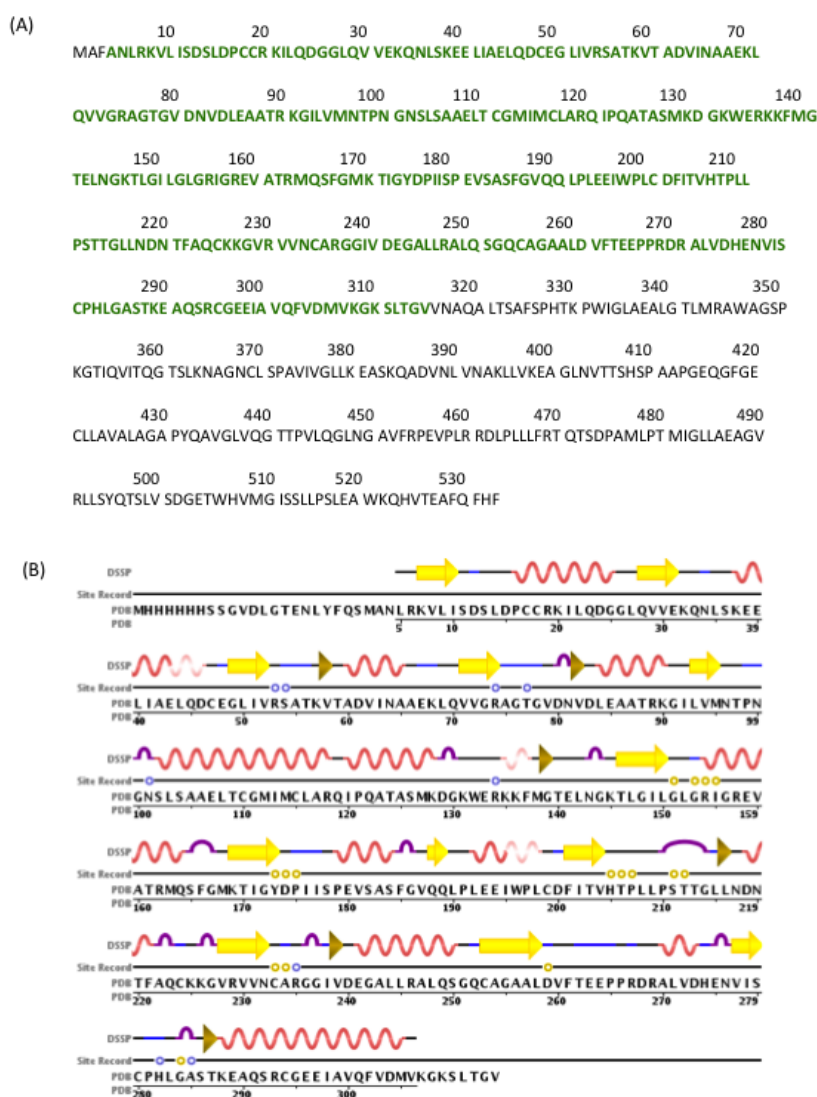
Figure 4.13 – Sequence of PHGDH and sPHGDH.

(A) Amino acid sequence of PHGDH with sequence of sPHGDH highlighted in green. (B) Secondary structure elements of sPHGDH are shown above the amino acid sequence with β-strands indicated by yellow arrows, β-bridges by khaki triangles, α-helices by red lines and turns by purple lines. The binding sites for NAD$^+$ and D-malate are marked with yellow and blue circles, respectively. Picture was taken from the RCSB protein databank (PDB 2G76).

The plasmid DNA used for transformation of competent *E. coli* Rosetta (DE3) cells contained human sPHGDH incorporated into the pNIC28-Bsa4 vector and was kindly donated by Wyatt Yue (SGC Oxford, Oxford, UK). sPHGDH was overexpressed with an N-terminal His$_6$-tag and purified using Ni-affinity chromatography and size exclusion chromatography (SEC) as well as an additional anion exchange chromatography step to increase purity of the protein (Figure 4.14). The large-scale production of sPHGDH gave very good protein yields of 30 mg/L culture.
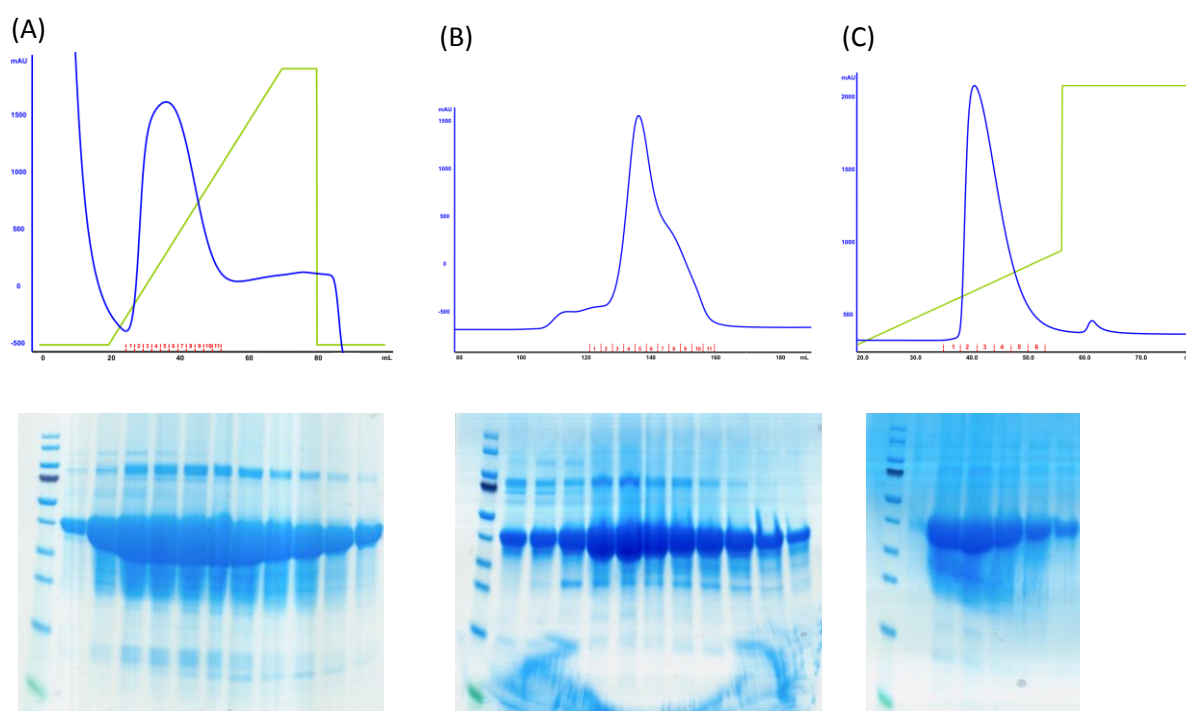
Figure 4.14 – Purification of sPHGDH.

Human sPHGDH was purified from *E. coli* Rosetta (DE3) cells. (A) After cell lysis, sPHGDH was purified by Ni-affinity chromatography, followed by (B) size exclusion chromatography (Superdex 75 26/60) and (C) anion exchange chromatography. A representative purification is shown with the UV traces of the respective purification steps together with the analysis of the obtained fractions by SDS-PAGE.

Initial crystallisation trials were aimed at reproducing the published crystallisation conditions (0.1 M MMT, pH 7.0, 30 % PEG1000, 5 mM NAD$^+$) using a home-made screen around these conditions (optimisation screen 1). Sitting drop vapour diffusion crystallisation trials were set up with 15 mg/mL protein and 200 and 300 nL drop size at 1:1 and 1:2 protein: reservoir ratios, respectively, at 4 °C did not produce any crystals. Therefore a wider screen against commercially available sparse matrix conditions was performed with 15 mg/mL sPHGDH mixed with 5 mM NAD$^+$. In

parallel, trials were also set up with sPHGDH plus 5 mM ADPR, as ADPR had been shown to be a stronger binder to PHGDH compared to NAD$^+$ in the FP assay (see section 4.4.1). Initial crystals were very small and formed in the presence of 5 mM ADPR and a crystallisation buffer comprising 0.1 M BIS TRIS, pH 5.5, 0.2 M ammonium acetate (NH$_4$Ac), 25 % (w/v) PEG3350. An optimisation trial was designed around this condition with variations in the pH (5.5-7.0), the amount of PEG3350 (21 – 31 % (w/v)) and the concentration of NH$_4$Ac (0.1-0.3 M) (optimisation screen 2) as well as exploring the temperature of the crystal trials (4 and 20 °C). Co-crystallisation with ADPR and obtained crystal structures will be further discussed in section 5.4.1.

4.4.5 Co-crystallisation of sPHGDH with substrate analogues

Using the crystallisation conditions described above for sPHGDH, co-crystallisation of the protein with cofactor (NAD$^+$) and substrate analogues, namely 3-PG, α-ketoglutarate and *L*-tartrate, was attempted. Co-crystallisation in the presence of α-ketoglutarate and *L*-tartrate gave protein crystals, although the diffraction pattern was generally of poor quality and a limited numbers of datasets were collected.

For a co-crystal of sPHGDH with NAD$^+$ and α-ketoglutarate grown in 0.1 M BIS-TRIS, pH 6.5, 0.2 M NH$_4$Ac and 27 % (w/v) PEG3350, the structure (sPHGDH 01) could be solved using chain A of the published structure as a search model (PDB 2G76). However, as the initial Fo-Fc electron density indicated that only NAD$^+$ was bound in the active site, the structure was not fully refined (Figure 4.15). Optimisation of the co-crystals of NAD$^+$ and α-ketoglutarate was attempted using higher concentrations of α-ketoglutarate, as well as increasing concentrations of the substrate analogue in the cryo-protectant to avoid diffusion of the substrate out of the crystal during crystal harvesting. Although additional crystals were obtained, none of the crystals tested showed good quality diffraction, and no datasets were collected.

	sPHGDH 01
Wavelength (Å)	0.98
Space Group	P 1
a, b, c (Å)	43.0, 59.2, 61.7
α, β, γ (°)	89, 89, 79
Resolution (Å)	61.7 – 3.86 (4.0 – 3.9)
$I/\sigma I$	5.0 (4.4)
R_{merge}	1.45 (1.26)
Completeness (%)	97.6 (94.5)
Multiplicity	1.9 (1.9)
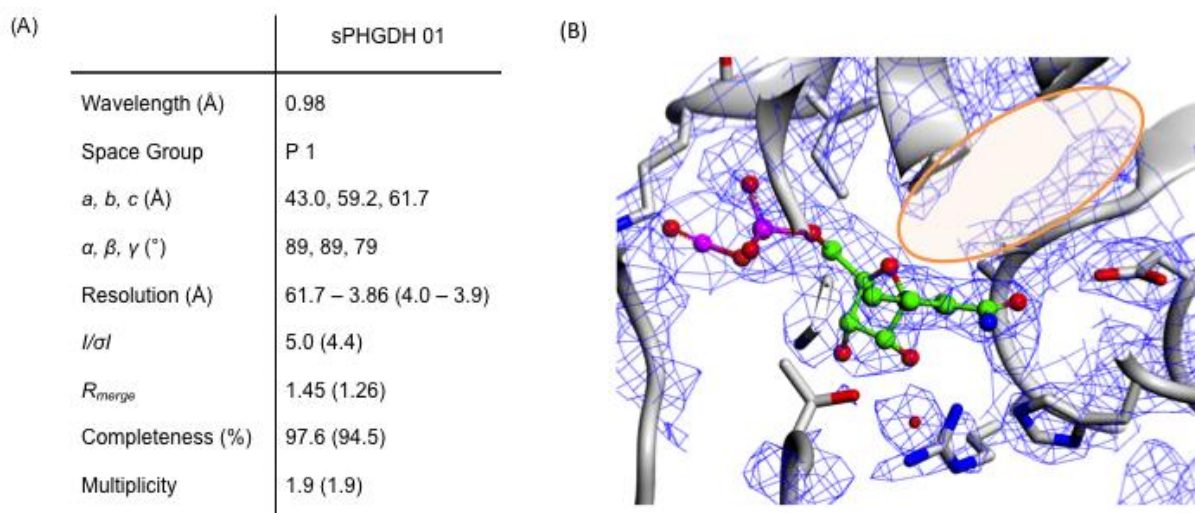
Figure 4.15 - Co-crystallisation of sPHGDH with NAD[+] and α-ketoglutarate (sPHGDH 01).

(A) Data collection statistics for sPHGDH 01 (co-crystal with α-ketoglutarate and NAD[+]). The data were collected on a single crystal at 100 K on beamline I02, Diamond Light Source (DLS). The data were processed using xia2 (3daii-run). The statistics are presented as averages with values for the highest resolution shell in parentheses. (B) Active site of sPHGDH 01 with NAD[+] bound. Protein secondary structure is shown in grey and NAD[+] is coloured by atom. No electron density was observed for α-ketoglutarate (orange circle); 2Fo-Fc electron density map contoured at 0.37 electrons/Å[3].

In addition, several crystals of sPHGDH grown in the presence of NAD[+] and *L*-tartrate were obtained. Of the measurable crystals, only one crystal grown in 0.1 M Tris, pH 8.5, 0.2 M MgCl$_2$ hexahydrate and 20 % (w/v) PEG 3350 showed evidence for bound *L*-tartrate. The dataset (sPHGDH 02) was collected to 2.24 Å resolution using synchrotron radiation (Figure 4.16). The measured crystal had the same symmetry (monoclinic, space group P 1 21 1) as the published structure (PDB 2G76).

	sPHGDH 02
Wavelength (Å)	0.92
Space Group	P 1 21 1
a, b, c (Å)	43.5, 109.7, 66.9
α, β, γ (°)	90, 94, 90
Resolution (Å)	57.0 - 2.24 (2.3 – 2.24)
$I/\sigma I$	12.6 (2.3)
R_{merge}	0.06 (0.59)
Completeness (%)	98.0 (98.8)
Multiplicity	3.8 (3.9)
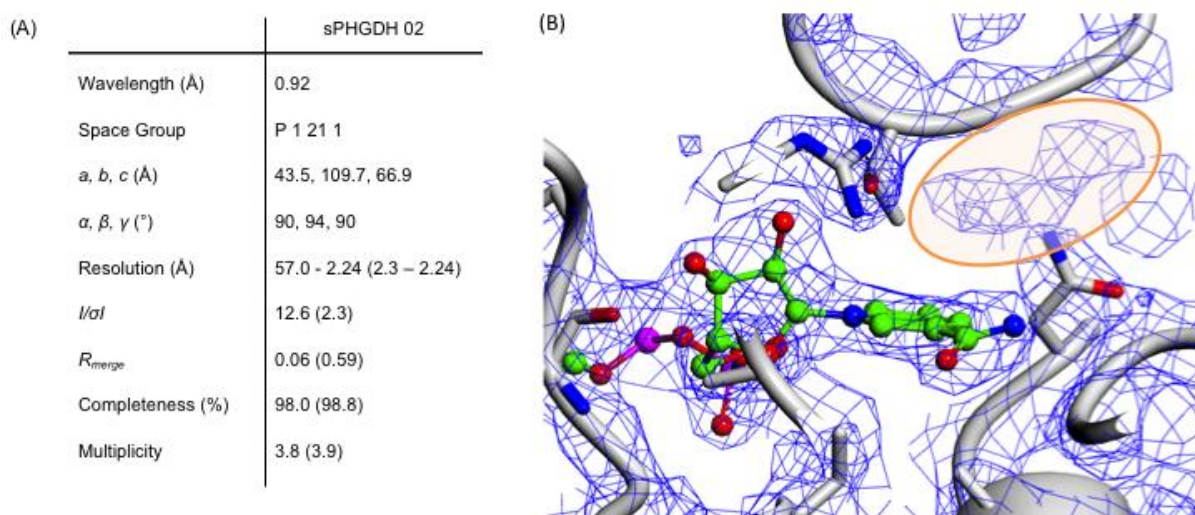
Figure 4.16 – Data collection statistics for sPHGDH co-crystallised with substrate analogue *L*-tartrate and NAD⁺ (sPHGDH 02).

(A) Data collection statistics for sPHGDH 02 (co-crystal with *L*-tartrate and NAD⁺). The data were collected on a single crystal at 100 K on beamline I04-1, Diamond Light Source (DLS). The data were processed using xia2 (3daii-run). The statistics are presented as averages with values for the highest resolution shell in parentheses. (B) Active site of sPHGDH 02 with NAD⁺ bound. Protein secondary structure is shown in grey and NAD⁺ is coloured by atom. Electron density observed for *L*-tartrate is highlighted with an orange circle; 2Fo-Fc electron density map contoured at 0.34 electrons/Å³.

Molecular replacement was performed with PHASER [106] as part of the CCP4 suite [105] using the published structure (PDB 2G76) as a search model. The model was refined with Refmac5 [148] and validated using MolProbity [109] (Table 4.2). Early refinement maps indicated a significant structural change in the orientation of the sPHGDH lid domain that was beyond the radius of convergence of REFMAC. An iterative autobuild protocol using BUCCANEER and REFMAC was used to help to reinterpret the electron density and reveal the conformational change.

(A) Refinement		(B) Molprobity Parameters	Value
R_{work}/R_{free}	0.21/0.25	Clashscore, all atoms	6.54[a]
Mean B-value ($Å^2$)		Poor rotamers	2.75 %
All atoms	26	Ramachandran outliers	0.5 %
Protein	8	Ramachandran favoured	96.0 %
Water	36	No of Cβ deviations (> 0.25 Å)	9
NAD^+	28	Molprobity Score	1.97[b]
L-Tartrate	33	Residues with bad bonds	19
No of atoms		Residues with bad angles	18
Protein	8809		
Water	74		
NAD^+	66		
L-Tartrate	12		
Rmsd			
Bond lengths (Å)	0.02		
Angles (°)	2.21		

Table 4.2 – Refinement and validation statistics for the co-crystal structure of sPHGDH with NAD^+ and *L*-tartrate.

(A) Refinement statistics generated with Refmac5 as part of CCP4i2 suite. (B) Validation statistics generated by MolProbity web server. The clashscore[a] was 98th percentile and the MolProbity score[b] was 91st percentile for a structure of 2.30 ± 0.25 Å.

As expected from the published structure, two molecules were present in the asymmetric unit (ASU). However, while both molecules had NAD^+ bound to the cofactor sites, *L*-tartrate was only present in one molecule. In the absence of *L*-tartrate, the lid domain, which otherwise closes over the substrate-binding site had undergone rigid domain movement. In order to characterise the movement, the two molecules in the ASU were compared using the DynDom webserver [149, 150]. Three domains could be distinguished: a fixed core domain (aa 100-287), the mobile lid domain (aa 9-93 and aa 289-301) and a hinge domain (aa 94-99 and aa 302) (Figure 4.17A-B). Superposition of the two conformations with and without bound *L*-

tartrate showed that the lid domain had rotated by 28.9° about an axis aligned along

upon substrate binding (Figure 4.17 C).
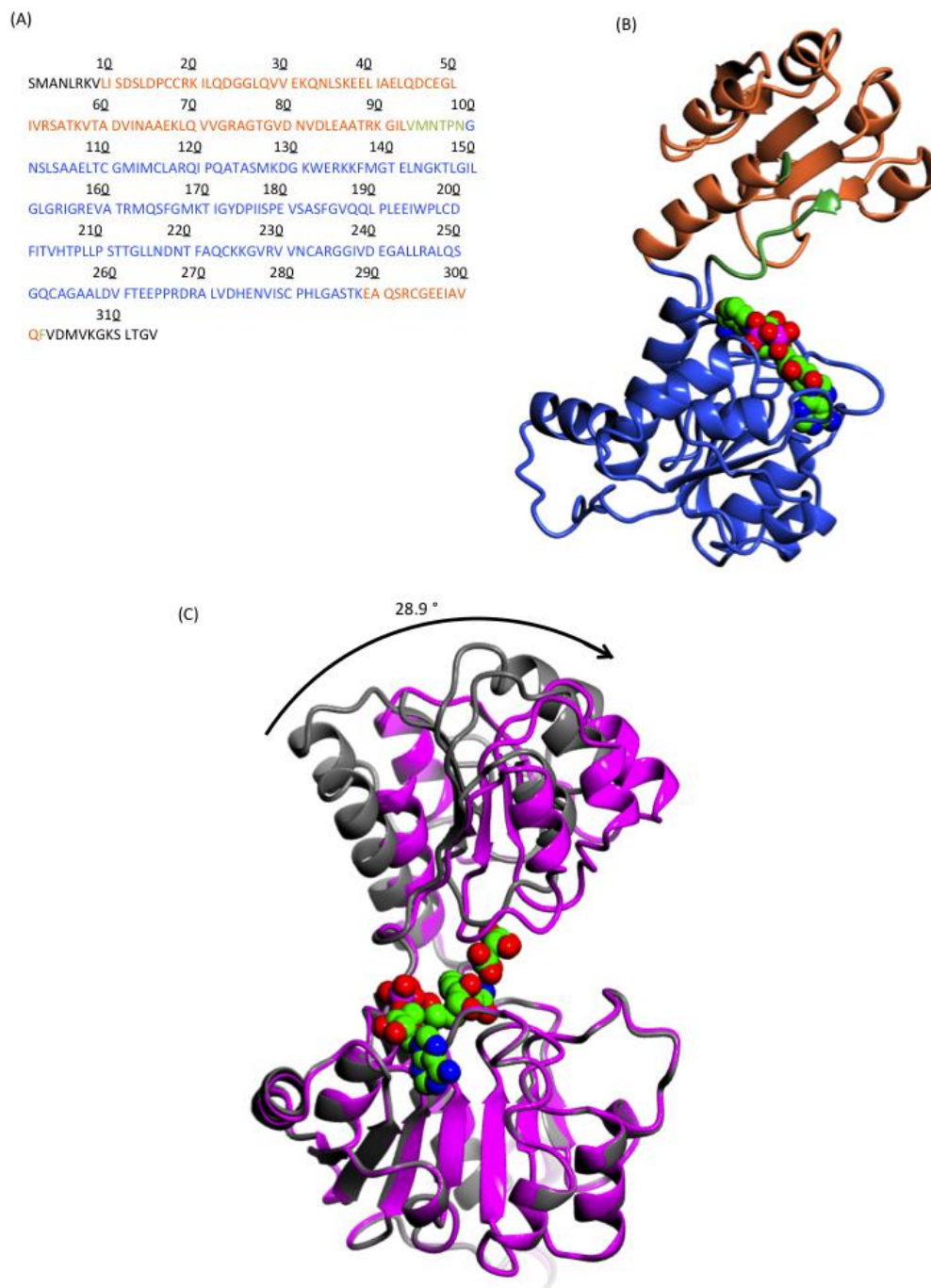
Figure 4.17 – Analysis of the domain movement in a co-crystal structure of sPHGDH with *L*-tartrate and NAD[+].

The protein crystal structure was analysed using the DynDom web server. Sequence (A) and structure of sPHGDH chain A (B) with different domains highlighted: fixed domain in blue, mobile domain in orange, hinge domain in green. The cofactor NAD[+] is shown as spheres coloured by atom type. (C) Superposition of the two molecules in the ASU with the structures containing *L*-tartrate and without *L*-tartrate coloured in purple and grey respectively. NAD[+] and *L*-tartrate are coloured by atom type and displayed as spheres.

A closer view of the substrate-binding site in both chains revealed distinct movements of the amino acids involved in *L*-tartrate binding that, in the absence of *L*-tartrate, either move out of the binding site (Arg74, Arg53) or adopt alternate rotamer that orient away from the binding site (Gln291) (Figure 4.18).
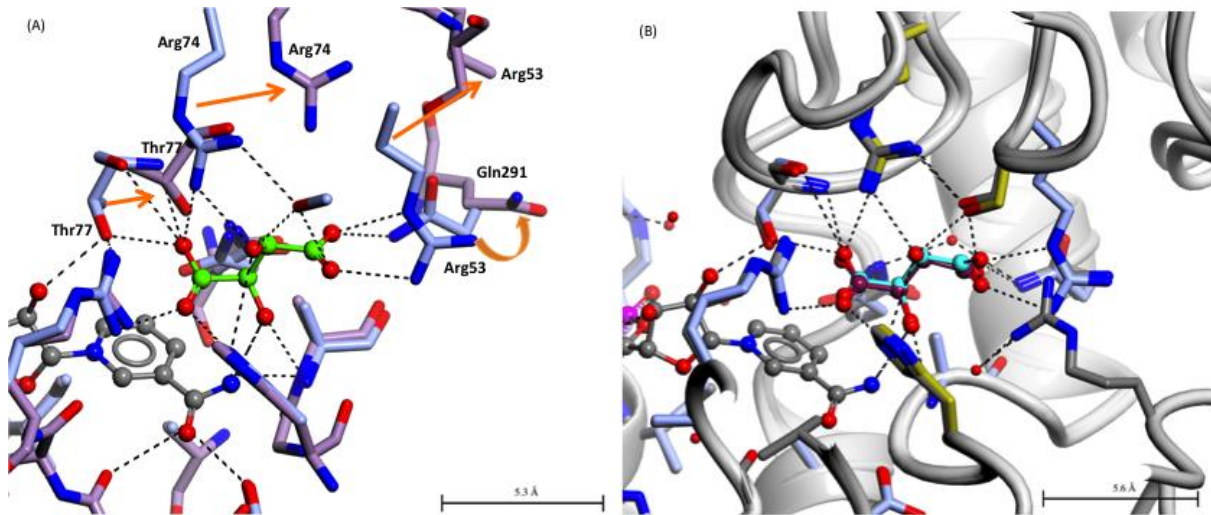
Figure 4.18 – Changes in substrate-binding site of sPHGDH observed on binding of *L*-tartrate

(A) Chain A (light blue) with *L*-tartrate (green) and NAD[+] (grey) bound to the active site superposed onto chain B (lilac) with only NAD[+] (grey) bound to the active site. Movement of individual amino acids involved in *L*-tartrate binding indicated with orange arrows. (B) Chain A (light blue/ gold) with *L*-tartrate bound (dark purple) superposed onto chain A (grey) of sPHGDH with NAD[+] and *D*-malate bound (cyan) (PDB 2G76). New hydrogen bonds are formed between the amino acids highlighted in gold and the additional OH-group of *L*-tartrate compared to binding of *D*-malate.

4.5 Discussion

To date there are no known inhibitors of PHGDH, thus precluding full validation of PHGDH as a potential target in cancer therapy. Knowledge of the structure of PHGDH is also limited with only one crystal structure available for the first 300 amino acids comprising the cofactor and substrate-binding domain. Although human PHGDH contains regulatory domains of incompletely understood function, their structure is not yet known. Therefore inhibitors were investigated and designed to target the substrate-binding site. With the cofactor-binding site being similar within the family of dehydrogenases, inhibitors binding specifically to the substrate-binding site of PHGDH were sought so as to provide the specificity for on-target inhibition.

In the crystal structure of sPHGDH (PDB 2G76), *D*-malate is bound to the substrate-binding site. A small set of commercially available malate containing fragments with malate attached through different linkers to an aromatic ring or, in one case, a carboxylic acid was tested for binding to PHGDH. This approach of linking a substrate analogue to an aromatic ring and thereby targeting simultaneously the cofactor and substrate-binding sites had been successfully used for the development of inhibitors of human lactate dehydrogenase A as well as in a study on the specificity of mitochondrial isocitrate dehydrogenase isozymes [97, 140]. To test the potential of the malate-containing fragments library, the compounds were investigated for their ability to displace NADH from the cofactor-binding site using a fluorescence polarisation assay with NADH as fluorophore. Although FP assays with NADH as fluorophore have been reported in the literature, NADH is a relatively weak fluorophore compared to fluorescent dyes such as fluorescein or rhodamine [142], thus the assay read-outs were not very robust and often suffered from a low signal-to-noise ratio. Despite these limitations, 3-indole-*DL*-aspartic acid (**3**) and diethyl 2-[(4-aminobenzyl) amino] succinate (**6**) were shown to be able to displace NADH to a significant extent and in a concentration-dependent manner. Testing of the compounds in a second assay, the DSF assay, confirmed binding of compound **3** but not **6**, thus compound **3** was taken forward for further investigation.

With the commercial 3-indole-*DL*-aspartic acid being a racemic mixture, it was necessary to synthesise the enantiomers separately as to find out if there was a difference in activity between them. Although enantiomers have the same physical and chemical properties in an achiral environment, this can change in a chiral

environment such as that presented by biological macromolecules [151]. Testing the enantiopure 3-indole-aspartic acids revealed the *L*-enantiomer as a significantly more potent compound compared to its *D*-analogue. Accordingly, displacement of NADH seen with 3-indole-*L*-aspartic acid was also increased compared to the racemic mixture. As a synthetic route had been established for 3-indole-aspartic acids, benzisoxazole and benzothiophene analogues were synthesised and investigated. Although changing the indole moiety deteriorated the induced NADH displacement, the 3-benzisoxazole- and 3-benzothiophene aspartic acids still displaced NADH from sPHGDH and PHGDH in a concentration-dependent manner.

In addition to malate, other substrate analogues of 3-PG and PHP were tested at 0.2 mM in a DSF screen. Of the tested substrate-analogues, *L*-tartrate and α-ketoglutarate (α-KG) stabilised PHGDH marginally, whereas the addition of the natural substrate 3-PG gave a significant increase in T_m. Tartrate has not been reported as a potential substrate for PHGDH. α-KG on the other hand can be used as a substrate by *E. coli* PHGDH, whereas human as well as rat and *M. tuberculosis* PHGDH have been reported not to share this trait [57]. A recent study, however, found human PHGDH to be able to catalyse the NADH-dependent reduction of α-KG to *D*-2-hydroxyglutarate, an oncometabolite associated with brain cancer and acute myeloid leukaemia [65].

Binding of substrate analogues was also investigated in combination with the cofactors, NAD^+ or NADH, as possible interdependence of substrate and cofactor binding is suggested by the close proximity of their respective binding sites. In addition, the catalytic mechanism of human PHGDH is not yet known and could proceed by an ordered Bi Bi mechanism requiring the binding of cofactor prior to substrate binding or *vice versa*. Indeed, the combination of 3-PG and NAD^+ increased the T_m significantly, and by more than the simple addition of the effects seen with the two compounds separately. The same was found to be true for *DL*-malate where the presence of NAD^+ appeared to be necessary for the substrate analogue to bind at all, whereas NAD^+ was able to bind on its own. Based on these findings it is, however, difficult to definitely determine the reaction mechanism, as according to the data for *DL*-malate and NAD^+, cofactor binding seems necessary to initiate binding of the substrate analogue, whereas the T_m shift seen with 3-PG alone does not support this hypothesis. The other tested substrate analogues did not show

synergistic behaviour in combination with NAD^+. In the case of PHP analogues, where the reaction partner is NADH, NADH failed to enhance substrate binding. However, NADH itself binds very strongly to PHGDH, making it difficult to detect any additional and probably small changes in the T_m following substrate binding.

Of the compounds investigated, *DL*-malate was found to be the most promising starting point for an inhibitor targeted to the substrate-binding site as supported by the published crystallographic data (PDB 2G76). Crystallisation of sPHGDH with additional compounds was attempted in order to provide a wider understanding of substrate and substrate analogue binding and provide additional start points for inhibitor design. Although the published crystallisation conditions could not be reproduced, crystals were obtained in 0.1 M BIS-TRIS buffer, pH 5-7, with different salts and in a range of concentrations of PEG3350. In general, the crystals grew as thin, stacked plates, making it difficult to obtain single crystals that would diffract well in all directions (*i.e.* isotropically). Optimisation did not result in better-diffracting crystals making it necessary to test many crystals before a measurable one was found. Co-crystals of substrate analogues and NAD^+ were only obtained for α-KG and *L*-tartrate. However, in the case of α-KG, the substrate was not visible in the electron density. For *L*-tartrate the substrate was present in one molecule of the crystal dimer whereas the other molecule contained only NAD^+. Comparison of the two molecules showed that in the presence of *L*-tartrate a movement of the lid domain towards a more closed substrate pocket had occurred.

Chapter 5. Investigating the cofactor binding pocket

5.1 Introduction

In addition to substrate, many enzymes require cofactors to assist catalysis, e.g. metal ions or coenzymes. Dehydrogenases usually require NAD^+/NADH as a cofactor. The hydride-transfer reaction catalysed by dehydrogenases requires that the nicotinamide group of NAD^+ be in close proximity to the hydrogen donor atom of the substrate. The substrate and coenzyme binding pockets are therefore adjacent. A common structural domain for nucleotide-binding proteins, the Rossmann fold, largely determines the structure of the coenzyme-binding site [152]. The Rossmann fold is among the most common structural motifs in proteins [153]. It is characterised by alternating β-strands and α-helices, and is found in dehydrogenases from different organisms (Figure 5.1).
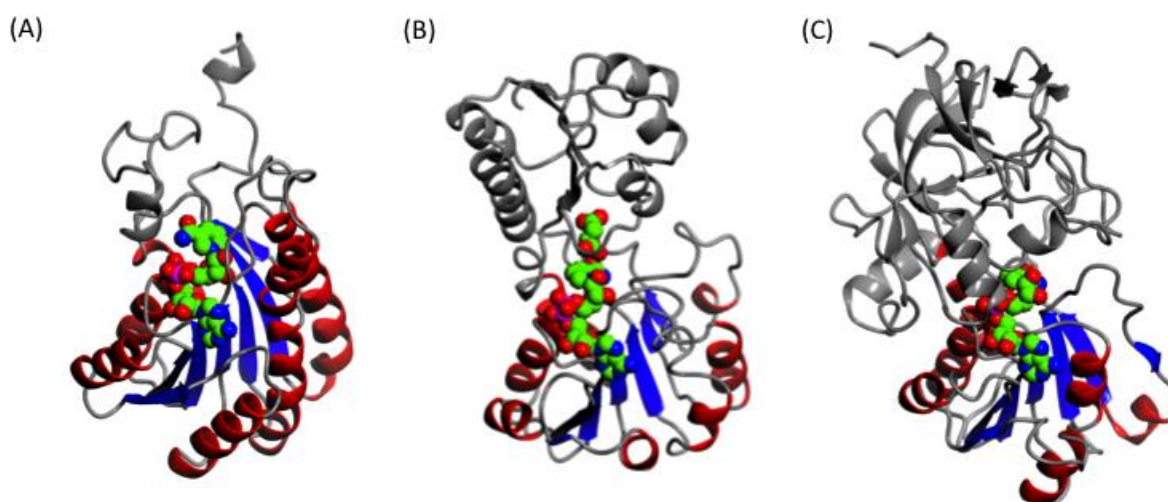
Figure 5.1 – Ribbon diagram illustration of different dehydrogenases.

The Rossmann-fold is depicted with β-strands in blue and α-helices in red. Additional secondary structure elements are coloured in grey. Cofactor NAD^+ is rendered as spheres and coloured by atom type**.** (A) Bacterial 3α-20β-hydroxysteroid dehydrogenase from *Streptomyces exfoliatus* (PDB 2HSD). (B) Truncated form of human 3-phosphoglycerate dehydrogenase (sPHGDH) (PDB 2G76). (C) Human class II alcohol dehydrogenase (PDB 3COS).

Not only might the conservation of the nucleotide-binding regions become problematic for the development of specific inhibitors, but also the pocket itself can be challenging to target due to its hydrophilic nature. For example one component of the pocket is the diphosphate-binding site, which is ideally constructed to bind a highly charged unit such as the diphosphate linker. Mimicking the properties of the

phosphate group with its tetrahedral geometry and dianionic charge at physiological pH is not trivial.

NAD$^+$binding also involves various hydrogen bonds and van der Waals interactions with the protein, so that mimicking the complete interaction pattern with a drug-like molecule will be difficult (Figure 5.2).
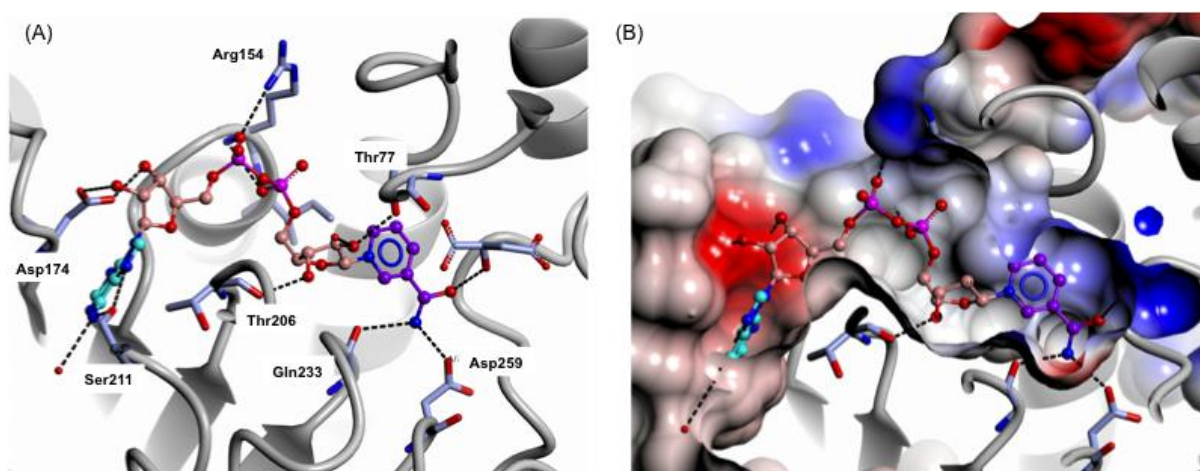
Figure 5.2 – Cofactor-binding site in PHGDH (PDB 2G76)

Protein chains are coloured in grey. Amino acids involved in cofactor binding are highlighted in light blue with oxygen and nitrogen coloured in red and blue, respectively. Cofactor NAD$^+$ is rendered as stick and balls and the different moieties coloured in cyan (adenine), pink (phosphate) and purple (nicotinamide).

However, NAD$^+$-requiring enzymes have been successfully targeted in the past through (parts of) the NAD$^+$-binding pocket, which is the case for the sirtuin inhibitor Ex-527 that binds to the nicotinamide part of the NAD$^+$-binding pocket and thus prevents the enzymatic reaction [154]. Another example was presented with the development of LDHA inhibitors that were shown to simultaneously target the adenine and nicotinamide subsites of the cofactor-binding pocket, but interestingly, did circumvent the phosphate-binding region [140].

5.2 Aims and Objectives

This chapter focuses on the investigation of the cofactor-binding site of PHGDH through separate examination of the different areas of this pocket in more detail. The investigated areas include: (1) the nicotinamide pocket where NAD^+ analogues with modified nicotinamide moieties were evaluated, (2) the phosphate binding region for which the impact of variations in the size of the phosphate linker were analysed and (3) the adenine-binding subsite for which analogues of adenine were tested.

5.3 Materials and Methods

All chemicals listed in the materials and methods section were of analytical grade and obtained from Sigma Aldrich, St. Louis, MO, USA, unless otherwise stated.

5.3.1 Differential scanning fluorimetry

DSF experiments were performed in 96 well plates according to section 2.9 with the following protein and SYPRO Orange concentrations. The assay contained 25 mM HEPES, pH 7.5 with 1 x Sypro Orange, 1 µM PHGDH and 200 µM ligand. All ligands were prepared fresh prior to the assay by dissolving in 25 mM HEPES, pH 7.5 and the pH of the solution was adjusted to about 7.5. The concentration of solutions containing NAD^+ and analogues thereof was determined spectrophotometrically by measuring the absorbance at λ = 260 nm. From the absorbance the respective concentrations were calculated using the Beer-Lambert law as described in section 4.3.2 with the specific extinction coefficients listed in Table 5.1, followed by dilution of the stocks to give the final assay concentration.

Compound	Extinction coefficient (ε) at λ = 260 nm (M^{-1} cm^{-1}) [Reference]
Nicotinamide adenine dinucleotide (NAD^+)	18,000 [155]
Thionicotinamide adenine dinucleotide (TAD)	19,700 [156]
Pyridinealdehyde adenine dinucleotide (PAD)	18,300 [157]
Acetylpyridine adenine dinucleotide (APAD)	16,400 [157]
Adenosine 5'-diphosphoribose (ADPR)	15,400 [158]
Adenosine triphosphate (ATP)	15,400 [155]
Adenosine diphosphate (ADP)	15,400 [155]
Adenosine monophosphate (AMP)	15,300 [155]

Table 5.1 – Specific extinction coefficients of NAD^+, NAD^+ analogues and fragments of NAD^+.

5.3.2 Isothermal titration calorimetry

ITC experiments were performed as described in section 2.10. Briefly, PHGDH as well as ADPR were in 25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP. Changes in enthalpy and entropy were measured at 25 °C using a MicroCal iTC200 (GE Healthcare, Little Chalfont, UK). All thermodynamic parameters of the reaction were determined using the ORIGIN software package, version 7.0.

5.3.3 Enzyme activity assay

In vitro enzymatic activity of PHGDH or sPHGDH was measured spectrophotometrically as described in section 2.8. The concentration range of NAD$^+$ and TAD used was 0.1 – 200 μM and for APAD and PAD was 0.5 – 700 μM.

5.3.4 Fluorescence polarisation assay

Fluorescence polarisation assays were carried out as described in section 2.11. Briefly, the assay mixture contained 40 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP or 2 mM DTT, 1 μM NADH and 2 μM PHGDH or sPHGDH. AMP, ADP, ADPR and NAD$^+$ were dissolved in 40 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP or 2 mM DTT and screened at a maximum concentration of 3 mM for AMP and ADP and mM for ADPR and 2 mM for NAD$^+$. The reaction mixture with compound was incubated for 30 min at RT and fluorescence polarisation was measured using a PHERAstar FS plate reader (BMG Labtech, Ortenberg, Germany) with a FP filter module (λ_{ex} = 340 nm and λ_{em} = 440 nm).

5.3.5 Crystallography

Co-crystallisation of sPHGDH with NAD$^+$ or its analogues (TAD, PAD, APAD) or fragments (AMP, ADP) was performed by screening against sparse matrix screens as well as optimisation screen I as described in section 2.12. Crystals for which datasets were collected were analysed as described in section 2.13.

5.4 Results

5.4.1 ADPR as a moderate inhibitor of PHGDH

In order to identify inhibitory hits of a possible drug target, it is necessary to develop biochemical and/ or biophysical assays of sufficient sensitivity. To validate those assays, one would ideally use a known and relatively potent inhibitor of the target. However, for PHGDH no such inhibitor has been reported to date. From the literature adenosine 5'-diphosphoribose (ADPR), a truncated analogue of NAD^+ lacking the nicotinamide moiety is known to bind strongly to rabbit glyceraldehyde dehydrogenase and function as potent competitive inhibitor [159]. Therefore, ADPR was investigated in an optimised *in vitro* enzyme activity assay based on a protocol developed during a preceding project [160]. Optimisation included i) the addition of Triton-X-100 to the assay mixture to decrease any surface adsorption of the protein, ii) the change of buffering agent from TRIS to BIS TRIS propane, as this buffer system can be used over a wider pH range (6.3 to 9.5), and iii) the reduction of 3-PG concentration from 2.5 mM to 140 μM, which was determined as the K_m of 3-PG under these reaction conditions. In the *in vitro* enzyme activity assay, ADPR was able to inhibit full-length PHGDH activity by 50 % at the maximum tested dose of 120 μM. In addition, ADPR also inhibited the activity of the truncated sPHGDH that contains only the substrate and cofactor-binding domains, by about 40 %, indicating that ADPR functions as a competitive inhibitor of the cofactor, which would be expected due to the structural homology of NAD^+ and ADPR (Figure 5.3).
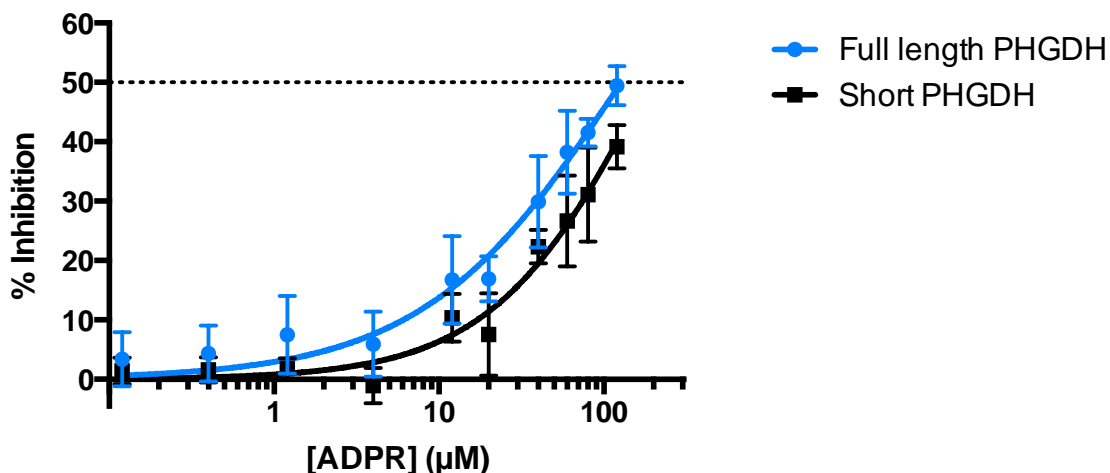
Figure 5.3 – Inhibition of PHGDH/sPHGDH activity with ADPR.

Enzymatic activity of PHGDH or sPHGDH was measured in the presence of increasing concentrations of ADPR (0.5 – 120 µM). % Inhibition was calculated by comparison to control samples containing no inhibitor (0 % inhibition) and a control without enzyme (100 % inhibition). Graph represents mean and standard deviation of three independent experiments with two intra-assay repeats. Data were analysed using nonlinear regression (log(inhibitor) vs. response – variable slope) in GraphPad Prism.

In addition, ADPR binding to PHGDH was investigated by ITC as well as in the DSF assay (Figure 5.4). ITC experiments were performed by titrating 1 mM ADPR into 0.05 mM PHGDH. ADPR was found to be a moderately potent ligand of PHGDH with a K_D of 40.0 ± 5.4 µM. Binding was mainly driven through changes in enthalpy (i.e. an exothermic reaction) with ΔH = -35.6 ± 14.1 kJ/mol with only a minor contribution from entropic factors, with a ΔS = -0.13 J/(mol x K).
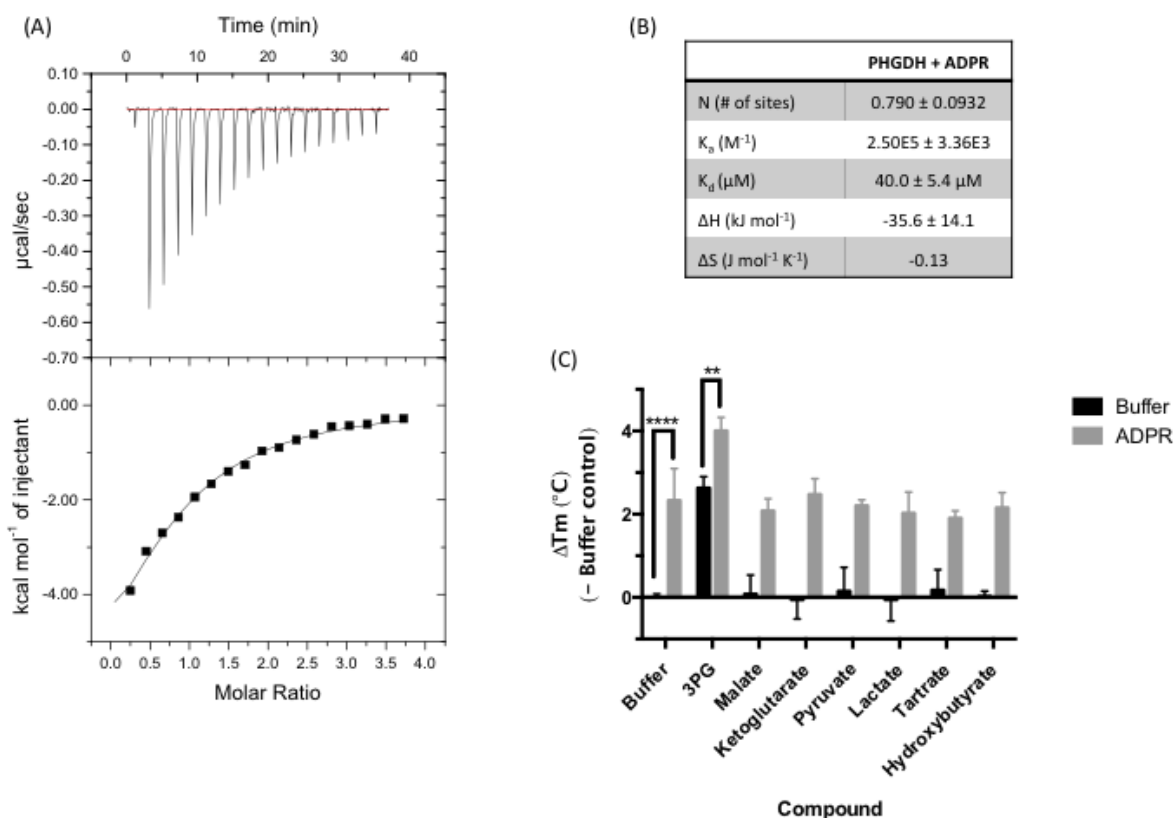
Figure 5.4 – Binding of ADPR to PHGDH as investigated by ITC and DSF.

(A) 1 mM ADPR was titrated into 0.05 mM PHGDH in 25 mM HEPES, pH 7.5, 100 mM NaCl and 0.5 mM TCEP. Top panel: raw data for 1 x 0.5 µL, followed by 17 x 2 µL injections of 1 mM ADPR into the isothermal cell containing 0.05 mM PHGDH. Data were corrected for heat of ADPR dilution by subtracting the heats from ADPR to buffer titration. Bottom panel: integrated heats from the peaks in the top panel plotted against the molar ratio of ADPR to PHGDH. The line of best fit to the data was plotted as obtained by non-linear regression using the built-in one-site fit model of the ORIGIN software. (B) Data obtained from the one-site fit model in ORIGIN. Stoichiometry of the reaction (= number of binding sites, N), changes in entropy (ΔS) and enthalpy (ΔH) as well as the association constant (K_a) were determined. The binding constant (K_D) was calculated by taking the reciprocal of K_a. (C) Thermal denaturation of 1 µM PHGDH in the presence of 0.2 mM ADPR in combination with 0.2 mM substrate analogue. Graphs represent mean and standard deviation of at least two independent experiments with two replicates per experiment. Statistical analysis was performed in GraphPad Prism (two way ANOVA). **** = $p < 0.0001$, ** = $p = 0.0022$.

Due to its structural similarity with NAD$^+$, together with the finding that ADPR was able to inhibit not only full-length, but also sPHGDH activity, ADPR was assumed to bind to the cofactor-binding site in a similar manner as NAD$^+$. In order to prove this, co-crystallisation of sPHGDH with ADPR was attempted. Initial crystal trials of sPHGDH in the presence of ADPR yielded crystalline hits from 0.1 M BIS TRIS, pH 5.5, 0.2 M NH$_4$Ac, 25 % (w/v) PEG3350 and this condition was used as start point for further optimisation (see section 4.4.4). From the optimisation, several crystals were

127

obtained, of which one diffracted sufficiently well to collect a dataset. This crystal was grown from 0.1 M BIS TRIS, pH 6, 0.3 M NH$_4$Ac, 23 % (w/v) PEG3350, and a dataset (sPHGDH 03) was collected to 2.2 Å resolution using synchrotron radiation (Table 5.2).

	sPHGDH 03
Wavelength (Å)	0.98
Space Group	P 1
a, b, c (Å)	43.2, 59.5, 63.7
α, β, γ (°)	89, 90, 80
Resolution (Å)	58.6 - 2.16 (2.22 – 2.16)
I/σI	3.7 (1.9)
R$_{merge}$	0.02 (0.53)
Completeness (%)	97.2 (95.3)
Multiplicity	2.0 (1.9)

Table 5.2 – Data collection statistics for sPHGDH co-crystallised with ADPR (sPHGDH03).

Data were collected from a single crystal at 100 K on beamline I02, DLS. The data were processed using xia2 (3daii-run). The statistics are presented as averages with values for the highest resolution shell in parentheses.

Analysis of the data showed that neither ADPR nor NAD$^+$ completely satisfied the observed electron density in the cofactor-binding site. This was best illustrated by modelling NAD$^+$ into the active site. The weighted electron density map from the subsequent refinement shows no evidence for the presence of the amide function (Figure 5.5 A). In addition, the nicotinamide moiety refines with much higher B-factors than the ADPR moiety, indicating either a large degree of flexibility or that this moiety is partly missing. A partial occupancy with some cofactor-binding sites containing ADPR and others NAD$^+$ provides a good fit to the experimental data.
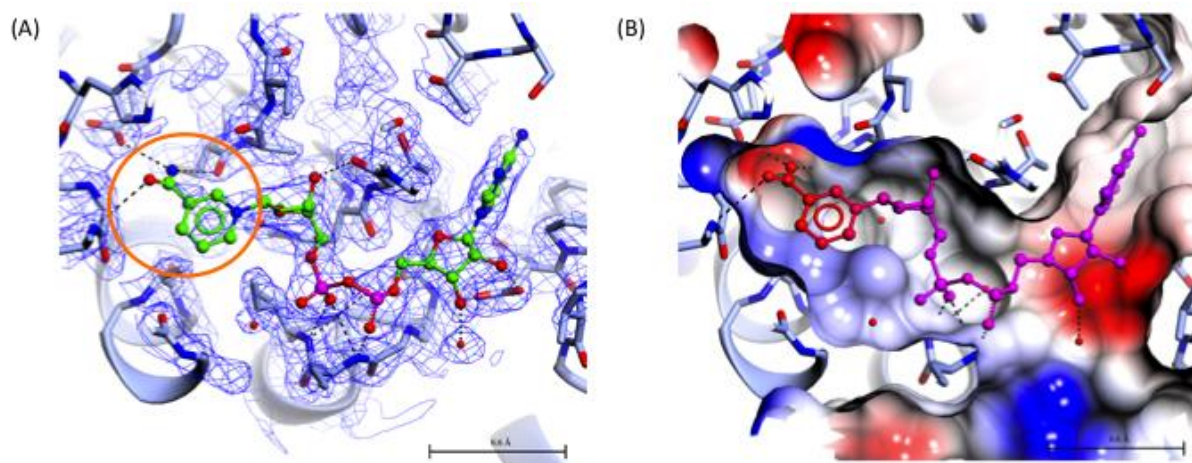
Figure 5.5 – Cofactor-binding site of crystal sPHGDH03 with NAD$^+$ modelled into the active site.
(A) Weighted electron density map from refinement (2Fo-Fc) contoured in blue at 0.39 electrons/$\bar{A}^3$. Position of the nicotinamide-moiety highlighted by an orange circle. (B) Detail of the cofactor-binding site with NAD$^+$ coloured according to temperature factors with increasing temperature factors from pink to red. The protein surface was coloured by electrostatic potential.

This interpretation of the electron density maps for crystal sPHGDH03 suggests that sPHGDH is likely to co-purify with its cofactors. This was supported from measured OD_{260}/OD_{280} ratios of purified sPHGDH, which were around 0.8 whereas 0.57 is reported for pure protein. Treatment of sPHGDH with charcoal, which is known to adsorb nucleotides, resulted in a reduction of the OD_{260}/OD_{280} ratio to 0.65 (see section 6.4.1). By contrast, full length PHGDH co-purified with NAD$^+$ or NADH to a lesser extent (if indeed at all) according to a measured OD_{260}/OD_{280} ratios of 0.66. Measurements of binding affinities of NAD at both oxidation states (NAD$^+$ and NADH) to full-length PHGDH by ITC, showed that especially NADH had a very high affinity for PHGDH with a K_D of 0.22 µM, whereas the K_D measured for NAD$^+$ was 2000-fold lower at 0.44 mM (Figure 5.6).
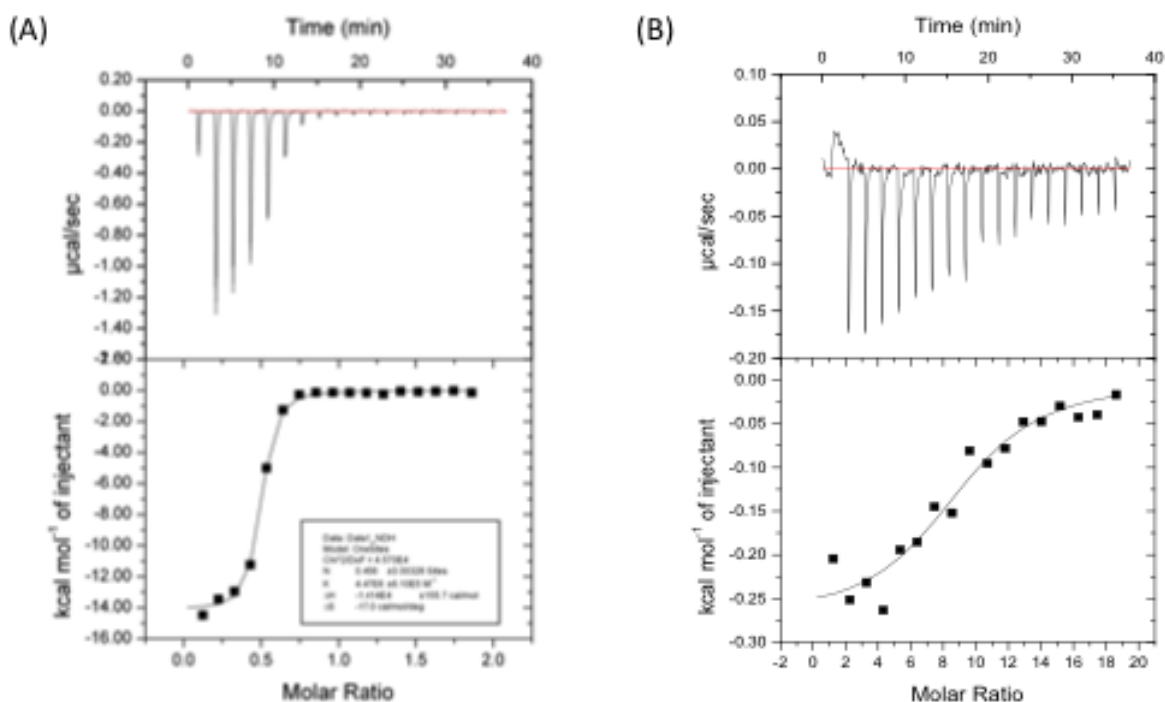
Figure 5.6 – Binding of NAD$^+$ and NADH to PHGDH as investigated by ITC.

