Structure and function of the eukaryotic ADP-dependent glucokinase

A dissertation presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy
in
Biochemistry

Massey University, Manawatu,
New Zealand

Jan Paul Richter
2014

Abstract

The ADP-dependent glucokinase enzymes (ADPGK) are the first new glycolytic enzymes to be discovered in over 40 years. This class of enzymes was first described in thermophilic archaea in 1994. A decade later, an ADPGK from a eukaryote was also identified and characterised. The ADPGK enzymes catalyse a phosphorylation reaction converting glucose and ADP to glucose-6-phosphate and AMP. The enzyme is well studied in extremophilic archaea, where ADPGK is part of a set of glycolytic enzymes that use ADP instead of ATP for the phosphorylation of various sugars. However, ADPGK has also been found in the genomes of mesophilic species and higher eukaryotes, suggesting that the enzyme is not necessarily an adaption to high temperatures. In eukaryotes, ADPGK has been linked to a modified glycolysis pathway that is required for T-cell activation. While crystal structures of the archaeal ADPGKs are known, no structure of a eukaryotic ADPGK had been solved before the work undertaken in this thesis. In this thesis, the kinetic analysis of a recombinant form of *Homo sapiens* ADPGK and the crystal structure of a truncated form of *Mus musculus* ADPGK are presented. Both enzymes were expressed recombinantly in *E. coli* and purified in soluble form. The kinetic parameters determined for *H. sapiens* ADPGK proved to be comparable to the mouse enzyme, which had been published earlier. In addition, the phosphoryl acceptor specificity of *H. sapiens* ADPGK was extensively tested by ^{31}P-NMR, where the enzyme proved to be highly specific for D-glucose. Residues important for catalysis have been modified by site-directed mutagenesis and the variants of *H. sapiens* ADPGK were purified and kinetic parameters determined. A single crystal was obtained from a truncated variant of *M. musculus* ADPGK, which diffracted to 2.1 Å. The structure of *M. musculus* ADPGK could be solved by molecular replacement using the known crystal structures of the archaeal ADPGKs for initial phasing. It proved to be quite similar to the archaeal, ADPGKs, despite the low sequence identity. The combined data in this work improves our understanding of the conservation of the structure-function relationship of eukaryotic ADPGKs.

Acknowledgements

First of all, I would like to thank my supervisor Dr. Andrew Sutherland-Smith for giving me the opportunity to work on this project, for his ongoing support, many insightful discussions and helpful advice during this difficult time.

I would like to thank my co-supervisor Dr. Ron Ronimus whose work on mouse ADPGK inspired this PhD project and whose practical help came when most needed and who always believed that there would be a crystal in the end.

I would like to thank my co-supervisor Dr. Kathryn Stowell for sharing her expertise in numerous discussions and providing much-appreciated encouragement.

I am also grateful to all other members of the ADPGK team - Dr. Bill Wilson, Dr. Susan Richter, Amanda Gribble and Rebecca Hole – whose experimental ADPGK work informed and enriched mine.

A special thank you for support and advice in practical matters goes to Dr. Vince Carbone and Dr. Tom Caradoc-Davies for their help with the data collection of the precious ADPGK crystal, Dr. Alexander Goroncy for his help with everything NMR, Dr. Matthew Bennett for his help with protein expression in insect cell lines and Dr. Bernd Rehm and his team for their help with the bioreactor.

I would like to express gratitude to the funding sources which ultimately made this project possible: the Marsden Fund grant for providing my scholarship, IMBS and my supervisor Dr. Sutherland-Smith for extending my scholarship, the Royal Society of New Zealand, the Maurice and Phyllis Paykel Trust, IMBS and the NZSBMB for support with travel funding.

A big thanks to everyone in the lab, office, etc. at IMBS (now IFS) who supported me.

Finally, I would like to thank my family for their love, support and patience, especially my daughter Lea, my partner Claudia and my parents Elfriede and Peter.

Table of Contents

List of Figures

List of Tables

Abbreviations and symbols

˚C	Temperature in centigrade
Å	Ångström (10^{-10} m)
aa	Amino acid
ADP	Adenosine diphosphate
ADPGK	ADP-dependent glucokinase
ADPPFK	ADP-dependent phosphofructokinase
Amp	Ampicillin
AMP	Adenosine monophosphate
AMPCP	Adenosine 5′-(α,β-methylene) diphosphate
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pair (of DNA)
BSA	Bovine serum albumin
BTP	Bis-tris propane buffer
Cam	Chloramphenicol
CD	Circular dichroism
cDNA	Complementary DNA, experimentally derived from messenger RNA
cv	Column volume (geometric column volume of a chromatography column)
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetate
g	Gram, unit of weight
g	Gravitational force
GK	Glucokinase