(A) 0.5 mM NADH or (B) 5 mM NAD$^+$ was titrated into 0.05 mM PHGDH in 25 mM HEPES, pH 7.5, 100 mM NaCl and 0.5 mM TCEP. Top panel: raw data for 1 x 0.5 µL, followed by 17 x 2 µL injections of NADH or NAD$^+$ into the isothermal cell containing 0.05 mM PHGDH. Data were corrected for heat of cofactor dilution by subtracting the heats from cofactor to buffer titration. Bottom panel: integrated heats from the peaks in the top panel plotted against the molar ratio of cofactor to PHGDH. The line of best fit to the data was plotted as obtained by non-linear regression using the built-in one-site fit model of the ORIGIN software. (B) Data obtained from the one-site fit model in ORIGIN. Stoichiometry of the reaction (= number of binding sites, N), changes in entropy (ΔS) and enthalpy (ΔH) as well as the association constant (K$_a$) were determined. Binding constant (K$_D$) was calculated by taking the reciprocal of the K$_a$.

5.4.2 NAD+ analogues

NAD+ analogues were used to study the effect of changes around the nicotinamide-binding subsite on the enzymatic activity and stability of PHGDH and to understand what functional groups would be tolerated in this part of the cofactor-binding site.

TAD, APAD and PAD were tested in the enzyme activity assay instead of NAD+ and all of them were found to allow the enzymatic reaction to proceed (Figure 5.7). In particular, TAD was very similar to NAD+ in terms of maximum velocity (V_{max}) (10.0 and 7.2 pmol/min for TAD and NAD+ respectively) and k_{cat} (0.45 and 0.32 min^{-1} for TAD and NAD+ respectively). In contrast, V_{max} for APAD and PAD was higher than NAD+. APAD showed a 10-fold increase in k_{cat} compared to NAD+, whereas PAD showed a slightly reduced k_{cat}.
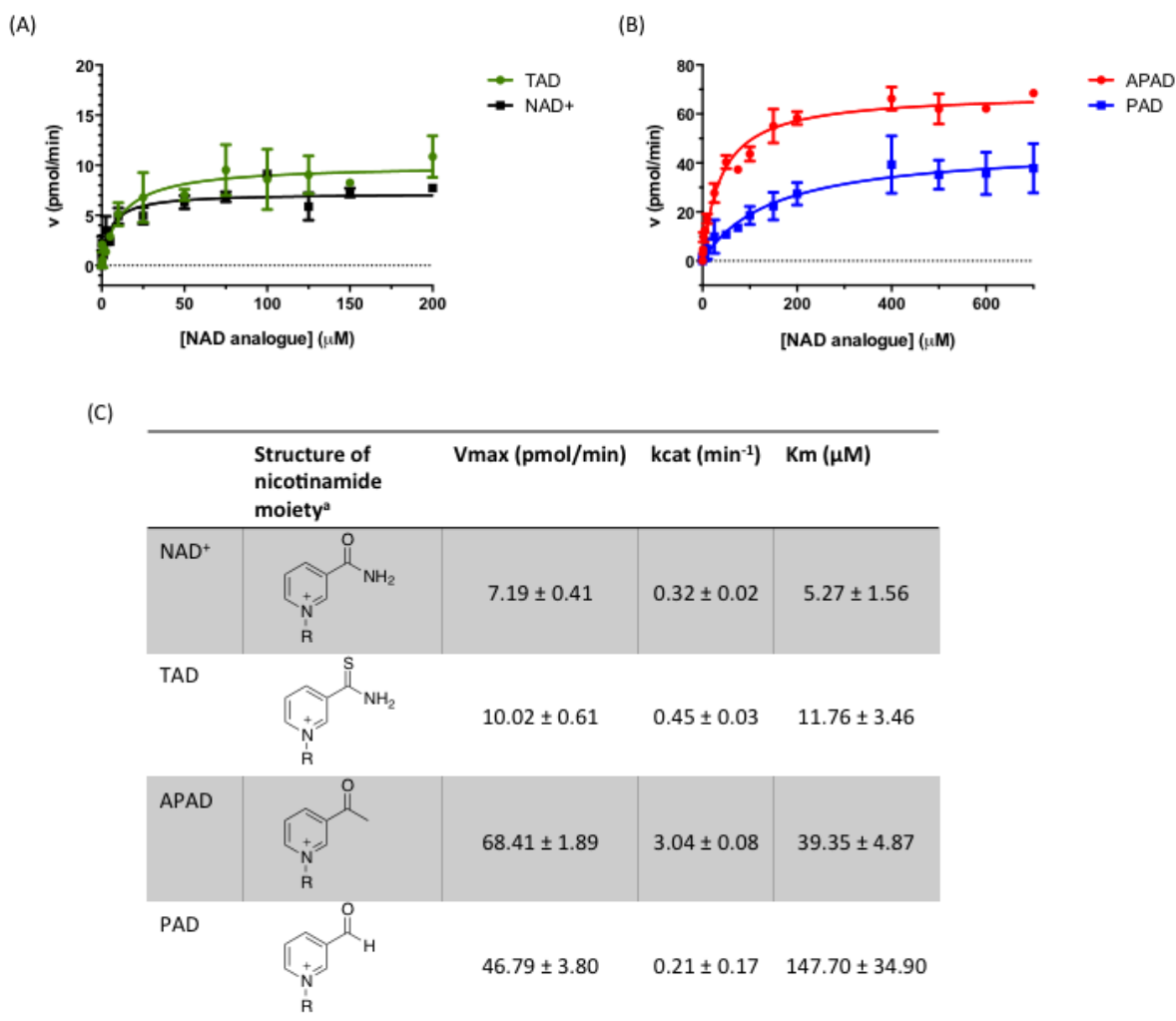
Figure 5.7 – Enzyme kinetic data for PHGDH with various substrate analogues.

In vitro PHGDH activity assay containing recombinant human PHGDH and increasing concentrations of NAD[+] or TAD (A) or the NAD[+] analogues PAD or APAD (B). The increase in fluorescence due to conversion of NAD[+] to NADH was measured and initial velocity determined by linear regression. The initial velocity was plotted against cofactor/ cofactor analogue concentration and data fitted to a Michaelis-Menten model. All curve fitting was performed using GraphPad Prism. Data plotted represents mean and standard deviation of two independent experiments with two replicates per experiment. (C) Summary of calculated parameters of PHGDH activity. [a] R is used to denote the ADP-ribose moiety.

All cofactor analogues tested could replace NAD[+] and sustain PHGDH activity. In order to compare their binding affinities with NAD[+], a DSF assay was performed. All the cofactor analogues stabilised PHGDH less than NAD[+], with APAD showing no increase in T_m, suggestive of a lack of binding (Figure 5.8). In combination with 3-PG, all analogues showed the same synergistic behaviour, increasing the T_m of PHGDH to the same extent as NAD[+]. This result is even more striking considering that the NAD[+] analogues yielded only small increases in T_m of PHGDH when used on their

own. Intriguingly, for NAD$^+$ analogues in combination with other substrate analogues, a significant change in T$_m$ was only seen for PAD together with α-KG.

Compound(s)	Δ Tm (- Buffer control) (°C)
3-PG	2.5 ± 0.6
3-PG + NAD+	4.5 ± 0.8
3-PG + APAD	4.8 ± 1.1
3-PG + PAD	5.1 ± 1.2
3-PG + TAD	4.6 ± 0.6
DL-Malate + NAD+	2.2 ± 0.8
α-Ketoglutarate + PAD	2.6 ± 0.7
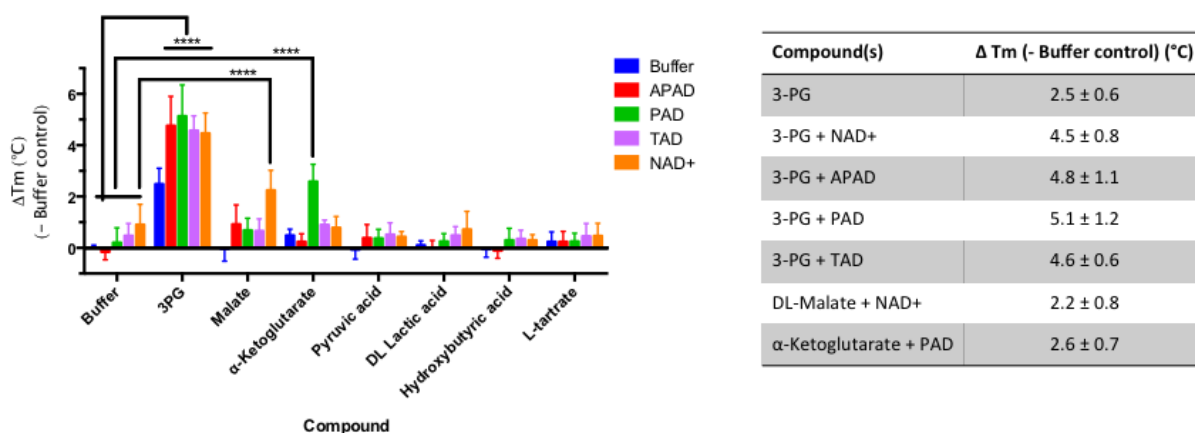
Figure 5.8 – Change in T$_m$ (ΔT$_m$) of PHGDH upon addition of NAD$^+$ analogues in combination with substrate/ substrate analogues.

Thermal denaturation of 1 μM PHGDH in the presence of 0.2 mM NAD$^+$ analogue in combination with 0.2 mM substrate analogue. Graphs represent mean and standard deviation of at least two independent experiments with two replicates per experiment. Statistical analysis was performed in GraphPad Prism (two way ANOVA). **** = $p < 0.0001$.

Co-crystallisation of sPHGDH with the cofactor-analogues was attempted by sitting drop vapour diffusion in 96-well plates. Optimisation trays around the published crystallisation conditions using MMT buffer (optimisation screen 1) and the in-house established conditions with BIS-TRIS buffer (optimisation screen 2) were set up with 15 mg/mL sPHGDH and 5 mM cofactor analogue at RT and 4 °C. Crystals were obtained with the cofactor analogues TAD and APAD only in conditions of the optimisation screen 2. Three datasets of co-crystals of sPHGDH with TAD were collected and analysed and an example dataset (sPHGDH04) collected to 1.95 Å resolution using synchrotron radiation is shown (Figure 5.9). NAD$^+$ and TAD differ only in the acetamide group that is a thioacetamide group in TAD. Although this means that the oxygen is changed for sulphur, the geometry of the respective adjacent carbon is trigonal in both molecules and due to the similar electronic configuration of oxygen and sulphur, sulphur is also able to form the same hydrogen bonds. Therefore from the electron density alone, it cannot be excluded that the active site of sPHGDH in the co-crystal with TAD is still partly occupied by NAD$^+$.

(A)	sPHGDH 04
Wavelength (Å)	0.98
Space Group	P 1 21 1
a, b, c (Å)	43.2, 125.2, 59.1
α, β, γ (°)	90, 100, 90
Resolution (Å)	37.67 – 2.10 (2.16 – 2.10)
I/σI	3.6 (0.9)
R_{merge}	0.17 (0.70)
Completeness (%)	96.1 (73.0)
Multiplicity	3.3 (1.9)
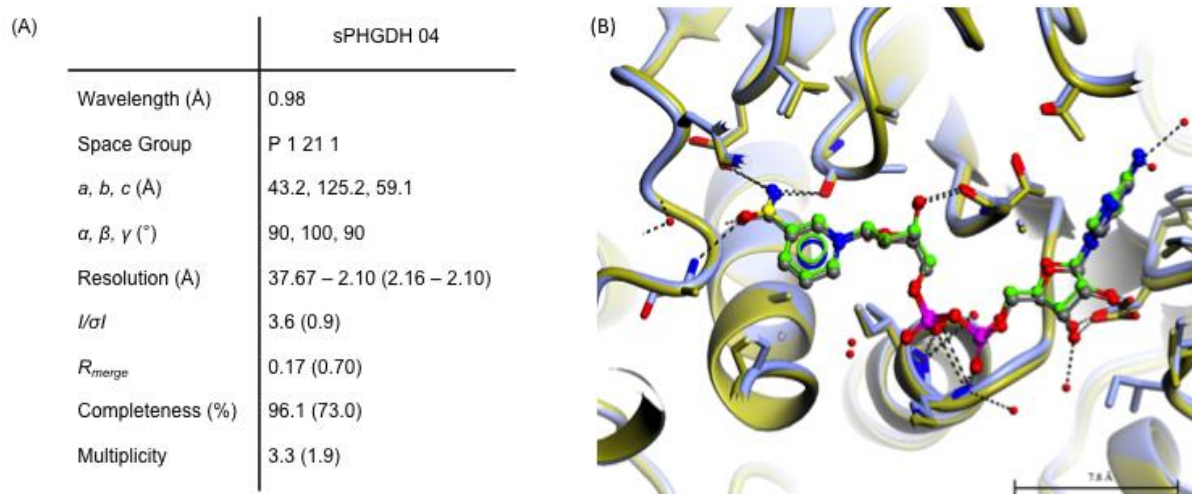
Figure 5.9 – Data collection statistics for sPHGDH co-crystallised with cofactor analogue TAD (sPHGDH 04).

(A) Data collection statistics for sPHGDH 04 (co-crystal with TAD). The data were collected on a single crystal at 100 K on beamline I04, Diamond Light Source (DLS). The data were processed, scaled and merged using xia2 (3daii-run), aimless, pointless and programs from the ccp4i2 suite. The statistics are presented as averages with values for the highest resolution shell in parentheses. (B) Superposition of the active site of sPHGDH 04 with TAD bound onto sPHGDH with NAD[+] bound (PDB 2G76). Protein secondary structure is shown in blue (sPHGDH 04) and gold (PDB 2G76) and TAD and NAD[+] are coloured by atom.

5.4.3 NAD[+] fragments

Phosphate groups are polar and negatively charged at physiological pH. The phosphate linker within the NAD[+] cofactor is therefore able to contribute substantially to ligand binding to the protein, provided complementary positive charges are suitable presented in the cofactor-binding site. In order to further investigate the effect of the phosphate groups on binding of NAD[+] to PHGDH, two cofactor fragments, namely ADP and AMP, as well as ATP, which has an additional phosphate group compared to NAD[+], were tested in the DSF assay. The results from the DSF assay clearly show an increase in ability to stabilise the T_m of PHGDH with increasing number of phosphate groups, making ATP the strongest binding partner with an increase in T_m of 4.7 ± 0.9 °C (Figure 5.10 A). Addition of substrate analogues to the NAD[+] fragments failed to further amplify the effect on the T_m of PHGDH. Direct comparison of NAD[+] with ADP shows that whilst ADP, in contrast to NAD[+], does not show a synergistic effect with any of the substrate analogues, ADP promotes a significantly larger increase in T_m of PHGDH than does NAD[+], and this effect is independent of the addition of substrate analogues (Figure 5.10 B).
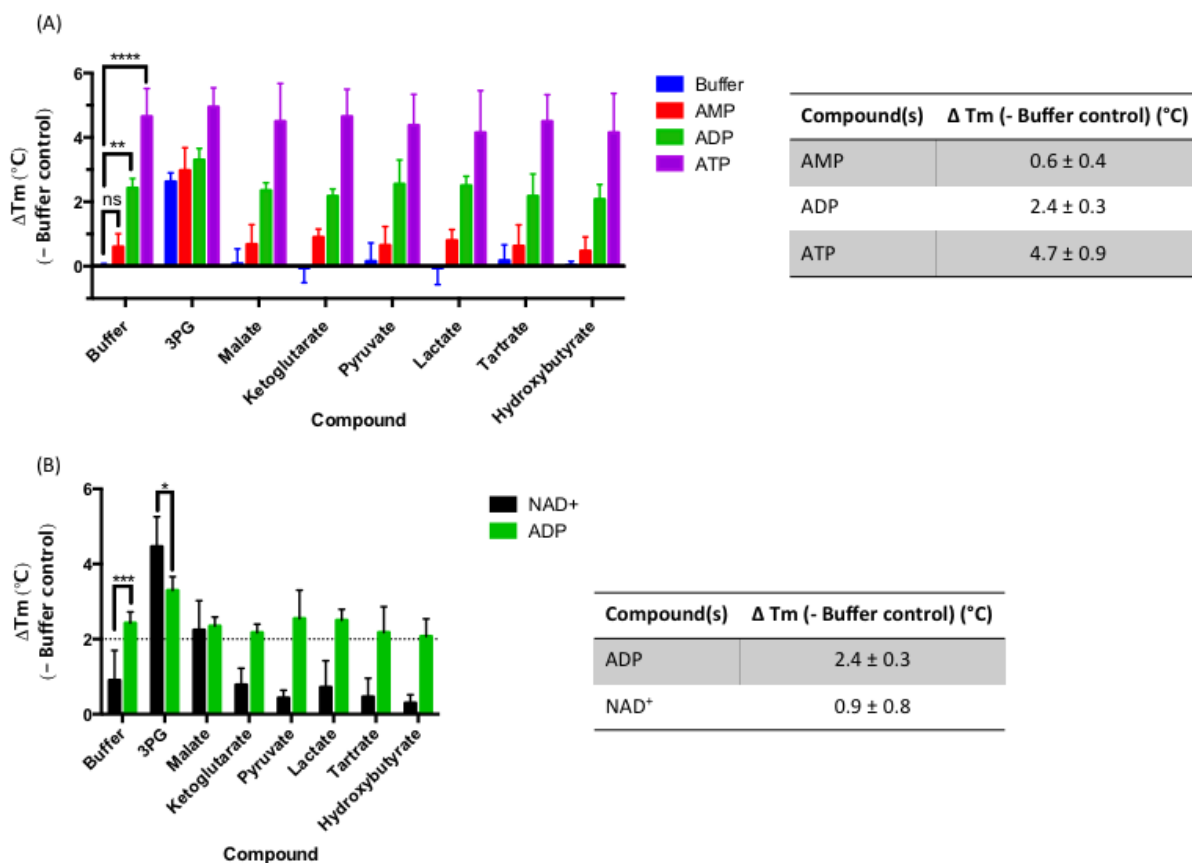
Figure 5.10 – Change in T_m (ΔT_m) of PHGDH through the addition of NAD$^+$ fragments in combination with substrate (analogues).

(A) Thermal denaturation of 1 µM PHGDH in the presence of 0.2 mM NAD$^+$ fragment and in combination with 0.2 mM substrate analogue. (B) Comparison of changes in T_m of PHGDH in the presence of NAD$^+$ or ADP. Graphs represent mean and standard deviation of at least two independent experiments with two replicates per experiment. Statistical analysis was performed in GraphPad Prism (two way ANOVA). * = p = 0.041, ** = p = 0.0003, **** = p < 0.0001.

In addition to the binding studies by DSF, the NAD$^+$ fragments were also investigated for their potential to displace NADH using the FP assay. Within the group of AMP, ADP and ADPR the potency for displacement of NADH from the protein increases with the size of the fragment (Figure 5.11 A-C). The difference between AMP, which results in 25 % displacement at the highest concentration of 3 mM, and ADP, which yields 70 % displacement at the same concentration, is particularly remarkable. The presence of the second phosphate group in ADP has a marked effect on potency. ADPR is the most active compound in this assay with EC$_{50}$ values of 236.3 ± 1.1 µM (PHGDH) and 188.7 ± 1.1 µM (sPHGDH), whereas NAD$^+$ shows a similar effect to ADP (**Figure 5.11** C-D).
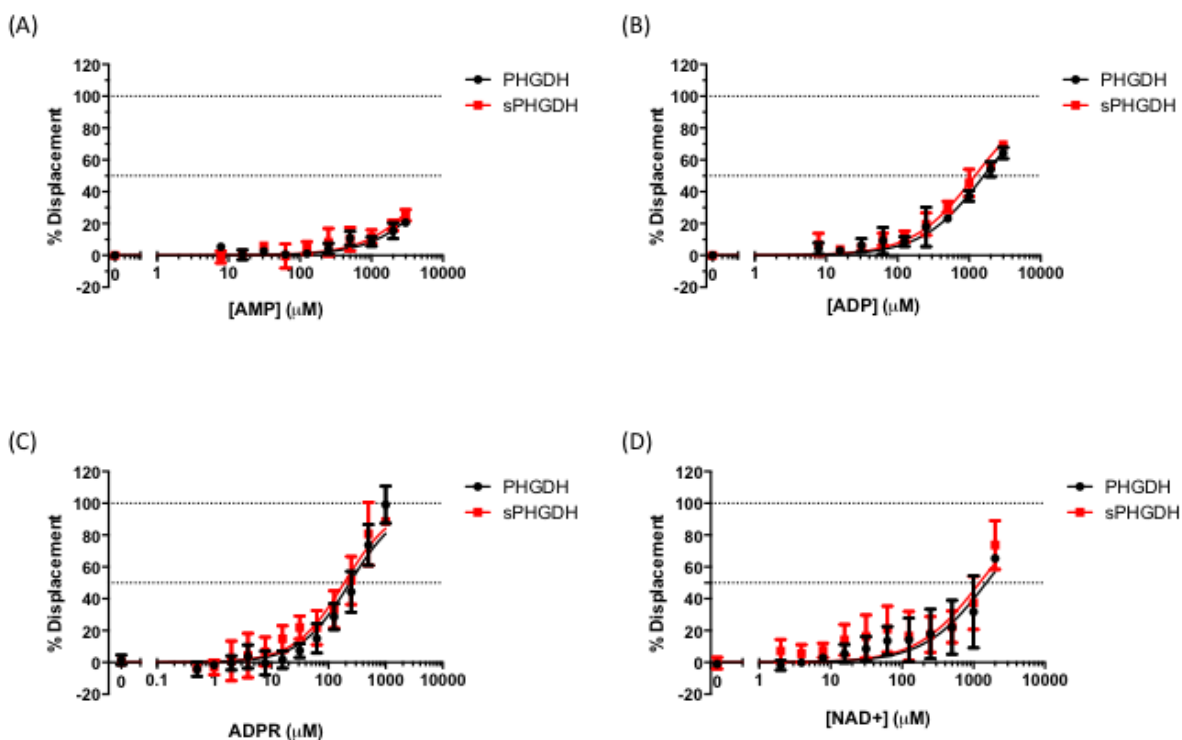
Figure 5.11 – Binding of NAD+ fragments to PHGDH and sPHGDH determined by FP.

FP of NADH was measured with 1 µM NADH and 2 µM PHGDH or sPHGDH in the assay mixture in the presence of NAD+ and fragments thereof. The compounds were tested at increasing concentrations over the range of 8-3000 µM for AMP (A) and ADP (B), and over the range of 0.5-1000 µM for ADPR (C), all three fragments of NAD+, as well as NAD+ itself over the range of 2-2000 µM (D).

Previous crystallisation experiments showed that NAD+ was often present in the crystal structure to some extent, even when not specifically added to the crystallisation conditions, which indicated co-purification of the protein with its cofactor (section 5.4.1). Fragments of NAD+ (AMP and ADP) were shown to bind to sPHGDH and to be able to displace NADH (Figure 5.11A and B). Therefore co-crystallisation of charcoal-treated sPHGDH with the NAD+ fragments AMP and ADP was hoped to yield crystals containing only the NAD+ fragments, thus leaving the nicotinamide pocket empty and accessible to inhibitors binding to this part of the active site.

Co-crystallisation of sPHGDH with AMP and ADP was therefore attempted to assess the degree of displacement of pre-bound cofactor and thus the potential to use these crystals for soaking of small inhibitors intended to bind to the nicotinamide subsite of the cofactor-binding pocket and/or the substrate-binding pocket. In addition, prior to crystallisation, sPHGDH was treated with activated charcoal onto which nucleotides

are known to adsorb as an additional means to remove the co-purified cofactor (see section 6.4.1).

Co-crystallisation of sPHGDH with 5 mM AMP resulted in a number of crystals, which grew from various conditions of both commercial sparse matrix screens as well as the optimisation screen II. A single dataset from one crystal grown in 0.1 M Sodium HEPES/ MOPS buffer, pH 7.5, 0.12 M monosaccharides, 30 % (w/v) PEG550-MME and PEG20K was obtained (sPHGDH 04) (Figure 5.12). Disappointingly, analysis of the data showed residual electron density for parts of nicotinamide and adjacent ribose (Figure 5.12). The same was found to be true for co-crystals of sPHGDH with ADP.

	sPHGDH 05 with AMP
Wavelength (Å)	0.92
Space Group	P 1 21 1
a, b, c (Å)	43.3, 127.4, 59.5
α, β, γ (°)	90, 101, 90
Resolution (Å)	63.7 – 1.85 (1.89 – 1.85)
$I/\sigma I$	8.4 (0.9)
R_{merge}	0.10 (1.46)
CC(1/2)	0.997 (0.463)
Completeness (%)	99.5 (98.9)
Multiplicity	3.5 (3.4)
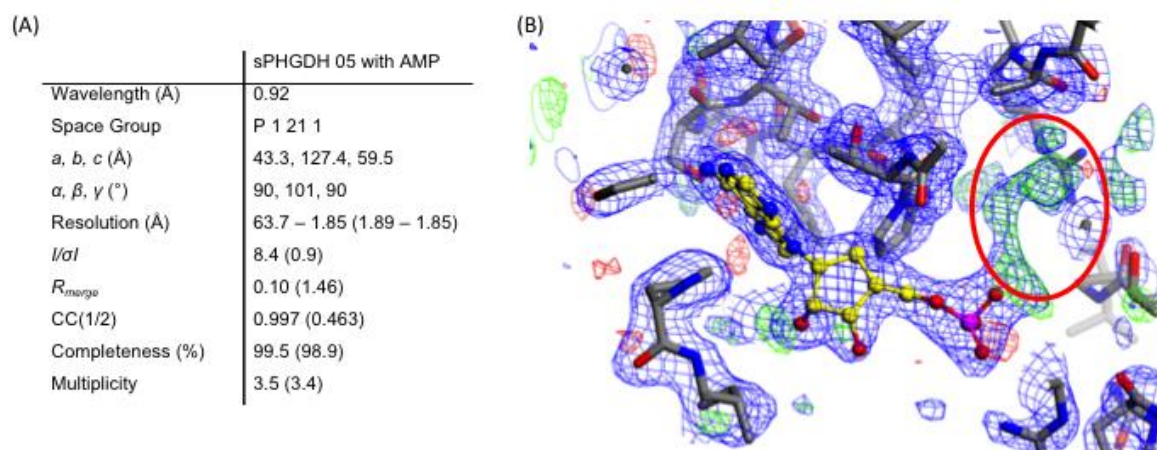
Figure 5.12 – Co-crystallisation of sPHGDH with AMP (sPHGDH 05)

A) Data collection statistics for sPHGDH 05 (co-crystal with AMP). The data were collected on a single crystal at 100 K on beamline I04-1, Diamond. The data were processed, scaled and merged using xia2 (3daii-run), aimless, pointless and programs from the ccp4i2 suite. The statistics are presented as averages with values for the highest resolution shell in parentheses. (B) Active site of sPHGDH 05 with AMP bound. Protein and AMP are coloured by atom. Additional electron density seen after AMP was built into the model indicating presence of NAD (red circle); electron density map (2Fo-Fc) shown as blue wire and the difference map (Fo-Fc) in red and green at contour levels of 0.29 electrons/$\bar{A}^3$.

5.4.4 Exploring the adenine moiety

Interactions in the adenine-binding pocket mainly occur through Ser211, which donates a hydrogen bond to the N-3 of the purine ring as well as accepting a hydrogen bond from the exocyclic amino-group (Figure 5.13A). A small pocket seemed available to accommodate a group at the 6-position of the purine ring. To investigate how well removal of the hydrogen-donor amino-group would be tolerated, nicotinamide hypoxanthine dinucleotide (NHD), which has a keto-group at the 6-position of the purine ring, and guanosine 5'-diphosphate, which has the keto-group in the 6-position and an additional amino-group in the 2-position, were tested in the DSF assay (Figure 5.13 B and C).
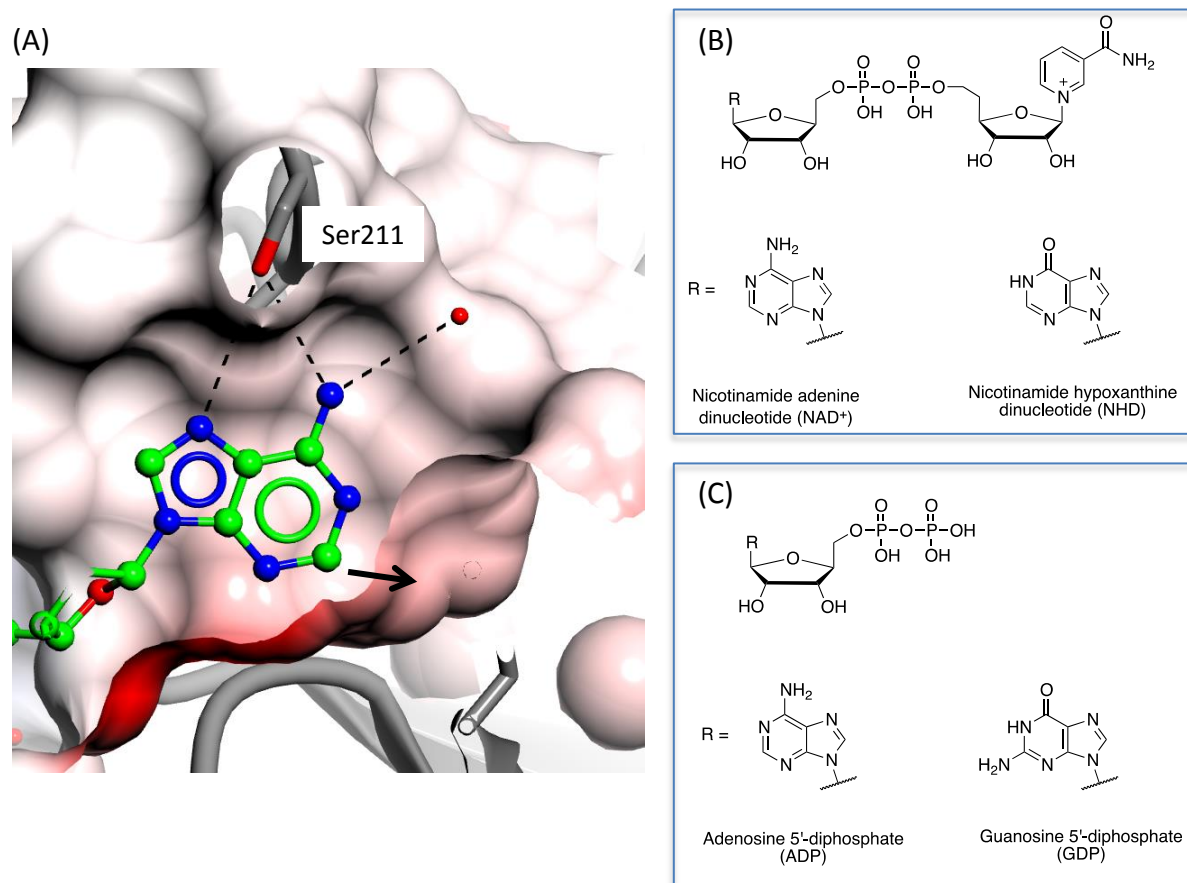
Figure 5.13 – Structures of nucleobase analogues.

(A) Detail of the cofactor-binding pocket of sPHGDH highlighting the interactions of the adenine moiety of NAD[+]. PHGDH backbone is shown in grey with protein surface shown according to its electrostatic potential. Structure of NAD[+] is coloured by atom type. (B-C) Structures of the NAD[+] analogue nicotinamide hypoxanthine dinucleotide (NHD) and (B) the ADP analogue, guanosine 5'-diphosphate (GDP).

Results from NHD and GDP addition to PHGDH in the DSF assay were compared to effects seen with their respective natural ligand counterparts, NAD[+] and ADP. In the DSF assay, GDP behaved very similarly to ADP, showing an identical increase in T_m of PHGDH of 2.2 ± 0.3 °C (Figure 5.14 A). As already observed for ADP, no additional increase in T_m was seen for GDP in combination with substrate/ substrate analogues. On the contrary, NHD was unable to stabilise PHGDH and did not assist malate binding in contrast to what was seen with NAD[+]. This indicates that exchanging the adenine moiety in NAD[+] for hypoxanthine, which is equivalent to replacing the amine group in the 4-position with a ketone group and thereby losing the hydrogen bond donor attributes of the amine, is detrimental for binding to PHGDH (Figure 5.14 B).
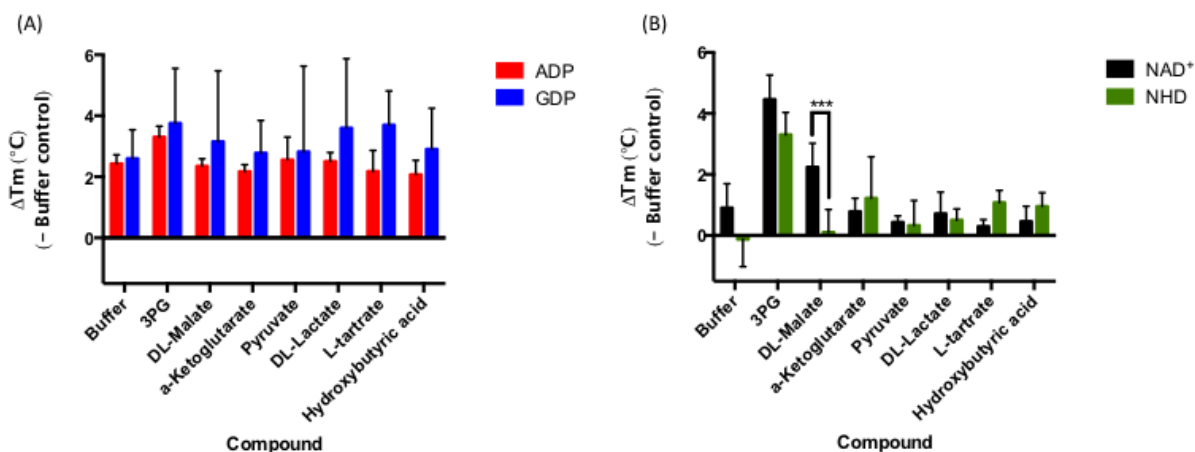
Figure 5.14 - Change in T_m (ΔT_m) of PHGDH through the addition of nucleobase analogues in combination with substrate/ substrate analogues.

Thermal denaturation of 1 µM PHGDH in the presence of 0.2 mM NAD$^+$, NHD, ADP or GDP in combination with 0.2 mM substrate analogue. Graphs represent mean and standard deviation of at least two independent experiments with two replicates per experiment. Statistical analysis was performed in GraphPad Prism (two way ANOVA). *** = $p < 0.001$.

During a previous study adenine analogues had been assessed for their potential to inhibit PHGDH activity using an *in vitro* enzyme activity assay (Figure 5.15). In this assay, which measures PHGDH activity in the physiological direction of 3-PG oxidation, compounds **14** and **15** were found to be moderate inhibitors of PHGDH, inhibiting its activity by about 50 % at the highest assessed concentration of 120 µM [160].
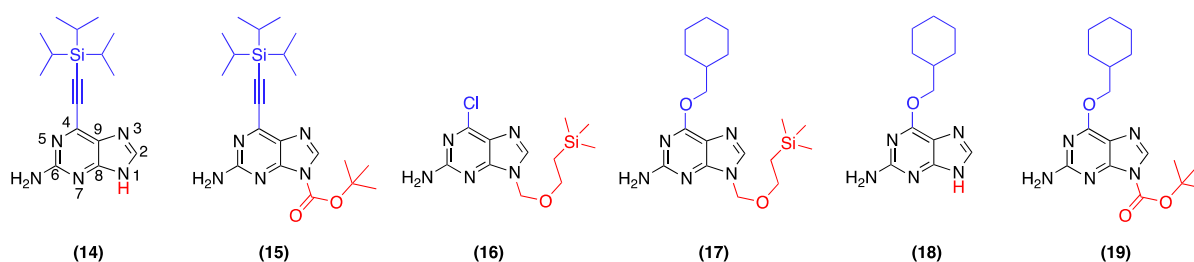
Figure 5.15 – Structures of available adenine analogues from the NICR compound library.

All adenine analogues investigated contain an additional amino group at the 6-position together with modifications at the 1- (highlighted in blue) or 4-position (highlighted in red).

To confirm their inhibitory effects, compounds **14** and **15** were re-assessed in the *in vitro* PHGDH activity assay using the optimised protocol (see section 5.4.1). Re-testing confirmed the inhibitory potential of **14** and **15** as similar to the inhibition seen with the pan-dehydrogenase inhibitor ADPR (Figure 5.16A). From an extrapolation of

140

the sigmoidal dose-response fit, the IC_{50} values for **14** and **15** were determined as 173.4 ± 1.2 µM and 276.5 ± 1.3 µM, respectively. Interestingly, and in contrast to ADPR, no inhibition was seen in the *in vitro* enzyme activity assay with truncated sPHGDH (Figure 5.16 B).
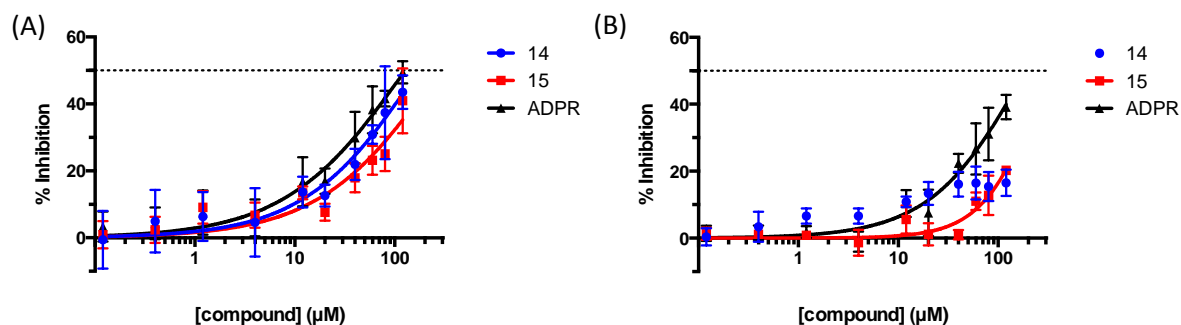
Figure 5.16 – Inhibition of PHGDH (A) and sPHGDH (B) with compounds 14 and 15 compared to ADPR.

Enzyme activity was measured in an in vitro activity assay using purified recombinant human proteins. Inhibitor concentrations ranged from 0.12 to 120 µM. Graphs represent averages and standard deviations of three independent experiments. Data were fitted to a dose-response model using GraphPad Prism.

The lack of inhibitory activity against sPHGDH, which comprises the active site and misses any regulatory domains, suggest that these compounds may not in fact target the adenine pocket. To investigate this further the mode of inhibition of compounds **14** and **15** was determined by kinetic studies and compared to that of ADPR. In the case of an NAD^+- competitive inhibitor, the same maximum velocity (V_{max}) should be achievable in the presence of the inhibitor as in the absence, with inhibition showing up as an apparently elevated K_m in the presence of the inhibitor. A noncompetitive inhibitor, which binds outside the active site and induces some conformational change that inhibits enzymatic activity, results in a decreased V_{max} (Figure 5.18 A). Equation 5.1, which describes noncompetitive binding, could be used to distinguish between competitive and noncompetitive inhibition by determining the parameter α. In the case where α >> 0, the equation can be simplified to that describing a competitive inhibitor, whereas for a noncompetitive inhibitor 0 < α < 1 (Equation 5.1).
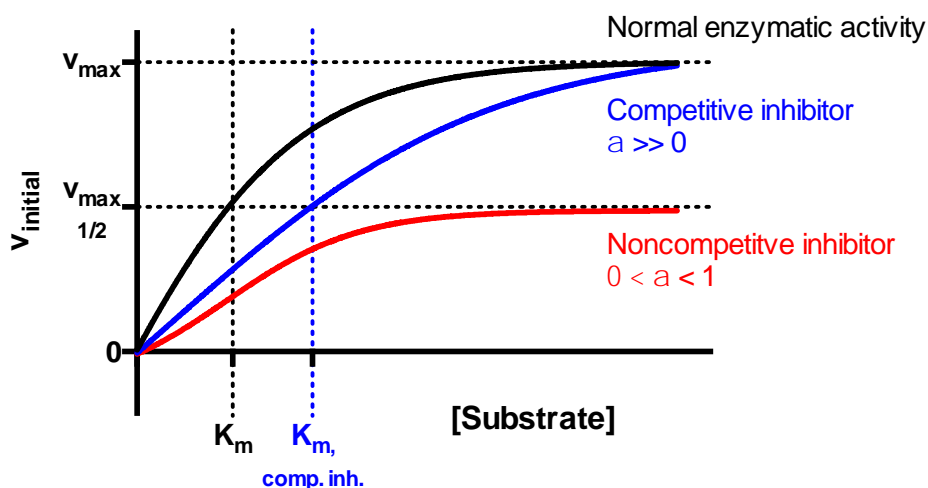
Figure 5.17 – Schematic Michaelis-Menten curve showing the effect of competitive and noncompetitive inhibitors.

(A)   (B)

$$v = \frac{v_{max}[S]}{[S]\left(1 + \frac{[I]}{\alpha K_i}\right) + K_M\left(1 + \frac{[I]}{K_i}\right)} \quad ; \quad for\ \alpha \gg 1:\ \ v = \frac{v_{max}[S]}{1 + K_M\left(1 + \frac{[I]}{K_i}\right)}$$

Equation 5.1 – Equation relating initial velocity to substrate concentration in the presence of an inhibitor for noncompetitive (A) and competitive (B) inhibition [161].

K_i = dissociation constant of the Enzyme-Inhibitor complex

αK_i = dissociation constant of the enzyme-inhibitor-substrate complex

V_{max} = maximum velocity at infinite substrate concentration

K_M = Michaelis-Menten constant = substrate concentration yielding half of V_{max}

Variability in the activity of PHGDH was observed between different purification batches. The independent repeats for each inhibitor were therefore analysed separately and a representative experiment is shown (Figure 5.18). As anticipated, ADPR was shown to act as a competitive inhibitor (α = 18.5-31.4), whereas compounds **14** and **15** were noncompetitive inhibitors (α = 0.16 – 0.67 and 0.75 – 1.36).
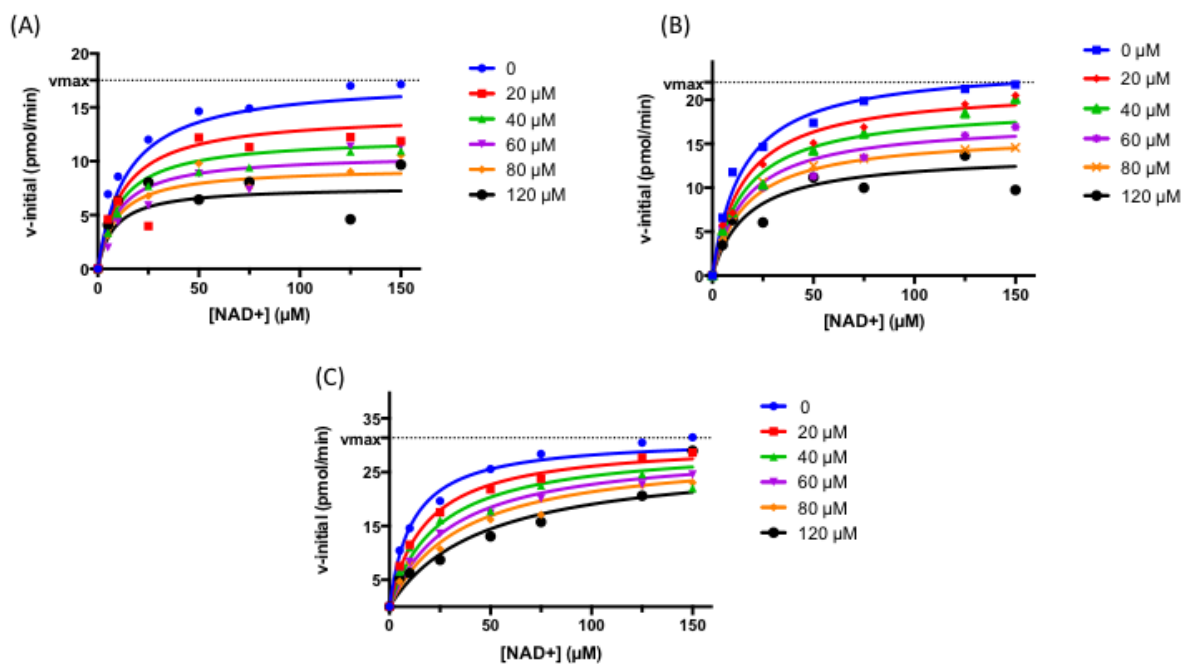
Figure 5.18 – Mode of inhibition for adenine analogues 14, 15 and ADPR.

In vitro activity of recombinant human PHGDH in the presence of varying concentrations of cofactor NAD$^+$ (0 – 150 µM) and constant substrate concentration (3-PG, 140 µM), and inhibitors 14 (A), 15 (B) or ADPR (C) (0 - 120 µM). Initial velocity was determined from the increase in fluorescence over time, and plotted against increasing concentrations of NAD$^+$. The data were fitted to the mixed-model equation in GraphPad Prism. Representative graphs of three independent experiments for each inhibitor are shown. The initial velocity was calculated from two intra-assay repeats.

Inhibitor	V_{max} (pmol/min)	K_M (µM)	K_i (µM)	α	Mode of inhibition
No inhibitor	7.2 ± 0.6	5.3 ± 3.5	-	-	-
14	6.7 ± 0.2	15.2 ± 2.2	146.1 ± 79.3	0.7 ± 0.4	Non-competitive
	23.5 ± 10.8	14.5 ± 1.9	1804 ± 6313	0.2 ± 0.6	
	17.6 ± 1.2	14.4 ± 3.8	465.8 ± 1157	0.2 ± 0.5	
15	30.0 ± 1.3	12.6 ± 2.1	149.3 ± 92.1	0.8 ± 0.6	Non competitive
	24.1 ± 0.8	15.0 ± 2.1	120.3 ± 51.6	1.4 ± 0.7	
	19.5 ± 0.7	14.0 ± 2.2	211.0 ± 143	1.0 ± 0.8	
ADPR	31.4 ± 1.2	11.4 ± 2.1	33.3 ± 9.5	28.6 ± 30.8	Competitive
	26.4 ± 0.7	10.4 ± 1.3	48.4 ± 10.8	8.9 ± 3.9	
	18.5 ± 0.4	10.4 ± 1.1	24.1 ± 3.9	15.4 ± 5.7	

Table 5.3 – Results of mixed-model inhibition fit in GraphPad Prism.

Three independent experiments were conducted with two intra-assay repeats. The data were analysed by fitting to the mixed-model inhibition equation (Equation 5.1) in GraphPad Prism. The determined α value was used to distinguish between competitive and noncompetitive inhibition.

5.5 Discussion

PHGDH belongs to the family of dehydrogenases, which catalyse redox reactions. Although different dehydrogenases catalyse reactions with a wide variety of substrates, most of them use NAD^+ as a common cofactor. The binding of NAD^+ is mediated by a conserved structural domain, the Rossmann fold, which suggests that specific inhibition of a dehydrogenase by targeting its NAD^+-binding site will be challenging. Therefore work in this chapter was focused on gaining a better understanding of the cofactor-binding site of PHGDH as well as probing for small compounds that could provide a starting point for potent inhibitors.

In the previous chapter, adenosine 5'-diphosphoribose (ADPR) was shown to be able to displace NADH from its binding site in an FP assay. Further characterisation of the interaction of PHGDH with ADPR using an *in vitro* enzymatic activity assay showed that ADPR was able to inhibit nearly 50 % of enzymatic activity at 120 μM not only of full-length PHGDH, but also using the truncated form sPHGDH. sPHGDH lacks the regulatory domains compared to full-length PHGDH, so that ADPR must bind to the catalytic domain and is expected to function as a competitive inhibitor of its structural analogue NAD^+. Although ADPR is only a moderate inhibitor of PHGDH, it represents an important tool compound for validation of biochemical and biophysical assays. ADPR is, however, also likely to have a general affinity for NAD^+-binding proteins and is a highly charged nucleotide analogue with difficulties to cross cell membranes. As such, it is not suitable as a chemical probe for target validation studies or for cell-based assays. Co-crystallisation studies of ADPR with sPHGDH revealed partial occupancy of the active site with a mix of cofactor and ADPR indicating co-purification of sPHGDH with bound cofactors. The pre-bound cofactor would make it challenging for weak inhibitors to bind to the active site of sPHGDH, and although partial removal of cofactor was achieved through charcoal treatment, this technique did not provide completely empty active sites (see section 6.4.1).

The NAD^+ analogues APAD, TAD and PAD are known to be able to replace NAD^+ in reactions catalysed by a variety of enzymes [162, 163]. This was also found to be true for PHGDH. In particular, APAD showed a ten-fold higher k_{cat} than NAD^+, which is likely due to its higher oxidation potential, which supports a faster reduction of APAD compared to NAD^+ [164]. Interestingly, analysis of the binding affinities of the cofactor analogues by DSF showed that none of the analogues stabilised PHGDH to

the same extent as NAD^+, and APAD, for example, did not stabilise PHGDH at all. However, all cofactor analogues were able to exert a stabilising effect in combination with 3-PG. This could be an indication that the catalytic reaction mechanism follows an ordered Bi Bi mechanism with substrate binding occurring before cofactor binding, as is the order of reaction found for PHGDH from *M. tuberculosis* [57]. In contrast, effective binding of *DL*-malate was only detected in the presence of NAD^+ and to a weaker extent with NAD^+ analogues, implying that in the absence of cofactor/ cofactor analogue, this substrate analogue was not able to bind to the enzyme. With the exception of α-ketoglutarate in combination with PAD, this effect was not observed with other substrate analogues. The significant increase in the T_m of PHGDH in the presence of PAD and α-KG, but not seen when combining one of the other cofactor/ cofactor analogues with α-KG is unlikely to be due to a difference in the electron distribution around the aromatic ring. The aldehyde, like the amide substituent on the aromatic ring, has an electron withdrawing inductive and resonance effect on the aromatic ring. Therefore it is more likely due to the difference in size with hydrogen replacing the NH_2 group in PAD [165]. This would also explain why the effect is only seen with PAD and not with APAD and TAD with the latter having a methyl group or an NH_2 group at the position of the hydrogen in the aldehyde group, respectively. Co-crystallisation of NAD^+ analogues in combination with substrate analogues was attempted to validate this hypothesis. A crystal structure with TAD was obtained, however it was not clear whether TAD or NAD^+/NADH was bound to the active site due to the co-purification of sPHGDH with cofactors.

The phosphate-binding region in the cofactor-binding site is another challenging area of this pocket to target as phosphate groups are charged at physiological pH and in the case of PHGDH show multiple interactions with the protein. In addition to charge, the phosphate groups possess a difficult to mimic sp^3 geometry, making them doubly difficult to replace with functional moieties typically found in bioavailable compounds. The NAD^+ fragments AMP, ADP and ATP were investigated in the DSF assay and highlighted that the addition of beta and gamma phosphate groups to AMP potentiated binding affinity. No synergistic effect was seen in combination with substrate analogues, which was expected, as the NAD^+ fragments do not contain the nicotinamide moiety, which is the part of cofactor that makes interactions with the

substrate. Interestingly, there was no additive effect with 3-PG (i.e. the T_m shift induced by 3-PG was of similar size as that of 3-PG together with AMP or ADP). The ability of AMP and ADP to displace NADH was also investigated in the FP assay. ADP was shown to displace NADH as effectively as NAD^+ whereas ADPR is more effective in displacing NADH than NAD^+ as well as ADP. Taking these findings together shows that in the absence of substrate, the nicotinamide moiety, although showing some specific interactions with the protein, makes a minimal contribution to the affinity for PHGDH.

Finally, the adenine-binding subsite was investigated, as it is assumed to be a pocket that may readily be targeted by heterocycles other than adenine. To investigate the effect of the amine function in the 4-position on the purine ring, NAD^+ was compared to NHD, which contains a keto group at the 4-position. NHD alone failed to stabilise PHGDH in the DSF assay, and was also unable to stabilise PHGDH in combination with *DL*-malate. This indicates that the interaction of the amine group with Ser211 of PHGDH makes an important contribution to binding affinity.

In addition, some guanine analogues with bulky side chains at the 1- and 4- positions of the purine ring were investigated. Compounds containing a triisopropylsilyl function in the 4-position showed potency in the enzymatic activity assay. However, in contrast to ADPR, inhibition was only seen with PHGDH and not sPHGDH indicating that these compounds probably bound outside the active site. This was confirmed by investigating their mode of inhibition, which was non-competitive for NAD^+. Although these compounds could inhibit PHGDH, their potency was probably driven by the somewhat un-drug-like triisopropylsilyl group, and together with the lack of information on their binding sites, they did not provide a good starting point for further optimisation.

Chapter 6. Site-directed mutants to investigate substrate- and cofactor-binding sites

6.1 Introduction

3-Phosphoglycerate dehydrogenase (PHGDH) was found to have a nanomolar affinity for its cofactor NADH, as described in chapter 5. This is in accordance with published data on other dehydrogenases where NADH binding is generally favoured over NAD$^+$ binding represented in lower free energies of binding [166]. In general, the thermodynamics of dehydrogenase-catalysed reaction reveal larger negative enthalpy changes accompanied with small positive entropy changes in the direction of NADH oxidation [166]. This is in accordance with literature findings showing that the equilibrium of the PHGDH catalysed reaction lies far in the direction of 3-PG with over 95 % of substrate/ product in the form of 3-PG [57]. This means that the reaction catalysed by PHGDH occurs predominantly in the direction of reduction of 3-phosphohydroxypyruvate (PHP) to 3-PG thereby oxidising NADH to NAD$^+$. Under physiological conditions the reaction is forced in the direction of oxidation of 3-PG through consumption of the reaction product PHP in the subsequent transamination reaction (Figure 6.1).
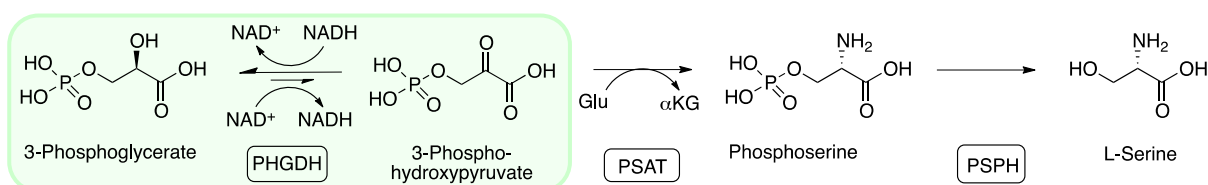
Figure 6.1 - Reaction catalysed by PHGDH.

Under physiological conditions the equilibrium of the PHGDH-catalysed reaction lies far in the direction of 3-phosphohydroxypyruvate (PHP) consumption. Only when 3-phosphohydroxypyruvate is further modified by phosphoserine aminotransferase (PSAT) followed by dephosphorylation to yield *L*-serine, will PHGDH catalyse the oxidation of 3-phosphoglycerate to compensate for the removal of PHP.

Due to the high affinity of PHGDH for NAD$^+$ and NADH, the artificially engineered truncated form sPHGDH always co-purified with its cofactors. This could be observed by measuring the OD_{260}/OD_{280} ratio after purification, showing an elevated ratio of 0.8-0.9, instead of 0.57 as could be expected for pure protein, indicating a contaminant that absorbs light at 260 nm [167]. In addition, the contamination observed spectrophotometrically was confirmed by analysis of the crystal structures

of different sPHGDH co-crystals with NAD$^+$ fragments, such as AMP or ADP, which contained more or less prominent electron density for the cofactor in the respective binding pocket (Figure 6.2). Although neither the OD_{260}/OD_{280} measurement nor the crystallographic data could give evidence whether NAD$^+$ or its reduced form NADH is bound to the protein, the 2000-fold higher affinity of NADH compared to NAD$^+$ for PHGDH suggests that sPHGDH co-purified with NADH.
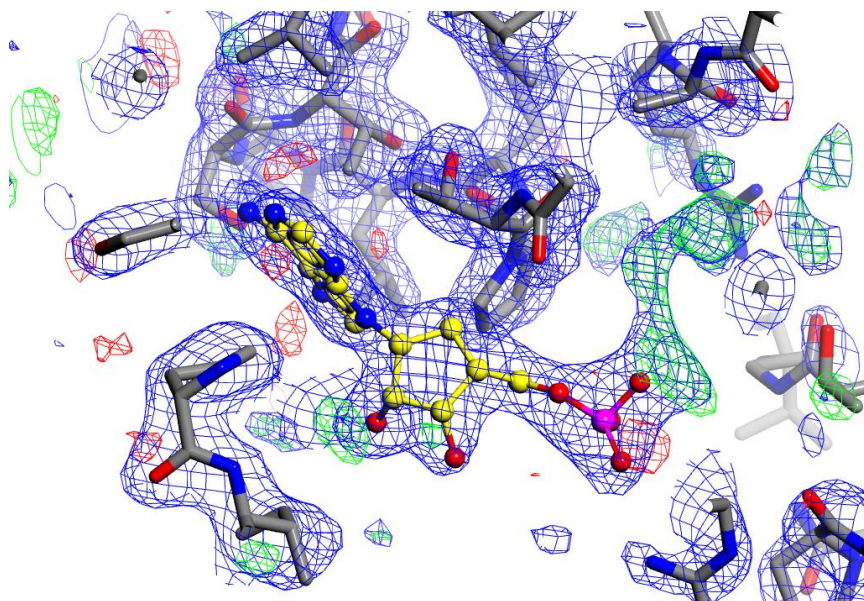
Figure 6.2 – Active site of sPHGDH with AMP

Active site of sPHGDH with AMP modelled into the active site. Protein and AMP are coloured by element. Additional electron density seen after AMP was built into the model indicating presence of NAD; electron density map (2Fo-Fc) shown as blue wire and the difference map (Fo-Fc) in red and green at contour levels of 0.29 electrons/$\bar{A}^3$.

To be able to search for potential inhibitors of PHGDH by targeting the cofactor- and substrate-binding sites, it is essential that these sites should be free and accessible. One possible strategy to remove the cofactor is through treatment of the protein with activated charcoal. This method was attempted and will be further explained in section 6.4.1.

Another option to remove the cofactor could be through unfolding of the protein followed by extensive dialysis and refolding. However, *in vitro* protein refolding can be difficult to achieve due to the differences between authentic folding and folding in an extracellular environment. These differences include differences in pH, temperature, macromolecular crowding, and chaperoning through co-translational interactions with the ribosome and post-translational interaction with dedicated

cellular protein-folding systems. In addition, it can be challenging to find suitable buffer conditions to perform protein folding *in vitro* as the buffer composition, ionic strength and pH can all effect the hydrophobic and electrostatic forces acting on the protein.

Against this background a novel approach was taken where key residues in the cofactor-binding pocket were mutated to prevent binding of the cofactors. A similar strategy of protein engineering of kinases is reported for the examination of biological signalling pathways [168]. As the binding pocket for NAD^+/NADH is quite large, it should be possible to mutate amino acids at one end of the pocket, e.g. the adenine-binding region, while leaving the other end, the nicotinamide-binding subsite, unchanged and *vice versa* (Figure 6.3).
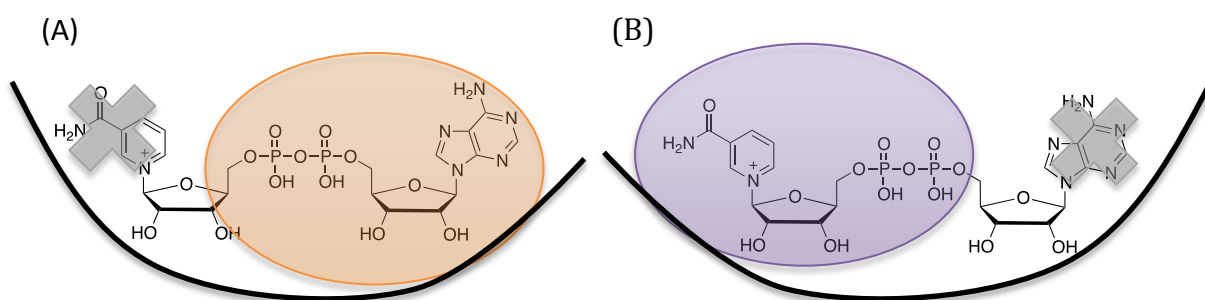
Figure 6.3 - Schematic representation of the mutation strategy

Mutants were designed in order to prevent cofactor binding while retaining free, accessible space for the binding of small molecule inhibitors.
(A) Mutation interfering with nicotinamide-binding subsite whilst the remaining part of the binding pocket (highlighted in orange) including the adenine-binding subsite is unchanged. (B) Opposite approach where the mutation is introduced at the adenine-binding subsite, whereas the nicotinamide-binding subsite (highlighted in purple) stays unmodified.

6.2 Aims and Objectives

In order to probe potential inhibitors for their binding to PHGDH, a variant of the protein with an empty and accessible cofactor-binding site was sought. Therefore two strategies were investigated, (i) removal of the bound cofactor from the purified protein through charcoal treatment, and (ii) prevention of cofactor binding to the protein through introduction of specific mutations. In the following sections an investigation of both approaches for their potential to yield a protein with free, accessible cofactor-binding site that can be used for structure-based drug design is described.

6.3 Materials and methods

All chemicals listed in the materials and methods section were of analytical grade and obtained from Sigma Aldrich, St. Louis, MO, USA, unless otherwise stated.

6.3.1 Charcoal treatment

sPHGDH was mixed with 0.5 to 1.5 weight equivalents of activated charcoal and the mixture incubated under agitation for 30 to 120 minutes at room temperature (RT). Charcoal was removed through centrifugation at 16000 g, 10 min, 4 °C. Any remaining charcoal was removed by running the supernatant through a PD-10 column (GE Healthcare, Little Chalfont, UK). The purity of the proteins was determined by measuring the OD_{260}/OD_{280} ratio on a NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA).

6.3.2 Site-directed mutagenesis

Point mutations were introduced by polymerase chain reaction (PCR) using the oligonucleotides outlined in Table 6.1. PCRs were performed with High-Fidelity Phusion DNA polymerase (New England Biolabs, Ipswich, MA, USA) using 20 ng purified sPHGDH DNA (see section 2.5.1), 1 µg of each primer, 0.5 mM dNTPs and 1 x GC buffer (New England Biolabs, Ipswich, MA, USA) and run according to the thermal cycling profile in Figure 6.4. PCR product separation by size was performed by agarose gel electrophoresis and visualised by GelRed nucleic acid gel stain (Biotium, Hayward, CA, USA). Where the amplicon was detected, the remaining template was digested with DpnI (New England Biolabs, Ipswich, MA, US) for 1 h at 37 °C, followed by transformation of the reaction mixture into *E. coli* DH5α cells.
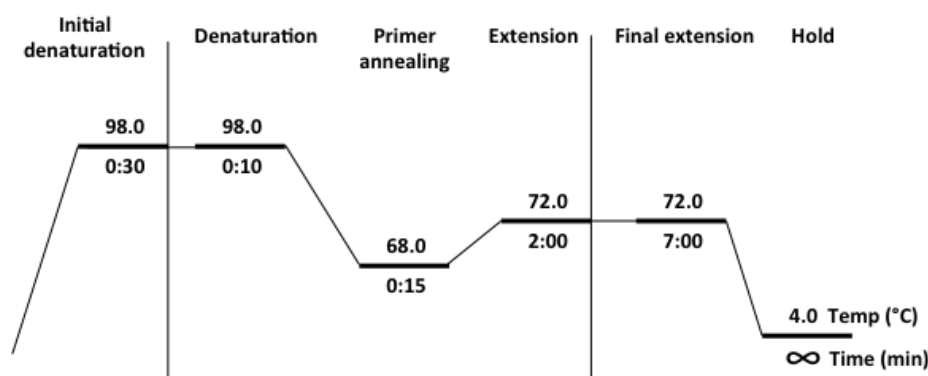
Figure 6.4 - Thermal profile of PCR reaction.

The PCR was performed using a GeneAmp PCR System 2700 (Applied Biosystems, Waltham, MA,USA). The cycle of denaturation, primer annealing and extension was repeated 18 times.

Mutation	Orientation	Primer sequence (5'-3')
R154E	Forward	cttggcctgggc<u>gag</u>attgggagagagg
	Reverse	cctctctcccaat<u>ctc</u>gcccaggccaag
T206W	Forward	catcactgtgcact<u>gg</u>cctctcctgccctcc
	Reverse	ggagggcaggagagg<u>cc</u>agtgcacagtgatg
T212W	Forward	cctctcctgccctcc<u>tgg</u>acaggcttgctgaatg
	Reverse	cattcagcaagcctgt<u>cc</u>aggagggcaggagagg
P175A/T206A/T212A	Forward	(1) tcatcactgtgcac<u>gcg</u>cctctcctgccctcc<u>gcg</u>acaggcttgc
		(2) ctatagggtatgac<u>gcg</u>atcatttccccagagg
	Reverse	(1) gcaagcctgt<u>cgc</u>ggagggcaggagagg<u>cgc</u>gtgcacagtgatga
		(2) cctctggggaaatgat<u>cgc</u>gtcataccctatag
D259N	Forward	gccggggctgcactg<u>aat</u>gtgtttacggaagagcc
	Reverse	ggctcttccgtaaacac<u>att</u>cagtgcagccccggc
R235A/H282A	Forward	(1) gtggtgaactgtgccg<u>gcg</u>ggagggatcgtggacg
		(2) gtcatcagctgtccc<u>gcg</u>ctgggtgccagcaccaagg
	Reverse	(1) cgtccacgatccctcc<u>cgc</u>ggcacagttcaccac
		(2) ccttggtgctggcacccag<u>cgc</u>gggacagctgatgac
A285E	Forward	gtccccacctgggt<u>gag</u>agcaccaaggaggctc
	Reverse	gagcctccttggtgct<u>ctc</u>acccaggtggggac

Table 6.1 - Primer pairs for the generation of site-directed mutants of sPHGDH.

Forward and Reverse primers are shown with the mutated positions underlined. All primers were manufactured by Eurofins Genomics (Ebersberg, Germany).

Figure 6.5 – Amino acids mutated in sPHGDH.

Amino acid sequence of sPHGDH with the amino acids mutated in the different site-directed mutants highlighted in red.

6.3.3 Transformation

Transformation into competent *E. coli* DH5α subcloning efficiency cells (Life Technologies, Thermo Fisher Scientific, Waltham, MA, USA) was performed according to the manufacturer's instructions. Briefly, 1 μL of DpnI-digested plasmid was added to 20 μL of competent cells and incubated on ice for 30 minutes. Cells underwent heat shock for 20 seconds at 42 °C, followed by cooling on ice for 2 minutes. After addition of 180 μL pre-warmed medium, the cells were incubated at 37 °C for 1 hour while shaking at 250 rpm, followed by plating on an LB/Agar plate containing 50 μg/mL kanamycin. Plates were incubated for 18 h at 37 °C and single colonies used to inoculate aliquots of 5 mL LB medium containing 50 μg/mL kanamycin. The cultures were grown under agitation at 200 rpm at 37 °C over night. DNA was isolated from the cells using the QIAprep Spin Miniprep Kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions. The mutations were confirmed by DNA sequencing (Source Bioscience, Nottingham, UK). The mutated DNA was used to transform the *E. coli* expression strain Rosetta (DE3) (Novagen, Darmstadt, Germany) grown in TB medium containing 50 μg/mL kanamycin. The transformation protocol followed was similar to the one described above with the exception of a 30 second heat shock at 42 °C. Glycerol stocks of the transformed cells in 15 % (v/v) glycerol were prepared from overnight cultures and stored at -80 °C.

6.3.4 Protein expression purification

Large-scale protein expression was performed in *E. coli* Rosetta (DE3) cells in 2 L culture volume (see section 2.5.1). Site-directed mutants were purified as described in section 2.6.1 with a final yield of 10 – 20 mg depending on the mutant.

6.3.5 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry was performed as described in section 2.10. 50 μM protein was exchanged into ITC buffer (25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP) using a PD-10 desalting column (GE Healthcare, Little Chalfont, UK). 50 μM wt sPHGDH or mutated protein or buffer alone was used as receptor in the cell and 0.2 – 0.5 mM NADH as ligand in the syringe. After an initial test injection of 0.5 μL ligand, a further 17 injections of 2 μL every 120 seconds were performed. For data analysis, the heat of dilution effect of titrating ligand into buffer was subtracted

from the ligand to protein titration data. Data were analysed using the ORIGIN software.

6.3.6 *Differential scanning fluorimetry (DSF)*

DSF experiments were performed with 20 µM protein in 25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP and 10 x SYPRO Orange. The assay was run in a total volume of 15 µL in a 384-well format according to the procedure in section 2.9.

6.3.7 Enzyme activity assay

The enzymatic activity of sPHGDH and its mutants was measured in the direction of 3-phosphoglycerate oxidation according to the procedure in section 2.8. Each sample was run in duplicate in three independent experiments, and the percentage enzymatic activity of the mutants compared to wt sPHGDH was determined.

6.3.8 Circular dichroism (CD)

Circular dichroism (CD) is a spectroscopic technique used to study the secondary structure and conformation of chiral molecules, such as proteins. CD is the difference in absorption of left-handed and right-handed circularly polarised light, which occurs when chiral molecules interact with light. The characteristic CD spectra of proteins are defined through the amines of the polypeptide backbone resulting in distinct CD spectra for defined structural elements such as α-helices, which *e.g.* show negative bands at 222 nm and 208 nm and a positive band at 193 nm [169] .

CD was performed on a JASCO J-810 spectropolarimeter (Japan Spectroscopic, Gross-Umstadt, Germany), recording 185-250 nm with a bandwidth of 0.2 nm and a scan speed of 20 nm/min. All measurements were performed with 0.5 mg/mL protein in 20 mM NaH_2PO_4, pH 8. Each measurement comprised the average of 10 repeated scans in a 0.02 cm path-length quartz cuvette and was subtracted from a blank measurement of buffer alone. The data was processed in CDSSTR software and secondary structure analysis performed on the DICHROWEB server with the CONTINLL algorithm and reference set 7 [170-172]. Data were plotted using GraphPad Prism.

6.3.9 Crystallisation

Crystallisation experiments were performed as described in section 4.3.5. Briefly, site-directed mutants were dialysed into low salt buffer composed of 25 mM HEPES,

pH 7.5, 100 mM NaCl and 0.5 mM TCEP. For co-crystallisation experiments, proteins were mixed with ligands to yield a final concentration of 5 mM ligand and 15 mg/mL protein and incubated at 4 °C overnight. Sitting drop vapour diffusion crystallisation trials using sparse matrix screens were performed in 96-well format at 4 and 20 °C. Trials contained drops with 0.3 + 0.3 µL and 0.3 + 0.6 µL protein and precipitant.

6.4 Results

Part of the work presented in this chapter was performed by Sophie Bex and Nabila Aljufri, a summer project student and an M.Sc. student, working under my supervision.

6.4.1 Charcoal treatment

The literature contains various examples in which the adsorption properties of purine and pyrimidine derivatives on charcoal were used for the isolation of nucleotides from biological samples, the purification of nucleotide derivatives and the removal of diphosphopyridine dinucleotide from glyceraldehyde 3-phosphate [173-175]. On the basis of these data, purified sPHGDH was incubated with different amounts of activated charcoal for a maximum of 2 hours at RT. After removal of the charcoal, the absorbance of the samples at 280 nm and the OD_{260}/OD_{280} ratio were used to monitor the protein concentration and purity, respectively. The measurements showed that not only cofactors adsorbed on the charcoal, but also protein, resulting in a reduced protein concentration. This was especially evident in samples treated with 1 or 1.5 weight equivalents of activated charcoal, which resulted in approximately 45 % and 70 % of protein loss, respectively, compared to approximately 25 % protein loss when treated with 0.5 weight equivalent of charcoal (Table 6.2). In addition, the highest purity achieved was an OD_{260}/OD_{280} ratio of 0.65-0.7, which is still higher than the ratio of 0.57 reported for pure protein, indicating that some cofactor is still bound. Incubation for longer than 30 minutes did not increase the purity of the protein. For further experiments with charcoal-treated sPHGDH, the protein was treated with 0.5 weight equivalent of activated charcoal for 30 minutes at room temperature as these conditions yielded protein with the highest degree of purity achieved without losing too much protein by its adsorption on charcoal.

	Untreated		0.25 eq. charcoal		0.5 eq. charcoal	
Time	[protein] (mg/mL)	OD_{260}/OD_{280}	[protein] (mg/mL)	OD_{260}/OD_{280}	[protein] (mg/mL)	OD_{260}/OD_{280}
0 min	15.0	0.90	15.0	0.90	15.0	0.90
30 min	15.0	0.89	12.1	0.65	11.2	0.69
60 min	15.2	0.89	11.4	0.69	10.6	0.69
120 min	15.4	0.89	11.3	0.7	10.2	0.7

	1 eq. charcoal		1.5 eq. charcoal	
Time	[protein] (mg/mL)	OD_{260}/OD_{280}	[protein] (mg/mL)	OD_{260}/OD_{280}
0 min	15.0	0.90	15.0	0.90
30 min	8.6	0.69	5.0	0.67
60 min	7.7	1.04	3.9	0.66
120 min	4.4	1.21	2.7	0.73

Table 6.2 - Removal of cofactor through charcoal treatment.

15 mg/mL sPHGDH was mixed with different amounts of charcoal (0.25, 0.5, 1.0 and 1.5 weight equivalent to sPHGDH) for a maximum of 120 minutes at rt. Samples were taken after 30, 60 and 120 minutes, charcoal removed by centrifugation (10 min, 4 °C, 16000 *g*) and protein concentration and purity measured using a NanoDrop 2000.

NADH binding to untreated and charcoal-treated sPHGDH was explored by ITC. When titrating 0.5 mM NADH in the syringe into 0.05 mM wt sPHGDH in the cell, the binding site was saturated too quickly to obtain a complete, sigmoidal binding curve. Therefore the concentration of NADH in the syringe was reduced to 0.2 mM in the titration experiment (Figure 6.6 A). However, after charcoal-treatment a full binding curve was obtained using 0.5 mM NADH giving a first indication that more cofactor binding sites were now empty and accessible (Figure 6.6 B). This was also reflected by the calculated stoichiometry (n) rising from 0.3 to 0.8 for non-treated and charcoal-treated sPHGDH respectively. Analysis of the different protein batches by DSF showed that charcoal-treated sPHGDH had a significant lower T_m of 46.0 ± 0.1 °C compared to non-treated protein with a T_m of 47.8 ± 0.1 °C, whereas addition of 1 mM NADH resulted in an increase in T_m compared to that observed for non-treated protein, indicating that the higher T_m of non-treated sPHGDH was probably due to some pre-bound ligand (Figure 6.6 C). Despite the success in removing the cofactor from the protein to a certain extent through charcoal treatment, it was not clear what proportion of binding sites would still be occupied after charcoal-treatment. In

addition, a variation in efficiency of cofactor removal from experiment to experiment was found. Due to these disadvantages a more robust technique was sought.
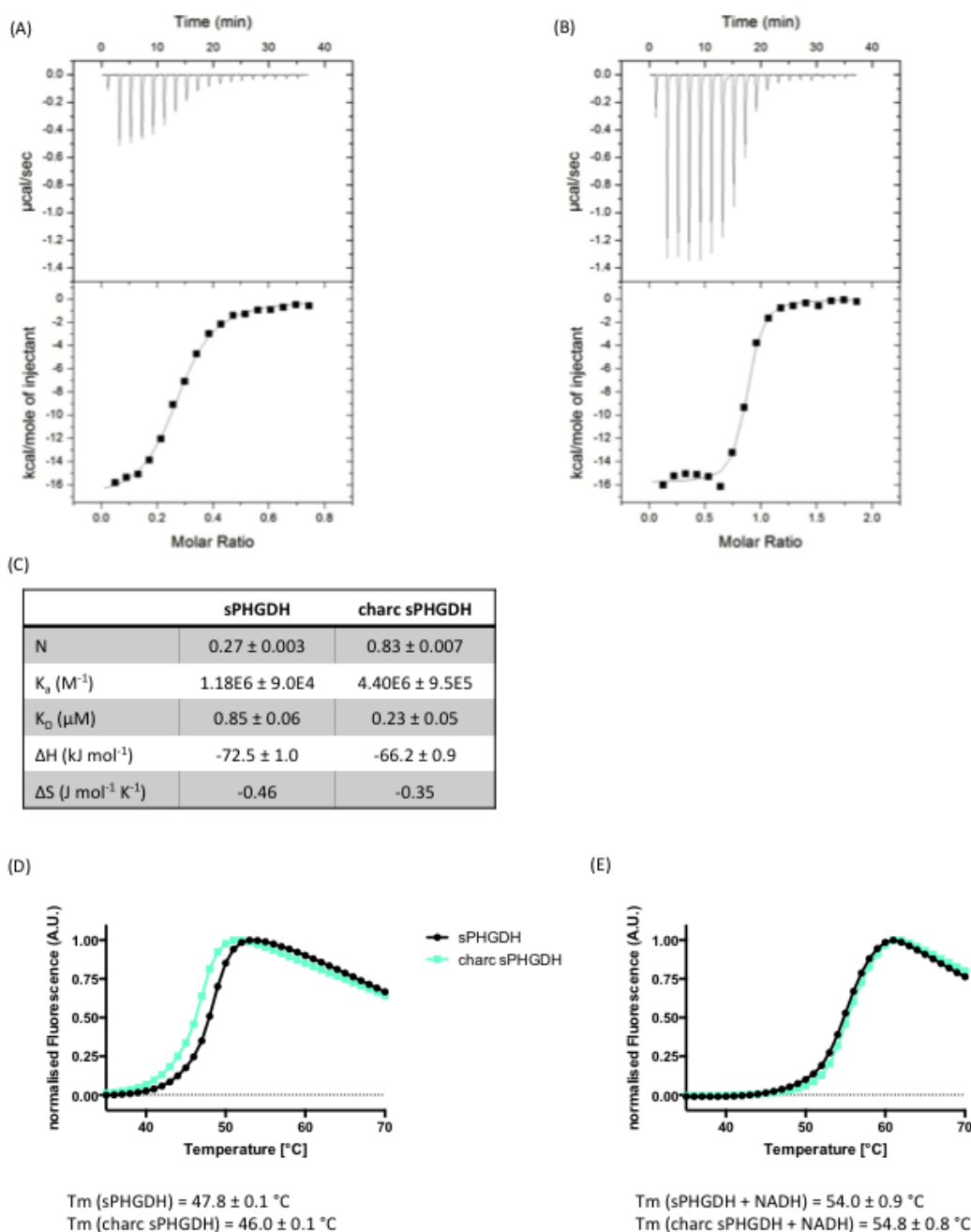
Figure 6.6 - Comparison of NADH binding to charcoal-treated and non-treated wt sPHGDH.

Calorimetric isothermal titration of (A) 0.05 mM non-treated wt sPHGDH with 0.2 mM NADH and (B) 0.05 mM charcoal-treated wt sPHGDH with 0.5 mM NADH. The upper panel shows the raw data for 1 x 0.5 μL, followed by 17 x 2 μL injections of NADH into the 0.05 mM protein. Data was corrected for the heat of NADH dilution by subtracting NADH to buffer titration heats. Bottom panel: integrated heats of from the raw data peaks in the top panel plotted against the molar ratio of NADH to protein. Shown is the line of best-fit data as obtained by a one-site fit model using the ORIGIN software. (C) Values of generated by one-site fit model. (D) and (E) DSF profiles of sPHGDH and charcoal-treated sPHGDH alone (D) or in the presence of 1 mM NADH (E).

6.4.2 Design, expression and purification of site-directed mutants

The cofactor-binding pocket is rather large, with NAD$^+$ or NADH maintaining many interactions with the surrounding amino acids (Figure 6.7). Therefore two mutation strategies were pursued to abrogate cofactor binding; (i) mutation of amino acid residues directly involved in binding the cofactor through hydrogen bonds or hydrophobic interactions into residues that would not be able to maintain those interactions and (ii) mutation of small amino acids around the cofactor-binding pocket into the bulky amino acid tryptophan so as to potentially occupy part of the pocket (Table 6.3, Figure 6.8). The mutants were designed based on prior investigation of the likely conformation in COOT [107].
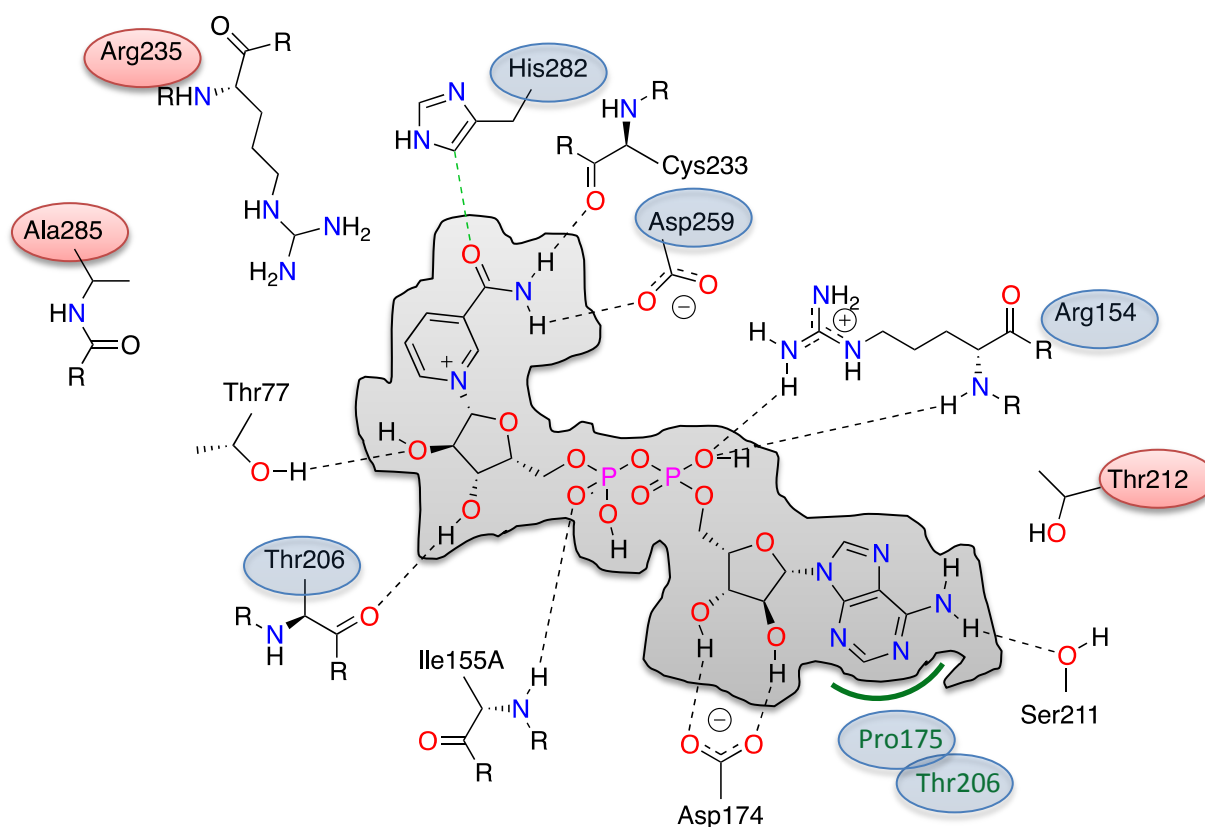
Figure 6.7 – Interactions of NAD$^+$ with sPHGDH

All interactions between NAD$^+$ (grey shape) and sPHGDH are shown with the amino acids targeted for mutation highlighted with blue circles. In addition amino acids that are not involved in direct binding of NAD$^+$, but were mutated in some of the site-directed mutants are depicted as well (red circles). Colour legend: black dashed lines – hydrogen bonds, salt bridges; green solid lines – hydrophobic interactions; green dashed lines – Pi-Pi, Pi-cation interactions.

Site-directed mutant	Category	Mutated residues	Mutated to	Concept
R154E	A	Arg154	Glu	Mutation of key binding residue involved in H-bonding
T206W	A	Thr206	Trp	Filling up adenine-binding subsite
T212W	A	Thr212	Trp	Filling up adenine-binding subsite
P175A/T206A/T212A	A	Pro175, Thr206, Thr212	3 x Ala	Mutation of key binding residue involved in H-bonding and hydrophobic interactions (Pro175)
D259N	N	Asp259	Asn	Mutating key binding residue involved in H-bonding
R235A/H282A	N	Arg235, His282	2 x Ala	Mutation of key binding residue involved in H-bonding
A285E	N	Ala285	Glu	Mutation of key binding residue

Table 6.3 - Summary of designed mutants.

The mutants are divided into two categories depending on whether the adenine-binding subsite (A) or nicotinamide-binding subsite (N) was mutated.
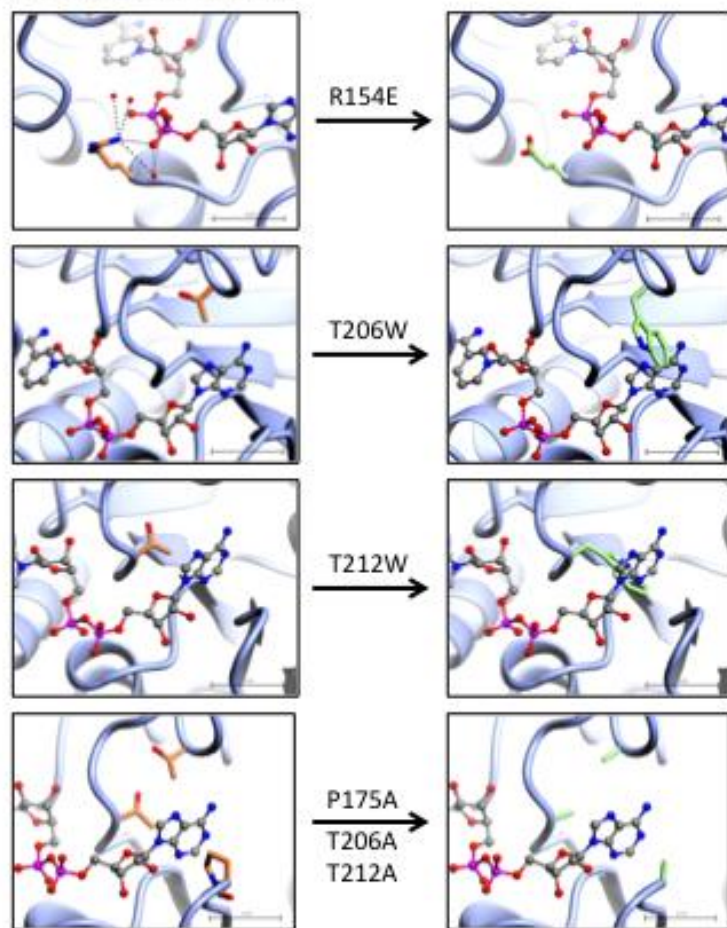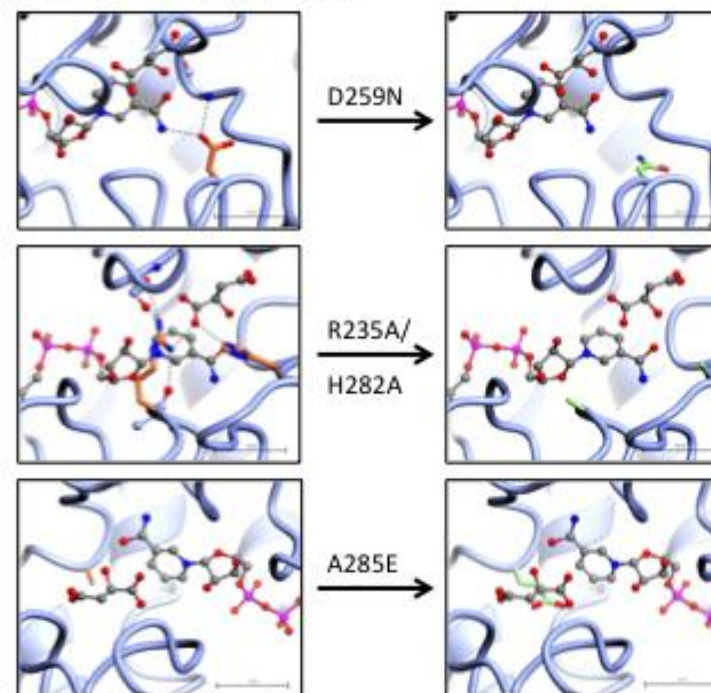
Figure 6.8 - Illustration of adenine-subsite (A) and nicotinamide-subsite (B) mutants.

Wild-type sPHGDH structures are shown on the left hand sides with key amino acids highlighted in orange. The right hand side pictures present models of the site-directed mutants assuming the side chains of mutated amino acids occupy the most likely conformation according to *COOT* [107] (mutated amino acids highlighted in green).

The mutants were generated by site-directed mutagenesis using specifically designed primers for each mutation. After transformation into competent *E. coli* DH5α cells, DNA was isolated and verified by sequencing. For large-scale expression, plasmid DNA was used to transform *E. coli* strain Rosetta (DE3). The general purification scheme for the mutants comprised an initial Ni-affinity chromatography step followed by cleavage of the His$_6$-tag with TEV protease, size exclusion chromatography and finally anion exchange chromatography to increase the purity of the protein. As an example, purification of mutant A285E is illustrated in Figure 6.9. As the mutations introduce only minor changes, the purification of the various mutants was very similar to wt sPHGDH with yields of about 10 mg/L bacterial culture. Initial determination of the purity of the mutants by measuring the OD$_{260}$/OD$_{280}$ ratio indicated that all proteins showed a remarkably lower ratio of 0.6-0.7, which is in the range of the purity seen for sPHGDH after charcoal treatment.
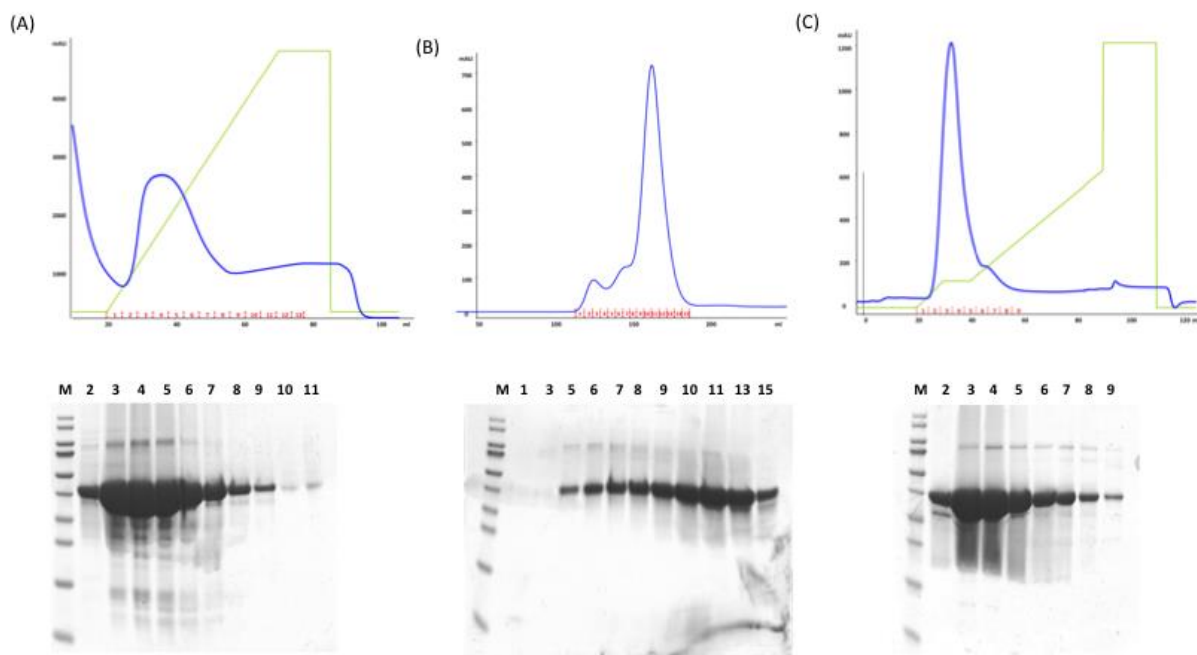
Figure 6.9 - Typical purification result for mutants of sPHGDH illustrated using mutant A285E.

sPHGDH mutants were purified from *E. coli* Rosetta (DE3) cells. A representative purification is shown with the UV traces of the different purification steps (upper panel) together with the analysis of the obtained fractions (red numbers) by SDS-PAGE and subsequent staining with a Coomassie-based stain. (A) First purification step by immobilised metal affinity chromatography (IMAC) using a HisTrap Ni-Sepharose column. Protein was eluted with a linear gradient up to a maximum concentration of 250 mM imidazole. There was a single peak in the UV trace comprising mainly the desired protein as determined by SDS-PAGE analysis. (B) Size exclusion chromatography of affinity-purified protein after TEV cleavage, using a Superdex^{TM} 75 26/60 gel filtration column equilibrated in 25 mM HEPES, pH 7.5, 100 mM NaCl and 0.5 mM TCEP. A major peak of the desired protein was obtained, which often contained some impurities of high molecular weight as found by SDS-PAGE. (C) Additional anion exchange purification step with the intention to improve protein purity; bound protein was eluted with a non-linear salt gradient.

6.4.3 Site-directed mutants fold properly but have altered secondary structure

To assess the proper folding of the engineered mutants they were investigated by differential scanning fluorimetry (DSF). All site-directed mutants exhibited the expected DSF profile for properly folded proteins with a gradually increasing content of unfolded protein as the temperature increased (Figure 6.10). However, there was some variation in the intrinsic fluorescence of the mutants, which ranged from 4 x 10^5 to 1.3 x 10^6 FU, as well as in the maximum fluorescence observed reflecting the differences in amino acid composition and possible structural differences.

Adenine-subsite mutants
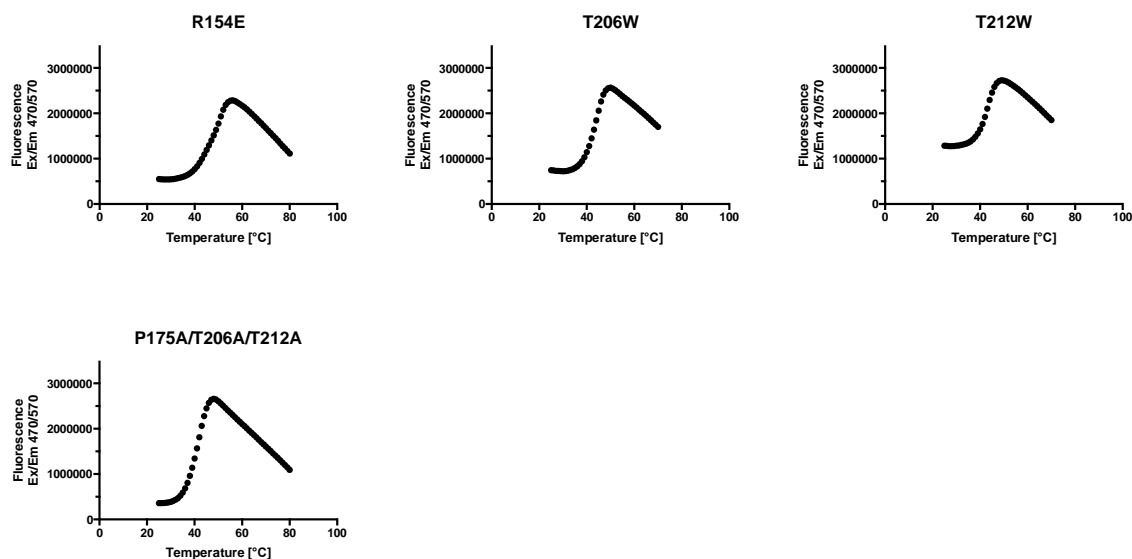
Nicotinamide-subsite mutants
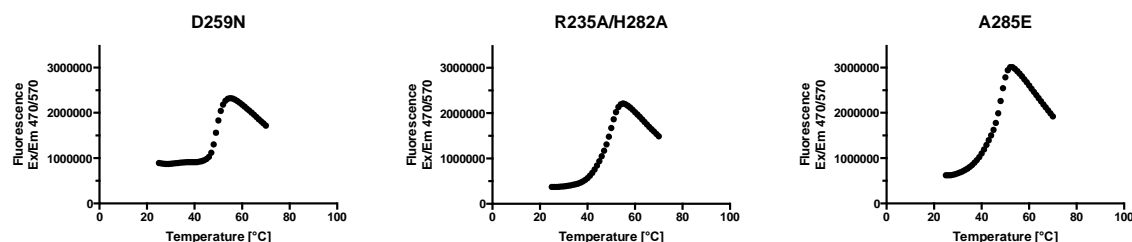
Figure 6.10 - DSF profiles of site-directed mutants.

Mutants are mixed with the fluorescent dye Sypro Orange and excited at 470 nm. Fluorescence emission was measured at 570 nm.

Compared to wt sPHGDH, the introduced mutations affected the protein stability to various degrees as shown by DSF when comparing the normalised datasets (Figure 6.11). Three of the adenine-subsite mutants, namely T206W, T212W and the triple mutant P175A/T206A/T212A, had a significantly lower T_m than the wt sPHGDH, with P175A/T206A/T212A displaying the largest reduction in stability of 7.3 °C. This decrease in T_m could actually represent the absence of bound cofactor as well as missing intramolecular interactions within the protein. Point mutations around the nicotinamide-binding subsite had less influence on protein stability. Remarkably for mutant D259N an increase in T_m of 1.3 °C was detected.

(A) Adenine-site mutants

Construct	Tm (°C)
R154E	47.3 ± 0.3
T206W	43.1 ± 0.3
T212W	44.1 ± 1.3
P175A/T206A/T212A	40.5 ± 0.2

(B) Nicotinamide-site mutants

Construct	Tm (°C)
D259N	49.1 ± 0.2
R235A/H282A	48.3 ± 0.5
A285E	46.8 ± 1.1
sPHGDH	47.8 ± 0.1
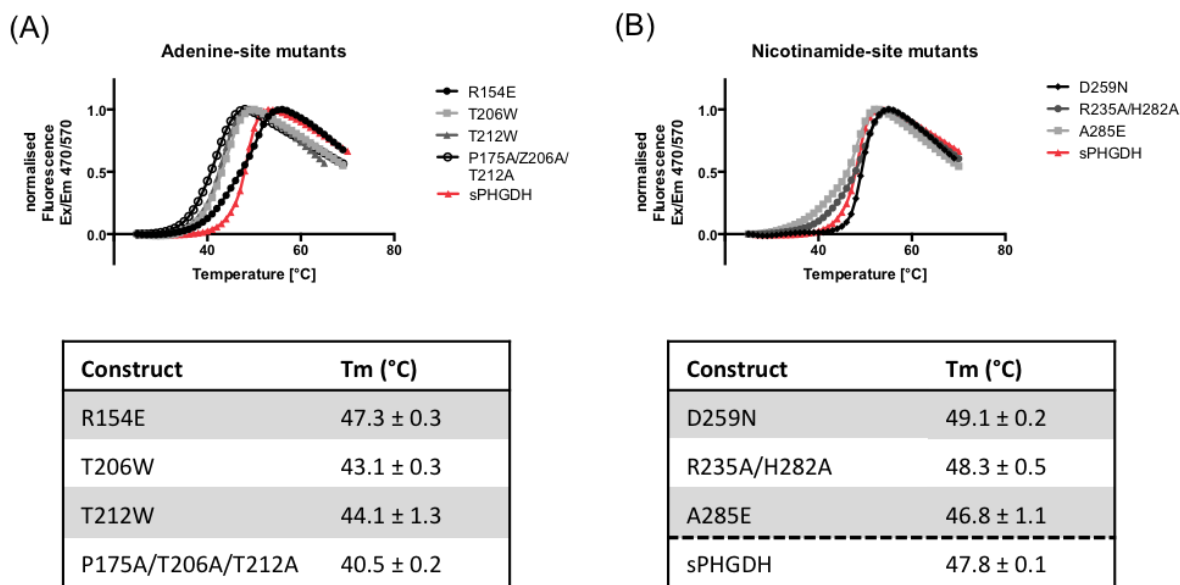
Figure 6.11 - Comparison of DSF curves for wt and site-directed mutants of sPHGDH.

Graphs showing the normalised fluorescence measurements of adenine-subsite mutants (A) and nicotinamide-subsite mutants (B) compared to wt sPHGDH under conditions of increasing temperature. The melting temperatures are calculated from three independent experiments with four replicates per experiment.

In addition to measurement of thermal stability, an initial investigation of the mutant structures was performed using circular dichroism (CD). CD is a spectroscopic technique that measures the differences in absorption of left-handed and right-handed circularly polarised light after the light has passed through a chiral molecule [176]. In the case of proteins the chromophoric groups are the backbone amide groups, which are oriented towards each other in a manner that depends on the position of the polypeptide backbone. As a consequence, the different structural elements seen in proteins, such as α-helixes or β-sheets, display distinct CD spectra (Figure 6.12). Although CD measurements cannot give structural information for specific residues within the protein, as may be obtained from X-ray crystallographic or NMR spectroscopic measurements, it allows for investigation of the secondary structure in aqueous buffer using a small amount of protein. As the structure of wild type sPHGDH was known, CD measurements were used as a reference to assess the effects of the point mutations on the protein's overall secondary structures.
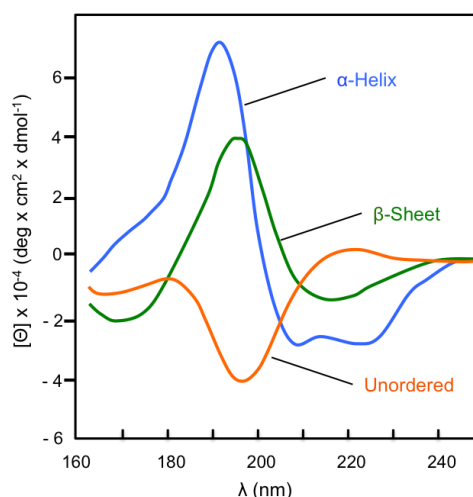
Figure 6.12 - Characteristic far UV CD spectra for different secondary structure elements.

Typical spectra associated with α-helix (blue line), β-sheet (green line) and unordered protein (orange line) in the far UV region (165-250 nm) are depicted (adapted from [176]).

Overall, the CD measurements show that all introduced mutations altered the secondary structure of the protein (Figure 6.13A and B). In order to quantify these changes, the data were further analysed using the DICHROWEB web server with the CONTINLL algorithm and reference set 7 [170-172]. The analysis showed that all mutants had an increase in unordered protein fraction of at least 26 % mainly at the expense of helicity (Figure 6.13). Most of the mutants also showed prominent reduction in the proportion of β-sheets and turns with only the nicotinamide-subsite mutant R235A/H282A and the adenine-subsite mutant P175A/T206A/T212A having a higher content of β-sheets. This decrease in secondary structure could result from the non-binding of the cofactor leading to a less ordered protein structure.

(A) Adenine-site mutants
- R154E
- T206W
- T212W
- P175A/T206A/T212A
- sPHGDH

(B) Nicotinamide-site mutants
- D259N
- R235A/H282A
- A285E
- sPHGDH

(C)

	Adenine-site mutants				Nicotinamide-site mutants		
	R154E	T206W	T212W	P175A/T206A/T212A	D259N	R235A/H282A	A285E
α-Helix	-23%	-34%	-23%	-44%	-31%	-39%	-23%
β-Sheet	-25%	9%	-37%	31%	-12%	22%	-21%
Turns	-13%	-4%	-15%	14%	-12%	-5%	-13%
Unordered	43%	35%	50%	26%	46%	35%	42%
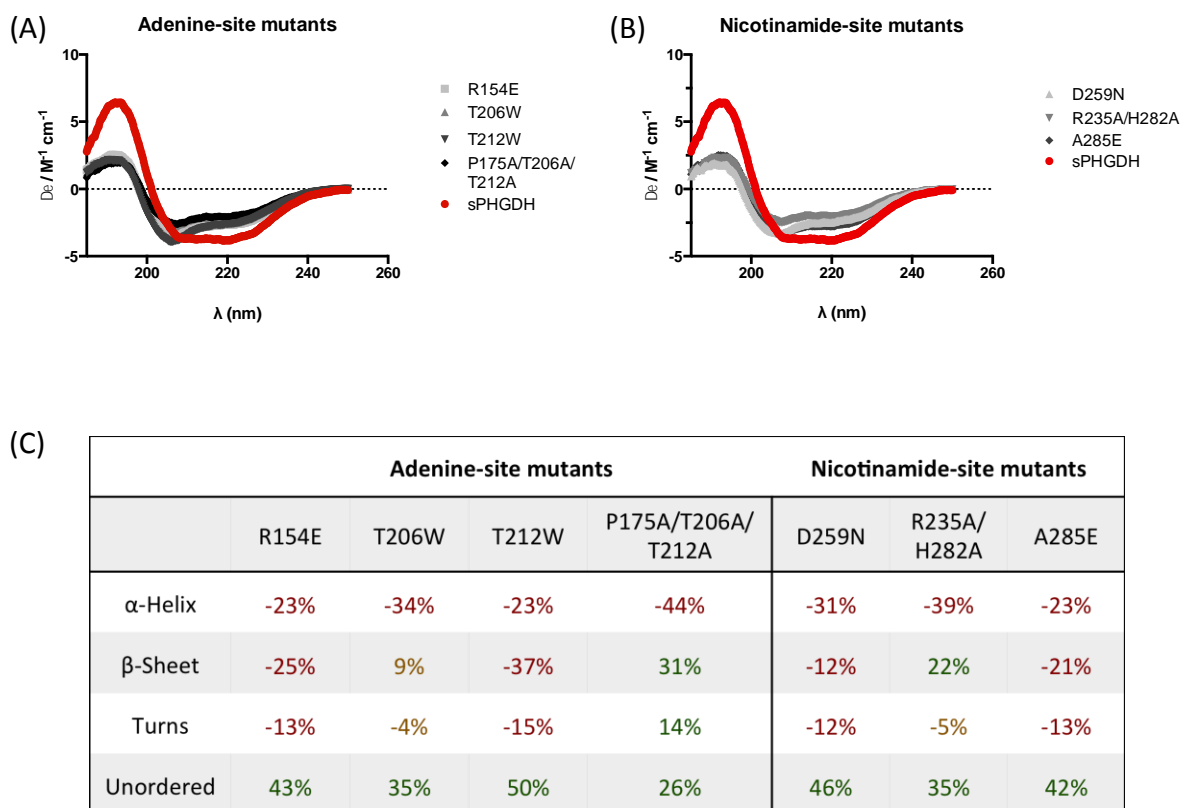
Figure 6.13 - Structural changes adopted from introduction of point mutations in sPHGDH as measured by circular dichroism.

Graph of wavelength blotted against molar elipticity of 0.5 mg/mL adenine-subsite (A) or nicotinamide-subsite mutant (B) in 20 mM NaH_2PO_4, pH 8 at 20 °C. (C) Per cent changes in secondary structure elements through introduction of mutations compared to wt sPHGDH. CD data analysis was performed using the CONTINLL algorithm (reference set 7) on dichroweb [170-172].

6.4.4 Evaluation of NADH binding by ITC, DSF and enzymatic activity

In order to investigate whether the site-directed mutants had abrogated cofactor binding and if so to what extent, binding experiments were performed with NADH as ligand. As NADH was found to bind much more tightly to PHGDH or its truncated form, sPHGDH, than NAD^+, NADH binding was easier to monitor in all of the biophysical experiments performed. Firstly, NADH binding was investigated by DSF by measuring the melting temperature (T_m) of the protein alone and in the presence of 1 mM NADH. According to the DSF data, all mutations reduced NADH binding, but to different extents (Figure 6.14). The adenine-subsite mutations P175A/T206A/T212A, T206W and T212W abrogated NADH binding completely whilst the mutant R154E retained some affinity for NADH, displaying a ΔT_m of 0.8 ± 0.4 °C. In contrast, the nicotinamide-subsite mutants all retained some ability to bind NADH with the strongest binding seen with mutant A285E with ΔT_m of 3.3 ± 0.6 °C.

However, this was still a two-fold lower ΔT_m value than was found with wt sPHGDH. Considering that the T_m value calculated for NADH binding to wt sPHGDH is an underestimate due to co-purification of the protein with cofactor, the abrogation of NADH binding seen with mutant A285E is significant.

(A) Adenine-subsite mutants

$\Delta Tm = 0.8 \pm 0.4\ °C$ $\Delta Tm = 0.1 \pm 0.2\ °C$ $\Delta Tm = 0.0 \pm 0.0\ °C$

$\Delta Tm = 0.3 \pm 0.0\ °C$

(B) Nicotinamide-subsite mutants

$\Delta Tm = 1.0 \pm 0.2\ °C$ $\Delta Tm = 1.6 \pm 0.2\ °C$ $\Delta Tm = 3.0 \pm 0.3\ °C$

(C) wt sPHGDH

$\Delta Tm = 6.1 \pm 0.8\ °C$
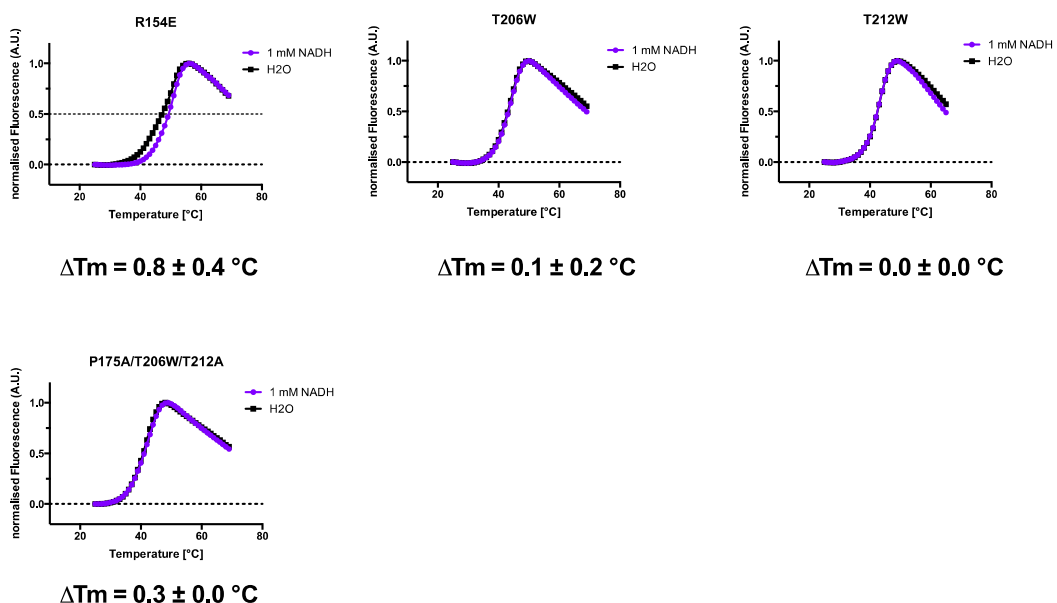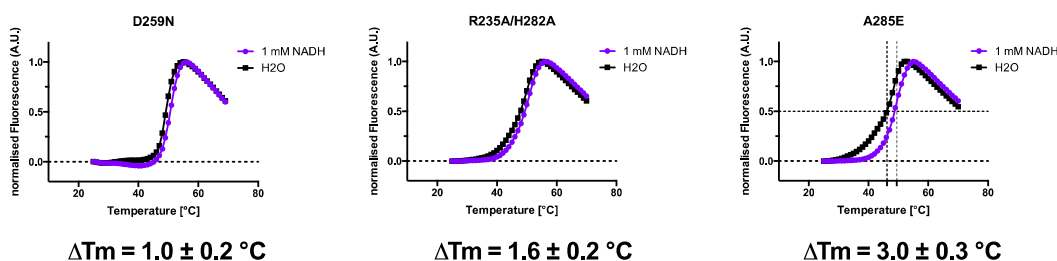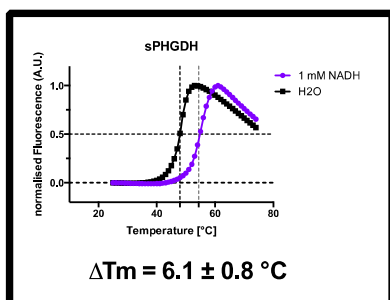
Figure 6.14 - NADH binding to wt and site-directed mutants of sPHGDH investigated by DSF.