h	Hour
HK	Hexokinase
HRP	Peroxidase from horseradish
IMAC	Immobilised metal ion affinity purification
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Temperature in Kelvin
Kan	Kanamycin
kDa	Kilodalton
K_i	Dissociation constant for an enzyme inhibitor
K_m	Michaelis-Menten constant
LB broth	Luria-Bertani broth
LB/Amp	Luria-Bertani broth with ampicillin selection marker
m	Metre
M	Molar
min	Minute
MTT	Thiazolyl blue tetrazolium bromide
NAD^+ or NADH	Nicotinamide adenine dinucleotide (oxidised or reduced)
$NADP^+$ or NADPH	Nicotinamide adenine dinucleotide phosphate (oxidised or reduced)
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at wavelength of 600 nm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pdb	Protein data bank file format (*.pdb)
PDB	Protein Data Bank for protein structures
pH	Measurement of acidity or basicity of an aqueous solution
P_i	Inorganic phosphate

pI	Isoelectric point
PMS	Phenazine methosulfate
PP_i	Pyrophosphate
psi	Pounds per square inch
RMSD	Root-mean-square deviation
RNAi	RNA interference
rpm	Revolutions per minute
SAXS	Small angle X-ray scattering
SDS	Sodium dodecyl sulphate
sec	Second
Str	Streptomycin
TAE	Tris-acetate-EDTA buffer
TBS	Tris-Buffered Saline solution
TBST	Tris-Buffered Saline solution with additional Tween 20 detergent
TCEP	Tris(2-carboxyethyl)phosphine
Tet	Tetracycline
U	Enzyme unit, 1 U = 1 μmol min^{-1}
U/mg	Specific enzyme activity, 1 U/mg = 1 μmol min^{-1} mg^{-1}
V	Enzyme velocity
(v/v)	Volume per volume
V_{max}	Maximum enzyme velocity
(w/v)	Weight per volume

1 Introduction

1.1 ADP-dependent glucokinases

1.1.1 Introduction to ADPGK

ADP-dependent glucokinases (ADPGK, EC 2.7.1.147) catalyse the phosphorylation of D-glucose using ADP, instead of ATP, producing glucose-6-phosphate and AMP. ADPGK enzymes are found in thermophilic archaea, mesophilic methanogens and metazoa. The enzyme was first discovered in *Pyrococcus furiosus*, a hyperthermophilic archaeon with optimum growth temperature above 80 ˚C (Kengen et al. 1994, 1995). A decade later a homologue in eukaryotic organisms was described (Ronimus and Morgan 2004). The specificity for ADP of these kinases is curious, since most kinases use ATP as phosphoryl donor. At first it was thought that the superior thermostability of ADP compared to ATP confers a fundamental advantage for certain hyperthermophilic organisms as an adaption to survival at high temperatures (Kengen et al. 1994). However, ATP-dependent variants have been identified in thermophile archaea and ADPGK is found in mesophilic organisms as well (Guixé and Merino 2009). Therefore, the advantage of thermal stability may not be the major reason for the existence of these specialised enzymes. Several functions have been proposed for ADPGK in eukaryotes. For one, ADPGK could be advantageous during hypoxic conditions to extend cell survival (Ronimus and Morgan 2004; Richter et al. 2012, 2013). Following a different direction, recent work has also shown a link between reactive oxygen species production in T-cells and ADPGK in eukaryotes (Kamiński et al. 2012, 2013). Additionally, recent advances in proteomics have helped identify a cholesterol binding site in ADPGK of eukaryotes (Hulce et al. 2013).

1.1.2 ADPGK and the evolution of glycolysis

The Embden-Meyerhof-Parnas pathway, or glycolysis, is a central pathway for glucose degradation and is conserved among all domains of life – archaea, eubacteria and eukarya (Woese et al. 1990; Ronimus and Morgan 2001; Voet and Voet 2011). All together, there are ten enzymatic steps in glycolysis, during which glucose is first

activated and then in multiple steps oxidised to regenerate ATP from inorganic phosphate. The overall reaction of the classical glycolysis pathway is:

$$\text{Glucose} + 2NAD^+ + 2ADP + 2P_i \rightarrow 2NADH + 2\ \text{pyruvate} + 2ATP + 2H_2O + 4H^+$$