Thermal denaturation of (A) adenine- and (B) nicotinamide-subsite mutants as well as (C) wt sPHGDH was monitored in the presence and absence of 1 mM NADH.

In order to collect more information on the nature of NADH binding, ITC experiments were performed by titrating 0.5 mM NADH into 0.05 mM site-directed mutant. ITC data confirmed the results of the DSF experiments with the adenine-subsite mutants T206W, T212W and P175A/T206A/T212A completely lacking affinity for NADH and reduced binding seen with the remaining mutants (Figure 6.15). Of the mutants still able to bind NADH, only A285E showed the expected binding stoichiometry (n) of about 1, whereas for the other mutants n was between 0.2 and 0.4. Against this background, the determined association constant (K_a) is only accurate for A285E, while for the other mutants an overestimate was calculated with the actual K_a being much weaker. However, even the mutant with the strongest remaining NADH affinity, A285E, had a 5-fold lower binding constant (K_D) compared to charcoal-treated sPHGDH (Table 6.4).
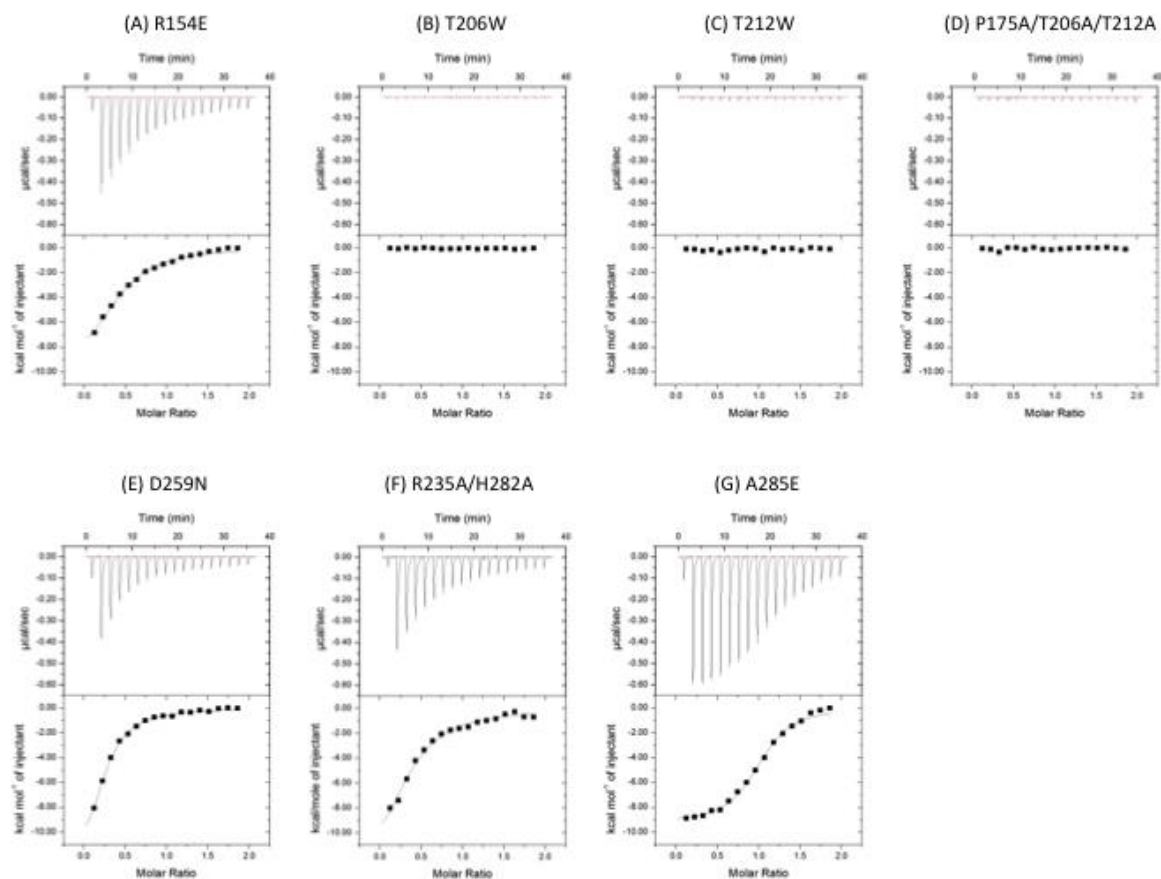
Figure 6.15 - Measurement of NADH binding to PHGDH mutants by isothermal titration calorimetry.

0.5 mM NADH was titrated into 0.05 mM protein in 25 mM HEPES, pH 7.5, 100 mM NaCl and 0.5 mM TCEP. Top panel: raw data for 1 x 0.5 µL, followed by 17 x 2 µL injections of 0.5 mM NADH into the isothermal cell containing 0.05 mM protein. Data were corrected for heat of NADH dilution by subtracting the heats from NADH to buffer titration. Bottom panel: integrated heats from the peaks in the top panel plotted against the molar ratio of NADH to protein. The line of best fit to the data was plotted as obtained by non-linear regression using a one-site fit model using the ORIGIN software.

	R154E	D259N	R235A/H282A	A285E	charc. sPHGDH
N (# of sites)	0.41 ± 0.03	0.22 ± 0.02	0.34 ± 0.05	0.99 ± 0.01	0.83 ± 0.01
K_a (M^{-1})	1.00E6 ± 1.57E4	1.40E5 ± 1.28E4	7.90E4 ± 1.50E4	4.99E5 ± 5.73E4	4.40E6 ± 9.50E5
K_D (µM)	1.00 ± 0.02	7.12 ± 0.65	12.66 ± 2.40	2.00 ± 0.23	0.23 ± 0.05
ΔH (kJ/mol)	-45.6 ± 4.5	-67.4 ± 5.4	-68.3 ± 11.3	-38.7 ± 0.6	-66.2 ± 0.9
ΔS (J/(mol x K)	-0.21	-0.47	-0.49	-0.08	-0.35

Table 6.4 - Thermodynamic parameters obtained from ITC experiments for binding of NADH to sPHGDH wt and mutants.

Data was analysed using the one-site fit model in ORIGIN. Stoichiometry of the reaction (= number of binding sites, N), changes in entropy (ΔS) and enthalpy (ΔH) as well as the association constant (K_a) were determined. Binding constant (K_D) was calculated by taking the reciprocal of the K_a. As no binding of NADH to mutants T206W, T212W and P175A/T206A/T212A was detected, no parameters could be calculated.

Not all of the mutants lost cofactor binding completely, as determined by DSF and ITC. For those mutants that were still able to bind to cofactor it was not clear how the binding occurred. It is possible that whilst the mutations prevented binding to the mutated part of the pocket, the cofactor, in this case NADH, retained binding to the other end of the binding pocket. For example, the nicotinamide-subsite mutants might still allow the adenosine moiety of NADH to bind to its cognate pocket while preventing binding to the nicotinamide-binding region. Taking into account that the cofactor-binding site in sPHGDH is quite open to the solvent, such partial-binding seems feasible. However, if the binding of cofactor to the mutants occurred in a different orientation or was hindered, the catalytic activity of the proteins should be impaired. The mutants and wt sPHGDH were therefore tested in an enzyme activity assay measuring the oxidation of 3-phosphoglycerate and simultaneous reduction of NAD$^+$. All mutants, irrespective of their ability to bind NADH, were inactive in the assay. This indicates that even in the cases in which NADH binding could still be measured, namely all nicotinamide-subsite mutants and the adenine-subsite mutant R154E, the binding must occur in such a way that the catalytic reaction can no longer

be performed or that the structure around the active site of the protein is less ordered and unable to form productive interactions with cofactor and substrate. In alignment with the design strategy of the mutations to inhibit binding at the different sub pockets (adenine- or nicotinamide-subsite) whilst maintaining integrity of the remaining pocket, the NADH-binding studies carried out so far could indicate that the measured NADH-binding is due to part of NADH still being able to bind to the unmutated subsites, *e.g.* the adenine moiety of NADH likely binds to the unmutated adenine-subsite of nicotinamide-subsite mutant D259N. To prove this hypothesis the protein crystal structures of the mutants were investigated.
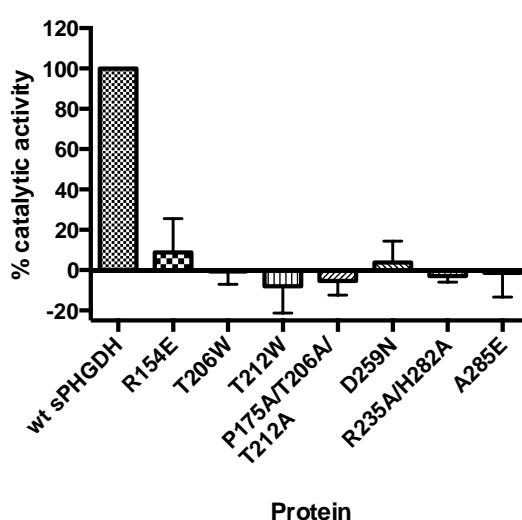
Protein

Figure 6.16 - Site-directed mutants are unable to catalyse the oxidation of 3-phosphoglycerate.

Site-directed mutants were run alongside wt sPHGDH in an enzymatic activity assay measuring protein activity based on NAD$^+$ reduction. Each protein was run in duplicate in three independent experiments.

6.4.5 Crystallisation

In addition to the overall 2D structural characterisation of the mutants, thermal stability measurements and the demonstrated abrogation of NADH binding, the determination of the structure, *e.g.* through X-ray crystallography, will allow a detailed analysis of the effects of the mutations on the cofactor-binding site as well as the overall protein structure.

All purified mutants were screened against several sparse matrix screens at 4 °C and 20 °C with only R154E crystallising in some of the conditions tested. R154E crystallised in different morphologies, but mainly as plates similar to crystals obtained

from wt sPHGDH (Figure 6.17 A-C). The crystals were tested using synchrotron radiation and were found to give very smeary diffraction in one direction of the two test shots separated by a rotation of 90 °. This smeary diffraction pattern in one direction is indicative of multiple layers of plates growing on top of each other. The same behaviour had been seen with crystals from wt sPHGDH, and could not be improved upon optimisation of the growth conditions. In view of successful co-crystallisation of wt sPHGDH in the presence of cofactor and substrate analogues (chapters 5 and 6), the co-crystallisation of R154E with nicotinamide mononucleotide (NMN) alone and in the presence of tartrate was attempted. NMN is a truncated version of NAD$^+$, lacking the adenine moiety and one phosphate group. As R154E had been designed as an adenine-site mutant, NMN ought to still be able to bind to R154E. Although more cubic crystals of R154E were obtained under these conditions, the diffraction was still smeary and a structure was not obtained (Figure 6.17 D-E).
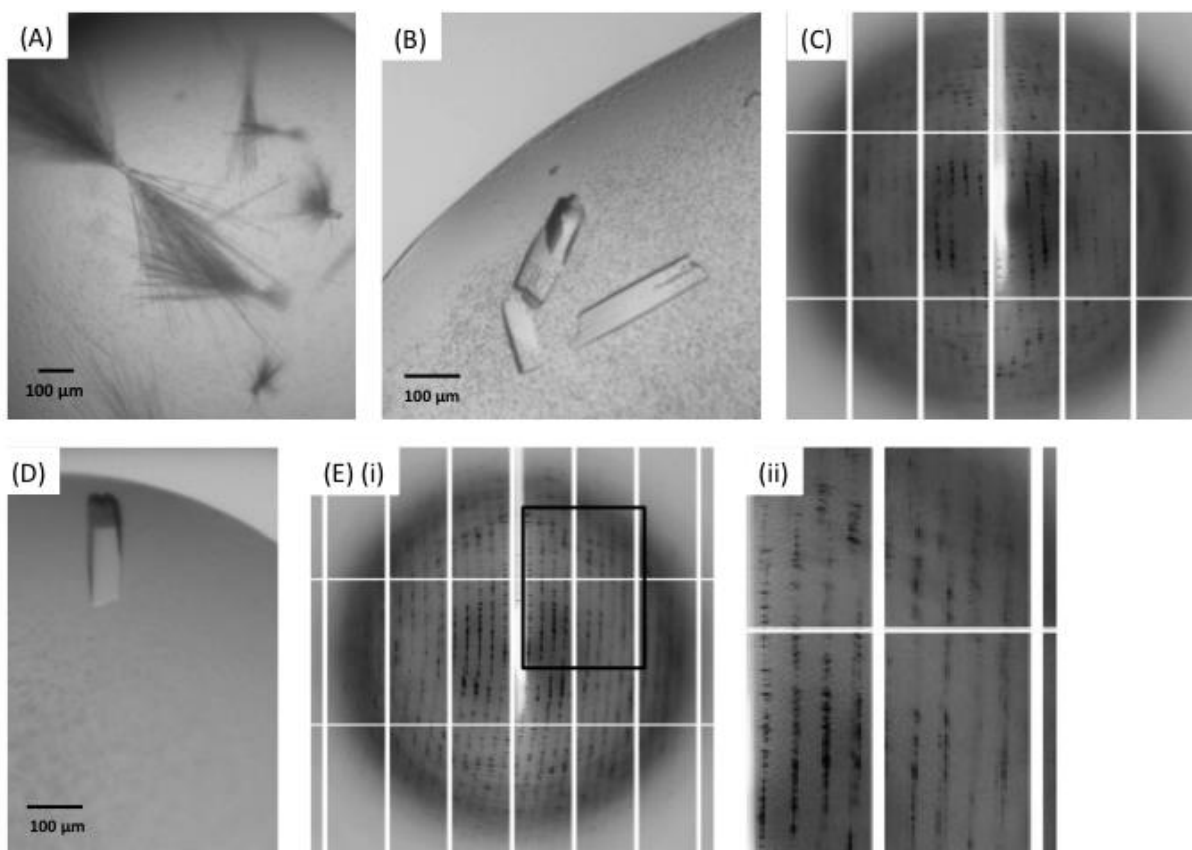
Figure 6.17 - Crystallisation attempt of R154E.

(A) and (B) Apo-crystals of R154E obtained in 0.1 M Bis-Tris propane pH 6.5, 0.2 M sodium nitrate, 20 % (w/v) PEG 3350 (PACT screen) and 0.1 M Bis-Tris pH 5.5, 0.2 M ammonium sulfate, 25 % (w/v) PEG 3350 (JCSG+ screen) respectively. (C) Diffraction pattern of an apo-crystal of R154E using synchrotron radiation (D) Co-crystal of R154E with 5 mM nicotinamide mononucleotide (NMN) in 0.8 M succinic acid pH 7.0 (JCSG+ screen) (E) (i) Diffraction pattern of R154E co-crystallised with NMN using synchrotron radiation. (ii) Closer view of the diffraction pattern in the black rectangle of (i).

6.5 Discussion

PHGDH as well as sPHGDH were found to co-purify with cofactor bound to the active site, which limits their utility in searching for potential PHGDH inhibitors targeting the cofactor/substrate-binding pockets. Therefore the focus of the experiments described in this chapter was to remove the cofactor from the protein and thus make the binding site accessible. Two strategies were pursued: (i) to remove the cofactor after protein purification through treatment with charcoal and (ii) to mutate the cofactor-binding pocket and thus prevent binding in the first place.

Treatment with activated charcoal was shown to remove cofactor from the protein, however this was accompanied by loss of protein, which was likely to be also absorbed onto the charcoal to a certain extent. According to the minimum OD_{260}/OD_{280} ratio of 0.65 measured after charcoal treatment, completely pure protein (OD_{260}/OD_{280} = 0.57) was never obtained. In addition, due to the nature of the experiment, batch-to-batch variability in the amount of cofactor that was removed from the protein added a further variability to subsequent experiments. ITC could have been used to determine the amount of free binding sites per charcoal-treated protein batch so as to take this into account for further experiments, but would have added a further time-consuming step to the purification process and would measured rather than solved the issue. In addition, although charcoal treatment of proteins is reported to successfully remove cofactors from proteins, the treatment was also shown to reduce the specific activity of the treated protein, raising the possibility that charcoal treatment might have led to partial inactivation of the protein [173].

Accordingly, a site-directed mutagenesis strategy was attempted, based on the idea that mutating the cofactor-binding site at one end would prevent binding of such large molecules as the cofactors, but at the same time retain a smaller pocket that would be targetable by low molecular weight inhibitors. Protein engineering approaches have been successfully used in the past in particular with kinases to modulate the ATP-binding site in order to bind a specific compound, such as an inhibitor or an analogue of ATP [168, 177]. In the case of protein kinase engineering, specific ligands and cofactors were designed in parallel that would only interact with the mutant protein. In contrast, the strategy presented herein was attempted in order to find mutants that would differ from wt protein in respect of cofactor binding, but at the same time allow identification of inhibitors that ultimately should target wt PHGDH.

The designed mutants were successful in reducing NADH binding, in some cases to zero. With no measurable enzymatic activity in any of the mutants, one could speculate that even in cases where cofactors were still able to bind, their binding mode must have changed so that the hydride-transfer reaction could no longer take place. As the hydride-transfer reaction likely requires very precise alignment of cofactor and substrate, even a small disturbance to the binding pocket may interfere with enzyme activity without completely preventing binding. The structural analysis of the site-directed mutants by DSF and CD suggested that the mutations had introduced changes in the 2D structure. In particular, a reduced α-helical content compared to wt sPHGDH was observed which could be indicative for changes of the cofactor binding region which is composed of the conserved Rossmann fold, an alternation of β-sheets and α-helices. Such a loss in ordered structure of the catalytic side on the protein could have abrogated enzymatic activity too. However, despite those structural alterations, binding studies by DSF and ITC showed that some of the mutants had retained the ability to bind NADH. This would suggest that some of the introduced mutations only induced minor changes around the binding site allowing it to still accommodate NADH albeit in a different orientation. Another possibility could be that only a fraction of the cofactor, such as the adenine or nicotinamide moiety, is able to bind to the respective unmutated binding subsite, placing the rest of the molecule outside the binding pocket. Given that the cofactor-binding site of the truncated sPHGDH is open to solvent, it would be possible to accommodate the cofactor in such a manner. However, the measured OD_{260}/OD_{280} ratios of all mutants showed that co-purification of the mutants with cofactor was less of an issue as the cofactor was not present in the binding pocket irrespective of the measured affinity of the mutant for NADH and at least to the same extent as after treatment of sPHGDH with charcoal. Comparison of the different mutants against each other shows that replacement of an amino acid with tryptophan, which contains a bulky side chain, results in complete abrogation of cofactor binding, but also led to a large increase in structural disorder. Replacement of amino acids involved in hydrogen bonding resulted in a broader spectrum of reduced NADH binding and less impact on structural integrity, as might be expected for substitutions that selectively remove specific interactions with the ligand.

However, without precise knowledge of the mutants' structures, it is not known how well the binding site is preserved. Crystallisation attempts so far have proven challenging as the only crystals obtained were with mutant R154E and these grew in layers of multiple plates, which precluded analysis. Crystallisation of sPHGDH had also previously been shown to result in the formation of layers of plates (Chapter 4). Introduction of the site-directed mutations increased the amount of unordered protein regions as monitored by CD and DSF and this will further interfere with crystallisation as well as enzymatic activity. According to the CD data, the portion of unordered structure was increased in all mutants by 26-50 % compared to wt sPHGDH, with triple mutant P175A/T206A/T212A retaining most structural order. However, in this case the amount of disordered structure is unlikely to predict the protein's tendency to crystallise as no knowledge is gained on which regions of the protein are more or less disordered. For the crystallisation, time limitations prevented an investigation of additional optimisation strategies such as crystal seeding, testing further additives or optimisation of the protein buffer conditions.

Chapter 7. Fragment-based drug design

7.1 Introduction

In recent years, fragment-based drug design (FBDD) has emerged as a valuable technique for the discovery of new molecules that modulate biological functions. Pioneering work in the field has been done by Jencks, who provided the basic concept for FBDD. Jencks described the binding energy of a molecule-protein interaction as a function of the binding energies arising from the molecule dissected into its individual components or fragments [178]. Further studies on the inhibition of HMG-CoA reductase showed that inhibitors could also be seen as linked fragments binding to different sites of the enzyme's binding pocket [179]. The essential task of FBDD lies in identifying these fragments and building on them to obtain high-affinity binders of the target of interest. In general, fragments are considered to comply with an adapted form of Lipinski's rules, the so-called "rule of three", which specifies a molecular weight of less than 300 Da, a calculated logP below three, and a maximum of three hydrogen bond donors and acceptors [180].

FBDD has certain advantages over the otherwise often-used high throughput screening (HTS) of large libraries of molecules. In FBDD a significantly lower number of compounds compared to HTS are screened while still sampling an equivalent or even larger chemical space. Even by constraining the molecular weight of ligands used in HTS to 300-500 Da, there are still 10^{60}-10^{200} potential molecules to be screened whereas there are only 10^7 possible fragment-like molecules with up to eleven atoms of C, N, O and F that comply with the rule of three [180]. There is also no need to screen such large libraries as in HTS to obtain hits as, due to the lower complexity of the fragments, the likelihood of binding to the target is increased [181]. Another consequence of the small size of fragments is that for a fragment found to bind to the target, the proportion of atoms involved in target binding is larger than in the case of bigger molecules, therefore fragments can be considered as more efficient binders with high binding energy per unit molecular mass often referred to as high ligand efficiency (LE) [181]. The physicochemical properties of molecules to be developed into drugs are very important, and compounds with higher molecular weight and higher lipophilicity have been shown to be more likely to fail during clinical development than smaller and less lipophilic compounds [182]. Whereas fragments

with a molecular weight of below 300 Da are well within the margins of size for a drug-like molecule, HTS libraries are likely to include compounds that are not suitable for further development into drugs in terms of size and physicochemical properties. In addition, the small-sized fragments have more scope to include further optimisation of physicochemical properties into the potency improvement process.

Development of potent inhibitors, or binders to the target, can be started once one (or more) fragment(s) binding to the target have been identified as a starting point. In principle, there are three approaches to increase binding affinity and specificity of fragments for the target: (1) Growing the fragment to increase target interaction through additional functionalities, (2) in the case where several fragments binding to different sites of the target pocket have been identified, increased potency can be achieved by linking the fragments, and (3) if chemically distinct fragments have been found their key functionalities can be combined in one molecule through "fragment merging" (Figure 7.1).

(A) Fragment growing

7-Azaindole	3-Aminophenyl-7-azaindole	PLX4720	Zelboraf
9 heavy atoms	16 heavy atoms	27 heavy atoms	33 heavy atoms
IC50 > 200 μM (Pim-1 Kinase)	IC50 = 100 μM (Pim-1 Kinase)	IC50 = 0.013 μM (B-Raf V600 mutant)	IC50 = 0.031 μM (B-Raf V600 mutant)
LE < 0.56	LE = 0.34	LE = 0.40	LE = 0.31

(B) Fragment linking

Fragment 1	Fragment 2	Linked fragment
LE = 0.20	LE = 0.17	LE = 0.19
Kd (LDHA) = 280 μM	Kd (LDHA) = 770 μM	Kd (LDHA) = 0.13 μM
IC50 > 500 μM (LDHA enzyme)	IC50 > 500 μM (LDHA enzyme)	IC50 = 4.2 μM (LDHA enzyme)

(C) Fragment merging

Fragment 1	Fragment 2	Merged fragment
LE = 0.29	LE = 0.39	LE = 0.39
Kd = 1600 μM	Kd = 400 μM	Kd = 28 μM

Figure 7.1 - Illustration of the basic strategies to increase fragment potency.

(A) Example of fragment growing starting from 7-azaindole (IC$_{50}$ > 200 μM) with low affinity and potency to finally give the highly potent and selective B-Raf kinase (VE600 mutant) inhibitor zelboraf (IC$_{50}$ = 0.031 μM) [183]. (B) Linking of two fragments binding to the cofactor binding site of lactate dehydrogenase A. Fragment 1 (K$_D$ = 280 μM) binds to the nicotinamide/substrate-binding site whereas fragment 2 (K$_D$ = 770 μM) binds to the adenine-binding site of the cofactor/substrate binding pocket. The linked fragment has a 1000-fold increase in binding affinity (K$_D$ = 0.13 μM) over the individual fragments [140]. (C) Fragment merging strategy illustrated in the search of inhibitors for CYP121 by fragment 1 (K$_D$ = 1600 μM) and fragment 2 (K$_D$ = 400 μM) whose merging leads to the significantly more potent fragment (K$_D$ = 28 μM) [184].

Despite the advantages of FBDD, there are also challenges with this approach. Most difficulties arise from the fact that, due to the small size and low complexity of fragments, they are most likely to exhibit only weak affinity for the target. It can therefore be difficult to develop a suitable assay, so as to unambiguously detect hits and follow them up. Thus FBDD often relies at some point on knowing the structure of the target through e.g. X-ray crystallography or nuclear magnetic resonance (NMR). Evaluation of the fragment can then be performed in the context of this

structural information (co-crystallisation, modelling). A proposed screening cascade for finding and evaluating fragments is shown in Figure 7.2.
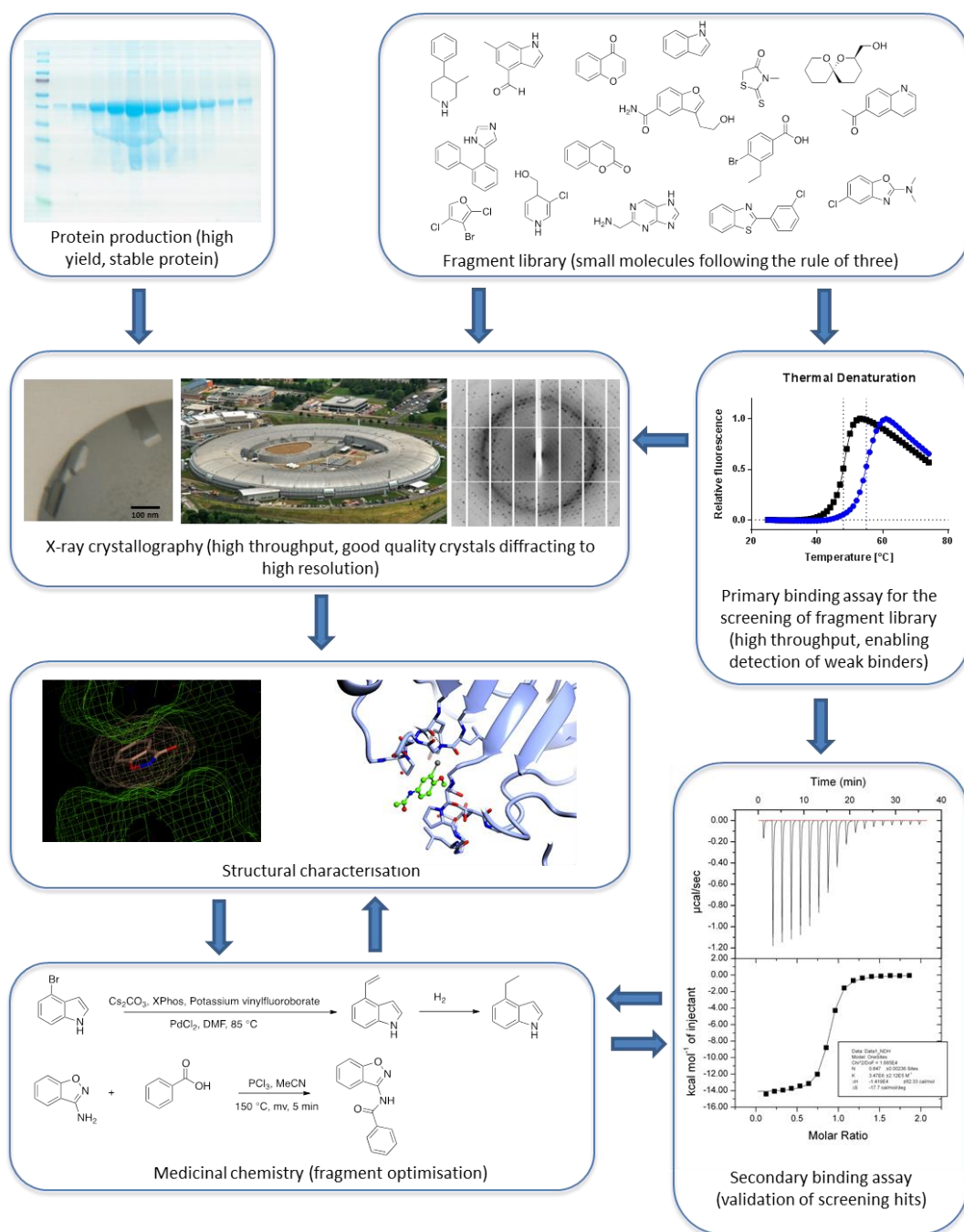
Figure 7.2 - Schematic representation of a typical fragment screening cascade.

Full-length protein or fragments thereof have to be produced in sufficient amounts for binding assays and crystallisation. A fragment library consisting of 300-1000 fragments that comply with the rule of three is screened directly by crystallography and/or using a biophysical method such as differential scanning fluorimetry. The primary screen has to be high-throughput and suitable for the detection of weak binders. In the case where crystallography is not used as the primary screen, hits from the primary assay are structurally characterised by soaking into protein crystals and/or co-crystallising with the target protein. In addition, hits are also confirmed using a second binding assay, which, in the case of isothermal titration calorimetry, also gives information on the thermodynamic and kinetic parameters of fragment binding. Structural and biophysical information gained on the fragment hits is used to guide medicinal chemistry to improve the potency and physicochemical properties of the initial hits. (Figure adapted from [185])

7.2 Aims and objectives

Prompted by the difficulties in finding inhibitors based on a rational substrate- and cofactor-mimic approach as outlined in previous chapters (chapter 4 and 5), a more general approach to hit finding was taken by screening a fragment library. The following objectives were formulated:

(1) Screening of fragment library against human PHGDH using a high throughput assay.

(2) Optimisation of the crystallisation of human PHGDH to allow for co-crystallisation with fragments or fragment soaking into empty crystals.

(3) Development of a secondary assay to validate the initial fragment hits.

(4) Optimisation of fragment hits based on crystal structures.

7.3 Materials and Methods

7.3.1 Differential scanning fluorimetry (DSF)

The DSF assay was used for screening of the fragment library and was performed according to the procedure described in section 2.9. Briefly, 150 nL of 100 mM compound in 100 % DMSO were dispensed in duplicate into 384-well plates using the Echo® liquid handler (Labcyte, Sunnyvale, CA, USA). 150 nL of 100 mM NADH was dissolved in assay buffer (25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP) and used as positive control. Reference wells contained 150 nL DMSO (reference for fragment containing wells) or 150 nL buffer only (reference for NADH containing wells). The final assay mixture contained 25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP, 10 x Sypro Orange (Thermo Fisher Scientific, Waltham, MA, USA), 3 % (v/v) DMSO, 15 µM full-length PHGDH and 1 mM compound in a final assay volume of 15 µL. Measurement and analysis of the heat denaturation curves was performed as described in section 2.9.

7.3.2 Isothermal titration calorimetry (ITC) competition experiments

For ITC competition experiments, sPHGDH was treated with charcoal as described in section 6.3.1 to remove as much co-purified cofactor as possible. To take into account batch-to-batch variability due to charcoal treatment, a reference titration of 0.5 mM NADH into 0.5 mM sPHGDH was performed for each charcoal-treated sPHGDH aliquot on the day of the competition experiments. For the competition experiments, sPHGDH at a final concentration of 0.05 mM was mixed with 100 mM fragment in 100 % (v/v) DMSO to yield a final fragment concentration of 5 mM (5 % (v/v) DMSO). NADH dissolved in ITC assay buffer was also mixed with 100 mM fragment in 100 % (v/v) DMSO to yield a final ligand mixture of 0.5 mM NADH, 5 mM fragment and 5 % (v/v) DMSO. The NADH-fragment mixture was used as titrant, and the sPHGDH-fragment mix was placed in the sample cell. ITC experiments were then performed at 25 °C in 25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP using the injection pattern described in section 2.10 on a MicroCal iTC200 (GE Healthcare, Little Chalfont, UK). All thermodynamic parameters of the reaction were determined using the ORIGIN software, version 7.0. K_a and ΔH for the fragments were calculated using Equation 7.1 and Equation 7.5 with the respective error propagation (Equation 7.2 and Equation 7.6) [186].

$$K_x = (\frac{K_1}{K_{app}} - 1) \times \frac{1}{L_{x,tot}}$$

Equation 7.1 – Calculation of K_a for fragment x (K_x)

K_1 = binding constant determined in direct ITC control experiment with high affinity ligand (NADH) (M^{-1})

K_{app} = binding constant calculated in the presence of fragment x in displacement ITC (M^{-1})

$L_{x,tot}$ = total ligand concentration of fragment x (M)

$$\Delta K_x = \left| \frac{c}{K_{app}} \right| \times \Delta K_1 + \left| \frac{-cK_1}{\left(K_{app} \right)^2} \right| \times \Delta K_{app}$$

Equation 7.2 – Error propagation for the calculation of K_x

ΔK_1 = error for K_1, calculated by Origin software

ΔK_{app} = error for K_{app}, calculated by Origin software

$$K_{D,x} = \frac{1}{K_x} = K_x^{-1}$$

Equation 7.3 – Calculation of K_D

$$\Delta K_{D,x} = \left| \frac{\partial K_{D,x}}{\partial K_x} \right| \times \Delta K_x = \left| \frac{-1}{K_x^2} \right| \times \Delta K_x$$

Equation 7.4 – Error propagation for the calculation of K_D

$$\Delta H_x = (\Delta H_1 - \Delta H_{app}) \times \left(1 + \frac{1}{K_x \times L_{x,tot}} \right)$$

Equation 7.5 – Calculation of ΔH_x for the fragment x

K_x = binding constant for fragment determined in displacement ITC (see Equation 7.1) (M^{-1})

ΔH_1 = change in enthalpy for high affinity ligand (NADH) as determined in direct ITC control experiment

ΔH_{app} = change in enthalpy in the presence of fragment x in displacement ITC

$$\Delta\Delta H_x = \left| 1 + \frac{1}{K_x \times L_{x,tot}} \right| \times \Delta\Delta H_1 + \left| -1 - \frac{1}{K_x \times L_{x,tot}} \right| \times \Delta\Delta H_{app} + \left| \frac{\Delta H_{app} - \Delta H_1}{K_x^2 \times c} \right| \times \Delta K_x$$

Equation 7.6 – Error propagation for the calculation of ΔH_x for the fragment x

$\Delta\Delta H_1$ = error in ΔH_1, calculated by Origin software

$\Delta\Delta H_{app}$ = error in ΔH_{app}, calculated by Origin software

7.3.3 Limited proteolysis

Limited proteolysis uses proteases with defined cleavage sites to digest proteins into smaller, more stable fragments that are expected to have a higher probability of crystallising. Limited proteolysis can be performed *in situ* by using the protein-protease mixture directly in crystal trials, or the digested fragments can be analysed, re-cloned and purified prior to crystallisation.

For *in situ* proteolysis 54 mg/mL sPHGDH was mixed with protease in a molar ratio of 1:100 or 1:1000 and incubated at rt for 10 minutes. The mixture was used in homemade crystallisation trials in a 1:1 and 1:2 protein:reservoir ratio using sodium citrate or MMT buffer over a range of pH in combination with different PEGs (400 – 1000) as the reservoir.

To determine if stable fragments were produced by proteolysis, 100 µg sPHGDH or PHGDH were mixed with different proteases at a molar ratio of 1:100 and 1:1000. Proteolysis was performed for one hour at RT or 37 °C before the reaction mixture was separated by SDS-PAGE. Characterisation of stable protein fragments was performed by LC-MS using the liquid protein and protease mixture prior to gel electrophoresis (Mass Spectrometry Facility, University of Leeds, Leeds, UK).

7.3.4 Cloning of new PHGDH fragments

Following on from the determination of new stable PHGDH fragments by limited proteolysis and subsequent LC-MS analysis, constructs corresponding to those fragments were obtained through removal of N- and C-terminal amino acids by deletion mutagenesis.

Mutagenesis of sPHGDH (aa 3-314) was performed in two rounds of PCR with the first round generating the N-terminal and the second round the C-terminal truncation (Figure 7.4). For the generation of the N-terminal truncations the NcoI restriction site was kept in the reverse primer, whereas the forward primer contained the NcoI restriction site followed by an alanine and the bases corresponding to the desired N-terminal amino acids of the new N-terminus (Table 7.1). The purified DNA obtained after the first round of PCR, followed by DpnI and NcoI digestion and ligation, was used for the second round of PCR. For removal of the C-terminal bases 5'-phosphorylated primers were used. In this case, the reverse primers determined the C-terminal truncation of the sequence, whereas a common forward primer for all

three constructs was used containing a stop codon followed by the LIC cloning site of the pNIC-28-BsaI vector. By performing two rounds of PCR mutagenesis, the need for subcloning into empty pNIC-28-BsaI vector was avoided. The PHGDH fragments generated were constructs 93 (aa 93-298), 92-2 (aa 92-299) and 92-3 (aa 92-301) (Figure 7.3).

1st round of PCR

Final construct name	Construct after PCR	Template DNA	Direction	Primer Sequence (5' → 3')
93	aa 93-314	sPHGDH	Forward	GCCTGAGC*CCATGG*CTTTGGTT ATGAACACCCCC
92-2 and 92-3	aa 92-314	sPHGDH	Forward	GCCTGAGC*CCATGG*CTATCTTG GTTATGAACACCCCC
93, 92-2 and 92-3	aa x-314	sPHGDH	Reverse	GAGCACTTTCCGCAGATTTGCC ATGGATTGGAAGTA

2nd round of PCR

93, 92-2 and 92-3	-	-	Forward	Phos-**TGA**CAGTAAAGGTGGAT ACGGATCCG
93	aa 93-298	aa 93-314	Reverse	Phos-AATTTCCTCCCCACAGCG GCTCTGAGCCTCC
92-2	aa 92-299	aa 92-314	Reverse	Phos-AGCAATTTCCTCCCCACA GCGGCTCTGAGC
92-3	aa 92-301	aa 92-314	Reverse	Phos-CTGAACAGCAATTTCCT CCCCACAGCGGCTCTGAGCC

Table 7.1 – Primers used for PCR mutagenesis to generate new PHGDH fragments

Primers for 1st round of PCR: NcoI restriction site is underlined with the codon for the starting alanine highlighted in blue, followed by the codon for the first amino acid of the new PHGDH fragments highlighted in yellow.
Primers for 2nd round of PCR: Stop codon highlighted in bold; codon for the respective final amino acid residue is highlighted in pink; Phos = phosphorylated residue.

```
            10          20          30          40          50          60
SMANLRKVLI SDSLDPCCRK ILQDGGLQVV EKQNLSKEEL IAELQDCEGL IVRSATKVTA

            70          80          90         100         110         120
DVINAAEKLQ VVGRAGTGVD NVDLEAATRK GILVMNTPNG NSLSAAELTC GMIMCLARQI

180         130         140         150         160         170
PQATASMKDG KWERKKFMGT ELNGKTLGIL GLGRIGREVA TRMQSFGMKT IGYDPIISPE

           190         200         210         220         230         240
VSASFGVQQL PLEEIWPLCD FITVHTPLLP STTGLLNDNT FAQCKKGVRV VNCARGGIVD

           250         260         270         280         290         300
EGALLRALQS GQCAGAALDV FTEEPPRDRA LVDHENVISC PHLGASTKEA QSRCGEEIAV

           310
QFVDMVKGKS LTGV
```

Figure 7.3 – Amino acid sequences of sPHGDH and shorter fragments.

Scheme representation of the amino acid sequence of PHGDH with the lid domain, hinge region and core domain underlined in orange, green and blue, respectively. The green square highlights the sequences of the shorter fragments 93, 92-2 and 92-3. Amino acids involved in substrate and cofactor binding, according to the crystal structure (PDB 2G76), are highlighted in red and magenta, respectively.

The PCR reactions followed the method and thermal profile described in section 6.3.2, except using sPHGDH DNA or DNA generated after the 1st round of PCR as template DNA. DpnI as well as DpnI and NcoI double digests were performed for 1 h at 37 °C, followed by ligation with T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). Ligation was achieved using 17 µL of the PCR restriction digested PCR reaction, 2 µL ligation buffer and 1 µL T4 DNA ligase for 2 hours at RT. Transformation of *E. coli* DH5α, DNA purification and sequencing was performed as described in sections 2.3.
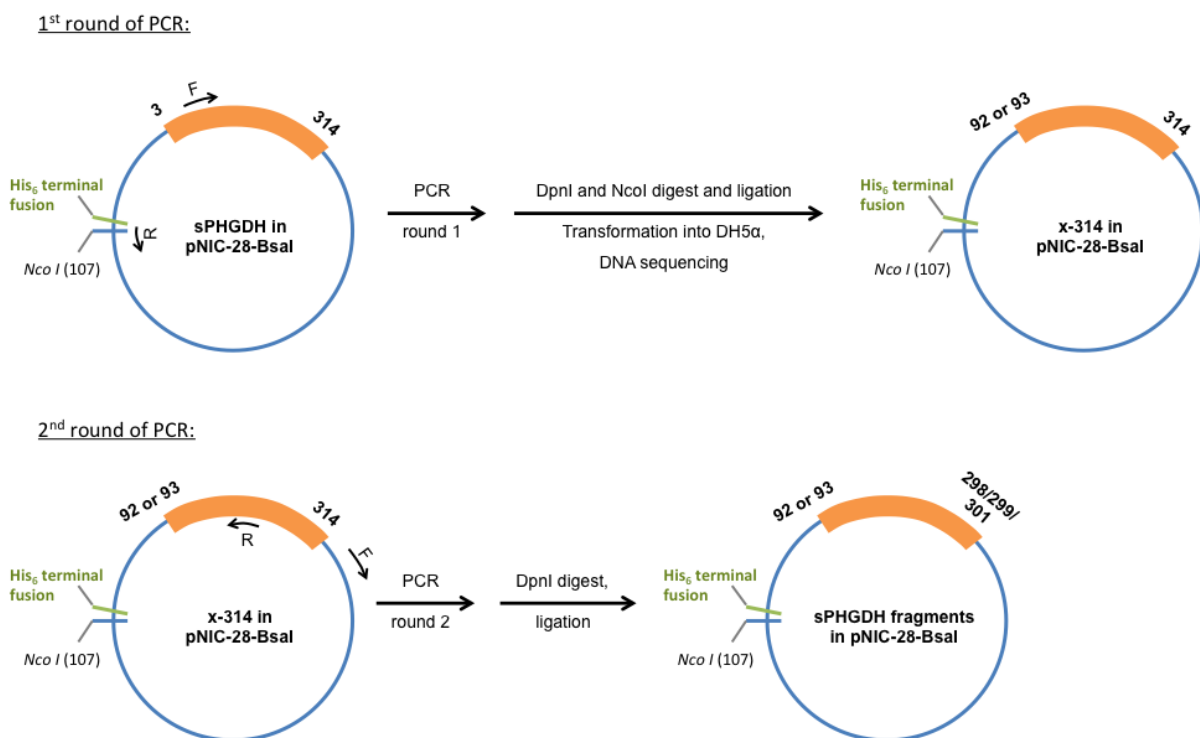
Figure 7.4 – Scheme of PCR-based mutagenesis method to generate PHGDH fragments.

7.3.5 Expression and purification

PHGDH fragments 93, 92-2 and 92-3 were transformed in to E. coli Rosetta (DE3) cells and clones selected on LB Agar containing 50 µg/mL kanamycin for 18 h at 37 °C. Single colonies were picked and used to inoculate aliquots of 5 mL LB medium containing 50 µg/mL kanamycin. From these cultures, glycerol stocks (15 % (v/v) glycerol) were prepared and stored at - 80 °C.

Glycerol stocks were taken to inoculate starter cultures for large-scale growth and expression of PHGDH fragments as described in sections 2.5.1.

7.3.6 Crystallisation of PHGDH fragment 93

Initial sitting drop vapour diffusion crystallisation trials were set up against commercially available sparse matrix screens as described in section 2.12 and stored at 4 or 20 °C. After first crystals were obtained, the growth conditions were optimised by screening different buffer systems in combination with different precipitant concentrations, and resulted in 0.1 M PCTP buffer, pH 7 and 23-28 % (w/v) PEG 1500 as the best growth conditions at 20 °C. Deep-well blocks of the optimised condition were prepared using a Biomek (Beckman Coulter, Brea, CA, USA) and drops pipetted as described in section 2.12.

7.3.7 Soaking of compounds into crystals of 93

For soaking experiments, compounds were dissolved at 80 mM in soaking buffer which was a mixture of the reservoir buffer (25 % PEG 1500, 0.1 M PCTB pH 7) and the gel filtration buffer used for the protein purification (25 mM HEPES, pH 7, 100 mM NaCl, 0.5 mM TCEP) at a 3:2 ratio supplemented with 20 % (v/v) DMSO. Crystals were transferred into drops containing the dissolved compounds, and soaked at 20 °C for 24 hours. Crystals were harvested using CryoLoops (Hampton Research, Aliso Viejo, CA, USA) matched in size to the crystal and cryo-protected in soaking buffer containing 80 mM compound and supplemented with 15 % (w/v) PEG400 before being flash cooled in liquid nitrogen.

Crystals were measured and data collected either at the Diamond Light Source (DLS) synchrotron or on an in-house Rigaku Micromax 007 (2005) fitted with an R-Axis IV[++] image plate detector. Data analysis was performed as described in section 2.13 using a truncated model of sPHGDH for molecular replacement until a high-resolution structure of 93 was obtained and could be used as a model for subsequent analysis.

7.4 Results

7.4.1 Fragment screening using DSF

The fragment library used within this study was kindly provided by CRT Cambridge and was a selection of compounds following the rule of three further selected on solution behaviour and suitability for NMR fragment screening.

The fragment library was screened at 1 mM compound concentration against full-length PHGDH in a DSF assay in two independent experiments. Hits were determined based on an average Z-score ≥ 1, which is equivalent to a change in T_m of PHGDH in the presence of compound of ≥ one standard deviation greater than the mean. For the Z-score calculation, the control wells were excluded. Using this cutoff, 42 out of 661 compounds were classified as hits, which is equivalent to a hit rate of 6.4 % (Figure 7.5A). The two independent screening runs performed correlated with a Pearson correlation coefficient of 0.36 (r^2 = 0.13) (Figure 7.5B).
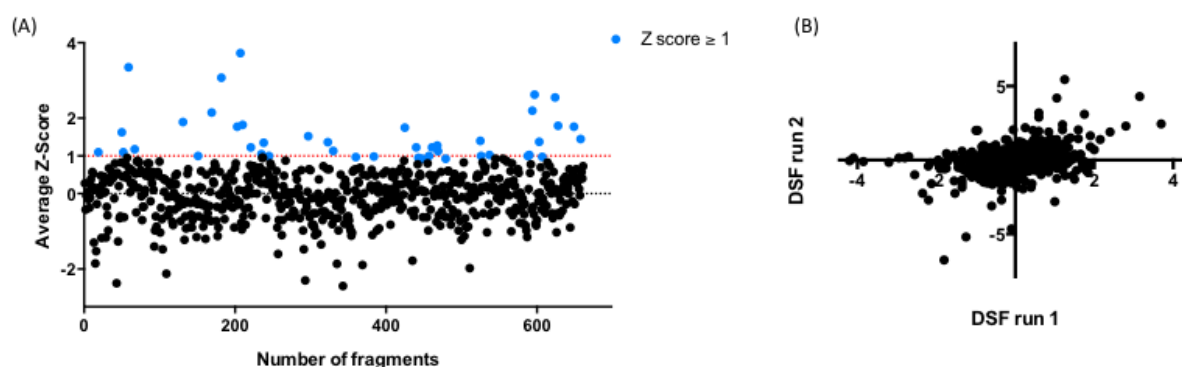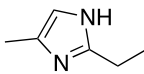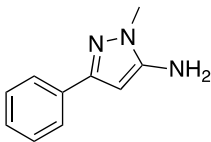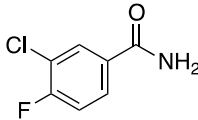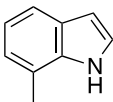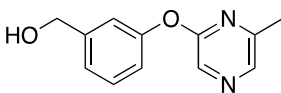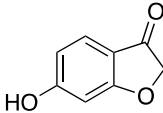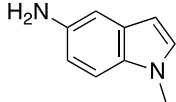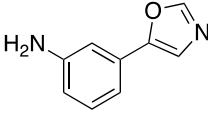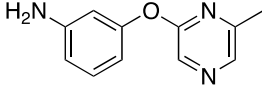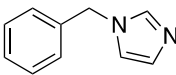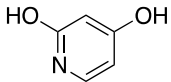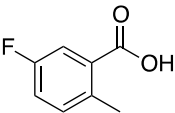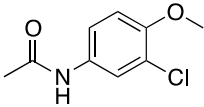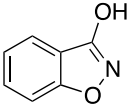
Figure 7.5 – Results of fragment screening using DSF.

(A) Thermal denaturation profiles of 10 μM PHGDH in the presence of 1 mM fragment (3 % (v/v) DMSO) were measured. Graph shows the mean of two independent experiments with two replicates per experiment. The average Z-score was calculated, and fragments with a Z-score ≥ 1 defined as hits (highlighted in blue). (B) Pearson correlation graph of the two independent DSF screens with the Pearson r = 0.36 and r^2 = 0.13, p < 0.0001.

From the 42 initial hits, 15 were commercially available and were purchased (Table 7.2). All purchased fragments were re-evaluated in the DSF screen and interestingly only seven were able to increase the T_m of PHGDH at all, indicating that the DSF screens are prone to false positive results.

ID	Name	Structure	Z score[a]	$\Delta T_m{}^a$ (°C)	$\Delta T_m{}^b$ (°C)
A	2-Ethyl-4-methylimidazole		1.0 ± 1.0	2.0 ± 2.1	-0.1 ± 0.2
B	1-Methyl-3-phenyl-*1H*-pyrazol-5-amine		1.0 ± 0.4	2.2 ± 0.8	-0.2 ± 0.2
C	5-Fluoro-4-hydroxy quinazoline		1.0 ± 0.8	2.4 ± 1.1	0.2 ± 0.3
D	3-Chloro-4-fluorobenzamide		1.0 ± 0.8	2.3 ± 1.0	-0.1 ± 0.2
E	7-Methyl-*1H*-indole		1.1 ± 0.7	2.4 ± 1.2	0.2 ± 0.2
F	[3-(6-methylpyrazin-2-yl)oxyphenyl] methanol		1.2 ± 0.9	1.8 ± 1.7	-0.3 ± 0.2
G	6-Hydroxy-2,3-dihydrobenzo[b] furan-3-one		1.5 ± 0.4	1.7 ± 0.6	0.4 ± 0.4
H	5-Amino-1-methyl-*1H*-indole		1.8 ± 1.4	2.0 ± 1.2	1.2 ± 0.4
I	3-(1,3-oxazol-5-yl)aniline		1.0 ± 0.9	1.4 ± 0.6	-0.8 ± 0.3
J	3-(6-Methylpyrazin-2-yl)oxyaniline		1.1 ± 0.6	1.7 ± 0.7	-0.6 ± 0.6
K	1-Benzylimidazole		1.1 ± 0.2	1.3 ± 0.9	-0.6 ± 0.4

ID	Name	Structure	Z score[a]	ΔT_m^b (°C)	ΔT_m^c (°C)
L	2,4.Dihydroxy pyridine		1.4 ± 0.2	0.8 ± 0.1	0.3 ± 0.1
M	5-Fluoro-2-methylbenzoic acid		1.1 ± 0.2	1.4 ± 0.5	0.4 ± 0.3
N	N-(3-chloro-4-methoxyphenyl) acetamide		1.2 ± 0.3	1.3 ± 0.2	-0.2 ± 0.4
O	3-Hydroxy benzisoxazole		1.4 ± 0.7	2.5 ± 0.7	0.2 ± 0.3

Table 7.2 – Repurchased fragment hits from DSF screening.

[a] = Average Z score of primary fragment screens by DSF; [b] = average ΔT_m of primary fragment screens by DSF; [c] = average ΔT_m of purchased fragments. Data represent average and standard deviation of two independent experiments with two intra-assay repeats.

7.4.2 Validation of hits by ITC competition assay

To validate the initial hits from the DSF screen, an orthogonal assay was needed, ideally with a non- fluorescent read-out to avoid repeating false positives that interfere with the fluorescence signal rather than binding to PHGDH. In this regard, isothermal titration calorimetry (ITC) is a valuable technique that measures the heat energy changes associated with a molecular interaction, and allows determination of the enthalpic and entropic contributions to the interaction as described in section 2.10. Wiseman c-parameters in a range from $10 \leq c \leq 500$ are thought to define a suitable range of binding constants that can be reliably determined by ITC [102]. For a one-site binding model, the binding constant is determined from the inflection point of the sigmoidal binding curve. As c approaches zero the shape of the binding curve approaches a horizontal line, thus complicating an error-free determination of the binding constant (Figure 7.6).
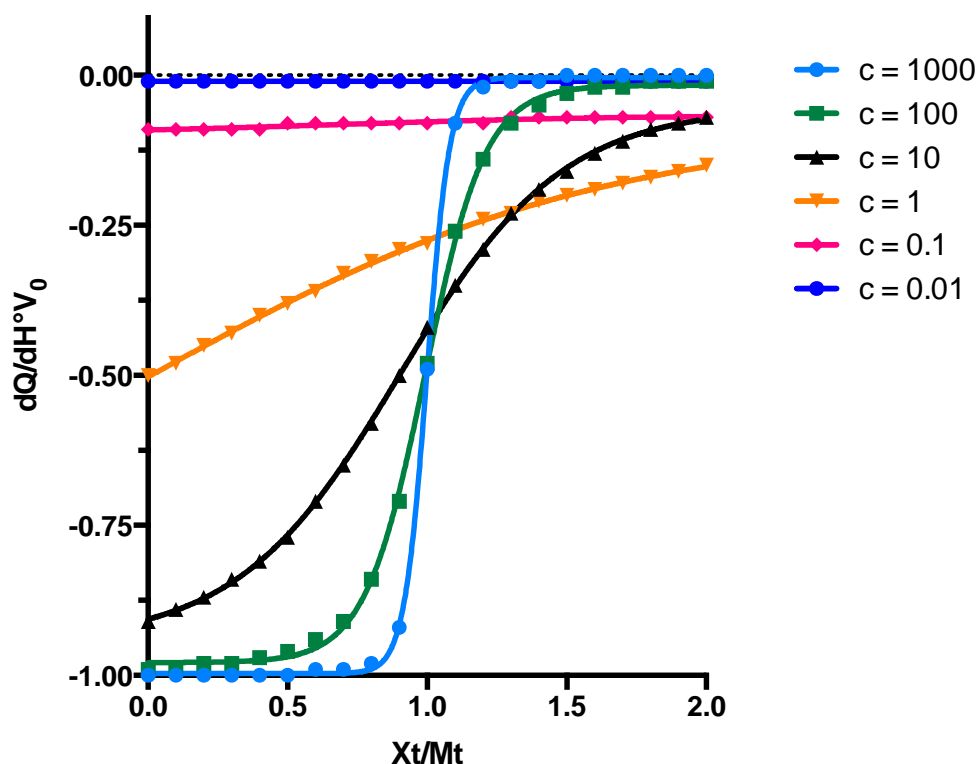
Figure 7.6 – Variations in the ITC titration curve in relation to the c-parameter.
Figure adapted from [102].

In cases where the binding affinity of the ligand for the receptor is very low (i.e. the experiment has a low c value), and an ideal c value cannot be achieved through increasing receptor/ ligand concentrations due to limited receptor and/or ligand solubility, displacement experiments can provide a solution. In these experiments, a high affinity ligand must be available that binds to the same binding site as the ligand under investigation. As the high affinity ligand competes for the binding site, it will bind more slowly, hindered by the low affinity ligand, resulting in a different shape of the binding curve due to differences in the thermodynamic parameters of the reaction. ITC displacement experiments have been mainly used to deconvolute the contribution of individual fragments of a natural, high affinity ligand to receptor binding [186-188], but is finding increasing application in fragment-based drug design [189, 190].

PHGDH was shown to bind to NADH with very high affinity, with a K_D of 0.22 μM as determined by direct ITC (section 6.4.1), so that NADH was a suitable high-affinity ligand for ITC competition experiments. For these experiments, human sPHGDH was used rather than full-length PHGDH as sPHGDH could be produced in higher yields,

and as ITC competition experiment would only allow the characterisation of ligands that compete for the cofactor-binding site, the regulatory domains of the protein were not required. As noted earlier (section 6.3.1), sPHGDH was found to co-purify with cofactor, so the protein was treated with charcoal to remove bound cofactor prior to the ITC experiments (see section 6.3.1 and 6.4.1). In order to be able to analyse the competition experiments, the binding thermodynamics of NADH to sPHGDH were characterised. Each set of experiments performed with a particular batch of charcoal-treated protein was compared to an NADH titration into the same batch of sPHGDH to account for variability in the efficiency of cofactor removal through charcoal treatment.

For the ITC competition experiment itself the NADH titration was repeated with the low-affinity ligand (e.g. fragment) added to sPHGDH in the calorimetric cell, as well as to the NADH, to avoid interference from any additional effects of ligand dilution. In order to validate the experimental set up, ADPR, which has a K_D for PHGDH of 40 µM as determined by direct ITC, was used at different concentrations in combination with 0.5 mM NADH (Figure 7.7).
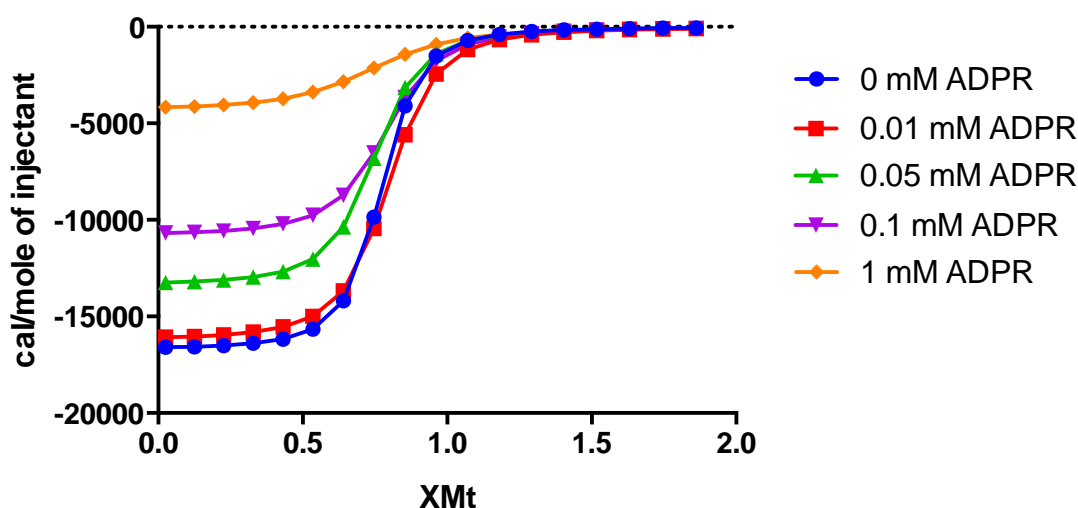
Figure 7.7 – ITC competition experiment with ADPR at different concentrations

0.5 mM NAD$^+$ mixed with different concentrations of ADPR (0-1 mM) was titrated into 0.05 mM PHGDH mixed with different concentration of ADPR (0-1 mM) in 25 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mM TCEP and 5 % (v/v) DMSO. Data were analysed by non-linear regression using the built-in one-site fit model of the ORIGIN software. Data showing the heat generated per mole of injected ligand against the molar ratio of receptor to ligand was plotted in GraphPad prism.

[ADPR] (mM)	0	0.01	0.05	0.1	1
K_{a1} (μM^{-1})	3.74 ± 0.30	2.45 ± 0.29	2.38 ± 0.17	1.85 ± 0.13	0.68 ± 0.12
K_{a2} (μM^{-1})	-	0.05 ± 0.03	0.01 ± 0.004	0.01 ± 0.003	0.005 ± 0.001
K_D (mM)	-	0.02 ± 0.01	0.09 ± 0.04	0.10 ± 0.03	0.22 ± 0.07
ΔH_1 (kJ/mol)	-1.63E4				
ΔH_2 (kJ/mol)		-4.84E6	-7.59E6	-1.20E7	-1.13E7

Table 7.3 – ITC displacement experiment with ADPR

The affinity constant K_{a1} was determined from the one-site fit model of the ORIGIN software as an analysis of a direct titration and used to calculate the actual affinity constant of ADPR K_{a2} in each competition titration experiment. The dissociation constant (K_D) was obtained from the reciprocal of K_{a2}. ΔH_1 refers to the chance in enthalpy measured in the titration experiment and ΔH_2 was calculated therefrom.

The ITC competition experiment with ADPR showed that the technique could be used to measure the binding of low-affinity fragments to sPHGDH, and the repurchased fragments A-O were thus evaluated in this assay (Table 7.4).

Fragment	ΔH (kJ/mol)	ΔS (kJ/(mol x K)	ΔG (kJ/mol)	K_D (mM)	LE (kJ/(mol x #))
A	No competition.				
B	-467.2	-1.61	12.8	5.8 ± 2.0	0.23
C	8.3	-0.01	10.8	12.8 ± 2.7	0.21
D	-5.4	-0.07	16.0	1.6 ± 0.2	0.35
E	-1593.9	-5.40	16.1	1.5 ± 0.2	0.38
F	15.8	0.02	10.2	16.6 ± 10.4	0.15
G	-292.7	-1.02	12.7	5.9 ± 1.5	0.28
H	-102.0	-0.38	11.2	11.1 ± 5.0	0.24
I	-260.8	-0.91	11.6	9.3 ± 3.8	0.23
J	-11.5	-0.08	11.1	11.5 ± 8.4	0.18
K	-64.4	-0.25	11.0	11.8 ± 9.6	0.22
L	No competition.				
M	-47.0	-0.19	9.0	26.2 ± 31.8	0.20
N	-1753.2	-5.93	14.5	2.8 ± 1.6	0.27
O	-382.6	-1.33	12.5	6.5 ± 4.7	0.30

Table 7.4 – Assessment of fragment hits by ITC competition assay

ITC displacement experiments were analysed using the one-site fit model in ORIGIN together with the calculation outlined in section 7.3.2. Changes in entropy (ΔS) and enthalpy (ΔH) as well as the binding constant (K_D) were determined. LE was calculated as the coefficient of ΔG per number of non-hydrogen atoms (= #).

7.4.3 Identification of a more optimal protein construct for X-ray crystallography by limited proteolysis

For the evaluation of the fragment hits, it was highly desirable to investigate their binding mode using X-ray crystallography. Therefore a protein construct was needed that resulted in robust diffraction-quality crystals and could be used for co-crystallisation or fragment soaking. The sPHGDH construct used for crystallisation thus far was not ideal in this context, because it often formed plate-like, multiple layered crystals that diffracted to a rather low resolution of 2.7 Å. In addition, sPHGDH co-purified with its natural cofactors, which were shown to be difficult to remove from the cofactor-binding pocket (see section 6.3.1) that, together with the substrate-binding site, forms the active site on the protein that was the intended target of potential inhibitors.

In this context, limited proteolysis was used to identify suitable protein constructs for crystallisation experiments by mixing the protein with catalytic amounts of protease. Proteases have been shown to remove primarily solvent accessible and flexible or unstructured domains of proteins such as disordered termini and loops. Through removal of those unstructured regions, the loss of conformational entropy associated with crystallisation should be reduced, resulting in a more stable protein core that should crystallise more easily, as well as resulting in well-diffracting crystals [191, 192]. In principle, there are two different ways to perform limited proteolysis, either *in situ*, which is the direct use of the protease-protein mixture in crystallisation trials or by construct-directed proteolysis which involves identification of the proteolytic fragment and subsequent re-cloning, expression and purification prior to its use in crystallisation trials (Figure 7.8).
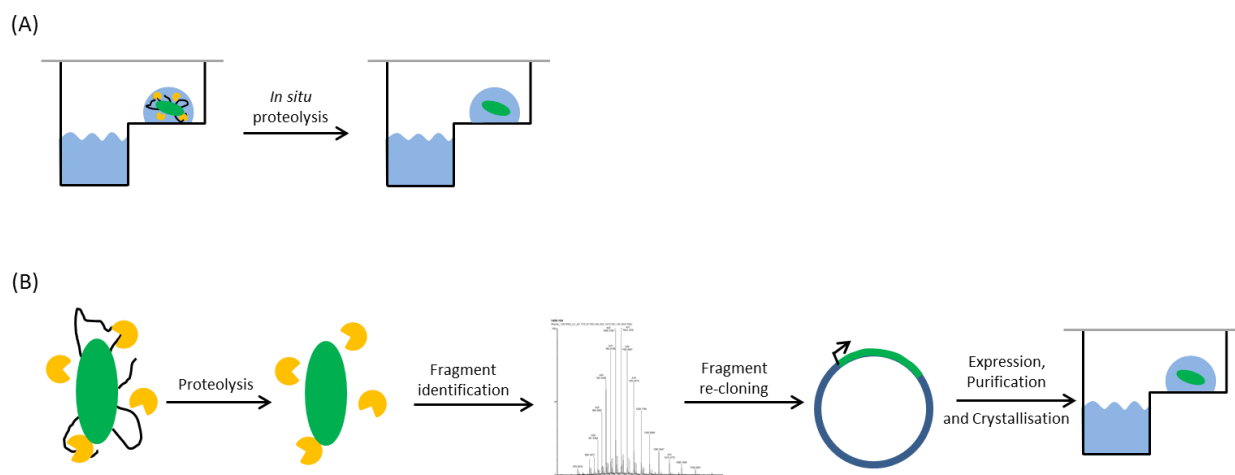
Figure 7.8 - Schematic representation of two different limited proteolysis strategies

(A) *In situ* proteolysis of protein consisting of core (green) and flexible (black) regions mixed with protease (yellow) directly in the crystallisation drop. (B) Construct-directed proteolysis whereupon the protein core obtained is characterised by LC-MS and N-terminal sequencing before being re-cloned and expressed for use in crystallisation trials (Picture adapted from [193]).

In situ proteolysis was attempted using 100 µg sPHGDH combined with different proteases (trypsin, α-chymotrypsin, thermolysin and endoproteinase GluC) at a molar ratio of 1:100 and/or 1:1000. Samples were used in crystallisation screens, prepared with malate- or citrate-containing buffers over a range of pH and various PEGs (400 – 10000). Screens in citrate-containing buffer yielded only plate shaped crystals as seen before with the parental sPHGDH construct (Figure 7.9 A). Trypsin-treated sPHGDH resulted in one small crystal in MMT buffer, pH 7, 25 % w/v PEG 1000 (Figure 7.9 B).
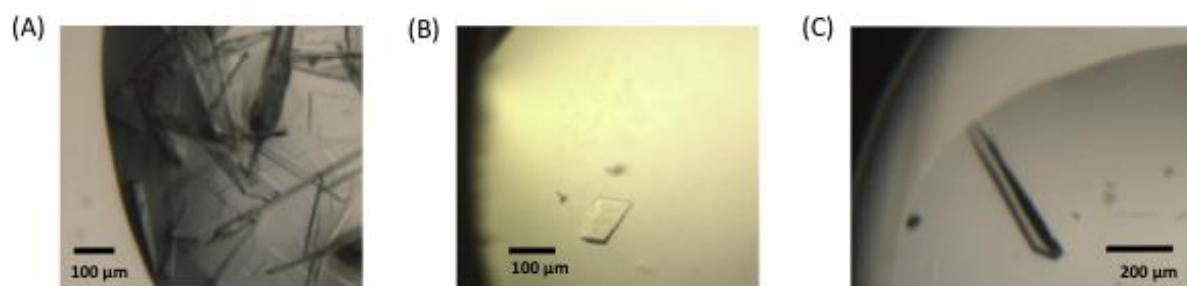
Figure 7.9 - Crystals obtained through *in situ* proteolysis.

(A) Crystallisation conditions: 0.1 M sodium citrate, pH 5.6, 30 % w/v PEG 3350, 1:1000 molar ratio of Chymotrypsin:Protein. (B) Crystallisation conditions: 0.1 M MMT buffer, pH 7, 25 % w/v PEG 1000, Trypsin: Protein in a molar ratio of 1:100. (C) Crystallisation conditions: 0.1 M HEPES, pH 7.5, 0.02 M MgCl$_2$ hexahydrate, 22 % (w/v) Poly(acrylic acid sodium salt) 5100, Chymotrypsin: Protein in a molar ratio of 1:1000.

Crystal trials using chymotrypsin and sPHGDH at a molar ratio of 1:1000 were expanded to include a set of commercial screens with a subset of screens also including fragment E at 5 mM to test the hypothesis that a ligand binding to the protein might exert a stabilising effect and thus facilitate protein crystallisation. Crystals were obtained from different crystallisation conditions in the presence and absence of fragment E (for example Figure 7.9 C). The two best crystals were harvested and measured at DLS synchrotron (Table 7.5).

(A)	sPHGDH 06 with chymotrypsin (1:1000)		(B)	sPHGDH 07 with chymotrypsin (1:1000) and 5 mM fragment E
Wavelength (Å)	0.92		Wavelength (Å)	0.92
Space Group	P 1 21 1		Space Group	P 1 21 1
a, b, c (Å)	42.9, 118.7, 59.2		a, b, c (Å)	43.2, 126.2, 59.7
α, β, γ (°)	90, 100, 90		α, β, γ (°)	90, 101, 90
Resolution (Å)	59.4 – 1.85 (1.89 – 1.85)		Resolution (Å)	58.7 – 2.10 (2.16 – 2.10)
$I/\sigma I$	6.2 (0.8)		$I/\sigma I$	4.6 (0.7)
R_{merge}	0.12 (1.69)		R_{merge}	0.13 (2.26)
CC(1/2)	0.996 (0.487)		CC(1/2)	0.990 (0.381)
Completeness (%)	97.3 (97.5)		Completeness (%)	84.1 (71.4)
Multiplicity	3.9 (3.8)		Multiplicity	3.2 (2.7)

Table 7.5 – Data collection statistics for crystals resulting from *in-situ* proteolysis of sPHGDH with chymotrypsin.

Data were collected from single crystals at 100 K on beamline I04, DLS. The data were processed, scaled and merged using xia2 (3daii-run), aimless, pointless and programs from the ccp4i2 suite. The statistics are presented as averages with values for the highest resolution shell in parentheses. (A) Crystallisation of sPHGDH with chymotrypsin at 1:1000 molar ratio, (B) crystallisation of charcoal-treated sPHGDH with chymotrypsin at molar ratio 1:1000 with 5 mM Fragment E.

Analysis of the crystals obtained from *in situ* proteolysis of sPHGDH with chymotrypsin showed that the crystals still comprised sPHGDH. Although the electron density for the lid domain was very weak, this had been observed in previous crystals of untreated sPHGDH (Figure 7.10). In addition, the crystals measured were still of the same symmetry (space group P 1 2_1 1) as sPHGDH crystals, which might not change for crystals of a proteolysed sPHGDH, but would have been a strong indicator that the crystals were not due to template sPHGDH.
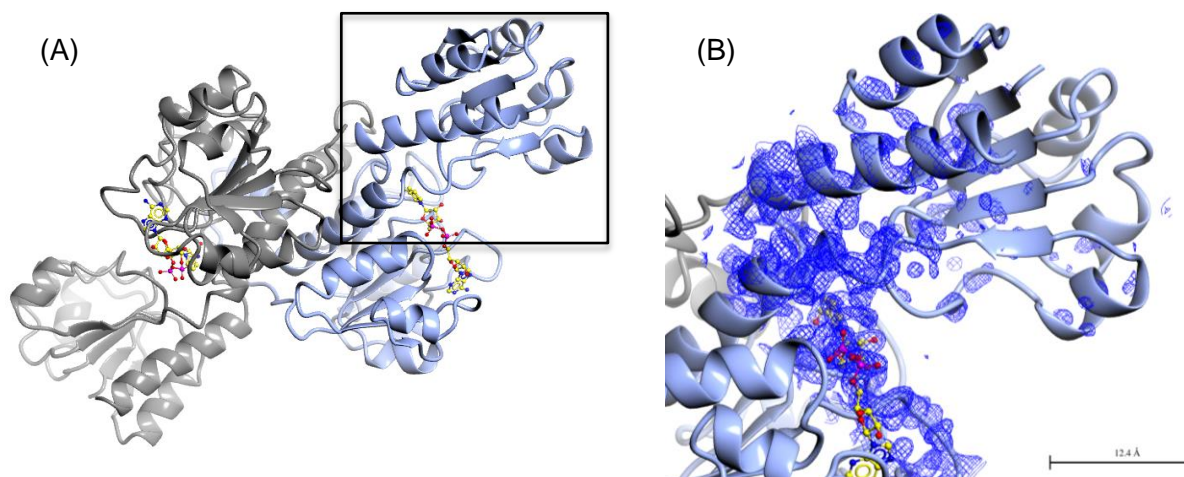
Figure 7.10 – Crystal structure of sPHGDH after *in-situ* proteolysis with chymotrypsin.

(A) Crystal structure of sPHGDH after chymotrypsin treatment with the different chains coloured in grey and blue and the cofactor NAD^+ coloured by atom type with carbon atoms in yellow. (B) Close up view of the region in the black rectangle of (A) with electron density map (2Fo-Fc) shown in blue at a contour level of 0.40 electrons/$\bar{A}^3$.

In parallel, construct-directed proteolysis was performed, whereupon sPHGDH or PHGDH was mixed with five different proteases separately (papain, endoproteinase Glu-C, thermolysin, trypsin and chymotrypsin) in 1:100 and 1:1000 molar ratios. The protein-protease mixtures were incubated at RT or 37 °C for one hour and analysed by SDS-PAGE (Figure 7.11). In general, proteolysis at RT showed evidence of partial cleavage. Treatment of sPHGDH with catalytic amounts of endoproteinase Glu-C or thermolysin at 37 °C yielded a new fragment at about 26 kDa (Figure 7.11 B, lanes 2 and 3). Thermolysin also cleaved PHGDH at 37 °C into two fragments, one at 36 kDa, which is expected to correspond to sPHGDH and one at 26 kDa (Figure 7.11 D, lane 3).

Figure 7.11 - Limited proteolysis of sPHGDH (A and B) and PHGDH (C and D) at RT (A and C) or 37 °C (B and D).

100 µg or 50 µg sPHGDH (A and B) or PHGDH (C and D) respectively was mixed with five different proteases at molar ratios of 1:100 and 1:1000. The digestion was performed for 1 h at RT (A and C) or 37 °C (B and D) and analysed by SDS-PAGE.

In order to identify the new fragment obtained through thermolysin treatment, a fresh sample of sPHGDH was mixed with thermolysin at 1:100 molar weight ratio and the reaction left to proceed at 37 °C for 1 hour. As the sample contained only trace amounts of thermolysin it was directly subjected to ESI-MS (Biomolecular Mass Spectrometry Facility, Leeds University) (Figure 7.12 D). Through mass spectrometry analysis under denaturing conditions, a major protein fragment of 22116.93 Da was identified together with two minor species of 22301.00 and 22528.31 Da (Figure 7.12A and B). The mass of the fragments together with the known cleavage sites of thermolysin, which cleaves before bulky and aromatic residues such as Ile, Leu, Val, Ala, Met and Val, but not after acidic amino acids such as Asp or Glu, allowed the potential amino acid sequences of the new PHGDH fragments to be narrowed down to five potential fragments (Table 7.6) [194].

Figure 7.12 - Identification of new PHGDH fragment 93.

(A) Positive ESI-MS m/z spectrum of new fragment 93, (B) Molecular mass profile of m/z spectrum shown in (A), (C) Amino acid sequence of sPHGDH (with His_6-tag) with potential thermolysin cleavage sites after the amino acids highlighted in red. (D) SDS-PAGE analysis of thermolysin treated His_6-sPHGDH sample used for mass spectrometry.

ID	Residues (aa #)	Corresponding mass (Da)	Deviation from measured mass
93	93-298	22116.60	-0.33
22	22-227	22118.66	1.73
92-2	92-299	22300.83	-0.17
94	94-301	22301.78	0.78
92-3	92-301	22528.10	-0.21

Table 7.6 – Potential PHGDH fragments concordant with mass spectrometry data and thermolysin cleavage sites.

7.4.4 Crystallisation of construct 93

Cloning, expression and purification

Of the five potential PHGDH fragments, three, namely 93, 92-2 and 92-3 were chosen for construct design and expression, each of them being closely associated with one of the masses detected.

In order to keep the expressed sequence in the pNIC28-BsaI vector, two rounds of PCR-based mutagenesis were performed with the first round generating the N-terminal truncation and the second round the C-terminal truncation as described in section 7.3.4. The first construct used for transformation into _E. coli_ Rosetta (DE3) was fragment 93 as this corresponded to the major species detected by mass spectrometry.

Fragment 93 was purified by Ni-affinity chromatography and the His_6-tag cleaved with TEV protease. Although the cleaved protein was further purified by size exclusion chromatography, two bands were evident on SDS-PAGE analysis. Therefore an additional anion exchange purification step was performed and this yielded pure protein (Figure 7.13). The measured OD_{260}/OD_{280} ratio of 0.71 for purified protein 93 indicated that protein 93 co-purified with the cofactors NAD^+/NADH, but to a lesser extent than sPHGDH, which had an OD_{260}/OD_{280} ratio of 0.8-0.9. Charcoal treatment of purified 93 performed as described for sPHGDH decreased the OD_{260}/OD_{280} ratio to 0.6, very close to the predicted ratio of 0.57 for pure protein.
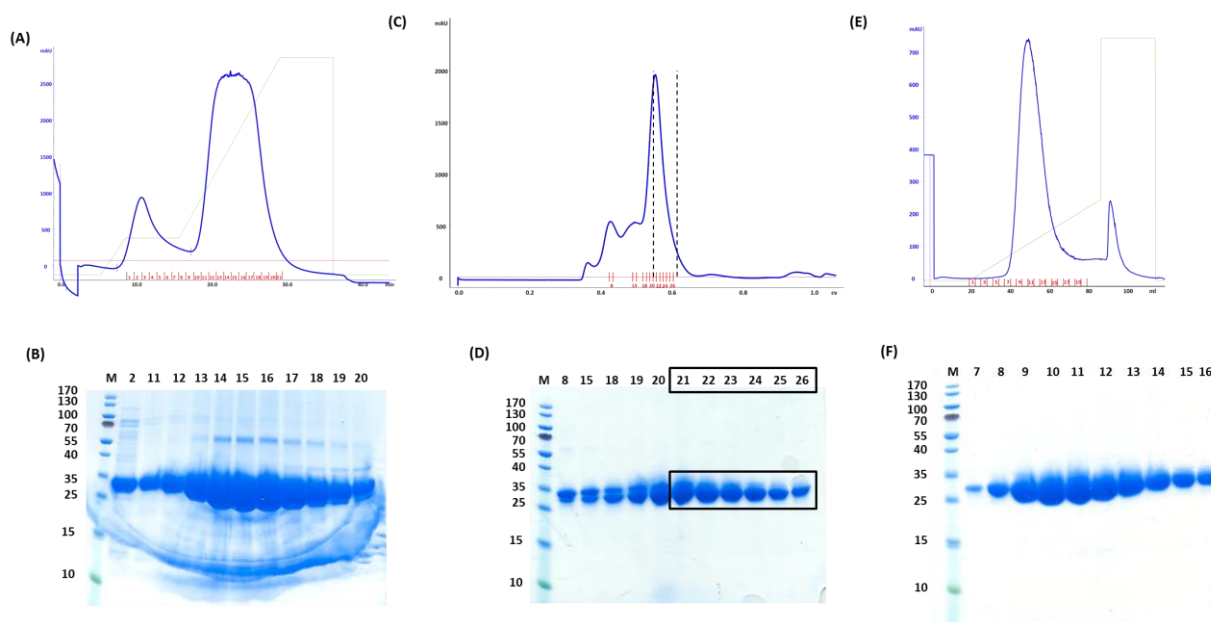
Figure 7.13 - Purification of protein 93.

Protein construct 93 was purified from *E. coli* Rosetta (DE3) cells. After cell lysis, 93 was purified by Ni-affinity chromatography (A and B), followed by size exclusion chromatography (Superdex 75 26/60) (C and D) and anion exchange chromatography (E and F). A representative purification is shown with the UV chromatograms of the respective purification steps together with the SDS-PAGE analysis of the eluted fractions.
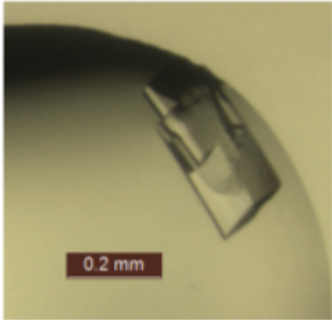
Crystallisation of 93

Crystallisation of charcoal-treated fragment 93 alone or in combination with fragment E or 3-indole-*DL*-aspartic acid (compound 3, section 4.4.2) was attempted using sparse matrix screens at 4 °C or 20 °C. Crystals were obtained from a variety of conditions and at both temperatures. Of nine crystals tested, diffraction data could be collected from five crystals and therefore those conditions were taken forward for optimisation trials at 20 °C (Figure 7.14). From the optimisation process, condition consisting of 0.1 M PCTP buffer, pH 7.0, 22-26 % (w/v) PEG 1500, gave the most robust crystals to use for soaking experiments or to set up co-crystallisation experiments.

Crystal ID	Reservoir composition
93-1	0.1 M SPG buffer, pH 8.0, 25 % (w/v) PEG 1500
93-2/3	0.1 M MIB buffer, pH 7.0, 25 % (w/v) PEG 1500
93-4	0.1 M PCTP buffer, pH 7.0, 25 % (w/v) PEG 1500
93-5	0.1 M MMT buffer, pH 7, 25 % (w/v) PEG 1500
93-6	0.1 M Sodium citrate, pH 5.5, 20 % (w/v) PEG 3000
93-7	0.1 M Citrate, pH 5.0, 20 % (w/v) PEG 6000
93-8	24 % (w/v) PEG 1500, 20 % (v/v) Glycerol
93-9	0.1 M MES buffer, pH 6.5, 15 % (w/v) PEG 6000, 5 % (v/v) MPD

(B)

Figure 7.14 – Crystallisation of construct 93

(A) Reservoir compositions giving rise to crystals of 93. (B) Crystal 93-3.

In general, crystals of construct 93 diffracted to high resolution of 1.2 Å, and after crystal optimisation the crystal quality was improved further, so that for nearly every crystal tested a dataset could be collected. Crystals of 93 belonged to spacegroup P1, whereas sPHGDH was mostly found to crystallise in spacegroup P 1 2_1 1. Among the first crystals of construct 93 to be measured was a co-crystal with 5 mM NAD$^+$ (crystal 93-10) for which data collection, processing, refinement and model validation statistics are shown as a representative example (Table 7.7).

(A)	Data collection and phasing	
Wavelength (Å)	0.92	
Space Group	P 1	
a, b, c (Å)	43.4, 50.3, 53.2	
α, β, γ (°)	98, 113, 104	
Resolution (Å)	39.0-1.35 (1.37-1.35)	
I/σI	6.5 (1.1)	
CC(1/2)	0.992 (0.517)	
R_{merge}	0.06 (0.73)	
Completeness (%)	87.4 (86.3)	
Multiplicity	2.0 (2.0)	

(B)	Refinement	
R_{work}/R_{free}	0.18/0.21
Mean B-value (Å^2)	
All atoms	37
Protein	32
Water	46
NAD$^+$	32
No of atoms	
Protein	5884
Water	242
NAD$^+$	66
Rmsd	
Bond lengths (Å)	0.02
Angles (°)	1.88

(C) Molprobity Parameters	Value
Clashscore, all atoms	5.76[a]
Poor rotamers	0.94 %
Ramachandran outliers	0.0 %
Ramachandran favoured	97.5 %
No of Cβ deviations (> 0.25 Å)	1.1 %
Molprobity Score	1.42[b]
Residues with bad bonds	0.20 %
Residues with bad angles	0.02 %

Table 7.7 – Data collection, processing, refinement and model validation statistics for the co-crystal structure of 93 with NAD$^+$ (93-10).

(A) Data collection statistics for 93-10, collected on a single crystal at 100 K on beamline I04-1, DLS. The data were processed, scaled and merged using xia2, aimless, pointless and programs from the CCP4i2 suite. Values for high resolution shell in parentheses. (B) Refinement statistics generated with Refmac5 as part of CCP4i2 suite. (C) Validation statistics generated by MolProbity web server. The clashscore[a] was 84th percentile and the MolProbity score[b] was 87th percentile for a structure of 1.35 ± 0.25 Å.

The co-crystal structure of 93 and NAD$^+$ was used to compare the overall structure of 93 to that of sPHGDH as well as the NAD$^+$ binding site (Figure 7.15). Despite the

missing lid domain, construct 93 folded properly and overlaid well onto sPHGDH. The binding site for NAD$^+$ remained unchanged in 93 with NAD$^+$ making the same interactions with the protein as with sPHGDH, apart from new hydrogen bonds between Arg235 and one ribose. This new hydrogen bond was possible, as the side chain of Arg235 adopted a different conformation in crystal 93, probably because amino acids from the lid domain, which formed interactions with Arg235 in crystal sPHGDH, were missing in 93.
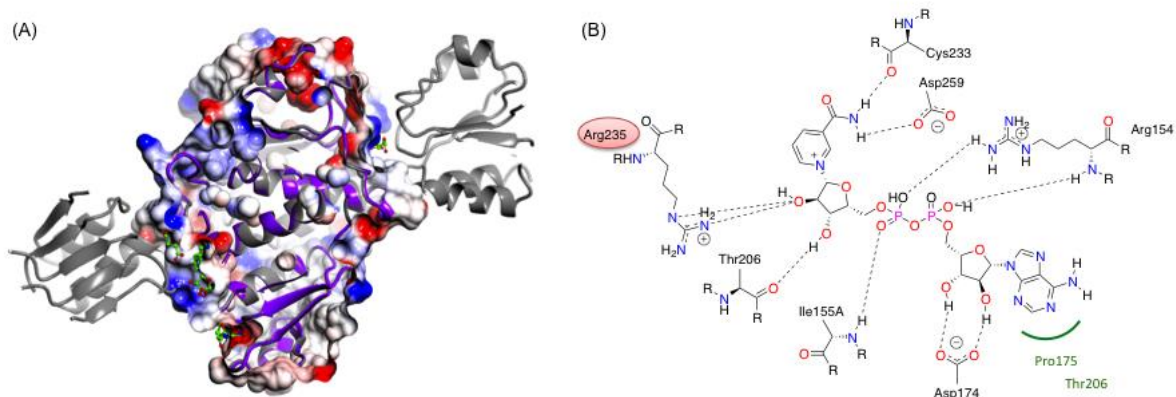
Figure 7.15 – Comparison of cofactor binding sites in crystals of sPHGDH and 93.

(A) Ribbon diagram of 93 (purple) superposed on sPHGDH (grey). The surface of 93 is shown and coloured by electrostatic potential and outlines the domain common to 93 and sPHGDH. (B) 2D image of cofactor binding site of 93 showing the interactions of NAD$^+$ and 93.

7.4.5 Soaking of fragments into crystals of 93

Although construct 93 lacks the lid domain that covers the substrate-binding site in sPHGDH, the good quality diffraction of the 93 crystals was ideal for use in soaking experiments. In addition, the crystals tolerated up to 20 % (v/v) DMSO, which allowed complete dissolution of the majority of the fragments investigated. Furthermore, structures of apo-crystals confirmed that the cofactor-binding site was free of cofactor and therefore accessible for binding of potential ligands.
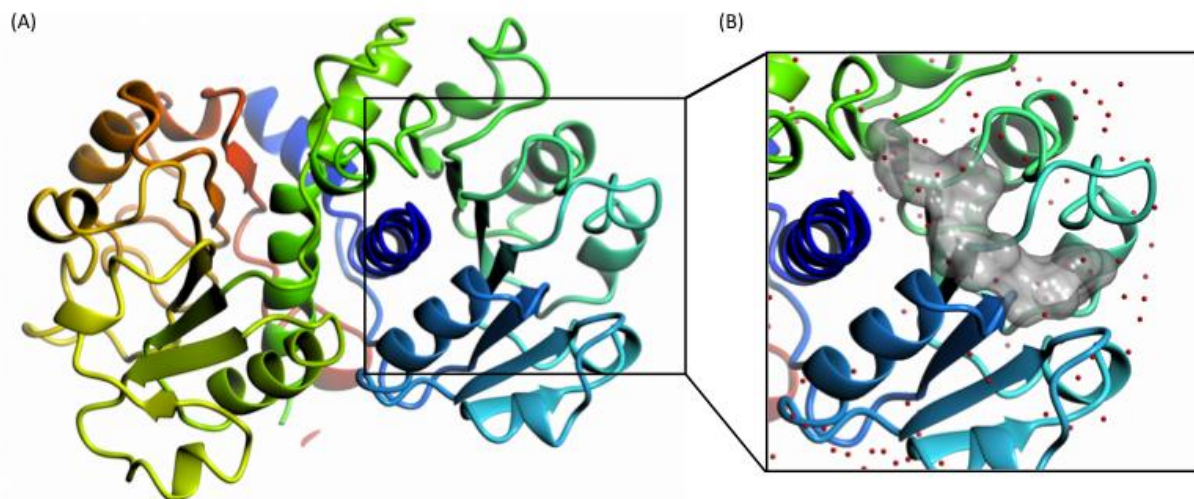
Figure 7.16 – Crystal structure of apoenzyme 93.

(A) Apo-structure of 93 coloured as rainbow model with the C-terminus in red and the N-terminus in blue. (B) Focus on cofactor-binding site of 93 showing the hypothetical position of NAD^+ in grey and the actually positioned water molecules in red. Position of cofactor was estimated through overlay of crystal 93 with NAD^+ bound onto apo-structure.

Purchased fragments **A-O** were dissolved at 80 mM in soaking buffer, which contained 20 % DMSO to increase solubility of fragments and potentially function as a cryo-protectant. As 20 % (v/v) DMSO was found to be an insufficient amount of cryo-protectant, 15 % PEG 400 was added to the soaking buffer. The fragments were then soaked individually into crystals for 24 hours at 20 °C. Under these conditions, 7 out of the 15 fragments investigated were found in the respective crystals (Figure 7.17). All of the fragments bound to the adenine-pocket of the cofactor-binding site. A subset formed specific hydrogen bonds with the protein, such as fragment N that formed a hydrogen bond with a backbone amine of amino acid residue Gly151 (Figure 7.17).
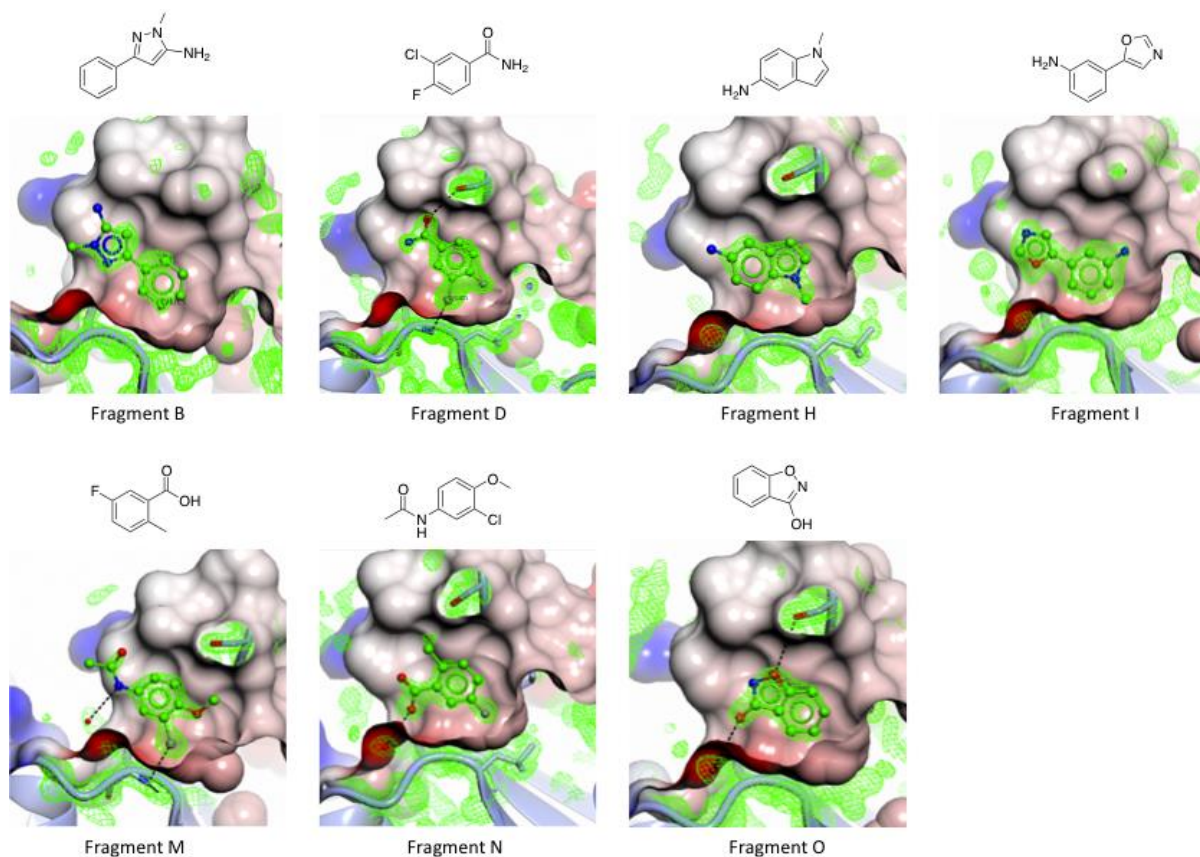
Figure 7.17 – Fragments detected after soaking into crystals of construct 93

Chemical structure and position in the adenine-binding site of 93 are shown. The protein surface around the binding site is coloured by electrostatic potential: red is negatively charged and blue is positively charged. Ligands are coloured by atom type with green = carbon, red = oxygen, blue = nitrogen and halogen = grey. The electron density maps (2Fo-Fc) is shown in green at a contour level of 0.5-0.75 electrons/Å^3.

7.4.6 Fragment optimisation

Of the fragments that could be soaked into crystals of 93, a subset was taken forward for optimisation. Commercial suppliers were sought for analogues of the hit fragments that would allow further exploration of the binding pocket. Additionally, some initial structure-activity relationship (SAR) studies were initiated around a number of scaffolds as described in the following sections.

Benzamide series

Among the purchased fragments for which binding was observed in crystals of 93, three fragments (fragments **D**, **M**, **N**) were substituted benzene rings. Overlay of these three fragments showed that the benzene ring of fragments **N** and **M** bound in the same part of the adenine-site, whereas fragment **D** bound more deeply into the

pocket and thus filling a small hydrophobic pocket that was not accessed by fragments **M** and **N** (Figure 7.18 A). Of the three fragments with a substituted benzene ring, 3-chloro-4-fluorobenzamide (fragment **D**) and *N*-(3-chloro-4-methoxyphenyl)acetamide (fragment **N**) had similar binding affinities for sPHGDH according to ITC experiments. Fragment **N** was chosen for further investigation, as the 4-methoxy-group was oriented favorably for growth in an additional direction, compared to fragment **D**. Fragment **N** formed one halogen and one hydrogen bond, the first one through its chloro-substituent with the protein backbone (Gly151) and the latter one through the NH of the amide group with a water molecule present in the pocket (Figure 7.18 B).
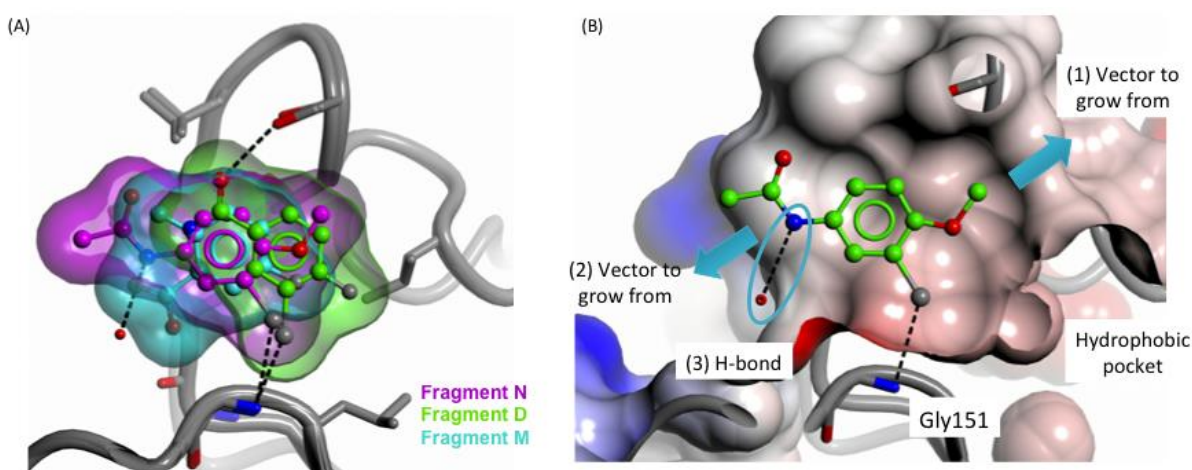
Figure 7.18 – Evaluation of the binding mode of N-(3-chloro-4-methoxyphenyl)acetamide (fragment N)

(A) Overlay of fragments **N** (magenta), **D** (green) and **M** (cyan) in the adenine-binding site of 93. (B) Binding mode of fragment N to the cofactor-binding site with the fragment coloured by atom type. The protein backbone is shown in grey with the protein surface coloured by electrostatic potential. Hydrogen and halogen bonds are shown as dashed lines.

Starting from fragment **N**, three compounds were initially investigated: the *N*-methylated form of *N*-(3-chloro-4-methoxyphenyl)acetamide (**27**) to evaluate the importance of the hydrogen bond, and two compounds with larger alkyl ether functions (Et and iPr) in the 4-position of the benzenecarboxamide (**28, 29**) to start growing further out of the pocket (Table 7.8). N-methylation resulted in a three-fold decrease in K_D and the extended alkyl ether functional groups further decreased affinity for sPHGDH. However, in order to fully understand the influence of the ethoxy- and isopropoxy-groups, the respective matched pairs, i.e. the 3-chloro

analogues of **28** and **29** and/or the des-chloro analogue of **N** would need to be made and tested.
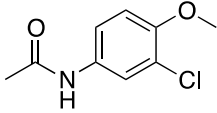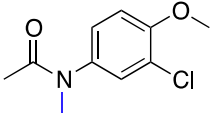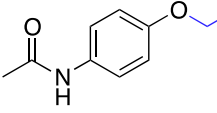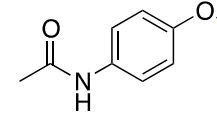
ID	N	27	28	29
Structure				
K_D (sPHGDH)	2.8 ± 0.7	8.3 ± 4.4	18.8 ± 14.2	21.1 ± 21.8
LE	0.27	0.20	0.18	0.16

Table 7.8 – Analogues of *N*-(3-chloro-4-methoxyphenyl)acetamide (fragment N)

Isoxazole series

The initial fragment hit, 3-hydroxybenzisoxazole (fragment **O**), was found to orient in the adenine-binding pocket in such a way that fragment growth in two directions could be initiated by either growing along the cofactor-binding pocket (2) or out of the pocket (1) (Figure 7.19). In addition, the fragment formed two hydrogen bonds with side chains of Ser211 and Asp174, which ideally should be retained during fragment growing. A small hydrophobic pocket (3) adjacent to the fragment was not occupied and could be filled with a small group by substitution on the 5-position of the benzisoxazole so as to increase interactions with the protein (Figure 7.19).
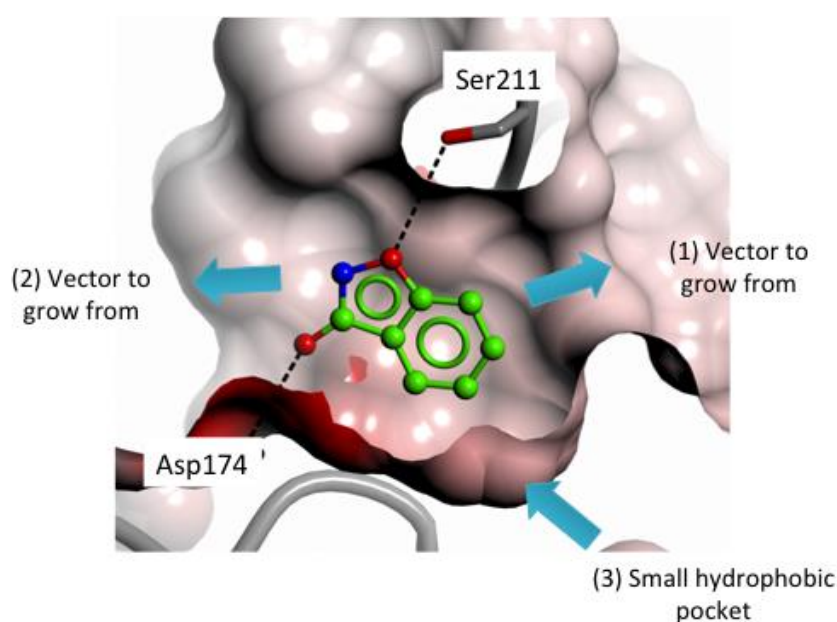
Figure 7.19 – Evaluation of binding mode of 3-hydroxybenzisoxazole (fragment O).

The alpha trace of the protein is shown in grey ribbons with amino acids involved in hydrogen bonds and the bound fragment shown as sticks and coloured by atom type. The protein surface is coloured by electrostatic potentials and hydrogen bonds are indicated by dashed lines.

From an analogue by catalogue search, the 6-chloro (**O2**) and 7-methyl (**O1**) substituted 3-hydroxybenzisoxazoles were chosen to probe fragment growth in the direction of vector (1) (Figure 7.19). Both fragments were investigated by competitive ITC with fragment **O1** showing weaker and **O2** improved binding affinities for sPHGDH compared to the unsubstituted 3-hydroxybenzisoxazole (Table 7.9). Crystal structures of the fragments showed that fragment analogue **O1** overlaid in principle with the original fragment **O**. However, small shifts in the binding mode resulted in the formation of different hydrogen bonds with the protein. The hydroxyl-group of

fragment **O1** formed a hydrogen bond with Thr206 whereas in the parental fragment **O** the same group formed a hydrogen bond with Ser211 (Figure 7.20 A and B). In contrast, fragment **O2** bound in a different orientation so that the chloro group in the 5-position filled into a small hydrophobic pocket instead of providing a vector to grow out of the pocket (Figure 7.20 C).
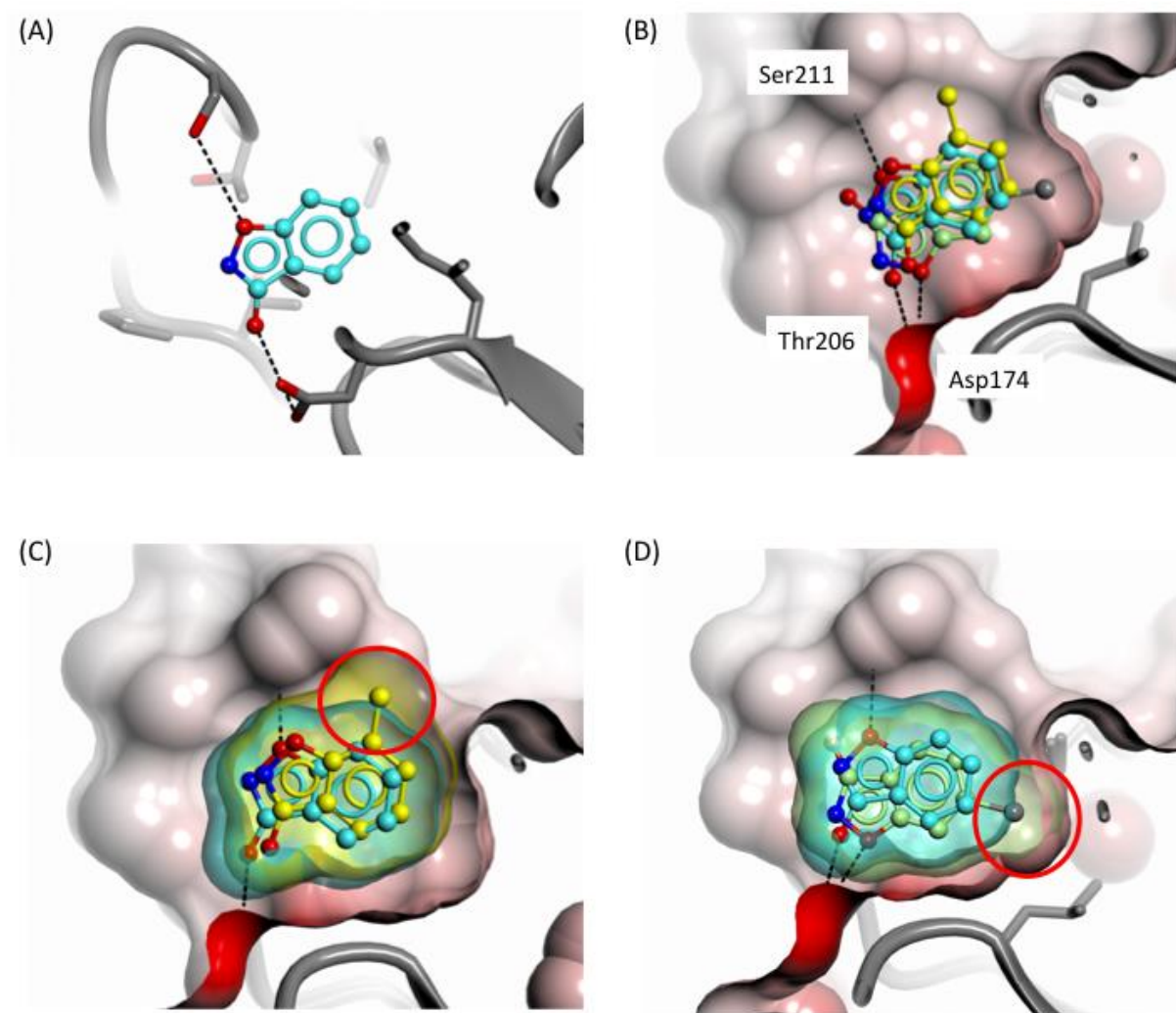
Figure 7.20 – Binding of 3-hydroxybenzisoxazole and analogues with substituents at positions 5 and 7.

Binding of fragments **O** (cyan), **O1** (yellow) and **O2** (light green) to the adenine-binding pocket of 93. (A) Fragment **O**, (B) Fragment **O**, **O1** and **O2**, (C) Fragment **O** and **O1**, (D) Fragment **O** and **O2**. The protein is shown as ribbon diagram in grey together with the protein surface coloured by electrostatic potential. The compound surfaces are coloured according to the respective fragment. Hydrogen bonds are shown as dashed lines. Red circles in pictures (C) and (D) highlight the position of the introduced side chains compared to parental fragment **O**.

Investigation of the two analogues showed that the additional methyl-group in the 7-position decreased binding affinity resulting in a lower ligand efficiency (LE) as well.

On the other hand the chloro-group in the 6-position increased binding affinity by filling the adjacent hydrophobic pocket, and thereby improved LE (Table 7.9).

ID	O	O1	O2
Structure			
K_D (sPHGDH) [mM]	6.5 ± 4.7	15.9 ± 7.4	2.4 ± 0.7
LE	0.30	0.22	0.33

Table 7.9 – Binding affinity of 3-hydroxybenzisoxazole analogues O1 and O2 as determined by displacement ITC.

The previously mentioned small lipophilic pocket (3) in Figure 7.19, which was not occupied by the original fragment **O**, was occupied by fragment **O2** by virtue of binding to the adenine-site in a different orientation from the unsubstituted parental compound. In order to fill the hydrophobic pocket, whilst retaining the binding mode of fragment **O**, a substituent in the 5-position would be needed. Therefore 5-fluoro-3-hydroxybenzisoxazole (**O5**) was investigated. Although the binding affinity decreased compared to 3-hydroxybenzisoxazole the crystal structure showed that the fluoro-group was binding in the expected hydrophobic pocket (Figure 7.21). Following on from this result, the 5-bromo-3-hydroxybenzisoxazole (**O6**) was also investigated to see if a bigger functional group would be equally well tolerated in the hydrophobic pocket. Indeed, **O6** bound in the expected orientation, although due to the increased size of the bromo-group the compound was shifted compared to **O5** (Figure 7.21 B). Through this shift, the hydroxyl-group of fragment **O6** was brought in proximity to Asp174 and the backbone of His205 so that two additional hydrogen bonds were formed (Figure 7.21 C).
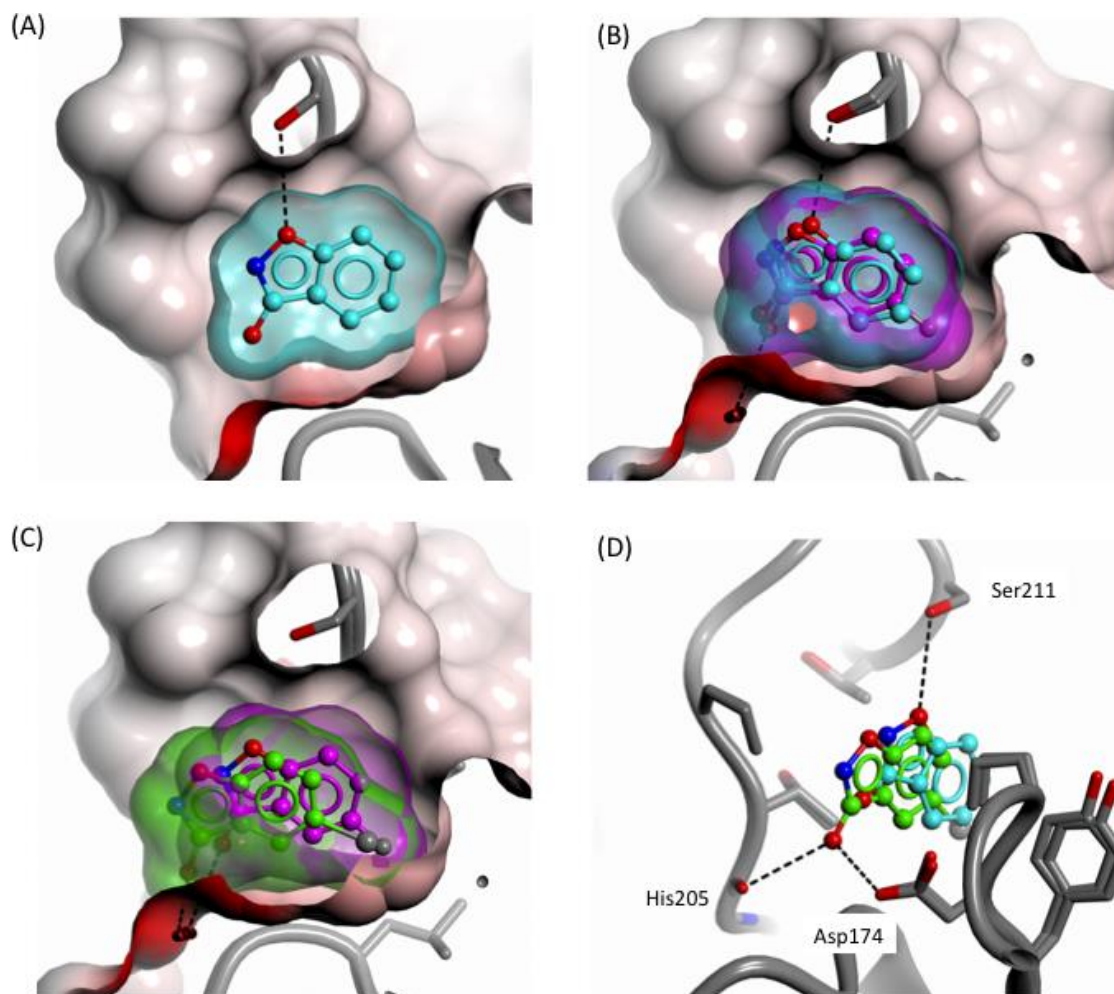
Figure 7.21 – Binding mode of 5-fluoro/bromo-3-hydroxybenzisoxazole to adenine-binding pocket.
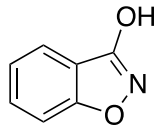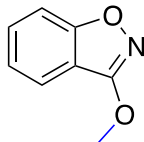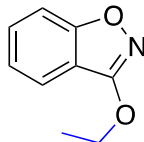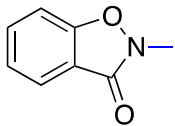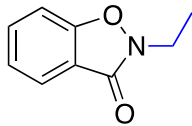
Binding of fragments **O** (cyan), **O5** (magenta) and **O6** (green) to the adenine pocket of 93. (A) Fragment **O**, (B) Fragment **O** and **O5**, (C) Fragment **O5** and **O6**. The protein is shown as ribbon diagram in grey together with the protein surface coloured by electrostatic potential. The compound surfaces are coloured according to the respective fragment. (D) Hydrogen bond formation of fragment **O5** and **O6** with PHGDH 93. The protein is shown as ribbon diagram together with the amino acids around the adenine pocket as sticks (grey) and fragment **O5** and **O6** are coloured by atom type with carbons in cyan and green, respectively. Hydrogen bonds are shown as dashed lines.

ID	O	O5	O6
Structure			
K$_D$ (sPHGDH) [mM]	6.5 ± 4.7	15.4 ± 10.2	Precipitated
LE	0.30	0.22	-

Table 7.10 - Binding affinity of 3-hydroxybenzisoxazole analogues O5 and O6 as determined by displacement ITC.

To probe for growth in the direction of vector (2), as well as the importance of the hydrogen bond to Ser211, 3-alkylated hydroxybenzisoxazoles were synthesised. 3-Methoxy- (**25**) and 3-ethoxyisoxazole (**26**) were synthesised by etherification using alkyl iodide with a moderate yield of 43 and 51 %, respectively. The synthesis also yielded the alkylated amines as side products, and these were tested as control compounds. The alkoxy-compounds showed reduced or no binding affinity for sPHGDH and none of the alkylated compounds were seen to bind in the crystal structure, indicating that hydrogen bond formation with Asp174 is indeed important for the benzisoxazole binding (Table 7.11). Interestingly, of the alkylated amines only the ethylated amine **26-sp** was able to bind to sPHGDH.

ID	O	25	26	25-sp	26-sp
Structure					
K_D (sPHGDH) [mM]		No competition.	Precipitated.	No competition.	10.6 ± 8.2

Table 7.11 – Binding affinities of alkylated 3-hydroxybenzisoxazole fragments.

As the hydrogen bond donor and acceptor attributes of the hydroxyl function were important for binding of fragments with the benzisoxazole core, 3-aminobenzo[*d*]isoxazole was investigated. The amine provided a potential starting point for growth in the direction of vector (2) allowing retention of the H-bond to Asp174, while growing from this position (compound **O4**). 3-Aminobenzo[*d*]isoxazole was an equally potent ligand for sPHGDH as the hydroxyl analogue (fragment **O**), and both compounds bound in the same orientation to the adenine-binding pocket (Figure 7.22).
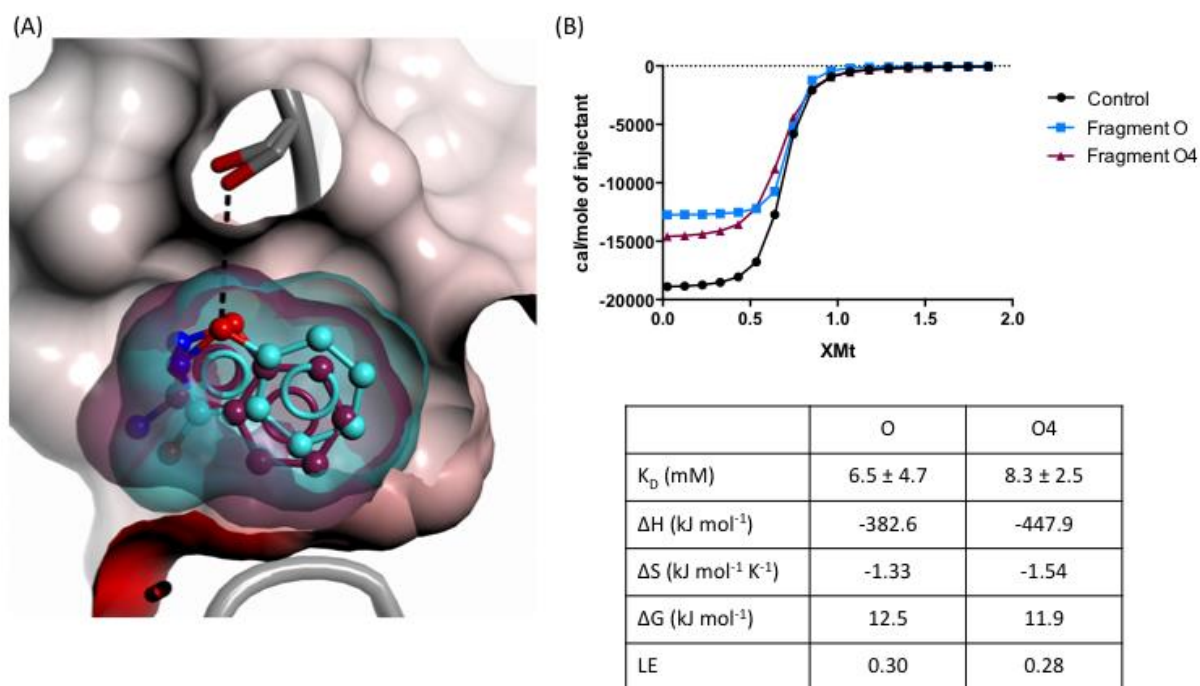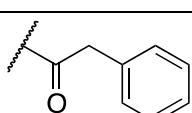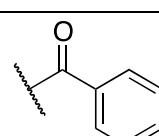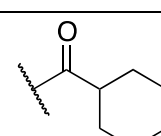
Figure 7.22 – Binding mode of 3-aminobenzo[d]isoxazole

(A) Overlay of 3-amino (purple) (**O4**) and 3-hydroxybenzo[*d*]isoxazole (cyan) (**O**) bound to the adenine-pocket of 93. (B) ITC competition experiment and derived thermodynamic parameters.

The 3-amine function allowed for further growth of the fragment using different peptide-coupling methods. Unfortunately none of the compounds in this small amide library produced K_D values in the ITC competition experiment due to poor solubility and likewise compounds were not observed in the crystals after soaking for 24 hours (Table 7.12).

ID	17	18	19
R			
Yield (%)	20	24	76

Table 7.12 – Small amide library of 3-aminobenzisoxazole.
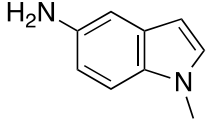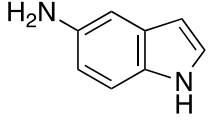
Indole, which is structurally similar to 1,2-benzisoxazole was also among the scaffolds identified in the fragment hits. 5-Amino-1-methyl-*1H*-indole was purchased and found to be significantly more effective in stabilizing sPHGDH (ΔT_m = 1.2 ± 0.4 °C) compared to all the other purchased fragments, but could not be investigated by competitive ITC due to poor solubility. Investigation of the binding mode in the crystal showed that 5-amino-1-methyl-*1H*-indole did not form any specific polar binding interactions with the protein. It was however in proximity to Ser211 such that the OH-group could be a potential hydrogen bond donor to target (Figure 7.23). Interestingly, the position of the indole-fragment in the adenine-binding site was rotated 180° compared to the 3-hydroxybenzisoxazole (Figure 7.23).
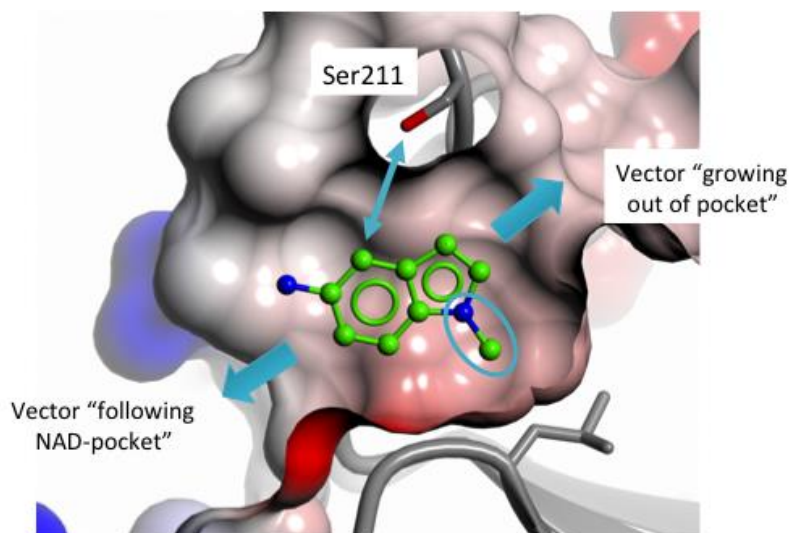
Figure 7.23 - Evaluation of binding of 5-amino-1-methyl-1H-indole (fragment H) to 93.

The protein backbone is shown in grey as a ribbon diagram with amino acids involved in hydrogen bonds shown as sticks and coloured by atom type with carbons in green. The protein surface is coloured by electrostatic potential.

To probe the importance of the N-methylation, the unmethylated 5-amino-indole was investigated. The orientation of the des-methyl analogue **H5** could not be clearly determined as the electron density for the fragment observed in crystals of 93 was too weak.

ID	H	H5
Structure	H₂N — (5-amino-1-methylindole structure)	H₂N — (5-aminoindole structure)
K_D (mM)	11.1 ± 5.5	No competition.
LE	0.24	-

Table 7.13 – Binding affinities of methylated and non-methylated 5-aminoindole.

3- and 5-substituted indoles were investigated in order to probe the two vectors to grow out of the adenine-binding site. 3-Hydroxyindole (**H1**) and indole-3-acetic acid (**H2**) showed an orientation flipped by 180° along a vertical axis compared to fragment H (Figure 7.24 A). In this orientation, both compounds are able to form hydrogen bonds with side chains of the protein.
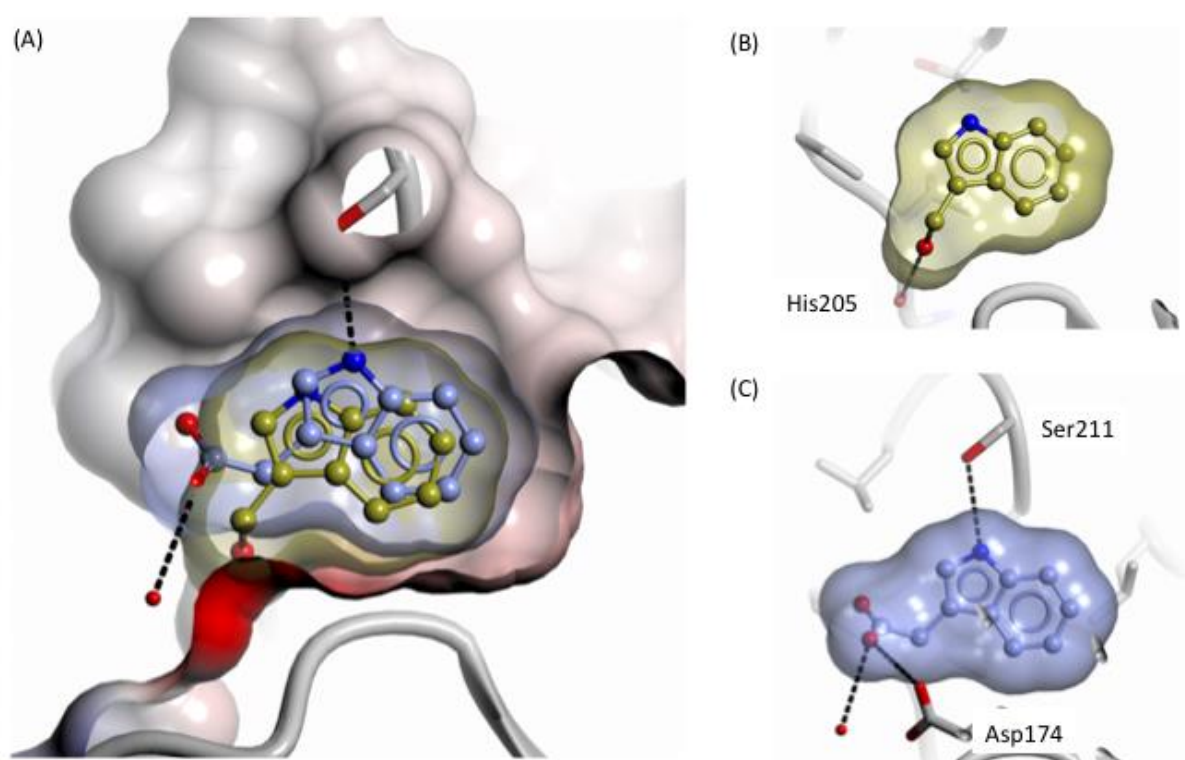
Figure 7.24 – Binding mode of fragments H1 and H2 to 93.

(A) Binding of fragments **H1** (gold) and **H2** (blue) to the adenine-binding pocket of 93 (shown in grey, protein surface coloured by electrostatic potential). Hydrogen bond interactions of (B) fragment **H1** and (C) **H2** with the protein. Hydrogen bonds are shown as dashed lines.

ID	H1	H2	H3
Structure	(3-indolylmethanol structure, CH$_2$OH)	(indole-3-acetic acid structure)	(5-acetamide-1H-indole structure, H$_2$N–C(=O))
K_D (mM)	5.1 ± 2.3	8.6 ± 5.8	Precipitated (ITC).
LE	0.28	0.22	-

Table 7.14 – Binding affinities of 3- and 5- substituted indoles.

Fragment H3 could not be investigated by ITC due to precipitation in the buffer conditions used. Soaking of H3 into crystals of 93 revealed a binding mode flipped by 180° around a horizontal axis when compared to fragment H (Figure 7.25 A). Together with the binding mode seen for H1 and H2, this indicates that the electron-rich hydrogen bond donor and acceptor groups within these compounds preferentially bind towards the region of the NAD-binding pocket normally occupied by the ribose moiety (Figure 7.25 B).
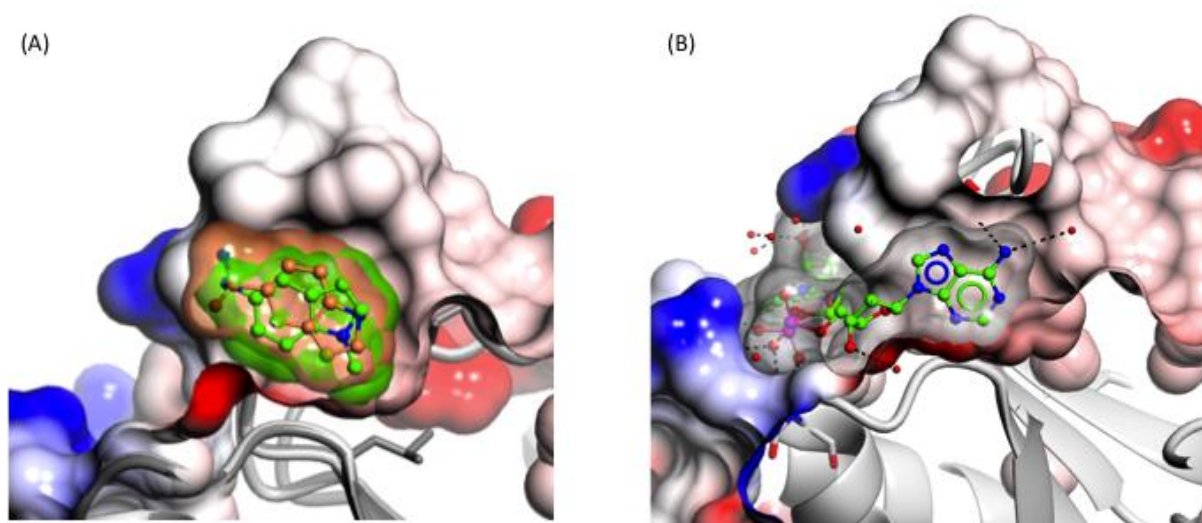
Figure 7.25 – Binding mode of 5-acetamide-1*H*-indole (H3) to 93.

(A) Binding of fragment **H3** (orange) and **H** (green) to the adenine-binding pocket of 93 (shown in grey, protein surface coloured by electrostatic potential). (B) NAD$^+$ (coloured by atom type with carbons in green) bound to the cofactor-binding site of 93. Water molecules are shown as red spheres and H-bonds as dashed black lines.

Apart from growing the initial fragment H further out of the pocket, an attempt was made to grow the fragment towards Ser211 and to provide a functional group for

hydrogen bond formation with the side chain of Ser211. Initially, 4-vinyl-1*H*-indole (**21**) was synthesised and soaked into crystals of 93 to see if the vinyl-group would orient towards the targeted amino acid. In addition, the carbon-carbon double bond allowed for derivatisation, such as oxidation to yield the diol (**23**). 4-Vinyl-1*H*-indole bound in the crystal with the vinyl group pointing towards Ser211, whereas the saturated analogue (**22**), which was prepared as a control compound, bound in a different orientation. The vinyl group was 2.7 Å from Ser211 so that oxidation of the alkene would place the terminal hydroxyl group approximately 1.4 Å closer to Ser211 and hydrogen bonding distance. Unfortunately, compound **23** could not be observed to bind upon soaking into crystals of 93.
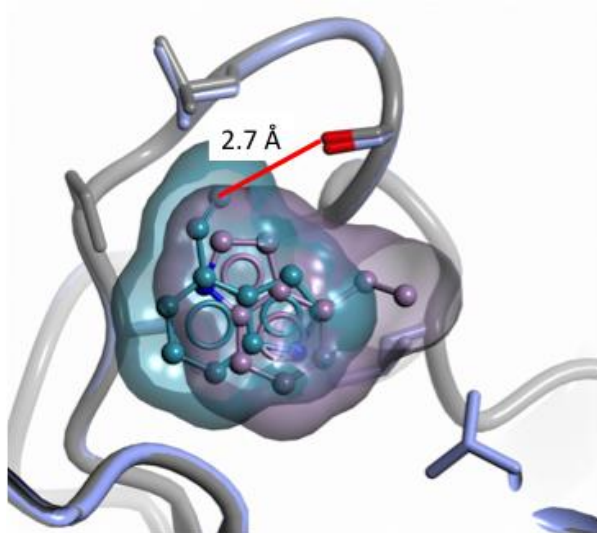
ID	21	22	23
Structure			
Yield (%)	57	52	57
K_D (mM)	precipitated	precipitated	2.7 ± 1.9
LE	-	-	0.27
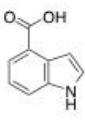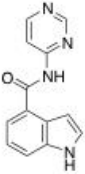
Figure 7.26 – Binding of compounds 21 and 22 to 93.

Binding of compounds **21** (blue) and **22** (purple) to the adenine-binding pocket of 93 (shown in grey and blue). Distance to Ser211 is indicated with a red line.

Similar to compound **23**, it was hypothesised that an indole with a carboxamide function in the 4-position would be able to form a hydrogen bond with Ser211 and additionally allow for further growth (Figure 7.27). Compound **24** was synthesised containing an aromatic ring bound to the carboxamide nitrogen. Due to the solubility issues previously observed with compounds synthesised in the benzisoxazole series containing a benzene ring, a pyrazine was investigated. However, compound **24** was not observed to bind into crystals of 93, despite prolonged soaking times of up to 7 days.

Figure 7.27 – Potential binding mode of compound 24 to PHGDH construct 93.

Compound **24** was modelled into the binding site of 93 through overlay onto the parent fragment, **H**. The preferred torsion angles of compound **24** were validated in Mogul [195] through comparison to similar compounds in small-molecule crystal structures.
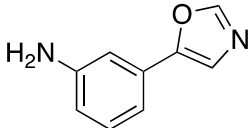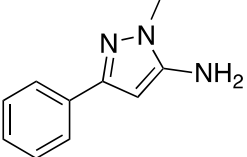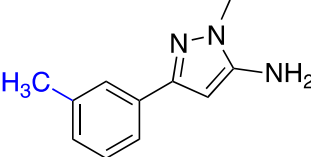
Among the commercially sourced fragment hits were two unfused heterocyclic systems consisting of a benzene ring linked to a five-membered unsaturated ring (pyrazole and oxazole). Both compounds were equipotent in the ITC competition experiment, and bound in the same orientation to 93 (Figure 7.28, Table 7.15). For compound **B**, an amide group provided one potential vector for growth out of the binding pocket.
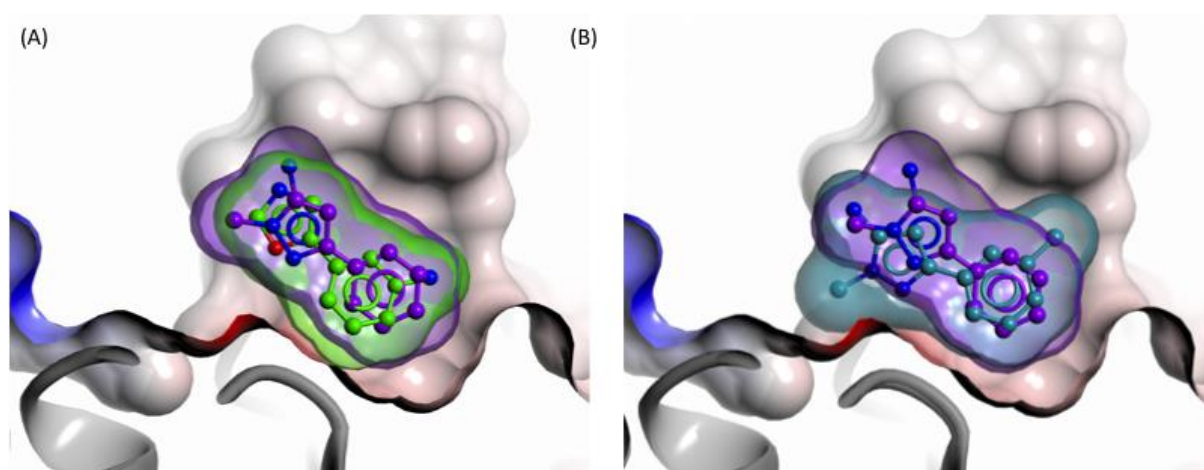
Figure 7.28 – Evaluation of binding of 3-(1,3-oxazol-5-yl)aniline (I) and 1-methyl-3-phenyl-*1H*-pyrazol-5-amine (B) to 93.

(A) Overlay of 3-(1,3-oxazol-5-yl)aniline (fragment **I**, green) and 1-methyl-3-phenyl-*1H*-pyrazol-5-amine (fragment **B**, purple) bound to the adenine-pocket of 93. (B) Overlay of fragment **B** (purple) and the methyl-substituent containing analogue **B1** (blue) bound to the adenine-pocket of 93.

ID	I	B	B1
Structure			
K_d (sPHGDH) [mM]	9.3 ± 3.8	5.8 ± 2.0	104.1
LE	0.23	0.21	0.10

Table 7.15 – Unfused heterocyclic fragments I, B and B1 investigated by ITC.

7.5 Discussion

After having pursued a targeted approach for finding inhibitors of PHGDH by investigating with compounds designed to interact with the cofactor and substrate-binding sites, a more general approach was undertaken by screening a fragment library. There are several advantages of fragment screens over the traditional screening of large libraries, and these have been outlined in section 7.1. One major advantage is that due to the smaller size and less functionality of fragments resulting in a higher probability of binding to targets a smaller library is screened so that the approach is more amenable for academic institutions that do not possess large HTS libraries or have the adequate screening facilities.

In this study, the fragment library was screened by DSF, which is an established screening technique for measuring binding of larger ligands, but has also been successfully used for fragment screening [196-198]. However, false positives and negatives appear commonly in DSF screens, so that a thorough hit validation with a different biophysical technique is necessary [199, 200]. For the determination of hits in the DSF a certain shift in T_m of between 0.5 and 1 °C is often used as a cut-off. In this screen, the z-score was used for hit evaluation, which takes into account a random error distribution within the plate, but also relies on means and standard deviations of all measurements which are influenced by outliers (potential false negatives and/or positives) [201]. In order to minimise the effect of outliers the fragment screen was performed twice, and the average z-score of both experiments used to validate hits. Comparison of the two screens gave a Pearson's correlation coefficient of 0.36, which is a moderate correlation, but was expected given that the changes in T_m due to ligand binding are very small. Validation of the performed DSF screens by comparison of the changes in T_m observed in the presence of 1 mM NADH gave intra- as well as inter-assay variations which were very low with a coefficient of variation (CV) of 1-8 % and 5 %, respectively. However, NADH bound very tightly to PHGDH resulting in large T_m shifts, so that small deviations in the detected T_m have less influence on standard error and CV.

Of the original fragment hits from the DSF screen only a limited number of fragments (15 out of 42 hits) was commercially available. These were re-tested in the DSF screen, and a positive T_m-shift was reproduced with half of the compounds. In order to further validate the initial hits, an orthogonal assay was sought, such as ITC in

225

competition mode. This technique does not rely on fluorescence like DSF, and still allows investigation protein-ligand interactions in solution. In addition, competitive ITC has been successfully used with low-affinity ligands [189, 190]. As expected, the calculated K_D values for the fragments spanned a range of affinities from 1.5 mM to no measurable affinity. The calculated errors in K_D were relatively large for most fragments, which reflects the fact that errors from two ITC titrations have to be taken into account; the control experiment and the displacement experiment. Nevertheless, ITC displacement experiments have been shown to determine K_D values as reliable as those obtained by direct ITC under low-c conditions [202]. However, the nature of the ITC competition experiment allows only the detection of ligands binding to the cofactor-binding site, whereas the DSF screen was performed with full-length PHGDH, which may present other potential ligand-binding sites.

With the small size of the fragments compared to the large cofactor-binding pocket, structural information on how the fragments bind to the protein was crucial to allow for optimisation of the hits. Although crystallisation of sPHGDH had been successfully performed within this project, the poor diffraction quality of sPHGDH crystals rendered them unsuitable for co-crystallisation or soaking of low-affinity fragments. Ideally crystals diffracting to high resolution were desired in order to facilitate detection of weakly bound ligands expected to be represented by weak difference electron density. As many crystallisation conditions for sPHGDH had been investigated without finding a suitable high resolution system, a different protein construct was more likely to yield better crystals than further attempts of optimisation of sPHGDH crystals. Limited proteolysis uses proteases to remove floppy regions of the protein to yield a more stable construct with increased likelihood of crystallising. This approach has been successfully used and has increased the success rate of crystallisation for proteins of different organisms [191, 192]. In this study, the *in-situ* proteolysis approach where protein and protease are mixed and subjected together to crystallisation, resulted in only a few crystals. Analysis of these crystals revealed that most likely the protein still contained all domains, although the electron density for the lid domain was very weak and difficult to interpret. In parallel, construct-directed proteolysis was performed and led to the identification of a new PHGDH fragment of 20-25 kDa in size. Mass-spectrometric analysis of the fragment together with the potential cleavage-sites in the protein sequence for the protease used,

narrowed down the sequence of the new fragment called '93'. This was then successfully cloned, expressed and crystallised. Comparison of the crystal structure of 93 to that of sPHGDH showed that the lid domain, which covers the substrate-binding site was completely removed, thus construct 93 would allow investigation only of ligands that bind to the cofactor-binding site, and similarly to the ITC competition experiment, allow only validation of ligands binding to the cofactor-binding site. Although protein construct 93 provided a very good crystal system that could be used for soaking experiments, one had to bear in mind that only part of the actual enzyme was present and as the ligands were optimised, they would ideally also be investigated in the structure of sPHGDH or even full-length PHGDH. Comparison of 93 to sPHGDH and full-length *M. tuberculosis* PHGDH highlights the large proportion of the enzyme which is missing (Figure 7.29).
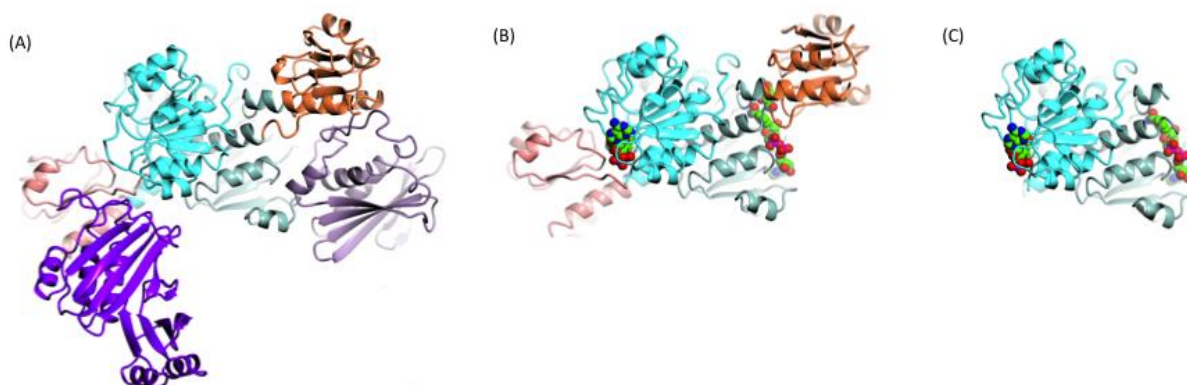
Figure 7.29 – Comparison of PHGDH structures.

(A) Full-length PHGDH from *M. tuberculosis* (PDB 1YGY) [60], (B) human sPHGDH (PDB 2G76) and (C) human PHGDH construct 93. Structures are shown as ribbon diagrams with the equivalent domains highlighted in the same colours in all three structures. If ligands were present in the crystal structure they are shown in spheres and coloured by atom type with carbons in green.

Of the soaked fragments, all fragments seen in the crystal bound to the adenine-binding site. As the removal of the lid domain in construct 93 mainly affects the nicotinamide binding site, which is no longer covered by the substrate binding site, but is open to solvent, it is not surprising to see enriched binding of ligands at the adenine-site which has not undergone structural changes. Therefore, the missing lid domain in the crystal could have biased the results of fragment soaking and once fragments have been improved to greater potency, they should be investigated in at least sPHGDH until crystal structures of the full-length protein became available.

227

The adenine-binding site is often very conserved among different enzymes that bind to adenine-containing cofactors such as ATP, CoA or NAD. General structural recognition features for adenine are a polar interaction, often with the backbone of the protein and a majority of non-specific hydrophobic interactions [203]. In this context, some of the fragments for which binding was observed in the crystal of 93, did not form specific hydrogen bonds with the protein. Therefore, increasing the specific protein contacts as well as exploring possibilities for fragment growth were the main aims of the initiated SAR studies.

Three of the fragments soaked into 93 were substituted benzene rings, namely 3-chloro-4-fluorobenzamide (fragment **D**), *N*-(3-chloro-4-methoxyphenyl)acetamide (fragment **N**) and 5-fluoro-2-methylbenzoic acid (fragment **M**). Interestingly, the benzene ring of fragment **D** did not overlay with the two others, showing that the adenine-binding pocket provided sufficient space for two-ring systems, such as the subsequently investigated benzisoxazoles and indoles. The SAR performed around fragment **N** revealed that (a) the hydrogen bond formed between the nitrogen of the amide function and a water molecule was not critical for binding and affinity of the fragment and (b) that the chloro-group in the 3-position was a key element for binding. The chloro-group bound into a small hydrophobic pocket of the adenine-binding site that, throughout the SAR studies, was seen to be successfully filled with small halogen atoms, such as chloro and fluoro as seen with 6-substituted 3-hydroxybenzisoxazoles.

In this series, 6-bromo-3-hydroxybenzisoxazole was shown to shift away from the position of the 6-fluoro-substituted compound due to the larger bromo-group. The importance of targeting this small hydrophobic pocket was further emphasized when investigating 5-chloro-3-hydroxybenzisoxazole (fragment **O2**). Although initially thought of as a starting point for further fragment growth from the 5-position, fragment **O2** was found to change its binding orientation in order to allow for the chloro-group to fill the hydrophobic pocket. SAR studies of 3-hydroxybenzisoxazole further showed that the initial hit could be substituted with 3-aminebenzisoxazole (**O4**) without losing binding affinity and whilst retaining the same binding mode. The 3-aminobenzisoxazole provided a starting point to grow the fragment further into the cofactor-binding pocket through peptide coupling. Unfortunately, the designed amides lacked sufficient solubility to determine K_D values.

Although structurally similar to benzisoxazole, 5-amino-1-methyl-*1H*-indole (fragment **H**) bound in a different orientation, which could be due to the methyl-group binding to the before mentioned small hydrophobic pocket. This was supported by the fact that the desmethyl analogue (**H5**) did not adopt a specific orientation in the adenine-binding site. Larger indole fragments containing charged functional groups were shown to improve binding affinity (**H1** and **H2**) through binding to the ribose-binding site of the cofactor-binding pocket, thus providing a suitable replacement for the ribose moiety. The indole-series was also grown towards Ser211, which could provide for an additional specific hydrogen bond. 4-Vinyl-1H-indole showed that the vinyl-linker reduced the distance to Ser211 to 2.7 Å, however, the final diol analogue, which was predicted to form a hydrogen bond to Ser211, was not shown to bind to 93, but could be confirmed to bind to PHGDH in an ITC displacement experiment.

7.6 Conclusion

The fragment-based approach provided some initial low-affinity ligands of PHGDH. Fragment hits were identified by DSF and competitive ITC experiments and binding further validated using crystals of PHGDH protein construct 93. Construct 93 was successfully engineered using limited proteolysis and was a key element in the validation and SAR studies of the fragment hits. The SAR studies conducted so far provide some insight into critical structural elements for compounds targeting the adenine-binding site of PHGDH: (a) hydrogen bond formation with Thr206, Ser211 and Asp174 has been seen with a subset of fragments, (b) filling the small hydrophobic pocket with a small halogen atom, such as fluorine or chlorine and (c) vectors for growth of fragments in two directions with growth towards the ribose-binding site of the cofactor pocket facilitated with charged groups (Figure 7.30).
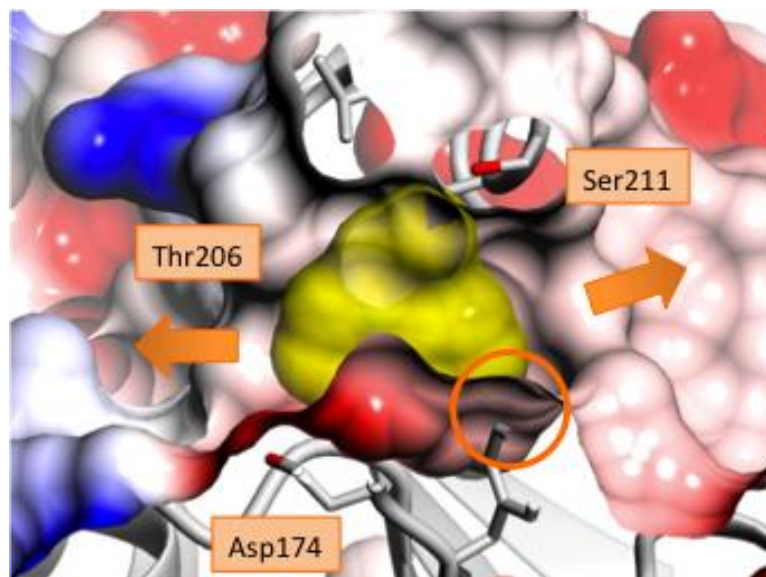
Figure 7.30 – Structural features of the adenine-binding pocket.

Residues around the adenine-binding site to target for hydrogen bond formation are indicated with orange squares. The small hydrophobic pocket in the adenine-binding site is highlighted with an orange circle. The two potential routes to grow the initial fragments out of the pocket are shown with orange arrows.

Chapter 8. Discussion, conclusions and future directions

The early discovery of aerobic glycolysis in cancer by Otto Warburg in the 1920s [11, 204] had been neglected until recently due to the focus on other drivers in cancer [5]. The importance of potential diagnostic and treatment opportunities that lie in cancer metabolism was only rediscovered in the last few years [12]. Whereas detection of early stage cancers through genetic screening is limited due to the somatic provenance of many cancers, screening methods based on the detection of oncometabolites, which are metabolites detected at significantly lower levels in healthy volunteers, might provide a better screening opportunity [205, 206]. In addition, inhibitors of certain metabolic pathways that show increased activity in cancer cells provide a new treatment opportunity. In this context, increased activity of PHGDH, the first enzyme in the serine synthesis pathway, has been reported [69, 72]. PHGDH oxidises 3-PG to PHP and thus diverts glycolytic flux into serine synthesis, which itself is implicated in the synthesis of various biomolecules. In a first evaluation of the relevance of PHGDH in breast cancer and melanoma, gene knockdown experiments showed that, in cancers expressing high levels of PHGDH, knockdown reduced cell growth *in vitro* [69, 72]. The translation of the reduction in cell growth in certain breast cancer cell lines following *PHGDH* knockdown into *in vivo* mice models was, however, ambiguous as two studies showed contradictory results [72, 115]. With the recognition of the increased importance of altered metabolism in cancers, together with initial, though controversial, reported target validation of PHGDH, the aim of this study was to investigate PHGDH as a potential target for cancer therapy by undertaking various steps of the drug discovery process, from target validation to screening and development of PHGDH inhibitors.

The initial observation that PHGDH was overexpressed in colon carcinoma [35], breast cancer [72] and melanoma [207] cell lines, has since been extended to cervical cancer [116, 208], lung squamous cell carcinoma [209] and astrocytoma [77]. In order to investigate if other cancer types could also be potential candidates for treatment with PHGDH inhibitors, a panel of various cancer cell lines spanning 15 different cancer types was investigated for their PHGDH expression levels. Apart from those cell lines reported to have *PHGDH* amplification and/or overexpression, a gastric cancer, a medulloblastoma and a choriocarcinoma cell line were among the cell lines with the highest PHGDH expression, indicating that PHGDH inhibitors could

be a treatment option in those cancer types as well (Figure 3.5). Surprisingly, a negative correlation of PHGDH protein expression with c-Myc expression was determined in contrast to reports of c-Myc positively regulating enzymes of the serine synthesis pathway [127]. However, visual inspection of the correlation graph did not convincingly confirm the calculated correlation coefficient. Therefore, increasing the panel of investigated cell lines would help to increase confidence in the correlation analysis. Similar correlation analysis of protein expression was also performed with PARP1, a major NAD^+-consuming enzyme that has also been linked to proliferation [210]. PARP1 expression did not correlate with PHGDH expression.

In order to further understand the effects of PHGDH knockdown in cancer cells, and to determine to what extent the target validation could be confirmed, siRNA-mediated knockdown experiments were performed in breast cancer and melanoma cell lines (Chapter 3). In agreement with the literature, PHGDH knockdown resulted in significantly reduced cell growth in the *PHGDH*-amplified breast cancer cell line MDA-MB-468 whereas no effect was seen with non-amplified MDA-MB-231 cells as determined by colony formation assays and live cell imaging (Figure 3.12, Figure 3.13) [72, 115]. On the other hand, Hs578T breast cancer cells, which have been reported to overexpress PHGDH and thus be susceptible to PHGDH knockdown despite not containing a PHGDH-amplification, and the *PHGDH*-amplified melanoma cell line, SKMel28 were not affected by PHGDH knockdown in our hands. However, the data presented in the literature were generated with stable transduced shRNA-containing cell lines, which allows for longer periods of protein knockdown compared to the transient siRNA-mediated knockdown and this could have been the reason for the different results obtained in those cell lines. shPHGDH-containing cell lines of the three breast cancer cell lines were therefore generated with two different on-target PHGDH shRNAs and one non-target control shRNA. Determination of efficiency of knockdown in the respective puromycin-selected populations showed that protein levels were only reduced to 50 % whereas siRNA-mediated knockdown resulted in 70-80 % protein reduction. Perhaps as a consequence of the less efficient knockdown, no reduction in cell growth was seen in Hs578T or MDA-MB-468 cells. The knockdown experiments would seem to indicate that levels of PHGDH have to be reduced to at least 70-80 %, as seen after siRNA treatment, to generate an observable response in cells.

Target validation is generally performed with knockdown models. These models may not, however, predict the effect that an inhibitor might have, as is seen in the case of PARP inhibitors where inhibition, but not depletion of the enzyme blocks DNA repair. These considerations highlight the importance of using chemical probes in target validation [211-213]. Because there is no inhibitor of PHGDH described to date, this work focused on developing such a small molecule. Due to the structural homology of NAD$^+$-binding pockets among different enzymes, the specificity of designed inhibitors was predicted to be provided through targeting the adjacent substrate-binding pocket. Because the natural substrate of PHGDH, 3-PG, being an acid, which is charged under physiological conditions, its binding-pocket on PHGDH is naturally complementary in character to a polar ligand. Polar ligands are likely to be challenging to transport across the cell membrane and might require administration as a prodrug, *e.g.* an ester instead of an acid. Combination of functional groups so as to simultaneously target the substrate- and cofactor-binding pockets in a single molecule may however help to increase the specificity and improve the PK/PD properties of a potential inhibitor. Previous studies have shown the effectiveness of these so-called bi-substrate inhibitors in targeting substrate- and cofactor-binding pockets of dehydrogenases in the case of IDH and LDHA [97, 140]. Malate-containing fragments were used as a starting point (Chapter 4), as the substrate-analogue *D*-malate had been observed to bind to PHGDH in a crystal structure of the catalytic subunit of human PHGDH (sPHGDH) (PDB 2G76), while human PHGDH has also been shown to catalyse the reduction of oxaloacetate to malate [65]. Compounds were tested in an established FP assay using NADH as fluorophore. Displacement of NADH accompanied by a decrease in polarised light could be detected with increasing concentrations of ADPR, thus validating the assay (Figure 4.6). Of the malate-containing compounds tested, a significant concentration-dependent displacement of NADH was measured with 3-indole-*DL*-aspartic acid when assessed with full-length PHGDH or the catalytic subunit only (sPHGDH). In the chiral environment of the protein, the stereochemistry of a chiral ligand can be an important determinant of the response to inhibitor binding, and the synthesis of the enantiopure compounds was undertaken following literature procedures [145, 146]. With the synthesis route to enantiopure 3-indole-aspartic acid in place, the procedure was extended to different heterocyclic cores, namely benzisoxazole and benzothiophene. Although the respective nitrophenylesters were synthesised to

same (benzisoxazole) or better (benzothiophene) yields, the subsequent amide coupling was remarkably lower in yield compared to the 3-indole-aspartic acid synthesis. All compounds within this small library were able to displace NADH in the FP assay, however the original 3-indole-aspartic acids were the most potent compounds in this assay (Figure 4.8). The original hit compound 3-indole-*DL*-aspartic acid was also investigated in the enzyme activity assay, however, the compound interfered with the fluorescent read-out, so that no conclusion could be drawn on how hits from the FP assay would translate into inhibition of PHGDH activity.

In contrast to the previous assumption that human PHGDH uses only 3-PG as a substrate [57], more recent literature data has shown that human PHGDH could also reduce α-ketoglutarate and oxaloacetate to 2-hydroxyglutarate and malate, respectively [65]. Therefore further substrate (3-PG) and product (PHP) analogues were assessed for their potential to stabilise the T_m by human PHGDH. These experiments confirmed 3-PG as the most stabilising ligand among the panel, being the only compound to increase the T_m of more than 1 °C. Interestingly, combination of 3-PG and NAD$^+$ stabilised the protein to a greater extent than each compound alone. This effect was even more prominent with *DL*-malate, which did not stabilise the protein unless in combination with NAD$^+$ (Figure 4.11). These data point towards a Bi-Bi ordered mechanism. In the case of *DL*-malate as a substrate, NAD$^+$ bound first, which would be the order of binding as seen with *E. coli* PHGDH and the opposite order to that of *M. tuberculosis* PHGDH [57]. This synergistic binding was only seen with NAD$^+$ and not with NADH. In addition to DSF-based binding studies, human sPHGDH was co-crystallised with some of the substrate and product analogues, yielding a crystal structure with *L*-tartrate bound to one chain of the dimeric sPHGDH. Comparison of the two chains in the crystal asymmetric unit revealed a lid domain movement upon *L*-tartrate binding that had not been previously described, although a similar movement is reported to occur in a variety of other dehydrogenases [214, 215].

Having explored the substrate-binding pocket of PHGDH, the next step was to focus on the cofactor-binding pocket, which provides a larger potential binding site for inhibitors (Chapter 5). Although selectively targeting the cofactor-binding pocket is expected to be challenging due to its structural conservation among different enzymes, NAD$^+$-requiring enzymes have been targeted through this site. Firstly, the

nicotinamide binding region was investigated using ADPR, which is essentially NAD$^+$ without the nicotinamide group. ADPR was found to be a moderate inhibitor of PHGDH in an established enzyme activity assay that, compared to activity assays described in the literature for PHGDH, was performed in the physiological direction of 3-PG oxidation and NAD$^+$ reduction (Figure 5.3). In addition, the binding affinity of ADPR was further analysed by direct ITC revealing a K_D for ADPR of 40.0 ± 5.4 µM, and an exothermic binding event with little contribution from entropic effects (Figure 5.4). Although ADPR was likely to bind to the cofactor-binding site, confirmation was sought through co-crystallisation of sPHGDH with ADPR. Interestingly, in the crystal obtained, modelling of ADPR into the active site could not completely satisfy the observed electron density, revealing partial occupancy of the binding site with the natural cofactors NAD$^+$ and/or NADH (Figure 5.5). Charcoal-treatment was undertaken with sPHGDH and full-length PHGDH, leading to the removal of the co-purified cofactor. No quantification method was available to investigate if all bound cofactor could be removed by charcoal treatment. However, direct ITC titration of NADH into full length PHGDH showed that NADH had very high affinity for the enzyme (K_D = 0.22 ± 0.03 µM) and this was 2000-fold higher than for NAD$^+$ (K_D = 444 ± 18 µM) (Figure 6.6). In addition, NAD$^+$ analogues TAD, PAD and APAD were investigated for their potential to function as cofactors for human PHGDH. In the enzyme activity assay, APAD and PAD showed significantly higher V_{max} than NAD$^+$ and TAD, as expected due to their higher redox potentials. However, the K_m values showed that NAD$^+$ was the preferred cofactor for this reaction, and this was further confirmed in DSF screens, where only NAD$^+$ stabilised PHGDH with an increase in T_m of 1 °C (Figure 5.7, Figure 5.8). Interestingly, except for NAD$^+$ and *DL*-malate, PAD in combination with α-ketoglutarate was the only NAD$^+$ analogue to exert a synergistic effect on binding with a substrate other than 3-PG. A co-crystal structure of sPHGDH with TAD was obtained and although there was clear electron density for bound cofactor, it is not possible to determine if this was TAD or one of the natural ligands due to the afore mentioned co-purification issues. After having looked into variations around the nicotinamide group, the importance of the phosphate linker was investigated through comparison of ATP, ADP and AMP binding by DSF. Stabilisation of PHGDH was increased with increasing numbers of phosphate groups with 0.2 mM ATP raising the T_m by 4.7 ± 0.9 °C, whereas AMP was about as effective as NAD$^+$ with an increase in T_m of 0.6 ± 0.4 °C. Addition of ADP resulted in

a significantly higher increase in the T_m of PHGDH compared to NAD^+, but slightly lower than with ADPR, thus showing that in the absence of the substrate, 3-PG, the additional nicotinamide function has rather a negative impact on cofactor binding (Figure 5.10). Finally, we focused on variations around the adenine moiety and showed by comparison of NAD^+ with nicotinamide hypoxanthine dinucleotide (NHD) by DSF that the presence of the amine-group at the 4-position of the purine, and hence its hydrogen bond with Ser211, were important for binding to PHGDH. However, the effect of the 4-NH_2 group seemed no longer relevant when comparing NAD^+ and GDP, suggesting that the additional 6-NH_2 group of GDP might compensate for the missing NH_2-group in the 4-position or that a shift in GDP binding mode could bring the 4-carbonyl function within hydrogen bond distance to Ser211. In addition, a small library of 1- and 4-substituted purines was investigated due to availability of these compounds within the NICR compound library. Of the six compounds investigated, two compounds, both containing a triisopropylsilyl group in the 4-position, showed similar potency to ADPR in the enzyme activity assay using full-length PHGDH, but did not inhibit sPHGDH activity in the same assay format (Figure 5.16). Kinetic analysis of their mechanism of action confirmed this result, with the inhibitors showing up as non-competitive with NAD. This suggested that these compounds were more likely to bind outside the nucleotide- and substrate-binding regions. These two compounds were not considered good starting points for inhibitor development and in the absence of a crystal structure of full-length PHGDH to aid determination of the binding mode and thus rationally design more optimal analogues, these compounds were not pursued.

The search for potential PHGDH inhibitors made clear that a major obstacle to the identification and optimisation of potential inhibitors lay in the co-purification of sPHGDH, and to a lesser extent PHGDH, with cofactor, especially as crystallisation experiments could only be performed with sPHGDH. Previously, *E. coli* PHGDH had been reported to co-purify with NADH due to its high affinity (estimated K_D ~50 nM) [216, 217]. In the case of *E. coli* PHGDH, removal of NADH was achieved by mixing with PHP, which resulted in the formation of the less-tightly binding cofactor NAD^+, which could then be separated from the protein through column chromatography [216]. However, as PHP was no longer commercially available, another strategy had to be employed. Therefore, charcoal-treatment of purified protein was attempted as

charcoal adsorbs nucleotides. The only means to monitor the effect of charcoal-treatment was through monitoring the OD_{260}/OD_{280} ratio. Despite extensive charcoal-treatment this could not be decreased below 0.65, which is still greater than the ratio reported for pure protein (0.57). Charcoal-treatment was also performed on purified full-length PHGDH, but did not alter the OD_{260}/OD_{280} ratio. Purified full-length PHGDH had an OD_{260}/OD_{280} ratio of 0.65-0.67, which indicates that PHGDH either does not co-purify with its cofactors or co-purifies to a lesser extent. This is an important difference as the previously described experiments with malate-containing compounds, and NAD^+ analogues and truncates were performed on full-length PHGDH and thus are more likely to provide a true readout of direct binding rather than of competition with prebound cofactor.

Although it was not clear from the OD_{260}/OD_{280} ratio if all binding sites were free of cofactor after charcoal treatment of sPHGDH, NADH-binding investigated by ITC showed that the estimate of binding sites on the protein increased from 0.3 to 0.8 after charcoal-treatment. Similarly, the T_m of the charcoal treated protein was lower than for the untreated sPHGDH. However, the charcoal-treatment had limitations, not only because the success of NADH removal could not be quantified accurately and, at least according to ITC experiments was not complete, but also a batch-to-batch variation in the outcome of charcoal-treatment was observed making a comparison of experiments run with different protein batches difficult. In addition, activated charcoal also adsorbed protein to a certain extent, thus negatively impacting on the yield of purified protein. In search of a different strategy for the generation of cofactor-free protein, site-directed mutants were developed that had point-mutations in amino acids that were either directly involved in cofactor binding through hydrogen bonds or hydrophobic interactions, or were in close proximity to the cofactor-binding pocket (chapter 6). Since attempts to crystallise full-length PHGDH within this study failed, mutations were performed on sPHGDH, to enable protein crystallisation studies. Four adenine- and three nicotinamide-site mutants of sPHGDH were designed, expressed and purified with all purified mutants displaying OD_{260}/OD_{280} ratio in the range 0.6-0.7, indicating that all mutants co-purified with cofactor to a lesser degree than sPHGDH. The mutants were subsequently analysed for proper folding by DSF and CD and both techniques confirmed proper folding. CD spectra also showed that the introduced mutations had affected the secondary structure, as the amount of

unordered structure had increased in all the mutants, putting the likelihood of successful crystallisation in question. This decrease in secondary structure could also result from non-binding of the cofactor, thus confirming the initial design hypothesis. Depending on the introduced mutation, the proteins showed weak to no affinity for NADH as measured by ITC and DSF, with the adenine-site mutants reducing the affinity more than the nicotinamide-site mutants (Figure 6.14, Figure 6.15). These results indicate that the protein-cofactor interactions around the nicotinamide-binding site are less relevant for cofactor binding. These findings are in agreement with the results obtained when investigating the binding affinities of fragments of NAD$^+$ (Chapter 5). Independently of NADH-binding detected by ITC and DSF, none of the mutants were able to catalyse 3-PG oxidation (Figure 6.16). Whether this means that the observed NADH binding does not reflect NAD$^+$ binding or that the cofactor might not be aligned correctly for catalysis, remains to be elucidated. Crystal structures of the mutants would provide further insight into the effects of the mutations on the active site of the protein, thus allowing full validation of their utility as surrogates in hit-finding strategies. Although crystals were obtained with one mutant (R154E), the diffraction pattern measured was not good enough to permit structure solution. Further optimisation of the initial crystallisation conditions might result in better crystals in this regard. However, due to the increasing unordered content within the protein as determined by CD, the mutations would appear to have introduced structural changes unfavourable to crystallisation.

After having probed the cofactor and substrate-binding sites with compounds specifically designed to target these sites, a more general approach to the identification of inhibitors was undertaken by performing a fragment screen, kindly provided by CRT Cambridge (chapter 7). Due to the small size of the fragments, a higher hit-rate than with larger compounds was anticipated. These compounds could then be further improved to eventually yield high affinity PHGDH inhibitors. Of the assays developed for PHGDH, the DSF assay was the most amendable format for high throughput screening, although this format is prone to false positive and negative results that interfere with the measured fluorescence [218]. The fragment library was screened twice and, as expected for this assay format, the correlation between the two runs was rather low, with a Pearson correlation coefficient of 0.36. In order to decrease the likelihood of obtaining false positive results, the Z-score was

used as a way of defining hits. Thus 42 hits were obtained (hit rate of 6.4 %) of which 15 were purchased and re-assessed by DSF. Binding to PHGDH was confirmed for a subset of these compounds. Looking for a second, orthogonal assay to validate these hits, an ITC-competition experiment was established, where NADH was titrated into a mixture of sPHGDH and the fragment. Although the competitive ITC could be validated by using ADPR, the accuracy of calculating K_D values from these experiments with lower affinity ligands is questionable. This is due to having to account for the errors in two titration experiments, one in the presence and one in the absence of fragment leading to very large errors in the calculated K_a, K_D, ΔH and ΔS values for the fragment. This is particularly significant in the case of fragments with very weak binding affinity. Nevertheless, the ITC competition experiment was useful in validating and (to a lesser extent) ranking the hits, as well as for comparing the initial fragment hits with further optimised fragment analogues. Fragment-based drug discovery requires information on how the fragments bind to the protein to properly guide fragment optimisation. Although sPHGDH had been successfully crystallised, co-purification with cofactors together with low resolution and unreliable diffraction quality of sPHGDH crystals made this construct unsuitable for use with low-affinity ligands at medium to high throughput. Seeking for an alternative construct, sPHGDH and PHGDH were subjected to limited proteolysis, which has been shown to be a successful strategy for obtaining more stable constructs for crystallisation due to removal of flexible or unstructured regions by proteases [191, 192]. By this means, the PHGDH truncate 93 was identified and shown to form crystals that diffracted to high resolution (1.1-1.7 Å) and that could be used for soaking experiments. Of the 15 purchased fragments, 7 were found to bind in crystals of truncate 93 after soaking for 24 hours, and were taken forward for fragment optimisation. For other fragments, longer soaking times of up to 1 week were used without resulting in crystals containing the ligand. All compounds seen in crystals of truncate 93 after soaking were found in the adenine-binding site. As construct 93, compared to sPHGDH, is missing the lid domain that contains the substrate-binding site and closes onto the nicotinamide-binding area of the cofactor-binding pocket, it might be that the fragments not seen to bind in crystals of 93 were actually binders to the nicotinamide-site.

Of the fragments seen in the adenine-binding site, four general features for optimisation could be derived: (1) there was a potential vector for growth out of the cofactor-binding site opposite to (2) a second vector that pointed in the direction of the ribose-binding site of the cofactor pocket, (3) the adenine-binding site contained a small hydrophobic pocket that could be filled with small halogens, such as chlorine or fluorine and (4) Ser211 was a potential hydrogen donor in close proximity to the adenine pocket that could be targeted to increase the number of specific interactions between the ligand and the protein.

In general, attempts to grow initial hits in the direction of vector (1) did not result in increased potency, and in the case of modified 3-hydroxybenzisoxazole, adding groups to grow out of the adenine-pocket resulted in reorientation of the fragments to fill the adjacent small lipophilic pocket instead. This lipophilic pocket was successfully targeted with chlorine and fluorine and also shown to be too small to accommodate bromine. Targeting of Ser211 was attempted in the indole series through adding groups in the 4-position and, although the 4-vinyl indole placed the vinyl group in the direction of Ser211, the respective diol was not observed to bind in the crystal structure and thus it could not be confirmed if an additional H-bond to Ser211 had been formed. Growth in the direction of the ribose pocket was possible, as shown with indoles with polar linkers (acidic groups, amide group, hydroxyl group). In this context, the malate analogue, 3-indole-aspartic acid, that had been investigated as potential binder to the substrate-site/nicotinamide-subsite, could also be envisaged to bind to the adenine-subsite of the cofactor-binding pocket.

8.1 Conclusion

The work presented here on PHGDH as a target in cancer therapy spans the early part of the drug discovery process from initial target validation to screening for potential inhibitors, including structural investigations of the target and the beginning of early SAR studies around initial hits.

PHGDH as a target was validated in *PHGDH*-amplified breast cancer cells, and analysis of a panel of cancer cell lines revealed high PHGDH protein expression in other cancers, which could be potentially targetable with PHGDH inhibitors.

Cofactor and substrate binding to PHGDH were shown to be synergistic and a crystal structure of sPHGDH with the substrate analogue, *L*-tartrate, bound to one chain of the crystallographic dimer revealed a distinct domain movement upon substrate binding.

For the inhibitor search, a variety of biochemical and biophysical assays were developed for human PHGDH and a truncated construct of the protein was generated that was particularly useful for obtaining protein crystals that diffracted to high resolution.

In addition, novel fragment binders to PHGDH were identified and crystals further optimised based on structural information from crystal soaking experiments.

8.2 Future directions

The investigations presented herein provide many opportunities for further extension:

1)   Strengthening the target validation through the use of *in vivo* models as well as investigation of PHGDH knockdown in cells under different conditions, such as serine starvation and/or hypoxia. The target validation could also be extended into other cancers apart from melanoma and breast cancer, as well as investigated in relation to genetic alterations, such as mutations in p53 or c-Myc [117, 127].

2)   Truncate 93 provides the possibility for a crystal-based high throughput fragment screen to allow identification of alternative scaffolds.

3)   Extension of testing hit compounds in all assay formats available to date to investigate if the results correlate and allow to rank the hits in the same order.

3)   Extension of the SAR studies around the obtained fragment hits. After increasing the potency of these compounds, their activity could be investigated in the *in vitro* enzyme activity assay, and subsequently cell-based assays could be developed and used for inhibitor validation.

4)   Ultimately, crystallisation of full-length PHGDH would allow observation of the potential ligands in a more relevant context as well as screening for allosteric binding sites as alternative targetable sites.

Chapter 9. Experimental Part – Chemistry

9.1 Summary of Generic Reactions, Analytical and Chromatographic Conditions

9.1.1 Chemicals and Solvents

All commercial reagents were purchased from Sigma-Aldrich Chemical Company, Alfa Aesar, Apollo Scientific or Tokyo Chemical Industry UK Ltd. The chemicals were of the highest available purity. Unless otherwise stated, chemicals were used as supplied without further purification. Anhydrous solvents were obtained from AcroSeal™ or Aldrich SureSeal™ bottles and were stored under nitrogen. Petrol refers to the fraction with a boiling point between 40 and 60 °C.

9.1.2 Chromatography

Thin layer chromatography performed to monitor reaction progress was conducted on plates pre-coated with silica gel (Merck $60F_{254}$). The eluent was as stated (where this consisted of more than one solvent, the ratio is stated as volume:volume) and visualisation was either by short wave (254 nm) ultraviolet light, or by treatment with the visualisation reagent potassium permanganate ($KMnO_4$) followed by heating. 'Flash' medium pressure liquid chromatography (MPLC) was carried out either on a Biotage SP4 automated purification system or a Varian 971-FP automated purification system, using pre-packed Varian or Grace silica cartridges.

9.1.3 Microwave Reactions

All reactions carried out in a microwave were performed in a Biotage Initiator with Sixty robot.

9.1.4 Analytical Techniques

Melting points were determined using a VWR Stuart SMP40 apparatus and are uncorrected.

^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectra were obtained as either $CDCl_3$ or MeOD solutions and recorded at 500 MHz and 125 MHz respectively, on a Bruker Avance III 500 spectrometer. Where ^{13}C NMR data are not quoted, insufficient material was available or problems obtaining high-resolution spectra were encountered. Chemical shifts are quoted in parts per million (δ) referenced to the

appropriate deuterated solvent employed. Multiplicities are given as s (singlet), d (doublet), t (triplet), q (quartet), m (multiplet) or combinations thereof. Coupling constant values are given in Hz. Homonuclear and heteronuclear two-dimensional NMR experiments were used where appropriate to facilitate assignment of chemical shifts.

LC-MS was carried out on a Waters Acquity UPLC system with PDA and ELSD employing positive or negative electrospray modes as appropriate to the individual compound. Where LRMS data is not quoted, the mass was not recognised for that compound. High-resolution mass spectrometry was performed by the EPSRC UK National Mass Spectrometry Facility, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP.

FTIR spectra were recorded on either a Bio-Rad FTS 3000MX diamond ATR or an Agilent Cary 630 FTIR as a neat sample.

UV spectra were obtained using a U-2001 Hitachi Spectrophotometer with the sample dissolved in ethanol.

9.2 Synthesis of PHGDH inhibitors – Experimental procedures

9.2.1 General procedures

General procedure A: The carboxylic acid (1 mol eq.) was dissolved in acetonitrile (15 mL/mmol acid) before 3-amino-1,2-benzoisoxazole (2.5 mol eq.) was added, followed by addition of PCl_3 (1 mol eq.). The reaction mixture was subjected to microwave irradiation at 150 °C for 5 to 25 minutes. The reaction was quenched with a few drops of water and the solvent removed *in vacuo*. The crude product was redissolved in EtOAc (60 mL/mmoL carboxylic acid) and washed with saturated $NaHCO_3$ (60 mL/mmoL carboxylic acid) solution. The aqueous layer was extracted with EtOAc (3 x 30 mL/mmoL carboxylic acid). The organic extracts were combined and dried over $MgSO_4$. The crude product was purified by MPLC, thereby recovering also the excess of 3-amino-1,2-benzisoxazole.

General procedure B: To the appropriate alcohol (1 mol eq.) in acetonitrile (6 mL/mmol of alcohol) were added Cs_2CO_3 (3 mol eq.) and the desired alkyl iodide (3 mol eq.). The reaction was heated to 80 °C and stirred for 6 h. The reaction mixture was cooled to RT and any insoluble precipitant filtered off. The solvent was removed *in vacuo* and the crude product purified by MPLC.

General procedure C [146]: The appropriate acetic acid (1 mol eq.) and 4-nitrophenol (1 mol eq.) were dissolved in EtOAc (4 mL/mmol acetic acid) and the mixture cooled to 0 °C. DCC (1 mol eq.) dissolved in EtOAc (1 mL/mmol of DCC) was added dropwise to the reaction mixture and left to stir at 0 °C for 1 h. The reaction mixture was warmed to RT and stirred for a further hour. The precipitated dicyclohexylurea was filtered off and the filtrate concentrated *in vacuo*. The crude product was purified by recrystallisation.

General procedure D [146]: The appropriate aspartic acid enantiomer (1 mol eq.) and tetramethylguanidine (2 mol eq.) were dissolved in MeOH (50 % (v/v) in water) (4 mL/mmol aspartic acid). Finely grounded appropriate *p*-nitrophenyl ester (1 mol eq.) was added and the reaction mixture stirred at RT for 48 h. The reaction was quenched by addition of water (20 mL/mmol aspartic acid) and extracted with Et_2O (2 x 30 mL). The aqueous fraction was acidified to pH 5 and extracted again with Et_2O (3 x 30 mL). The aqueous layer was further acidified to pH 1 and extracted with *n-*

butanol (1 x 40 mL, 1 x 10 mL). The butanol fractions were combined, washed with water (2 x 15 mL, 2 x 10 mL) and concentrated *in vacuo*. The crude product was purified by recrystallisation if necessary.

9.2.2 Synthetic procedures

4-Nitrophenyl 2-(1*H*-indol-3-yl)acetate, (8)
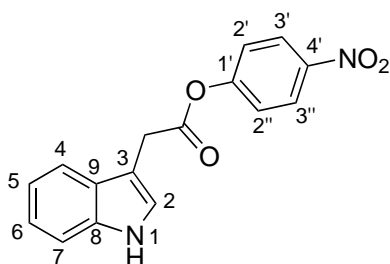
Compound **8** was synthesised according to general procedure C using the following reagents: 3-indole acetic acid (5.80 g, 33.1 mmol), 4-nitrophenol (4.60 g, 33.1 mmol), *N'N*-dicyclohexylcarbodiimide (6.82 g, 33.1 mmol) and EtOAc (125 mL). The pure product was obtained by crystallisation at 4 °C. The solid was filtered off and dried to yield the *title compound* as a yellow solid (5.41 g, 55 %); R_f = 0.6 (petrol:EtOAc, 3:2); m.p. 104-105 °C; λ_{max}(EtOH)/nm 272.4, 218.0; IR (neat) v_{max}/cm^{-1} 3398, 3322, 2927, 2849, 2114, 1769, 1624, 1571; ^{1}H NMR (500 MHz, CDCl$_3$) δ 4.1 (2H, s, CH$_2$), 7.21 (1H, dd, *J* = 7.5 Hz, H-5), 7.24 (1H, s, H-2), 7.31-7.24 (3H, m, H-2', H-2'', H-6), 7.40 (1H, d, *J* = 8.1 Hz, H-7), 7.71 (1H, d, *J* = 8.1 Hz, H-4), 8.25-8.20 (2H, m, H-3' and H-3''); ^{13}C NMR (125 MHz, CDCl$_3$) δ 31.6 (CH$_2$), 107.2 (C-2), 111.5 (C-7), 118.7 (C-4), 120.1 (C-5), 122.5 (C-2' and C-2''), 122.6 (C-6), 123.5 (C-1), 125.3 (C-3' and C-3''), 127.1 (C-3), 136.3 (C-8), 145.4 (C-4'), 155.7 (C-1'), 169.6 (CO); LRMS (ES$^+$) *m/z* 297.3 [M+H]$^+$; HRMS (ESI) calcd for C$_{16}$H$_{13}$N$_2$O$_4$ [M+H]$^+$ 297.0870, found 297.0868.

(2-(1*H*-Indol-3-yl)acetyl)-*D*-aspartic acid, (9)
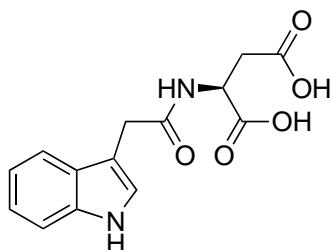
Compound **9** was synthesised according to general procedure D using the following reagents: D-aspartic acid (333 mg, 2.5 mmol), tetramethylguanidine (627 µL, 5.0 mmol), 4-nitrophenyl 2-(1*H*-indol-3-yl)acetate (**8**) (740 mg, 2.5 mmol) and aq. MeOH (50 % (v/v), 10 mL). Removal of *n*-butanol *in vacuo* yielded the *title compound* as a pale rose solid (692 mg, 95 %); R_f = 0.2 (DCM:MeOH, 2:1); m.p. 180-182 °C; λ_{max}(EtOH)/nm 280.6, 219.8; IR (neat) v_{max}/cm^{-1} 3382, 3056, 2931, 2553, 2085, 1896, 1716, 1618, 1523; ^{1}H NMR (500 MHz, MeOD) δ 2.77 (2H, d, *J* = 5.7 Hz, C*H$_2$*COOH), 3.69 (2H, d, *J* = 1.3 Hz, C*H$_2$*CONH), 4.73 (1H, t, *J* = 5.7 Hz, C*H*COOH), 6.98 (1H, dd, *J* = 7.5 Hz, H-5), 7.06 (1H, dd, *J* = 7.5 Hz, H-6), 7.17 (1H, s, H-2), 7.31 (1H, d, *J* = 8.1 Hz, H-7), 7.50 (1H, d, *J* = 8.1 Hz, H-4), 10.40 (2H, s, COOH); ^{13}C NMR (125 MHz, MeOD) δ 33.7 (CH$_2$CO), 36.8 (CH$_2$COOH), 50.2 (CHCOOH), 109.0 (C-3), 112.2 (C-7), 119.4 (C-5), 120.0 (C-4), 122.6 (C-6), 125.0 (C-2), 128.5 (C-9), 138.1 (C-8), 173.9 (CO), 174.0 (CO), 174.6 (CO); LRMS (ES$^+$) *m/z* 291.3 [M+H]$^+$; Compound degraded before HRMS (ESI) analysis could be performed.

(2-(1*H*-Indol-3-yl)acetyl)-*L*-aspartic acid, (10)
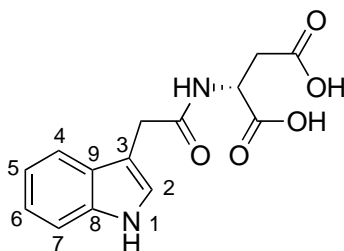
Compound **10** was synthesised according to general procedure D using the following reagents: *L*-aspartic acid (333 mg, 2.5 mmol), tetramethylguanidine (627 µL, 5.0 mmol), 4-nitrophenyl 2-(1*H*-indol-3-yl)acetate (**8**) (740 mg, 2.5 mmol) and aq. MeOH (50 % v/v) (10 mL). Removal of *n*-butanol *in vacuo* yielded the *title compound* as a pale rose solid (372 mg, 51 %); R_f = 0.2 (DCM:MeOH, 2:1); m.p. 180-182 °C;

λ_{max}(EtOH)/nm 280.4, 219.6; IR (neat) v_{max}/cm^{-1} 3382, 3054, 2930, 2530, 2114, 1896, 1715, 1616, 1523; ^{1}H NMR (500 MHz, MeOD) δ 2.77 (2H, d, J = 5.7 Hz, CH_2COOH), 3.69 (2H, d, J = 1.3 Hz, CH$_2$CONH), 4.73 (1H, t, J = 5.7 Hz, CHCOOH), 6.98 (1H, dd, J = 7.5 Hz, H-5), 7.06 (1H, dd, J = 7.5 Hz, H-6), 7.17 (1H, s, H-2), 7.31 (1H, d, J = 8.1 Hz, H-7), 7.50 (1H, d, J = 8.1 Hz, H-4); ^{13}C NMR (125 MHz, MeOD) δ 33.7 (CH$_2$CO), 36.8 (CH$_2$COOH), 50.2 (CHCOOH), 109.0 (C-3), 112.2 (C-7), 119.4 (C-5), 120.0 (C-4), 122.6 (C-6), 125.0 (C-2), 128.5 (C-9), 138.1 (C-8), 173.9 (CO), 174.0 (CO), 174.6 (CO); LRMS (ES$^+$) m/z 291.3 [M+H]$^+$; Compound degraded before HRMS (ESI) analysis could be performed.

4-Nitrophenyl 2-(benzo[*d*]isoxazol-3-yl)acetate, (11)
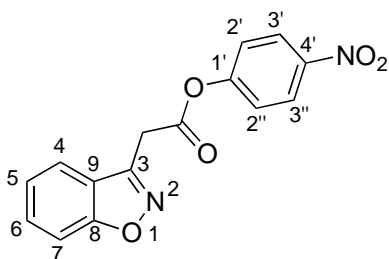
Compound **11** was synthesised according to general procedure C using the following reagents: 2-(1,2-benzisoxazol-3-yl) acetic acid (1.00 g, 5.6 mmol), 4-nitrophenol (785 mg, 5.6 mmol), *N'N*-dicyclohexylcarbodiimide (1.17 g, 5.6 mmol) and EtOAc (25 mL). The product was purified by recrystallisation from petrol:EtOAc (1:1) to yield the *title compound* as an orange solid (766 mg, 46 %); R_f = 0.4 (petrol:EtOAc, 4:1); m.p. 97-98 °C; λ_{max}(EtOH)/nm 275.8, 237.0; IR (neat) v_{max}/cm^{-1} 3117, 3076, 2921, 2851, 2078, 1912, 1751, 1612, 1521; ^{1}H NMR (500 MHz, CDCl$_3$) δ 4.35 (2H, s, CH$_2$), 7.34-7.28 (2H, m, H-2' and H-2''), 7.38 (1H, ddd, J = 8.0, 1.3 Hz, H-5), 7.70-7.57 (2H, m, H-6 and H-7), 7.76 (1H, d, J = 8.0 Hz, H-4), 8.21-8.13 (2H, m, H-3' ad H-3''); ^{13}C NMR (125 MHz, CDCl$_3$) δ 32.0 (CH$_2$), 110.4 (C-7), 121.4 (C-9), 122.4 (C-2' and C-2''), 124.2 (C-4), 125.5 (C-3' and C-3''), 126.3 (C-5), 130.5 (C-6), 145.7 (C-4'), 145.8 (C-3), 151.7 (C-1'), 155.0 (C-8), 166.1 (CO); LRMS (ES$^+$) m/z 299.2 [M+H]$^+$; Compound degraded before HRMS (ESI) analysis could be performed.

(2-(1*H*-Benzo[*d*]isoxazol-3-yl)acetyl)-*D*-aspartic acid, (12)

Compound **12** was synthesised according to general procedure D using the following reagents: D-aspartic acid (150 mg, 1.1 mmol), tetramethylguanidine (283 µL, 2.2 mmol), 4-nitrophenyl 2-(benzo[d]isoxazol-3-yl)acetate (**11**) (336 mg, 1.1 mmol) and aq. MeOH (50 % v/v) (10 mL). Removal of *n*-butanol *in vacuo* yielded the *title compound* as a white solid (34 mg, 11 %); R_f = 0.2 (DCM:MeOH, 2:1); m.p. 147-149 °C; λ_{max}(EtOH)/nm 281.8, 236.6; IR (neat) v_{max}/cm^{-1} 3264, 3083, 2927, 2598, 2114, 1726, 1609, 1561; ^{1}H NMR (500 MHz, MeOD) δ 2.84 (1H, dd, *J* = 6.9, 16.8 Hz, C*H*HCOOH), 2.90 (1H, dd, *J* = 5.5, 16.8 Hz, CH*H*COOH), 4.03 (2H, s, C*H₂*CONH), 4.79 (1H, dd, *J* = 5.5, 6.9 Hz, C*H*COOH), 7.38-7.32 (1H, m, H-5), 7.62-7.58 (2H, m, H-6 and H-7), 7.86 (1H, d, *J* = 8.0 Hz, H-4), 8.80 (1H, s, NH); ^{13}C NMR (125 MHz, MeOD) δ 33.2 (*C*H₂CONH), 36.8 (CH₂COOH), 50.6 (CHCOOH), 110.5 (C-7), 122.6 (C-9), 123.4 (C-4), 124.7 (C-5), 131.3 (C-6), 155.0 (C-3), 164.6 (C-8), 170.0 (CO), 173.7 (CO), 173.8 (CO); LRMS (ES$^+$) *m/z* 293.2 [M+H]$^+$; HRMS (ESI) calcd for $C_{13}H_{13}N_2O_6$ [M+H]$^+$ 293.0768, found 293.0774.

**(2-(1*H*-Benzo[*d*]isoxazol-3-yl)acetyl)-*L*-aspartic acid, (13)**

Compound **13** was synthesised according to general procedure D using the following reagents: L-aspartic acid (150 mg, 1.1 mmol), tetramethylguanidine (283 µL, 2.2 mmol), 4-nitrophenyl 2-(benzo[*d*]isoxazol-3-yl)acetate (**11**) (336 mg, 1.1 mmol) and aq. MeOH (50 % v/v) (10 mL). Removal of *n*-butanol *in vacuo* yielded the *title compound* as a white solid (25 mg, 8 %); R_f = (DCM:MeOH, 2:1) 0.2; m.p. 147-149

°C; λ_{max}(EtOH)/nm 281.8, 236.8; IR (neat) v_{max}/cm^{-1} 3261, 3093, 2934, 2600, 2111, 1726, 1609, 1561; ^{1}H NMR (500 MHz, MeOD) δ 2.84 (1H, dd, J = 6.9, 16.8 Hz, C*H*HCOOH), 2.90 (1H, dd, J = 5.5, 16.8 Hz, CH*H*COOH), 4.03 (2H, s, C*H$_2$*CONH), 4.79 (1H, dd, J = 5.5, 6.9 Hz, C*H*COOH), 7.38-7.32 (1H, m, H-5), 7.62-7.58 (2H, m, H-6 and H-7), 7.86 (1H, d, J = 8.0 Hz, H-4), 8.80 (1H, s, NH); ^{13}C NMR (125 MHz, MeOD) δ 33.2 (*C*H$_2$CONH), 36.8 (*C*H$_2$COOH), 50.6 (*C*HCOOH), 110.5 (C-7), 122.6 (C-9), 123.4 (C-4), 124.7 (C-5), 131.3 (C-6), 155.0 (C-3), 164.6 (C-8), 170.0 (CO), 173.7 (CO), 173.8 (CO); LRMS (ES$^+$) m/z 293.2 [M+H]$^+$; HRMS (ESI) calcd for C$_{13}$H$_{13}$N$_2$O$_6$ [M+H]$^+$ 293.0768, found 293.0768.
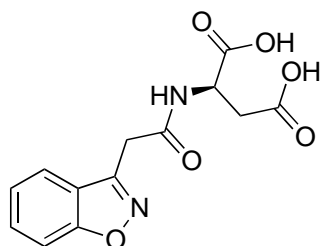
4-Nitrophenyl 2-(benzo[*b*]thiophen-3-yl)acetate, (14)
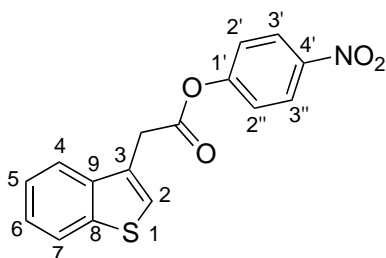
Compound **14** was synthesised according to general procedure C using the following reagents: benzo[*b*]thiophene-3-acetic acid (400 mg, 2.1 mmol), 4-nitrophenol (289 mg, 2.1 mmol), *N'N*-dicyclohexylcarbodiimide (429 mg, 2.1 mmol) and EtOAc (10 mL). The product was purified by recrystallization from petrol:EtOAc (1:1) to yield the *title compound* as a yellow solid (599 mg, 91 %); R$_f$ = 0.5 (petrol:EtOAc, 4:1); m.p. 99-101 °C; λ_{max}(EtOH)/nm 298.4, 266.6, 226.2; IR (neat) v_{max}/cm^{-1} 3318, 2925, 2848, 2118, 2081, 1758, 1615, 1571; ^{1}H NMR (500 MHz, CDCl$_3$) δ 4.16 (2H, s, CH$_2$), 7.25 (2H, d, J = 9.1 Hz, H-2' and H-2''), 7.39-7.47 (2H, m, H-5 and H-6), 7.49 (1H, s, H-2), 7.85 (1H, d, J = 7.7 Hz, H-4), 7.91 (1H, d, J = 7.7 Hz, H-7), 8.24 (1H, d, J = 9.1 Hz, H-3' and H-3''); ^{13}C NMR (125 MHz, CDCl$_3$) δ 34.7 (CH$_2$), 121.6 (C-2), 122.5 (C-2' and C-2''), 123.2 (C-4), 124.6 (C-7), 124.9 (C-5), 125.4 (C-3' and C-3''), 125.6 (C-6), 126.9 (C-3), 130.6 (C-9), 138.4 (C-8), 140.4 (C-4'), 155.4 (C-1'), 168.3 (CO); Compound degraded before HRMS (ESI) analysis could be performed.

(2-(1*H*-Benzo[*b*]thiophen-3-yl)acetyl)-*D*-aspartic acid, (15)
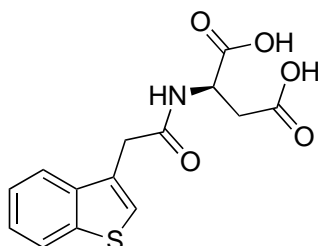
Compound **15** was synthesised according to general procedure D using the following reagents: D-aspartic acid (106 mg, 0.8 mmol), tetramethylguanidine (200 μL, 1.6 mmol), 4-nitrophenyl 2-(benzo[*b*]thiophen-3-yl)acetate (**14**) (250 mg, 0.8 mmol) and aq. MeOH (50 % v/v) (10 mL). Removal of *n*-butanol *in vacuo* yielded the *title compound* as a white solid (38 mg, 15 %); R_f = 0.25 (DCM:MeOH, 2:1); m.p. 156-158 °C; λ_{max}(EtOH)/nm 298.6, 290.0, 260.0; IR (neat) v_{max}/cm^{-1} 3318, 3156, 2920, 2087, 1722, 1649, 1603, 1562; ^{1}H NMR (500 MHz, MeOD) δ 2.81 (1H, dd, *J* = 7.0, 16.8 Hz, C*H*HCOOH), 2.87 (1H, dd, *J* = 5.3, 16.8 Hz, CH*H*COOH), 3.84 (2H, s, C*H$_2$*CONH), 4.75 (1H, dd, *J* = 5.3, 7.0 Hz, C*H*COOH), 7.36 (2H, m, H-5 and H-6), 7.46 (1H, s, H-2), 7.81 (1H, d, *J* = 7.8 Hz, H-4), 7.86 (1H, d, *J* = 7.8 Hz, H-7); ^{13}C NMR (125 MHz, MeOD) δ 35.8 (*C*H$_2$CONH), 36.8 (*C*H$_2$COOH), 48.6 (*C*HCOOH), 122.9 (C-4), 123.6 (C-7), 125.2 (C-5 or C-6), 125.4 (C-5 or C-6), 125.5 (C-2), 130.8 (C-3), 140.1 (C-8 or C-9), 141.6 (C-8 or C-9), 172.8 (CO), 172.9 (CO), 173.9 (CO); LRMS (ES$^+$) *m/z* 308.2 [M+H]$^+$; HRMS (ESI) calcd for C$_{14}$H$_{14}$NO$_5$S [M+H]$^+$ 308.0587, found 308.0585.
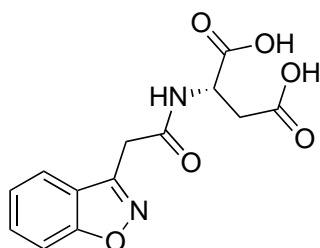
(2-(1*H*-Benzo[*b*]thiophen-3-yl)acetyl)-*L*-aspartic acid, (16)
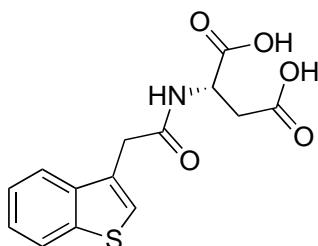
Compound **16** was synthesised according to general procedure D using the following reagents: D-aspartic acid (106 mg, 0.8 mmol), tetramethylguanidine (200 μL, 1.6 mmol), 4-nitrophenyl 2-(benzo[*b*]thiophen-3-yl)acetate (**14**) (250 mg, 0.8 mmol) and aq. MeOH (50 % v/v) (10 mL). Removal of *n*-butanol *in vacuo* yielded the *title compound* as a white solid (32 mg, 13 %); R_f = 0.25 (DCM:MeOH, 2:1); m.p. 156-168

°C; λ_{max}(EtOH)/nm 298.6, 289.8, 261.2; IR (neat) v_{max}/cm^{-1} 3318, 3156, 2920, 2087, 1722, 1649, 1603, 1562; ^{1}H NMR (500 MHz, MeOD) δ 2.81 (1H, dd, J = 7.0, 16.8 Hz, C*H*HCOOH), 2.87 (1H, dd, J = 5.3, 16.8 Hz, CH*H*COOH), 3.84 (2H, s, C*H$_2$*CONH), 4.75 (1H, dd, J = 5.3, 7.0 Hz, C*H*COOH), 7.36 (2H, m, H-5 and H-6), 7.46 (1H, s, H-2), 7.81 (1H, d, J = 7.8 Hz, H-4), 7.86 (1H, d, J = 7.8 Hz, H-7); ^{13}C NMR (125 MHz, MeOD) δ 35.8 (*C*H$_2$CONH), 36.8 (*C*H$_2$COOH), 48.6 (*C*HCOOH), 122.9 (C-4), 123.6 (C-7), 125.2 (C-5 or C-6), 125.4 (C-5 or C-6), 125.5 (C-2), 130.8 (C-3), 140.1 (C-8 or C-9), 141.6 (C-8 or C-9), 172.8 (CO), 172.9 (CO), 173.9 (CO); LRMS (ES$^+$) m/z 308.2 [M+H]$^+$; HRMS (ESI) calcd for $C_{14}H_{14}NO_5S$ [M+H]$^+$ 308.0587, found 308.0584.

N-(Benzo[d]isoxazol-3-yl)-2-phenylacetamide, (17) [219]
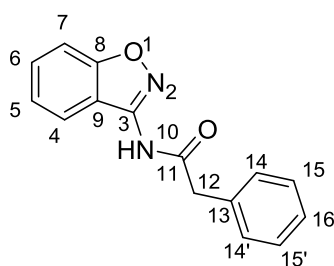
3-Amino-1,2-benzisoxazole (125 mg, 0.93 mmol) was dissolved in DCM (4 mL) before subsequent addition of NEt$_3$ (130 µL, 0.93 mmol) followed by dropwise addition of phenylacetylchloride (123 µL, 0.93 mmol). After stirring the reaction at RT for 18 h, the solvent was removed *in vacuo*. The crude product was purified by MPLC (silica gel, petrol:EtOAc, 95:5), followed by recrystallisation from isopropanol to yield the *title compound* as a white solid (47 mg, 20 %); R$_f$ = 0.5 (petrol:EtOAc, 95:5, UV); m.p. 143 °C; λ_{max} (EtOH)/nm 283.8; IR (neat) v_{max}/cm^{-1} 3253, 3212, 3153, 3032, 2343, 2116, 1998, 1804, 1712, 1621, 1569; ^{1}H NMR (500 MHz, MeOD) δ 3.85 (2H, s, H-12), 7.37 – 7.25 (4H, m, H-5, H-15 and H-15', H-16), 7.40 (2H, d, J = 7.5 Hz, H-14 and H-14'), 7.56 (1H, d, J = 8.2 Hz, H-7), 7.61 (1H, td, J = 1.1, 6.8 Hz, H-6), 7.97 (1H, d, J = 8.2 Hz, H-7); ^{13}C NMR (125 MHz, CDCl$_3$) δ 43.7 (C-12), 110.8 (C-7), 117.5 (C-9), 124.4 (C-5), 124.7 (C-4), 128.2 (C-16), 129.7 (C-15 and C-15'), 130.3 (C-14 and C-14'), 131.8 (C-6), 136.1 (C-3), 154.6 (C-8), 164.9 (CO); LRMS (ES$^+$) m/z 253.2 [M+H]$^+$; HRMS (ESI) calcd for $C_{15}H_{13}N_2O_2$ [M+H]$^+$ 253.0972, found 253.0965.

N-(Benzo[d]isoxazol-3-yl)benzamide, (18), [220]
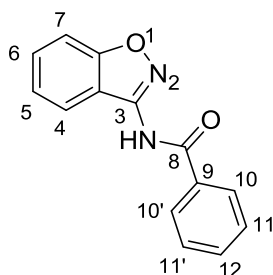
N-(Benzo[d]isoxazol-3-yl)benzamide was synthesised according to general procedure A, using benzoic acid (45 mg, 0.37 mmol), 3-amino-1,2-benzoisoxazole (125 mg, 0.93 mmol), PCl$_3$ (32 µL, 0.37 mmol) and acetonitrile (5 mL). The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 80:20) to yield the *title compound* as a white solid (21 mg, 24 %); R$_f$ = 0.8 (petrol:EtOAc, 4:1); m.p. 126 °C; λ$_{max}$(EtOH)/nm 285.0, 234.8; ν$_{max}$/cm^{-1} 3207, 3149, 3058, 2922, 2731, 2372, 2343, 2109, 1908, 1828, 1669, 1606, 1557; ^{1}H NMR (500 MHz, MeOD) δ 7.37 (1H, ddd, *J* = 1.2, 6.5, 8.3 Hz, H-5), 7.54-7.68 (5H, m, H-6, H-7, H-13 and H-13', H-14), 8.04 (3H, m, H-4, H-12 and H-12');^{13}C NMR (125 MHz, MeOD) δ 110.8 (C-7), 117.9 (C-9), 124.4 (C-5), 125.2 (C-4), 129.3 (C-12 and C-12'), 129.8 (C-13 and C-13'), 131.7 (C-6), 133.8 (C-14), 134.5 (C-11), 155.1 (C-3), 165.2 (C-8), 168.7 (CO); LRMS (ES$^+$) *m/z* 239.2 [M+H]$^+$; HRMS (ESI) calcd for C$_{14}$H$_{11}$N$_2$O$_2$ [M+H]$^+$ 239.0815, found 239.0808.

N-(Benzo[d]isoxazol-3-yl)cyclohexanecarboxamide, (19)
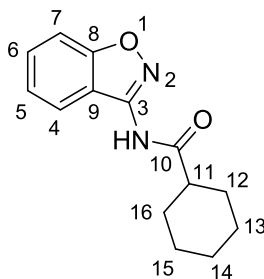
Compound **19** was synthesised according to general procedure A. The reaction mixture contained cyclohexane carboxylic acid (48 mg, 0.37 mmol), 3-amino-1,2-benzisoxazole (125 mg, 0.93 mmol) and PCl$_3$ (32 µL, 0.372 mmol) in acetonitrile (5 mL). The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 90:10) to yield the *title compound* as a white solid (69 mg, 76 %); R$_f$ = 0.7 (petrol:EtOAc,

4:1); m.p. 143-145 °C; λ_{max}(EtOH)/nm 283.6; IR (neat) v_{max}/cm^{-1} 3270, 2921, 2852, 2119, 1662, 1608, 1542; ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.35-1.24 (1H, m, H-14), 1.47-1.35 (2H, m, H-13, H-13', H-15 and H-15'), 1.66-1.55 (2H, m, H-12/H-12' and H-16/H-16'), 1.78-1.71 (1H, m, H-14'), 1.84-1.92 (2H, m, H-13/H-13' and H-15/H-15'), 2.08-2.01 (2H, m, H-12/H-12' and H-16/H-16'), 2.58-2.44 (1H, m, H-11), 7.30 (1H, ddd, J = 0.9, 7.0, 8.5 Hz, H-5), 7.50 (1H, dd, J = 0.9, 8.5 Hz, H-7), 7.56 (1H, ddd, J = 0.9, 7.0, 8.5 Hz, H-6), 8.28 (1H, d, J = 7.0 Hz, H-4), 8.85 (1H, s, NH); ^{13}C NMR (125 MHz, CDCl$_3$) δ 25.7 (C-13/13' and C-15/15'), 25.8 (C-14/14'), 29.6 (C-12/12' and C-16/16'), 45.7 (C-11), 109.9 (C-7), 116.0 (C-9), 123.4 (C-5), 126.2 (C-4), 126.3 (C-3) 130.8 (C-6), 154.2 (C-8), 164.5 (CO); LRMS (ES$^+$) m/z 245.2 [M+H]$^+$; HRMS (ESI) calcd for C$_{14}$H$_{17}$N$_2$O$_2$ [M+H]$^+$ 245.1271, found 245.1276.

4-Vinyl-1_H_-indole, (20)
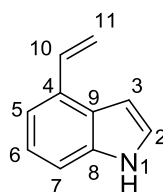
4-Bromo-1_H_-indole (500 mg, 2.55 mmol), caesium carbonate (2.49 g, 7.65 mmol), XPhos (61 mg, 0.128 mmol) and potassium vinyltrifluoroborate (376 mg, 2.81 mmol) were partly dissolved in dry DMF (10 mL). The PdCl$_2$ catalyst was added and the solution degased for 20 minutes. Water was added to the reaction mixture and the reaction subjected to microwave irradiation at 120 °C for 15 min. The reaction mixture was filtered through Celite and washed with methanol. The solvent was removed _in vacuo_ and the crude product purified by silica-based MPCL (PE: EtOAc = 95:5) to yield the _title compound_ as an orange oil (207.5 mg, 1.45 mmol, 57 %); R$_f$ = 0.5 (petrol:EtOAc, 3:2); λ_{max}(EtOH)/nm 220.4, 302.2; IR (neat) v_{max}/cm^{-1} 3400, 2095, 1894, 1829, 1703, 1601; ^{1}H NMR (500 MHz, CDCl$_3$) δ 5.42 (1H, dd, J = 2.1, 11.0 Hz, H-11$_{cis}$), 5.95 (1H, dd, J = 2.1, 17.6 Hz, H-11$_{trans}$), 6.78-6.81 (1H, m, H-3), 7.17 (1H, dd, J = 11.0, 17.6 Hz, H-10), 7.21 (1H, dd, J = 7.4, 8.0 Hz, H-6), 7.24 (1H, t, J = 2.8 Hz, H-2), 7.30 (1H, d, J = 7.4 Hz, H-5), 7.32 (1H, d, J = 8.0 Hz, H-7), 8.15 (1H, s, NH); ^{13}C NMR (125 MHz, CDCl$_3$) δ 101.2 (C-3), 110.7 (C-7), 114.7 (C-11), 117.5 (C-5), 122.2 (C-6), 124.5 (C-2), 126.1 (C-9), 130.2 (C-4), 135.5 (C-10), 136.3 (C-8);

LRMS (ES$^+$) *m/z* 144.1 [M+H]$^+$; Compound degraded before HRMS (ESI) analysis could be performed.

4-Ethyl-1*H*-indole, (21)
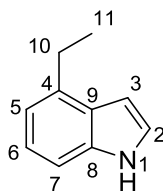
4-Vinyl-1*H*-indole (**21**) (75 mg, 0.52 mmol) in EtOAc:MeOH (5:1, 10 mL) was subjected to palladium-catalysed hydrogenation using a balloon filled with H2 and a 10 % Pd/C catalyst. The reaction was conducted at RT for 4 h. The reaction mixture was filtered through Celite and the filtrate concentrated *in vacuo*. The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 95:5) to yield the *title compound* as a pale rose oil (39 mg, 52 %). R$_f$ = 0.7 (petrol:EtOAc, 4:1); λ$_{max}$(EtOH)/nm 270.0, 218.8; IR (neat) v$_{max}$/cm^{-1} 3400, 3051, 2963, 2929, 2869, 2116, 1890, 1717, 1582, 1500; ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.37 (3H, t, *J* = 7.6 Hz, H-11), 2.96 (2H, q, *J* = 7.6 Hz, H-10), 6.60-6.63 (1H, m, H-3), 6.97 (1H, d, *J* = 7.4 Hz, H-5), 7.16 (1H, dd, *J* = 7.4, 8.2 Hz, H-6), 7.20 (1H, dd, *J* = 2.8, 3.2 Hz, H-2), 7.26 (1H, d, *J* = 8.2 Hz, H-7), 8.13 (1H, s, NH); ^{13}C NMR (125 MHz, CDCl$_3$) δ 14.9 (C-11), 26.5 (C-10), 101.1 (C-3), 108.9 (C-7), 118.3 (C-5), 122.4 (C-6), 123.6 (C-2), 127.1 (C-9), 135.8 (C-8), 136.7 (C-4); LRMS (ES$^+$) *m/z* 146.2 [M+H]$^+$; Compound degraded before HRMS (ESI) analysis could be performed.

1-(1*H*-Indol-4-yl)ethane-1,2-diol, (22)
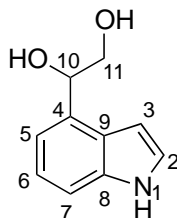
4-Vinyl-1*H*-indole (30 mg, 0.21 mmol) was dissolved in 7 mL of a THF:water (3:1) mixture and cooled to 0 °C. *N*-methyl morpholine-*N*-oxide (33 mg, 0.29 mmol) and osmium tetroxide solution (2.5 % wt in t-BuOH) (70 µL, 0.007 mmol) were added and the reaction mixture warmed to RT and stirred for 48 h. The reaction was quenched

with sat. aq. Na$_2$SO$_4$ solution (15 mL) and stirred at RT for 30 min. The organic fractions were extracted with EtOAc (3 x 20 mL) and the combined fractions dried over MgSO$_4$ and concentrated *in vacuo*. The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 80:20) to yield the *title compound* as a green oil (21 mg, 57 %); R$_f$ = 0.2 (petrol:EtOAc, 1:1); λ$_{max}$(EtOH)/nm 278.6, 218.8; IR (neat) v$_{max}$/cm^{-1} 3306, 2924, 2874, 2487, 2114, 1900, 1684, 1612, 1501; ^{1}H NMR (500 MHz, MeOD) δ 3.72 (1H, dd, *J* = 11.4, 8.1 Hz, C*H$_2$*OH), 3.80 (1H, dd, *J* = 11.4, 3.8 Hz, C*H$_2$*OH), 5.14 (1H, dd, *J* = 8.1, 3.8 Hz, C*H*OH), 6.58 (1H, dd, *J* = 3.2, 0.8 Hz, H-3), 7.11-7.08 (2H, m, H-5 and H-6), 7.23 (1H, d, *J* = 3.2 Hz, H-2), 7.33-7.28 (1H, m, H-7); ^{13}C NMR (125 MHz, MeOD) δ 68.2 (C-11), 75.0 (C-10), 100.5 (C-3), 111.6 (C-7), 117.1 (C-5), 122.1 (C-6), 125.4 (C-2), 127.1 (C-9), 134.1 (C-4), 137.8 (C-8); HRMS (ESI) calcd for C$_{10}$H$_{10}$NO$_2$ [M+H]$^-$ 176.0717, found 176.0719.

N-(Pyrimidin-4-yl)-1*H*-indole-4-carboxamide, (23)
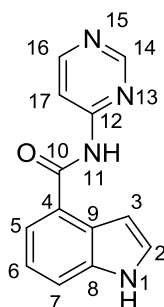
Compound **24** was synthesised according to general procedure A. The reaction mixture contained 1*H*-indole-4-carboxylic acid (50 mg, 0.31 mmol), 4-aminopyrimidine (74 mg, 0.78 mmol) and PCl$_3$ (27 µL, 0.31 mmol) in acetonitrile (5 mL). The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 90:10) to yield the *title compound* as a pale brown solid (10 mg, 14 %); R$_f$ = 0.1 (petrol:EtOAc, 1:1); m.p. 180-182°C; λ$_{max}$(EtOH)/nm 315.2, 266.2, 205.4; IR (neat) v$_{max}$/cm^{-1} 3319, 2920, 2851, 2474, 2113, 1667, 1571; ^{1}H NMR (500 MHz, MeOD) δ 6.96 (1H, d, *J* = 3.1 Hz, H-3), 7.25 (1H, *p*t, *J* = 7.8 Hz, H-6), 7.44 (1H, d, *J* = 3.1 Hz, H-2), 7.68-7.62 (2H, m, H-5 and H-7), 8.40 (1H, dd, *J* = 1.2, 5.5 Hz, H-17), 8.67 (1H, d, *J* = 5.5 Hz, H-16), 8.87 (s, 1H, H-14); ^{13}C NMR (125 MHz, MeOD) δ 102.6 (C-3), 111.8 (C-17), 117.0 (C-5), 121.1 (C-7), 121.6 (C-6), 126.1 (C-9), 127.8 (C-4), 128.2 (C-2), 138.6 (C-8), 158.7 (C-16), 159.3 (C-14), 160.1 (C-12), 170.5 (CO); LRMS

(ES⁺) *m/z* 239.3 [M+H]⁺; HRMS (ESI) calcd for $C_{13}H_{11}N_4O$ [M+H]⁺ 239.0927, found 239.0926.

3-Methoxybenzo[*d*]isoxazole, (24)
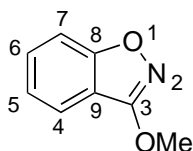
Compound **25** was synthesised according to general procedure B, using the following reagents: 3-hydroxybenzisoxazole (50 mg, 0.37 mmol), caesium carbonate (362 mg, 1.11 mmol), methyl iodide (69 μL, 1.11 mmol) and acetonitrile (2 mL). The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 0:1) to yield the *title compound* as a clear oil (24 mg, 43 %); R_f = 0.9 (petrol:EtOAc, 1:3); λ_{max}(EtOH)/nm 281.4, 236.6; IR (neat) v_{max}/cm⁻¹ 2943, 1919, 1613, 1544; ¹H NMR (500 MHz, CDCl₃) δ 4.17 (3H, s, OC*H₃*), 7.26 (1H, ddd, *J* = 0.8, 7.0, 8.5 Hz, H-5), 7.43 (1H, ddd, *J* = 0.8, 1.0, 8.5 Hz, H-7), 7.52 (1H, ddd, *J* = 1.2, 7.0, 8.5 Hz, H-6), 7.62 (1H, ddd, *J* = 1.0, 1.2, 8.5 Hz, H-4); Not sufficient material for ¹³C NMR; LRMS (ES⁺) *m/z* 150.1 [M+H]⁺; Compound degraded before HRMS (ESI) analysis could be performed.

3-Ethoxybenzo[*d*]isoxazole, (25)
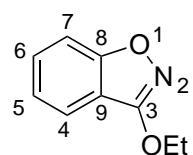
Compound **26** was synthesised according to general procedure B, using the following reagents: 3-hydroxybenzisoxazole (50 mg, 0.37 mmol), caesium carbonate (362 mg, 1.11 mmol), methyl iodide (83 μL, 1.11 mmol) and acetonitrile (2 mL). The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 0:1) to yield the *title compound* as a white solid (31 mg, 51 %); R_f = 0.8 (petrol:EtOAc, 1:3); m.p. 30-31 °C; λ_{max}(EtOH)/nm 281.8, 236.6; IR (neat) v_{max}/cm⁻¹ 3058, 2978, 2278, 2111, 1915, 1795, 1611, 1534; ¹H NMR (500 MHz, CDCl₃) δ 1.52 (3H, t, *J* = 7.1 Hz, C*H₃*), 4.51 (2H, q, *J* = 7.1 Hz, C*H₂*), 7.26 (1H, td, *J* = 0.7, 7.5 Hz, H-5), 7.43 (1H, ddd, *J* = 0.7, 1.0, 8.5 Hz, H-7), 7.52 (1H, ddd, *J* = 1.2, 7.0, 8.5 Hz, H-6), 7.62 (1H, ddd, *J* = 1.0, 1.2, 8.5 Hz, H-4); ¹³C NMR (125 MHz, CDCl₃) δ 14.8 (CH₃), 66.4 (CH₂), 110.3 (C-7),

114.6 (C-9), 121.0 (C-4), 123.0 (C-5), 130.5 (C-6), 164.0 (C-3), 166.6 (C-8); LRMS (ES$^+$) m/z 164.1 [M+H]$^+$; HRMS (ESI) calcd for $C_9H_{10}NO_2$ [M+H]$^+$ 164.1434, found 164.1431.

N-(3-Chloro-4-methoxyphenyl)-N-methylacetamide, (26)
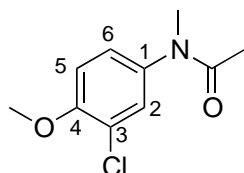
To a solution of *N*-(3-chloro-4-methoxyphenyl) (50 mg, 0.25 mmol) in dry THF was added NaH (60 % dispersion in mineral oil, 15 mg, 0.375 mmol) at 0 °C. After stirring for 10 min, methyl iodide (0.12 mL, 1.875 mmol) was added and the reaction stirred at RT for 2 h. The mixture was concentrated in vacuo and the remaining residue dissolved in water and the mixture extracted with Et$_2$O (3 x 10 mL). The organic fractions were combined, washed with brine, dried over MgSO4 and concentrated in vacuo. The crude product was purified by recrystallization from hexane to yield the *title compound* as white solid (23 mg, 43 %); R$_f$ = 0.3 (petrol:EtOAc, 1:3); m.p. 76-78 °C; λ_{max}(EtOH)/nm 280.8, 232.4; IR (neat) v$_{max}$/cm^{-1} 3034, 2930, 2837, 2321, 2086, 1982, 1890, 1646; ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.89 (3H, s, COC*H$_3$*), 3.24 (3H, s, NC*H$_3$*), 3.95 (3H, s, OC*H$_3$*), 6.96 (1H, d, *J* = 8.5 Hz, H-5), 7.09 (1H, dd, *J* = 2.6, 8.5 Hz, H-6), 7.26 (1H, d, *J* = 2.6 Hz, H-2); ^{13}C NMR (125 MHz, CDCl$_3$) δ 22.4 (CO*C*H$_3$), 37.3 (N*C*H$_3$), 56.4 (O*C*H$_3$), 112.4 (C-5), 123.1 (C-3), 126.5 (C-6), 129.1 (C-2), 137.8 (C-1), 154.6 (C-4), 170.7 (*C*O); LRMS (ES$^+$) m/z 214.2 [M+H]$^+$; Compound degraded before HRMS (ESI) analysis could be performed.

N-(4-Ethoxyphenyl)acetamide, (28)
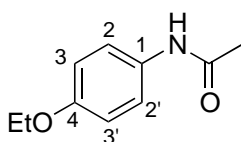
Compound **28** was synthesised according to general procedure B. The reaction mixture contained 4-acetamidophenol (100 mg, 0.66 mmol), caesium carbonate (645 mg, 1.98 mmol), ethyl bromide (148 µL, 1.98 mmol) and acetonitrile (4 mL). The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 0:1) to yield the

title compound as a white solid (96 mg, 81 %); R_f = 0.4 (petrol:EtOAc, 1:3); m.p. 133-134 °C; λ_{max}(EtOH)/nm 249.2; IR (neat) v_{max}/cm^{-1} 3279, 3193, 3131, 3074, 2983, 2928, 2886, 2809, 2118, 1882, 1657, 1605, 1553; ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.39 (3H, s, OCH$_2$C*H$_3$*), 2.18 (3H, s, COC*H$_3$*), 4.00 (2H, q, *J* = 8.9 Hz, OC*H$_2$*CH$_3$), 6.84 (2H, d, *J* = 8.9 Hz, H-3 and H-3'), 7.37 (2H, d, *J* = 8.9 Hz, H-2 and H-2'); ^{13}C NMR (125 MHz, CDCl$_3$) δ 15.0 (OCH$_2$*C*H$_3$), 24.3 (CO*C*H$_3$), 63.9 (O*C*H$_2$CH$_3$), 114.9 (C-3 and C-3'), 122.2 (C-2 and C-2'), 130.7 (C-1), 156.2 (C-4), 168.6 (*C*O); LRMS (ES$^+$) *m/z* 180.2 [M+H]$^+$; HRMS (ESI) calcd for C$_{10}$H$_{14}$NO$_2$ [M+H]$^+$ 180.1019, found 180.1011.

N-(4-Isopropoxyphenyl)acetamide, (29)
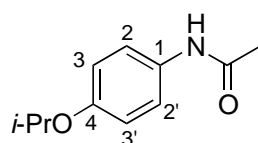
Compound **29** was synthesised according to general procedure B. The reaction mixture contained 4-acetamidophenol (100 mg, 0.66 mmol), caesium carbonate (645 mg, 1.98 mmol), ethyl bromide (148 μL, 1.98 mmol) and acetonitrile (4 mL). The crude product was purified by MPLC (silica gel, petrol:EtOAc, 1:0 → 1:1) to yield the *title compound* as a white solid (116 mg, 90 %); R_f = 0.5 (petrol:EtOAc, 1:3); m.p. 130-131 °C; λ_{max}(EtOH)/nm 250.2; IR (neat) v_{max}/cm^{-1} 3239, 3183, 3122, 3066, 2975, 2929, 2866, 2099, 2005, 1645, 1607, 1557, 1503; ^{1}H NMR (500 MHz, CDCl$_3$) δ 1.31 (6H, d, *J* = 6.1 Hz, 2 x CH$_3$), 2.18 (3H, s, COC*H$_3$*), 4.53-4.45 (1H, m, COC*H*), 6.84 (2H, d, *J* = 8.9 Hz, H-3 and H-3'), 7.36 (2H, d, *J* = 8.9 Hz, H-2 and H-2'); ^{13}C NMR (125 MHz, CDCl$_3$) δ 22.2 (2 x CH$_3$), 24.3 (CO*C*H$_3$), 70.5 (O*C*H), 116.5 (C-3 and C-3'), 122.2 (C-2 and C-2'), 130.7 (C-1), 155.1 (C-4), 168.6 (CO); LRMS (ES$^+$) *m/z* 194.2 [M+H]$^+$; HRMS (ESI) calcd for C$_{11}$H$_{16}$NO$_2$ [M+H]$^+$ 194.1176, found 194.1169.

Chapter 10. Appendix

10.1 Data collection and phasing statistics

The following section shows the data collection and phasing statistics. The data were processed, scaled and merged using xia2, aimless, pointless and programs from the CCP4i2 suite.

Benzamide series

(1) Fragment D (Data collected on a single crystal at 100 K on beamline I03, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.11	39.11	1.53
High resolution limit	1.5	8.22	1.5
R_{merge}**(within I+/I-)**	0.075	0.04	0.604
R_{merge}**(all I+ and I-)**	0.075	0.031	0.637
R_{meas} **(within I+/I-)**	0.105	0.056	0.855
R_{meas} **(all I+ & I-)**	0.102	0.041	0.88
R_{pim} **(within I+/I-)**	0.075	0.04	0.604
R_{pim} **(all I+ & I-)**	0.07	0.026	0.605
R_{merge} **in top intensity bin**	0.05		
Number of observations	114413	801	5577
Number unique	52769	359	2565
Mean((I)/sd(I))	5.5	14.5	1.3
Half-set correlation CC(1/2)	0.912	0.991	0.293
Completeness	88.1	98.4	87.2
Multiplicity	2.2	2.2	2.2
Anomalous completeness	76	98.1	70.7
Anomalous multiplicity	1	1.1	1.1
DelAnom CC(1/2)	-0.862	-0.586	0
Mid-Slope of Anom Probability	0.405		

(2) Fragment M (Data collected on a single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.19	39.19	1.42
High resolution limit	1.4	7.67	1.4
R_{merge}(within I+/I-)	0.168	0.05	1.807
R_{merge}(all I+ and I-)	0.117	0.028	1.091
R_{meas} (within I+/I-)	0.237	0.071	2.555
R_{meas} (all I+ & I-)	0.16	0.038	1.489
R_{pim} (within I+/I-)	0.168	0.05	1.807
R_{pim} (all I+ & I-)	0.108	0.025	1.008
R_{merge} in top intensity bin	0.096		
Number of observations	130615	842	4111
Number unique	68698	431	2450
Mean((I)/sd(I))	2.6	7.1	0.2
Half-set correlation CC(1/2)	0.927	0.993	0.554
Completeness	93.5	96.4	67.8
Multiplicity	1.9	2	1.7
Anomalous completeness	61.7	75.8	31
Anomalous multiplicity	0.7	1.1	1.1
DelAnom CC(1/2)	-0.818	-0.721	0
Mid-Slope of Anom Probability	0.394		

(3) Fragment N (Data collected on a single crystal at 100 K on beamline I04-1, DLS.)

	Overall	Inner	Outer
Low resolution limit	42	42	1.63
High resolution limit	1.6	8.76	1.6
R_{merge} (within I+/I-)	0.06	0.04	0.488
R_{merge} (all I+ and I-)	0.073	0.04	0.483
R_{meas} (within I+/I-)	0.085	0.057	0.69
R_{meas} (all I+ & I-)	0.099	0.054	0.653
R_{pim} (within I+/I-)	0.06	0.04	0.488
R_{pim} (all I+ & I-)	0.067	0.036	0.437
R_{merge} in top intensity bin	0.043		
Number of observations	92858	627	4685
Number unique	43080	283	2162
Mean((I)/sd(I))	8.1	19.1	1.8
Half-set correlation CC(1/2)	0.975	0.987	0.714
Completeness	89.1	97.3	90.6
Multiplicity	2.2	2.2	2.2
Anomalous completeness	77.5	96.6	74.6
Anomalous multiplicity	1	1.1	1.1
DelAnom CC(1/2)	-0.601	-0.594	0
Mid-Slope of Anom Probability	1.082		

Isoxazole series

(4) Fragment O (Data collected on a single crystal at 100 K on beamline I04-1, DLS.)

	Overall	Inner	Outer
Low resolution limit	38.78	38.78	1.32
High resolution limit	1.3	7.12	1.3
R_{merge}(within I+/I-)	0.056	0.028	0.822
R_{merge}(all I+ and I-)	0.045	0.024	0.635
R_{meas} (within I+/I-)	0.079	0.04	1.163
R_{meas} (all I+ & I-)	0.063	0.034	0.899
R_{pim} (within I+/I-)	0.056	0.028	0.822
R_{pim} (all I+ & I-)	0.045	0.024	0.635
R_{merge} in top intensity bin	0.026		
Number of observations	153124	1088	7334
Number unique	78176	545	3737
Mean((I)/sd(I))	9.2	25.7	1.3
Half-set correlation CC(1/2)	0.997	0.996	0.56
Completeness	86.3	98.6	83.9
Multiplicity	2	2	2
Anomalous completeness	63.9	94.6	57.4
Anomalous multiplicity	0.7	1	1
DelAnom CC(1/2)	0	0	0
Mid-Slope of Anom Probability	1.047		

(5) Fragment O1 (Data collected on a single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	42.51	42.51	1.6
High resolution limit	1.57	8.6	1.57
R_{merge}(within I+/I-)	0.127	0.037	0.309
R_{merge}(all I+ and I-)	0.105	0.047	0.228
R_{meas} (within I+/I-)	0.179	0.053	0.437
R_{meas} (all I+ & I-)	0.143	0.064	0.311
R_{pim} (within I+/I-)	0.127	0.037	0.309
R_{pim} (all I+ & I-)	0.096	0.044	0.211
R_{merge} in top intensity bin	0.086		
Number of observations	97115	598	4537
Number unique	50309	305	2468
Mean((I)/sd(I))	3.5	6.5	1
Half-set correlation CC(1/2)	0.946	0.978	0.922
Completeness	96.1	95.9	94.3
Multiplicity	1.9	2	1.8
Anomalous completeness	66.4	77.3	54.9
Anomalous multiplicity	0.8	1.1	1.1
DelAnom CC(1/2)	-0.567	0.47	0
Mid-Slope of Anom Probability	0.451		

(6) Fragment O2 (Data collected on a single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	50.25	50.25	1.42
High resolution limit	1.4	7.67	1.4
R_{merge}(within I+/I-)	0.067	0.027	1.183
R_{merge}(all I+ and I-)	0.051	0.02	0.817
R_{meas} (within I+/I-)	0.095	0.038	1.672
R_{meas} (all I+ & I-)	0.07	0.026	1.113
R_{pim} (within I+/I-)	0.067	0.027	1.183
R_{pim} (all I+ & I-)	0.047	0.017	0.751
R_{merge} in top intensity bin	0.03		
Number of observations	102892	831	1190
Number unique	55928	425	763
Mean((I)/sd(I))	6	15.9	0.5
Half-set correlation CC(1/2)	0.993	0.999	0.45
Completeness	77	96.6	21.1
Multiplicity	1.8	2	1.6
Anomalous completeness	48.1	77.7	7.8
Anomalous multiplicity	0.7	1.1	1.1
DelAnom CC(1/2)	-0.543	-0.608	0
Mid-Slope of Anom Probability	0.612		

(7) Fragment O4 (Data collected on a single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.03	39.03	1.73
High resolution limit	1.7	9	1.7
R_{merge}(within I+/I-)	0.108	0.047	0.242
R_{merge}(all I+ and I-)	0.073	0.031	0.272
R_{meas} (within I+/I-)	0.153	0.067	0.343
R_{meas} (all I+ & I-)	0.099	0.042	0.374
R_{pim} (within I+/I-)	0.108	0.047	0.242
R_{pim} (all I+ & I-)	0.067	0.028	0.255
R_{merge} in top intensity bin	0.045		
Number of observations	73697	503	3176
Number unique	39410	264	1997
Mean((I)/sd(I))	4.2	9.7	0.9
Half-set correlation CC(1/2)	0.985	0.994	0.404
Completeness	96.2	95	89.8
Multiplicity	1.9	1.9	1.6
Anomalous completeness	63.7	71.8	38
Anomalous multiplicity	0.7	1.1	1.1
DelAnom CC(1/2)	-0.674	-0.782	0
Mid-Slope of Anom Probability	0.371		

(8) Fragment O5 (Data collected on a single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.12	39.12	1.5
High resolution limit	1.47	8.05	1.47
R_{merge}(within I+/I-)	0.055	0.035	0.527
R_{merge}(all I+ and I-)	0.039	0.041	0.348
R_{meas} (within I+/I-)	0.077	0.049	0.745
R_{meas} (all I+ & I-)	0.053	0.056	0.475
R_{pim} (within I+/I-)	0.055	0.035	0.527
R_{pim} (all I+ & I-)	0.036	0.038	0.321
R_{merge} in top intensity bin	0.031		
Number of observations	118832	723	5212
Number unique	60640	372	2896
Mean((I)/sd(I))	7.8	16.1	1.6
Half-set correlation CC(1/2)	0.996	0.985	0.911
Completeness	95.9	95.9	93.1
Multiplicity	2	1.9	1.8
Anomalous completeness	67.4	74.9	51
Anomalous multiplicity	0.8	1.1	1.1
DelAnom CC(1/2)	-0.492	-0.322	0
Mid-Slope of Anom Probability	0.746		

(9) Fragment O6, (Data collected on a single crystal at 100 K on beamline I02, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.06	39.06	1.37
High resolution limit	1.35	7.39	1.35
R_{merge}(within I+/I-)	0.077	0.095	0.436
R_{merge}(all I+ and I-)	0.057	0.055	0.641
R_{meas} (within I+/I-)	0.109	0.134	0.616
R_{meas} (all I+ & I-)	0.076	0.073	0.866
R_{pim} (within I+/I-)	0.077	0.095	0.436
R_{pim} (all I+ & I-)	0.05	0.047	0.579
R_{merge} in top intensity bin	0.044		
Number of observations	150732	1056	5614
Number unique	76482	478	2975
Mean((I)/sd(I))	5.3	10.1	1.2
Half-set correlation CC(1/2)	0.994	0.977	0.363
Completeness	94.2	97	73.9
Multiplicity	2	2.2	1.9
Anomalous completeness	65.7	95.9	43.4
Anomalous multiplicity	0.8	1.1	1.1
DelAnom CC(1/2)	-0.693	-0.61	0
Mid-Slope of Anom Probability	0.731		

Indole series

(10) Fragment H (Data collected on a single crystal at 100 K on beamline I03, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.04	39.04	1.53
High resolution limit	1.5	8.22	1.5
R_{merge}(within I+/I-)	0.157	0.081	0.847
R_{merge}(all I+ and I-)	0.119	0.046	0.768
R_{meas} (within I+/I-)	0.222	0.114	1.197
R_{meas} (all I+ & I-)	0.162	0.062	1.046
R_{pim} (within I+/I-)	0.157	0.081	0.847
R_{pim} (all I+ & I-)	0.109	0.041	0.706
R_{merge} in top intensity bin	0.077		
Number of observations	113904	788	5685
Number unique	52816	357	2640
Mean((I)/sd(I))	2.4	5.3	1
Half-set correlation CC(1/2)	0.914	0.984	0.253
Completeness	88.1	98.1	88.1
Multiplicity	2.2	2.2	2.2
Anomalous completeness	75.4	98.1	71
Anomalous multiplicity	1	1.1	1.1
DelAnom CC(1/2)	-0.36	-0.472	0
Mid-Slope of Anom Probability	0.187		

(11) Fragment H1 (Data collected on single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	42.45	42.45	1.53
High resolution limit	1.5	8.22	1.5
R_{merge}(within I+/I-)	0.072	0.028	0.594
R_{merge}(all I+ and I-)	0.055	0.02	0.608
R_{meas} (within I+/I-)	0.102	0.039	0.84
R_{meas} (all I+ & I-)	0.075	0.026	0.829
R_{pim} (within I+/I-)	0.072	0.028	0.594
R_{pim} (all I+ & I-)	0.05	0.017	0.56
R_{merge} in top intensity bin	0.031		
Number of observations	110732	707	5229
Number unique	56869	347	2749
Mean((I)/sd(I))	4.8	12.3	1.1
Half-set correlation CC(1/2)	0.996	0.997	0.517
Completeness	96	96.7	94
Multiplicity	1.9	2	1.9
Anomalous completeness	67.3	81.8	58.8
Anomalous multiplicity	0.8	1.1	1.1
DelAnom CC(1/2)	-0.506	-0.461	0
Mid-Slope of Anom Probability	0.642		

(12) Fragment H2 (Data collected on a single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	50.41	50.41	1.32
High resolution limit	1.3	7.12	1.3
R_{merge}(within I+/I-)	0.039	0.018	0.953
R_{merge}(all I+ and I-)	0.032	0.023	0.686
R_{meas} (within I+/I-)	0.055	0.025	1.348
R_{meas} (all I+ & I-)	0.043	0.032	0.932
R_{pim} (within I+/I-)	0.039	0.018	0.953
R_{pim} (all I+ & I-)	0.029	0.022	0.628
R_{merge} in top intensity bin	0.022		
Number of observations	162937	1067	4669
Number unique	83440	539	2630
Mean((I)/sd(I))	9.8	31.4	0.8
Half-set correlation CC(1/2)	0.996	0.994	0.638
Completeness	92.1	97.6	59.1
Multiplicity	2	2	1.8
Anomalous completeness	62.7	79.5	29.4
Anomalous multiplicity	0.8	1.1	1.1
DelAnom CC(1/2)	-0.331	0.088	0
Mid-Slope of Anom Probability	0.78		

(13) Fragment H3 (Data collected on single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.07	39.07	1.37
High resolution limit	1.35	7.39	1.35
R_{merge}(within I+/I-)	0.077	0.027	0.605
R_{merge}(all I+ and I-)	0.061	0.025	0.692
R_{meas} (within I+/I-)	0.11	0.039	0.855
R_{meas} (all I+ & I-)	0.082	0.034	0.949
R_{pim} (within I+/I-)	0.077	0.027	0.605
R_{pim} (all I+ & I-)	0.055	0.023	0.646
R_{merge} in top intensity bin	0.042		
Number of observations	147861	958	5805
Number unique	77486	487	3474
Mean((I)/sd(I))	4.2	10.1	0.7
Half-set correlation CC(1/2)	0.994	0.994	0.292
Completeness	94.6	97.4	85.9
Multiplicity	1.9	2	1.7
Anomalous completeness	62.3	79.6	38
Anomalous multiplicity	0.7	1.1	1.1
DelAnom CC(1/2)	-0.799	-0.665	0
Mid-Slope of Anom Probability	0.626		

(14) Fragment 21 (Data collected on a single crystal at 100 K on beamline I04-1, DLS.)

	Overall	Inner	Outer
Low resolution limit	38.92	38.92	1.32
High resolution limit	1.3	7.12	1.3
R_{merge}(within I+/I-)	0.177	0.115	0.721
R_{merge}(all I+ and I-)	0.102	0.053	0.768
R_{meas} (within I+/I-)	0.251	0.162	1.019
R_{meas} (all I+ & I-)	0.134	0.068	1.032
R_{pim} (within I+/I-)	0.177	0.115	0.721
R_{pim} (all I+ & I-)	0.085	0.042	0.685
R_{merge} in top intensity bin	0.085		
Number of observations	172830	1120	8527
Number unique	86485	547	4172
Mean((I)/sd(I))	2.6	4.2	0.7
Half-set correlation CC(1/2)	0.983	0.988	0.495
Completeness	95.6	99.3	93
Multiplicity	2	2	2
Anomalous completeness	68.6	85.1	64.8
Anomalous multiplicity	0.8	1.1	1.1
DelAnom CC(1/2)	-0.716	-0.289	0
Mid-Slope of Anom Probability	0.64		

(15) Fragment 22 (Data collected on a single crystal at 100 K on beamline I04-1, DLS.)

	Overall	Inner	Outer
Low resolution limit	49.92	49.92	1.37
High resolution limit	1.35	7.14	1.35
R_{merge}(within I+/I-)	0.144	0.077	0.464
R_{merge}(all I+ and I-)	0.094	0.053	0.66
R_{meas} (within I+/I-)	0.203	0.108	0.656
R_{meas} (all I+ & I-)	0.126	0.07	0.897
R_{pim} (within I+/I-)	0.144	0.077	0.464
R_{pim} (all I+ & I-)	0.083	0.045	0.604
R_{merge} in top intensity bin	0.073		
Number of observations	152934	1095	7691
Number unique	77017	542	3974
Mean((I)/sd(I))	3.1	5	0.9
Half-set correlation CC(1/2)	0.982	0.983	0.531
Completeness	95.8	99.6	93.1
Multiplicity	2	2	1.9
Anomalous completeness	68.6	84.2	58.8
Anomalous multiplicity	0.8	1.1	1.1
DelAnom CC(1/2)	-0.56	-0.282	0
Mid-Slope of Anom Probability	0.696		

Other heterocyclic ring systems

(16) Fragment I (Data collected on a single crystal at 100 K on beamline I03, DLS.)

	Overall	Inner	Outer
Low resolution limit	42.57	42.57	1.44
High resolution limit	1.42	7.78	1.42
R_{merge}(within I+/I-)	0.092	0.04	1.37
R_{merge}(all I+ and I-)	0.089	0.042	2.046
R_{meas} (within I+/I-)	0.13	0.057	1.937
R_{meas} (all I+ & I-)	0.122	0.057	2.793
R_{pim} (within I+/I-)	0.092	0.04	1.37
R_{pim} (all I+ & I-)	0.083	0.038	1.891
R_{merge} in top intensity bin	0.056		
Number of observations	133955	941	6107
Number unique	62238	427	2968
Mean((I)/sd(I))	3.3	10	0.1
Half-set correlation CC(1/2)	0.985	0.988	0.363
Completeness	87.9	98.8	84.1
Multiplicity	2.2	2.2	2.1
Anomalous completeness	74.4	98.6	61.4
Anomalous multiplicity	0.9	1.1	1.1
DelAnom CC(1/2)	-0.247	0.061	0
Mid-Slope of Anom Probability	0.543		

(17) Fragment B (Data collected on a single crystal at 100 K on beamline I04-1, DLS.)

	Overall	Inner	Outer
Low resolution limit	42.52	42.52	1.4
High resolution limit	1.38	7.56	1.38
R_{merge}(within I+/I-)	0.069	0.045	0.876
R_{merge}(all I+ and I-)	0.059	0.041	0.637
R_{meas} (within I+/I-)	0.097	0.063	1.239
R_{meas} (all I+ & I-)	0.083	0.058	0.9
R_{pim} (within I+/I-)	0.069	0.045	0.876
R_{pim} (all I+ & I-)	0.059	0.041	0.637
R_{merge} in top intensity bin	0.036		
Number of observations	125963	900	6164
Number unique	64601	451	3149
Mean((I)/sd(I))	8.7	21.7	1.2
Half-set correlation CC(1/2)	0.988	0.989	0.496
Completeness	85.4	97.6	83.2
Multiplicity	1.9	2	2
Anomalous completeness	64.1	94.1	56.8
Anomalous multiplicity	0.8	1	1
DelAnom CC(1/2)	0	0	0
Mid-Slope of Anom Probability	0.997		

(18) Fragment B1 (Data collected on a single crystal at 100 K on beamline I04, DLS.)

	Overall	Inner	Outer
Low resolution limit	39.16	39.16	1.63
High resolution limit	1.6	8.76	1.6
R_{merge}(within I+/I-)	0.054	0.036	0.328
R_{merge}(all I+ and I-)	0.057	0.021	0.547
R_{meas} (within I+/I-)	0.076	0.052	0.465
R_{meas} (all I+ & I-)	0.078	0.029	0.755
R_{pim} (within I+/I-)	0.054	0.036	0.328
R_{pim} (all I+ & I-)	0.053	0.019	0.517
R_{merge} in top intensity bin	0.03		
Number of observations	89064	542	3225
Number unique	46883	283	2003
Mean((I)/sd(I))	8	22.8	1.2
Half-set correlation CC(1/2)	0.992	0.998	0.571
Completeness	95.6	96.3	82.9
Multiplicity	1.9	1.9	1.6
Anomalous completeness	64.6	73	36.1
Anomalous multiplicity	0.7	1.1	1.1
DelAnom CC(1/2)	-0.577	-0.745	0
Mid-Slope of Anom Probability	0.883		

10.2 Data refinement statistics

The following section contains the data refinement statistics. The refinement statistics were generated with Refmac5 as part of CCP4i2 suite.

Benzamide series

Refinement	Fragment D	Fragment M	Fragment N
R_{work}/R_{free}	0.27/0.29	0.35/0.37	0.19/0.24
Mean B-value ($Å^2$)			
All atoms	24	10	29
Protein	19	8	27
Water	23	9	32
Ligand	29	12	28
No of atoms			
Protein	5830	5811	5830
Water	167	97	122
Ligand	16	17	23
Rmsd			
Bond lengths (Å)	0.02	0.02	0.02
Angles (°)	1.80	1.71	2.05

Refinement	Fragment O	Fragment O1	Fragment O2	Fragment O4
R_{work}/R_{free}	0.21/0.23	0.32/0.34	0.26/0.29	0.29/0.32
Mean B-value (Å^2)				
All atoms	26	15	21	17
Protein	19	10	17	12
Water	20	14	19	14
Ligand	39	21	27	25
No of atoms				
Protein	5698	5812	5830	5830
Water	116	235	173	205
Ligand	15	18	16	16
Rmsd				
Bond lengths (Å)	0.02	0.02	0.02	0.02
Angles (°)	2.00	2.07	2.06	1.82

Refinement	Fragment O5	Fragment O6
R_{work}/R_{free}	0.25/0.28	0.25/0.27
Mean B-value (Å^2)		
All atoms	20	31
Protein	18	21
Water	21	26
Ligand	22	45
No of atoms		
Protein	5798	5840
Water	203	188
Ligand	14	14
Rmsd		
Bond lengths (Å)	0.03	0.02
Angles (°)	2.34	2.03

Indole series

Refinement	Fragment H	Fragment H1	Fragment H2	Fragment H3
R_{work}/R_{free}	0.31/0.33	0.26/0.29	0.23/0.27	0.25/0.28
Mean B-value (Å^2)				
All atoms	17	26	27	23
Protein	13	19	20	18
Water	15	25	25	24
Ligand	23	33	35	26
No of atoms				
Protein	5844	5830	5818	5830
Water	144	224	199	203
Ligand	21	20	21	20
Rmsd				
Bond lengths (Å)	0.01	0.02	0.02	0.02
Angles (°)	1.49	1.72	2.13	1.96

Refinement	Fragment 21	Fragment 22
R_{work}/R_{free}	0.24/0.26	0.24/0.27
Mean B-value (Å^2)		
All atoms	30	27
Protein	22	21
Water	27	26
Ligand	40	33
No of atoms		
Protein	5847	5829
Water	177	168
Ligand	20	22
Rmsd		
Bond lengths (Å)	0.02	0.02
Angles (°)	1.89	1.92

Other heterocyclic ring systems

Refinement	Fragment I	Fragment B	Fragment B1
R_{work}/R_{free}	0.33/0.35	0.23/0.27	0.27/0.31
Mean B-value ($\mathring{A}^2$)			
All atoms	6	19	21
Protein	7	16	17
Water	2	17	18
Ligand	8	25	28
No of atoms			
Protein	5758	5698	5830
Water	153	113	151
Ligand	20	24	27
Rmsd			
Bond lengths ($\mathring{A}$)	0.02	0.02	0.02
Angles (°)	1.81	1.93	2.10

Chapter 11. References

1. IACR and CRUK, *World Cancer Factsheet. Cancer Research UK, London.* 2014.
2. CRUK. accessed 08/15]; Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer.
3. Forbes, S.A., et al., *COSMIC: exploring the world's knowledge of somatic mutations in human cancer.* Nucleic Acids Res, 2015. **43**(Database issue): p. D805-11.
4. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* Cell, 2000. **100**(1): p. 57-70.
5. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-74.
6. DeBerardinis, R.J. and C.B. Thompson, *Cellular metabolism and disease: what do metabolic outliers teach us?* Cell, 2012. **148**(6): p. 1132-44.
7. Lunt, S.Y. and M.G. Vander Heiden, *Aerobic glycolysis: meeting the metabolic requirements of cell proliferation.* Annu Rev Cell Dev Biol, 2011. **27**: p. 441-64.
8. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.
9. Osthus, R.C., et al., *Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc.* J Biol Chem, 2000. **275**(29): p. 21797-800.
10. Warburg, O., *On respiratory impairment in cancer cells.* Science, 1956. **124**(3215): p. 269-70.
11. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**(3191): p. 309-14.
12. Hsu, P.P. and D.M. Sabatini, *Cancer cell metabolism: Warburg and beyond.* Cell, 2008. **134**(5): p. 703-7.
13. Moreno-Sanchez, R., et al., *Energy metabolism in tumor cells.* FEBS J, 2007. **274**(6): p. 1393-418.
14. Zu, X.L. and M. Guppy, *Cancer metabolism: facts, fantasy, and fiction.* Biochem Biophys Res Commun, 2004. **313**(3): p. 459-65.
15. Dang, C.V., *The interplay between MYC and HIF in the Warburg effect.* Oncogenes Meet Metabolism: From Deregulated Genes to a Broader Understanding of Tumour Physiology, 2008. **4**: p. 35-53.
16. Justus, C.R., E.J. Sanderlin, and L.V. Yang, *Molecular Connections between Cancer Cell Metabolism and the Tumor Microenvironment.* International Journal of Molecular Sciences, 2015. **16**(5): p. 11055-11086.
17. Harris, A.L., *Hypoxia--a key regulatory factor in tumour growth.* Nat Rev Cancer, 2002. **2**(1): p. 38-47.
18. Pfeiffer, T., S. Schuster, and S. Bonhoeffer, *Cooperation and competition in the evolution of ATP-producing pathways.* Science, 2001. **292**(5516): p. 504-7.
19. Wu, W. and S. Zhao, *Metabolic changes in cancer: beyond the Warburg effect.* Acta Biochim Biophys Sin (Shanghai), 2013. **45**(1): p. 18-26.
20. Burgess, D.J., *Metabolism: Glutamine connections.* Nat Rev Cancer, 2013. **13**(5): p. 293.
21. Cheong, H., et al., *Therapeutic targets in cancer cell metabolism and autophagy.* Nat Biotechnol, 2012. **30**(7): p. 671-8.
22. Hensley, C.T., A.T. Wasti, and R.J. DeBerardinis, *Glutamine and cancer: cell biology, physiology, and clinical opportunities.* J Clin Invest, 2013. **123**(9): p. 3678-84.
23. DeBerardinis, R.J., et al., *The biology of cancer: metabolic reprogramming fuels cell growth and proliferation.* Cell Metab, 2008. **7**(1): p. 11-20.
24. Reid, M.A., et al., *The B55alpha subunit of PP2A drives a p53-dependent metabolic adaptation to glutamine deprivation.* Mol Cell, 2013. **50**(2): p. 200-11.
25. Gameiro, P.A., et al., *In vivo HIF-mediated reductive carboxylation is regulated by citrate levels and sensitizes VHL-deficient cells to glutamine deprivation.* Cell Metab, 2013. **17**(3): p. 372-85.

26. Metallo, C.M., et al., *Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia.* Nature, 2012. **481**(7381): p. 380-4.

27. Tsun, Z.Y. and R. Possemato, *Amino acid management in cancer.* Semin Cell Dev Biol, 2015.

28. Wang, S., et al., *Metabolism. Lysosomal amino acid transporter SLC38A9 signals arginine sufficiency to mTORC1.* Science, 2015. **347**(6218): p. 188-94.

29. Ban, H., et al., *Arginine and Leucine regulate p70 S6 kinase and 4E-BP1 in intestinal epithelial cells.* Int J Mol Med, 2004. **13**(4): p. 537-43.

30. Efeyan, A., W.C. Comb, and D.M. Sabatini, *Nutrient-sensing mechanisms and pathways.* Nature, 2015. **517**(7534): p. 302-10.

31. Populo, H., J.M. Lopes, and P. Soares, *The mTOR signalling pathway in human cancer.* International Journal of Molecular Sciences, 2012. **13**(2): p. 1886-918.

32. Snell, K., *Enzymes of serine metabolism in normal, developing and neoplastic rat tissues.* Adv Enzyme Regul, 1984. **22**: p. 325-400.

33. Eagle, H., *Amino acid metabolism in mammalian cell cultures.* Science, 1959. **130**(3373): p. 432-7.

34. Snell, K., Y. Natsumeda, and G. Weber, *The modulation of serine metabolism in hepatoma 3924A during different phases of cellular proliferation in culture.* Biochem J, 1987. **245**(2): p. 609-12.

35. Snell, K., et al., *Enzymic imbalance in serine metabolism in human colon carcinoma and rat sarcoma.* Br J Cancer, 1988. **57**(1): p. 87-90.

36. de Koning, T.J., et al., *L-serine in disease and development.* Biochem J, 2003. **371**(Pt 3): p. 653-61.

37. Kim, H.Y., B.X. Huang, and A.A. Spector, *Phosphatidylserine in the brain: metabolism and function.* Prog Lipid Res, 2014. **56**: p. 1-18.

38. de Koning, T.J. and L.W. Klomp, *Serine-deficiency syndromes.* Curr Opin Neurol, 2004. **17**(2): p. 197-204.

39. Kalhan, S.C. and R.W. Hanson, *Resurgence of serine: an often neglected but indispensable amino Acid.* J Biol Chem, 2012. **287**(24): p. 19786-91.

40. de Koning, T.J., *Treatment with amino acids in serine deficiency disorders.* Journal of Inherited Metabolic Disease, 2006. **29**(2-3): p. 347-51.

41. Lowry, M., et al., *Renal metabolism of amino acids in vivo: studies on serine and glycine fluxes.* Am J Physiol, 1987. **252**(2 Pt 2): p. F304-9.

42. Narkewicz, M.R., et al., *Serine and glycine metabolism in hepatocytes from mid gestation fetal lambs.* Pediatr Res, 1996. **39**(6): p. 1085-90.

43. Pfendner, W. and L.I. Pizer, *The metabolism of serine and glycine in mutant lines of Chinese hamster ovary cells.* Arch Biochem Biophys, 1980. **200**(2): p. 503-12.

44. Fu, T.F., J.P. Rife, and V. Schirch, *The role of serine hydroxymethyltransferase isozymes in one-carbon metabolism in MCF-7 cells as determined by (13)C NMR.* Arch Biochem Biophys, 2001. **393**(1): p. 42-50.

45. Herbig, K., et al., *Cytoplasmic serine hydroxymethyltransferase mediates competition between folate-dependent deoxyribonucleotide and S-adenosylmethionine biosyntheses.* J Biol Chem, 2002. **277**(41): p. 38381-9.

46. Go, M.K., et al., *Glycine Decarboxylase Is an Unusual Amino Acid Decarboxylase Involved in Tumorigenesis.* Biochemistry, 2014. **53**(5): p. 947-956.

47. Snell, K., *The Duality of Pathways for Serine Biosynthesis Is a Fallacy.* Trends in Biochemical Sciences, 1986. **11**(6): p. 241-243.

48. Cheung, G.P., J.P. Cotropia, and H.J. Sallach, *The effects of dietary protein on the hepatic enzymes of serine metabolism in the rabbit.* Arch Biochem Biophys, 1969. **129**(2): p. 672-82.

49. Rowsell, E.V., et al., *Liver-L-alanine-glyoxylate and L-serine-pyruvate aminotransferase activities: an apparent association with gluconeogenesis.* Biochem J, 1969. **115**(5): p. 1071-3.

50. Ichihara, A. and D.M. Greenberg, *Pathway of Serine Formation from Carbohydrate in Rat Liver.* Proc Natl Acad Sci U S A, 1955. **41**(9): p. 605-9.

51. Tabatabaie, L., et al., *Novel mutations in 3-phosphoglycerate dehydrogenase (PHGDH) are distributed throughout the protein and result in altered enzyme kinetics.* Hum Mutat, 2009. **30**(5): p. 749-56.

52. van der Crabben, S.N., et al., *An update on serine deficiency disorders.* Journal of Inherited Metabolic Disease, 2013. **36**(4): p. 613-9.

53. Klomp, L.W., et al., *Molecular characterization of 3-phosphoglycerate dehydrogenase deficiency--a neurometabolic disorder associated with reduced L-serine biosynthesis.* Am J Hum Genet, 2000. **67**(6): p. 1389-99.

54. Ye, J., et al., *Serine catabolism regulates mitochondrial redox control during hypoxia.* Cancer Discov, 2014. **4**(12): p. 1406-17.

55. Ding, Z., et al., *Phosphoserine aminotransferase 1 (PSAT1) as a novel anti-tumor target in hepatocellular carcinoma [abstract].* Proceedings of the 106th Annual Meeting of the American Association for Cancer Research, 2015.

56. Dey, S., et al., *D-3-Phosphoglycerate dehydrogenase from Mycobacterium tuberculosis is a link between the Escherichia coli and mammalian enzymes.* J Biol Chem, 2005. **280**(15): p. 14884-91.

57. Grant, G.A., *Contrasting catalytic and allosteric mechanisms for phosphoglycerate dehydrogenases.* Arch Biochem Biophys, 2012. **519**(2): p. 175-85.

58. Schuller, D.J., G.A. Grant, and L.J. Banaszak, *The allosteric ligand site in the Vmax-type cooperative enzyme phosphoglycerate dehydrogenase.* Nat Struct Biol, 1995. **2**(1): p. 69-76.

59. Grant, G.A., *The ACT domain: a small molecule binding domain and its role as a common regulatory element.* J Biol Chem, 2006. **281**(45): p. 33825-9.

60. Dey, S., G.A. Grant, and J.C. Sacchettini, *Crystal structure of Mycobacterium tuberculosis D-3-phosphoglycerate dehydrogenase: extreme asymmetry in a tetramer of identical subunits.* J Biol Chem, 2005. **280**(15): p. 14892-9.

61. Burton, R.L., et al., *A novel mechanism for substrate inhibition in Mycobacterium tuberculosis D-3-phosphoglycerate dehydrogenase.* J Biol Chem, 2007. **282**(43): p. 31517-24.

62. Achouri, Y., et al., *Cloning, sequencing and expression of rat liver 3-phosphoglycerate dehydrogenase.* Biochem J, 1997. **323 (Pt 2)**: p. 365-70.

63. Cho, H.M., et al., *Nucleotide sequence and differential expression of the human 3-phosphoglycerate dehydrogenase gene.* Gene, 2000. **245**(1): p. 193-201.

64. *UniProt: a hub for protein information.* Nucleic Acids Res, 2015. **43**(Database issue): p. D204-12.

65. Fan, J., et al., *Human phosphoglycerate dehydrogenase produces the oncometabolite D-2-hydroxyglutarate.* ACS Chem Biol, 2015. **10**(2): p. 510-6.

66. Snell, K. and G. Weber, *Enzymic imbalance in serine metabolism in rat hepatomas.* Biochem J, 1986. **233**(2): p. 617-20.

67. Knox, W.E., A. Herzfeld, and J. Hudson, *Phosphoserine phosphatase distribution in normal and neoplastic rat tissues.* Arch Biochem Biophys, 1969. **132**(2): p. 397-403.

68. Beroukhim, R., et al., *The landscape of somatic copy-number alteration across human cancers.* Nature, 2010. **463**(7283): p. 899-905.

69. Locasale, J.W., et al., *Phosphoglycerate dehydrogenase diverts glycolytic flux and contributes to oncogenesis.* Nat Genet, 2011. **43**(9): p. 869-74.

70. Pollari, S., et al., *Enhanced serine production by bone metastatic breast cancer cells stimulates osteoclastogenesis.* Breast Cancer Res Treat, 2011. **125**(2): p. 421-30.

71. van de Vijver, M.J., et al., *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med, 2002. **347**(25): p. 1999-2009.

72. Possemato, R., et al., *Functional genomics reveal that the serine synthesis pathway is essential in breast cancer.* Nature, 2011. **476**(7360): p. 346-50.

73.   Jing, Z., et al., *Downregulation of phosphoglycerate dehydrogenase inhibits proliferation and enhances cisplatin sensitivity in cervical adenocarcinoma cells by regulating Bcl-2 and caspase-3.* Cancer Biol Ther, 2015. **16**(4): p. 541-8.

74.   Xu, W., et al., *Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases.* Cancer Cell, 2011. **19**(1): p. 17-30.

75.   Terunuma, A., et al., *MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis.* J Clin Invest, 2014. **124**(1): p. 398-412.

76.   Mullarky, E., et al., *PHGDH amplification and altered glucose metabolism in human melanoma.* Pigment Cell Melanoma Res, 2011. **24**(6): p. 1112-5.

77.   Liu, J., et al., *Phosphoglycerate dehydrogenase induces glioma cells proliferation and invasion by stabilizing forkhead box M1.* J Neurooncol, 2013. **111**(3): p. 245-55.

78.   Galluzzi, L., et al., *Metabolic targets for cancer therapy.* Nat Rev Drug Discov, 2013. **12**(11): p. 829-46.

79.   Sborov, D.W., B.M. Haverkos, and P.J. Harris, *Investigational cancer drugs targeting cell metabolism in clinical development.* Expert Opin Investig Drugs, 2015. **24**(1): p. 79-94.

80.   Farber, S. and L.K. Diamond, *Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid.* N Engl J Med, 1948. **238**(23): p. 787-93.

81.   Vander Heiden, M.G., *Targeting cancer metabolism: a therapeutic window opens.* Nat Rev Drug Discov, 2011. **10**(9): p. 671-84.

82.   Kourelis, T.V. and R.D. Siegel, *Metformin and cancer: new applications for an old drug.* Med Oncol, 2012. **29**(2): p. 1314-27.

83.   Zhan, T., et al., *Silybin and dehydrosilybin decrease glucose uptake by inhibiting GLUT proteins.* J Cell Biochem, 2011. **112**(3): p. 849-59.

84.   Stein, M., et al., *Targeting tumor metabolism with 2-deoxyglucose in patients with castrate-resistant prostate cancer and advanced malignancies.* Prostate, 2010. **70**(13): p. 1388-94.

85.   Lesk, A.M., *NAD-binding domains of dehydrogenases.* Curr Opin Struct Biol, 1995. **5**(6): p. 775-83.

86.   Chen, L., et al., *Nicotinamide adenine dinucleotide based therapeutics.* Curr Med Chem, 2008. **15**(7): p. 650-70.

87.   Pankiewicz, K.W., et al., *Synthesis of isosteric analogues of nicotinamide adenine dinucleotide containing C-nucleotide of nicotinamide or picolinamide.* J Med Chem, 1993. **36**(13): p. 1855-9.

88.   Li, H., et al., *Crystallographic studies of isosteric NAD analogues bound to alcohol dehydrogenase: specificity and substrate binding in two ternary complexes.* Biochemistry, 1994. **33**(39): p. 11734-44.

89.   Pankiewicz, K.W., *Novel nicotinamide adenine dinucleotide analogues as potential anticancer agents: quest for specific inhibition of inosine monophosphate dehydrogenase.* Pharmacol Ther, 1997. **76**(1-3): p. 89-100.

90.   Popsavin, M., et al., *Synthesis and antiproliferative activity of two new tiazofurin analogues with 2'-amido functionalities.* Bioorg Med Chem Lett, 2006. **16**(10): p. 2773-6.

91.   Curtin, N.J., *PARP inhibitors for cancer therapy.* Expert Rev Mol Med, 2005. **7**(4): p. 1-20.

92.   Banasik, M., et al., *Specific inhibitors of poly(ADP-ribose) synthetase and mono(ADP-ribosyl)transferase.* J Biol Chem, 1992. **267**(3): p. 1569-75.

93.   Suto, M.J., et al., *Dihydroisoquinolinones: the design and synthesis of a new series of potent inhibitors of poly(ADP-ribose) polymerase.* Anticancer Drug Des, 1991. **6**(2): p. 107-17.

94.   Griffin, R.J., et al., *Novel potent inhibitors of the DNA repair enzyme poly(ADP-ribose)polymerase (PARP).* Anticancer Drug Des, 1995. **10**(6): p. 507-14.

95.   Ferraris, D.V., *Evolution of poly(ADP-ribose) polymerase-1 (PARP-1) inhibitors. From concept to clinic.* J Med Chem, 2010. **53**(12): p. 4561-84.

96.   Parang, K. and P.A. Cole, *Designing bisubstrate analog inhibitors for protein kinases.* Pharmacol Ther, 2002. **93**(2-3): p. 145-57.

97.   Ehrlich, R.S., *Use of a bisubstrate inhibitor to distinguish between isocitrate dehydrogenase isozymes.* J Enzyme Inhib, 2000. **15**(3): p. 265-72.

98.   Brown, M.J., et al., *Rational design of femtomolar inhibitors of isoleucyl tRNA synthetase from a binding model for pseudomonic acid-A.* Biochemistry, 2000. **39**(20): p. 6003-11.

99.   Gileadi, O. *Vector information sheet - pNIC28-Bsa4*. [cited 2015 08-05]; Available from: http://apps.thesgc.org/structures/MM/Vectors/pNIC28-Bsa4/pNIC28-Bsa4_Vector_sheet.pdf.

100.   Gasteiger, E., et al., *Protein identification and analysis tools on the ExPASy server*. The proteomics protocols handbook2005.

101.   Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. **2**(9): p. 2212-21.

102.   Turnbull, W.B. and A.H. Daranas, *On the value of c: Can low affinity systems be studied by isothermal titration calorimetry?* Journal of the American Chemical Society, 2003. **125**(48): p. 14859-14866.

103.   Zhang, J.H., T.D. Chung, and K.R. Oldenburg, *A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays.* J Biomol Screen, 1999. **4**(2): p. 67-73.

104.   Kabsch, W., *Xds.* Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 125-132.

105.   Bailey, S., *The Ccp4 Suite - Programs for Protein Crystallography.* Acta Crystallographica Section D-Biological Crystallography, 1994. **50**: p. 760-763.

106.   McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

107.   Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

108.   Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.

109.   Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 1): p. 12-21.

110.   McNicholas, S., et al., *Presenting your structures: the CCP4mg molecular-graphics software.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 386-94.

111.   Hughes, J.P., et al., *Principles of early drug discovery.* Br J Pharmacol, 2011. **162**(6): p. 1239-49.

112.   Hoelder, S., P.A. Clarke, and P. Workman, *Discovery of small molecule cancer drugs: successes, challenges and opportunities.* Mol Oncol, 2012. **6**(2): p. 155-76.

113.   Smith, C., *Drug target validation: Hitting the target.* Nature, 2003. **422**(6929): p. 341, 343, 345 passim.

114.   Prinz, F., T. Schlange, and K. Asadullah, *Believe it or not: how much can we rely on published data on potential drug targets?* Nature Reviews Drug Discovery, 2011. **10**(9): p. 712-U81.

115.   Chen, J., et al., *Phosphoglycerate dehydrogenase is dispensable for breast tumor maintenance and growth.* Oncotarget, 2013. **4**(12): p. 2502-11.

116.   Jing, Z., et al., *Expression and clinical significance of phosphoglycerate dehydrogenase and squamous cell carcinoma antigen in cervical cancer.* Int J Gynecol Cancer, 2013. **23**(8): p. 1465-9.

117.   Ou, Y., et al., *p53 Protein-mediated regulation of phosphoglycerate dehydrogenase (PHGDH) is crucial for the apoptotic response upon serine starvation.* J Biol Chem, 2015. **290**(1): p. 457-66.

118.   Holliday, D.L. and V. Speirs, *Choosing the right cell line for breast cancer research.* Breast Cancer Res, 2011. **13**(4): p. 215.

119. Locasale, J.W., *Serine, glycine and one-carbon units: cancer metabolism in full circle.* Nat Rev Cancer, 2013. **13**(8): p. 572-83.

120. Petitjean, A., et al., *Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database.* Hum Mutat, 2007. **28**(6): p. 622-9.

121. von Bueren, A.O., et al., *Anti-proliferative activity of the quassinoid NBT-272 in childhood medulloblastoma cells.* Bmc Cancer, 2007. **7**.

122. Meco, D., et al., *Antitumor effect in medulloblastoma cells by gefitinib: Ectopic HER2 overexpression enhances gefitinib effects in vivo.* Neuro-Oncology, 2009. **11**(3): p. 250-259.

123. Muret, J., et al., *Attenuation of Soft-Tissue Sarcomas Resistance to the Cytotoxic Action of TNF-alpha by Restoring p53 Function.* PLoS One, 2012. **7**(6).

124. Chavez, K.J., S.V. Garimella, and S. Lipkowitz, *Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer.* Breast Dis, 2010. **32**(1-2): p. 35-48.

125. Chang, H., et al., *p53 mutations, c-myc and bcl-2 rearrangements in human non-Hodgkin's lymphoma cell lines.* Leuk Lymphoma, 1995. **19**(1-2): p. 165-71.

126. Drakos, E., et al., *Activation of the p53 pathway by the MDM2 inhibitor nutlin-3a overcomes BCL2 overexpression in a preclinical model of diffuse large B-cell lymphoma associated with t(14;18)(q32;q21).* Leukemia, 2011. **25**(5): p. 856-67.

127. Sun, L., et al., *cMyc-mediated activation of serine biosynthesis pathway is critical for cancer progression under nutrient deprivation conditions.* Cell Res, 2015. **25**(4): p. 429-44.

128. Miller, D.M., et al., *c-Myc and cancer metabolism.* Clin Cancer Res, 2012. **18**(20): p. 5546-53.

129. Bruggers, C.S., et al., *Expression of the c-Myc protein in childhood medulloblastoma.* J Pediatr Hematol Oncol, 1998. **20**(1): p. 18-25.

130. Gibson, B.A. and W.L. Kraus, *New insights into the molecular and cellular functions of poly(ADP-ribose) and PARPs.* Nature Reviews Molecular Cell Biology, 2012. **13**(7): p. 411-424.

131. Bai, P. and C. Canto, *The Role of PARP-1 and PARP-2 Enzymes in Metabolic Regulation and Disease.* Cell Metab, 2012. **16**(3): p. 290-295.

132. Weiss, W.A., S.S. Taylor, and K.M. Shokat, *Recognizing and exploiting differences between RNAi and small-molecule inhibitors.* Nat Chem Biol, 2007. **3**(12): p. 739-44.

133. Ferguson, R.E., et al., *Housekeeping proteins: a preliminary study illustrating some limitations as useful references in protein expression studies.* Proteomics, 2005. **5**(2): p. 566-71.

134. Li, R. and Y. Shen, *An old method facing a new challenge: re-visiting housekeeping proteins as internal reference control for neuroscience research.* Life Sci, 2013. **92**(13): p. 747-51.

135. van Tonder, A., A.M. Joubert, and A.D. Cromarty, *Limitations of the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay when compared to three commonly used cell enumeration assays.* BMC Res Notes, 2015. **8**: p. 47.

136. Nelson, D.L., M.M. Cox, and A.L. Lehninger, *Lehninger - Principles of biochemistry*. 5th ed 2008.

137. Rask-Andersen, M., M.S. Almen, and H.B. Schioth, *Trends in the exploitation of novel drug targets.* Nature Reviews Drug Discovery, 2011. **10**(8): p. 579-590.

138. Ballatore, C., D.M. Huryn, and A.B. Smith, 3rd, *Carboxylic acid (bio)isosteres in drug design.* ChemMedChem, 2013. **8**(3): p. 385-95.

139. Dragovich, P.S., et al., *Identification of substituted 3-hydroxy-2-mercaptocyclohex-2-enones as potent inhibitors of human lactate dehydrogenase.* Bioorg Med Chem Lett, 2014. **24**(16): p. 3764-71.

140. Ward, R.A., et al., *Design and synthesis of novel lactate dehydrogenase A inhibitors by fragment-based lead generation.* J Med Chem, 2012. **55**(7): p. 3285-306.

141. Beaumont, K., et al., *Design of ester prodrugs to enhance oral absorption of poorly permeable compounds: challenges to the discovery scientist.* Curr Drug Metab, 2003. **4**(6): p. 461-85.

142. Chang, Y.H., et al., *Contributions of active site residues to cofactor binding and catalysis of 3alpha-hydroxysteroid dehydrogenase/carbonyl reductase.* Biochim Biophys Acta, 2010. **1804**(1): p. 235-41.

143. Shaw, K.N.F., et al., *Preparation and Properties of b-3-Indolyl Compounds Related to Tryptophan Metabolism.* The Journal of Organic Chemistry, 1958. **23**(8): p. 1171-1178.

144. Weller, L. and H. Sell, *Synthesis of 3-Indoleacetamides.* The Journal of Organic Chemistry, 1958. **23**(11): p. 1776-1777.

145. Montalbetti, C.A.G.N. and V. Falque, *Amide bond formation and peptide coupling.* Tetrahedron, 2005. **61**(46): p. 10827-10852.

146. Mollan, R.C., D.M. Donnelly, and M.A. Harmey, *Synthesis of Indole-3-Acetylaspartic Acid.* Phytochemistry, 1972. **11**(4): p. 1485-&.

147. Fedorov, O., F.H. Niesen, and S. Knapp, *Kinase inhibitor selectivity profiling using differential scanning fluorimetry.* Methods Mol Biol, 2012. **795**: p. 109-18.

148. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallogr D Biol Crystallogr, 2011. **67**(Pt 4): p. 355-67.

149. Hayward, S. and H.J.C. Berendsen, *Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme.* Proteins-Structure Function and Genetics, 1998. **30**(2): p. 144-154.

150. Hayward, S. and R.A. Lee, *Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50.* J Mol Graph Model, 2002. **21**(3): p. 181-3.

151. McConathy, J. and M.J. Owens, *Stereochemistry in Drug Action.* Prim Care Companion J Clin Psychiatry, 2003. **5**(2): p. 70-73.

152. Rossmann, M.G., D. Moras, and K.W. Olsen, *Chemical and biological evolution of nucleotide-binding protein.* Nature, 1974. **250**(463): p. 194-9.

153. Hanukoglu, I., *Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites.* Biochem Mol Biol Educ, 2015. **43**(3): p. 206-9.

154. Gertz, M., et al., *Ex-527 inhibits Sirtuins by exploiting their unique NAD+-dependent deacetylation mechanism.* Proc Natl Acad Sci U S A, 2013. **110**(30): p. E2772-81.

155. Dawson, R.M.C., et al., *Data for biochemical research*. Third edition ed1986.

156. Stein, A.M., et al., *The Thionicotinamide Analogs of Dpn and Tpn. I. Preparation and Analysis.* Biochemistry, 1963. **2**: p. 1015-7.

157. Siegel, J.M., G.A. Montgomery, and R.M. Bock, *Ultraviolet absorption spectra of DPN and analogs of DPN.* Arch Biochem Biophys, 1959. **82**(2): p. 288-99.

158. Niviere, V., et al., *The NAD(P)H:flavin oxidoreductase from Escherichia coli. Evidence for a new mode of binding for reduced pyridine nucleotides.* J Biol Chem, 1999. **274**(26): p. 18252-60.

159. Shen, Y.Q., S.Y. Song, and Z.J. Lin, *Structures of D-glyceraldehyde-3-phosphate dehydrogenase complexed with coenzyme analogues.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 8): p. 1287-97.

160. Unterlass, J.E., *Inhibition of key-cancer associated NAD[+]-dependent enzymes by BSI-201 and nicotinamide analogues*, 2012, Universitaet Konstanz.

161. Copeland, R.A., *Enzyme reaction mechanisms*, in *Evaluation of enzyme inhibitors in drug discovery*2013.

162. Karsten, W.E., P.A. Tipton, and P.F. Cook, *Tartrate dehydrogenase catalyzes the stepwise oxidative decarboxylation of D-malate with both NAD and thio-NAD.* Biochemistry, 2002. **41**(40): p. 12193-9.

163. Hedstrom, L., *IMP dehydrogenase: structure, mechanism, and inhibition.* Chem Rev, 2009. **109**(7): p. 2903-28.

164. Ferrier, B., *An enzymatic cycling method for 3-acetylpyridine adenine dinucleotide to increase the sensitivity of enzymatic methods which employ this NAD analog.* Anal Biochem, 1990. **186**(2): p. 229-32.

165.   Brueckner, R., *Reaktionsmechanismen: Organische Reaktionen, Stereochemie, Moderne Synthesemethoden*2009: Spektrum Akademischer Verlag.

166.   Subramanian, S., *Thermodynamics of Dehydrogenase Reactions.* Trends in Biochemical Sciences, 1979. **4**(5): p. 102-105.

167.   Warburg, O. and W. Christian, *Isolation and crystallisation of enolase.* Biochemische Zeitschrift, 1942. **310**: p. 384-421.

168.   Bishop, A., et al., *Unnatural ligands for engineered proteins: new tools for chemical genetics.* Annu Rev Biophys Biomol Struct, 2000. **29**: p. 577-606.

169.   Greenfield, N.J., *Using circular dichroism spectra to estimate protein secondary structure.* Nat Protoc, 2006. **1**(6): p. 2876-2890.

170.   Lobley, A., L. Whitmore, and B.A. Wallace, *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra.* Bioinformatics, 2002. **18**(1): p. 211-212.

171.   Vanstokkum, I.H.M., et al., *Estimation of Protein Secondary Structure and Error Analysis from Circular-Dichroism Spectra.* Analytical Biochemistry, 1990. **191**(1): p. 110-118.

172.   Provencher, S.W. and J. Glockner, *Estimation of Globular Protein Secondary Structure from Circular-Dichroism.* Biochemistry, 1981. **20**(1): p. 33-37.

173.   Velick, S.F., J.E. Hayes, Jr., and J. Harting, *The binding of diphosphopyridine nucleotide by glyceraldehyde-3-phosphate dehydrogenase.* J Biol Chem, 1953. **203**(2): p. 527-44.

174.   Stambaugh, R.L. and D.W. Wilson, *The chromatography of nucleotides, nucleosides, and pyrimidines and purines on activated charcoal.* Journal of Chromatography A, 1960. **3**(0): p. 221-224.

175.   Jandorf, B.J., *A method for the isoloation of diphosphopyridine dinucleotide.* J Biol Chem, 1941(138): p. 305-309.

176.   Brahms, S. and J. Brahms, *Determination of Protein Secondary Structure in Solution by Vacuum Ultraviolet Circular-Dichroism.* Journal of Molecular Biology, 1980. **138**(2): p. 149-178.

177.   Hwang, Y.W. and D.L. Miller, *A mutation that alters the nucleotide specificity of elongation factor Tu, a GTP regulatory protein.* J Biol Chem, 1987. **262**(27): p. 13081-5.

178.   Jencks, W.P., *On the attribution and additivity of binding energies.* Proc Natl Acad Sci U S A, 1981. **78**(7): p. 4046-50.

179.   Nakamura, C.E. and R.H. Abeles, *Mode of interaction of beta-hydroxy-beta-methylglutaryl coenzyme A reductase with strong binding inhibitors: compactin and related compounds.* Biochemistry, 1985. **24**(6): p. 1364-76.

180.   Scott, D.E., et al., *Fragment-based approaches in drug discovery and chemical biology.* Biochemistry, 2012. **51**(25): p. 4990-5003.

181.   Rees, D.C., et al., *Fragment-based lead discovery.* Nat Rev Drug Discov, 2004. **3**(8): p. 660-72.

182.   Murray, C.W., M.L. Verdonk, and D.C. Rees, *Experiences in fragment-based drug discovery.* Trends Pharmacol Sci, 2012. **33**(5): p. 224-32.

183.   Tsai, J., et al., *Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity.* Proc Natl Acad Sci U S A, 2008. **105**(8): p. 3041-6.

184.   Hudson, S.A., et al., *Application of Fragment Screening and Merging to the Discovery of Inhibitors of the Mycobacterium tuberculosis Cytochrome P450 CYP121.* Angewandte Chemie-International Edition, 2012. **51**(37): p. 9311-9316.

185.   Larsson, A., et al., *Efficiency of hit generation and structural characterization in fragment-based ligand discovery.* Current Opinion in Chemical Biology, 2011. **15**(4): p. 482-488.

186.   Zhang, Y.L. and Z.Y. Zhang, *Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B.* Analytical Biochemistry, 1998. **261**(2): p. 139-48.

187. Turnbull, W.B., B.L. Precious, and S.W. Homans, *Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry.* Journal of the American Chemical Society, 2004. **126**(4): p. 1047-1054.

188. Daranas, A.H., H. Shimizu, and S.W. Homans, *Thermodynamics of binding of D-galactose and deoxy derivatives thereof to the L-arabinose-binding protein.* Journal of the American Chemical Society, 2004. **126**(38): p. 11870-11876.

189. Ciulli, A., et al., *Probing hot spots at protein-ligand binding sites: A fragment-based approach using biophysical methods.* J Med Chem, 2006. **49**(16): p. 4992-5000.

190. Taylor, J.D., et al., *Identification of novel fragment compounds targeted against the pY pocket of v-Src SH2 by computational and NMR screening and thermodynamic evaluation.* Proteins-Structure Function and Bioinformatics, 2007. **67**(4): p. 981-990.

191. Wernimont, A. and A. Edwards, *In situ proteolysis to generate crystals for structure determination: an update.* PLoS One, 2009. **4**(4): p. e5094.

192. Dong, A., et al., *In situ proteolysis for protein crystallization and structure determination.* Nat Methods, 2007. **4**(12): p. 1019-21.

193. Reeks, J., *Structural studies of CRISPR-associated proteins*, in *School of Chemistry*2013, University of St. Andrews.

194. Keil, B., *Specificity of Proteolysis*1992: Springer-Verlag.

195. Bruno, I.J., et al., *Retrieval of crystallographically-derived molecular geometry information.* J Chem Inf Comput Sci, 2004. **44**(6): p. 2133-44.

196. Surade, S., et al., *A structure-guided fragment-based approach for the discovery of allosteric inhibitors targeting the lipophilic binding site of transcription factor EthR.* Biochem J, 2014. **458**(2): p. 387-94.

197. Amaning, K., et al., *The use of virtual screening and differential scanning fluorimetry for the rapid identification of fragments active against MEK1.* Bioorg Med Chem Lett, 2013. **23**(12): p. 3620-6.

198. Hung, A.W., et al., *Application of fragment growing and fragment linking to the discovery of inhibitors of Mycobacterium tuberculosis pantothenate synthetase.* Angew Chem Int Ed Engl, 2009. **48**(45): p. 8452-6.

199. Meiby, E., et al., *Fragment screening by weak affinity chromatography: comparison with established techniques for screening against HSP90.* Anal Chem, 2013. **85**(14): p. 6756-66.

200. Hubbard, R.E., *Fragment-based lead discovery applied to protein-protein interactions*, in *Fragment-based drug discovery*, S. Howard and C. Abell, Editors. 2015.

201. Malo, N., et al., *Statistical practice in high-throughput screening data analysis.* Nat Biotechnol, 2006. **24**(2): p. 167-75.

202. Ruhmann, E., et al., *Thermodynamic signatures of fragment binding: Validation of direct versus displacement ITC titrations.* Biochim Biophys Acta, 2015. **1850**(4): p. 647-56.

203. Denessiouk, K.A., V.V. Rantanen, and M.S. Johnson, *Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins.* Proteins, 2001. **44**(3): p. 282-91.

204. Warburg, O., K. Posener, and E. Negelein, *Über den Stoffwechsel der Carcinomzelle.* Biochem Z, 1924. **152**: p. 319-344.

205. Xie, G., et al., *Plasma metabolite biomarkers for the detection of pancreatic cancer.* J Proteome Res, 2015. **14**(2): p. 1195-202.

206. Wang, H., et al., *Development and validation of a highly sensitive urine-based test to identify patients with colonic adenomatous polyps.* Clin Transl Gastroenterol, 2014. **5**: p. e54.

207. Locasale, J.W. and L.C. Cantley, *Genetic selection for enhanced serine metabolism in cancer development.* Cell Cycle, 2011. **10**(22): p. 3812-3.

208. Rotondo, J.C., et al., *Gene expression changes in progression of cervical neoplasia revealed by microarray analysis of cervical neoplastic keratinocytes.* J Cell Physiol, 2015. **230**(4): p. 806-12.

209. Zhang, W., et al., *Proteomic profiles of human lung adeno and squamous cell carcinoma using super-SILAC and label-free quantification approaches.* Proteomics, 2014. **14**(6): p. 795-803.

210. Kun, E., et al., *Quantitative correlation between cellular proliferation and nuclear poly (ADP-ribose) polymerase (PARP-1).* Int J Mol Med, 2006. **17**(2): p. 293-300.

211. Ma, W., et al., *Differential effects of poly(ADP-ribose) polymerase inhibition on DNA break repair in human cells are revealed with Epstein-Barr virus.* Proc Natl Acad Sci U S A, 2012. **109**(17): p. 6590-5.

212. Murai, J., et al., *Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors.* Cancer Res, 2012. **72**(21): p. 5588-5599.

213. Satoh, M.S. and T. Lindahl, *Role of poly(ADP-ribose) formation in DNA repair.* Nature, 1992. **356**(6367): p. 356-8.

214. Son, H.F., I.K. Kim, and K.J. Kim, *Structural insights into domain movement and cofactor specificity of glutamate dehydrogenase from Corynebacterium glutamicum.* Biochem Biophys Res Commun, 2015. **459**(3): p. 387-92.

215. Yasutake, Y., et al., *Crystal structure of the monomeric isocitrate dehydrogenase in the presence of NADP+: insight into the cofactor recognition, catalysis, and evolution.* J Biol Chem, 2003. **278**(38): p. 36897-904.

216. Burton, R.L., J.W. Hanes, and G.A. Grant, *A stopped flow transient kinetic analysis of substrate binding and catalysis in Escherichia coli D-3-phosphoglycerate dehydrogenase.* J Biol Chem, 2008. **283**(44): p. 29706-14.

217. Sugimoto, E. and L.I. Pizer, *The mechanism of end product inhibition of serine biosynthesis. I. Purification and kinetics of phosphoglycerate dehydrogenase.* J Biol Chem, 1968. **243**(9): p. 2081-9.

218. Howard, S. and C. Abell, *Fragment-based drug discovery.* 2015.

219. *JANSSEN PHARMACEUTICA N.V., Benzisoxazoles*, 2005, WO 2005/089753 A2.

220. Molyneux, L., *The discovery of small-molecule inhibitors of ERK5 for the treatment of cancer*, Northern Institute for Cancer Research 2014, Newcastle University.