A summary below compares the classical glycolysis pathway with the modified pathway from *P. furiosus*, since it is one of the most studied and well-understood archaea (see Figure 1). Only four of the ten enzymes involved in glycolysis, triosephosphate isomerase (TIM), phosphoglycerate mutase (PGM), enolase (ENO) and pyruvate kinase (PYK), are orthologues of the enzymes found in the classical glycolysis pathway (Verhees et al. 2003). The reaction of glycolysis in *P. furiosus* converts glucose to pyruvate and in the process ferredoxin (Fd) is reduced. The overall reaction is:

$$\text{Glucose} + 4Fd^{ox} \rightarrow \text{pyruvate} + 4Fd^{red} + 4H^+$$

After its uptake into the cell, the initial step in glycolysis is the phosphorylation of glucose, which is catalysed by a hexose kinase. In the classical version of the glycolysis pathway this is a hexokinase or glucokinase (HK/GK), which uses ATP as a phosphoryl-donor, producing glucose-6-phosphate and ADP. In mammalia, there are four isoforms of this enzyme present, which are generally referred to as hexokinase type I, II, III and IV. Hexokinase IV is also known as glucokinase (Wilson 2003; Iynedjian 2009). In the glycolytic pathway of *P. furiosus*, this first step is catalysed by an ADPGK, which uses ADP instead of ATP, producing AMP and glucose-6-phosphate (Kengen et al. 1995). However, ATP-dependent enzymes have been identified in other thermophilic archaea and are described in detail in *Aeropyrum pernix* (Hansen et al. 2002) and *Thermoproteus tenax* (Dörr et al. 2003; Siebers et al. 2004). Also polyphosphate-dependent variants have been described, such as the glucokinase from *Microlunatus phosphovorus* (Tanaka et al. 2003) and a less specific inorganic polyphosphate/ATP-dependent glucomannokinase (Mukai et al. 2004), the latter of which is able to utilise either ATP or inoraganic polyphosphate to phosphorylate either glucose or mannose. Inorganic polyphosphate is presumably an ancient form of phosphoryl donor (Kornberg et al. 1999).

During the next step, the glucose-6-phosphate product is converted to fructose-6-phosphate by phosphoglucose isomerase (PGI), which can be further phosphorylated

by phosphofructokinase (PFK) to fructose-1,6-bisphosphate. Phosphofructokinase is a highly regulated enzyme as it catalyses one of the rate determining reactions of glycolysis. Phosphofructokinase can be inhibited or activated by various metabolites in a complex manner. While the tetrameric phosphofructokinases of the more complex eukaryotes are ATP-dependent, there are many examples of ADP-dependent homologues in archaea (ADPPFK). The ADPPFK enzymes have been reported in different oligomerisation states, including monomers, dimers and tetramers (Tuininga et al. 1999; Labes and Schönheit 2003; Ronimus and Morgan 2003; Currie et al. 2009). It should be noted that ATPPFK enzymes have been reported in archaea as well (Siebers and Schönheit 2005). Furthermore, in *Methanococcus jannaschii,* a bifunctional ATP-dependent phosphofructokinase/glucokinase has been discovered. It has been shown that the glycolytic pathway is functional in this organism and that the bifunctional enzyme can substitute for the two separate enzymes found in other organisms (Verhees et al. 2001; Sakuraba et al. 2002). Aldolase (FBA) catalyses the cleavage of fructose-1,6-bisphosphate into two trioses, glyeraldehyde-3-phosphate and dihydroxyacetonephosphate. Phosphate triose isomerase can convert the two 3-carbon sugars into each other. Only glyceraldehyde-3-phosphate is further used in glycolysis. In the next step glyceraldehyde-3-phosphate is phosphorylated to 1,3-bisphosphoglycerate by glyceraldehyde-3-phosphate dehydrogenase (GAPDH), using inorganic phosphate and generating reducing equivalents in the form of $NADH/H^+$. Homologs of triosephosphate isomerase are conserved in archaea, bacteria and eukarya (Ronimus and Morgan 2001; Verhees et al. 2003). The dephosphorylation of 1,3-bisphosphoglyerate is catalysed by phosphoglycerate kinase (PGK). The phosphate group is transferred to ADP generating ATP and 3-phosphoglycerate. Subsequently 3-phosphoglycerate is converted to 2-phosphoglycerate and phosphoenolpyruvate by phosphoglycerate mutase and enolase, respectively. The glycolysis of *P. furiosus* differs here, as the enzyme glyceraldehyde-3-phosphate ferredoxin reductase (GAPOR) can catalyse the conversion of glyceraldehyde phosphate into 3-phosphoglycerate in a single step. The last reaction in glycolysis is catalysed by pyruvate kinase, using phosphoenolpyruvate and ADP as substrates, producing ATP and pyruvate. The pyruvate kinases are conserved in archaea, bacteria and eukarya (Verhees et al. 2003).

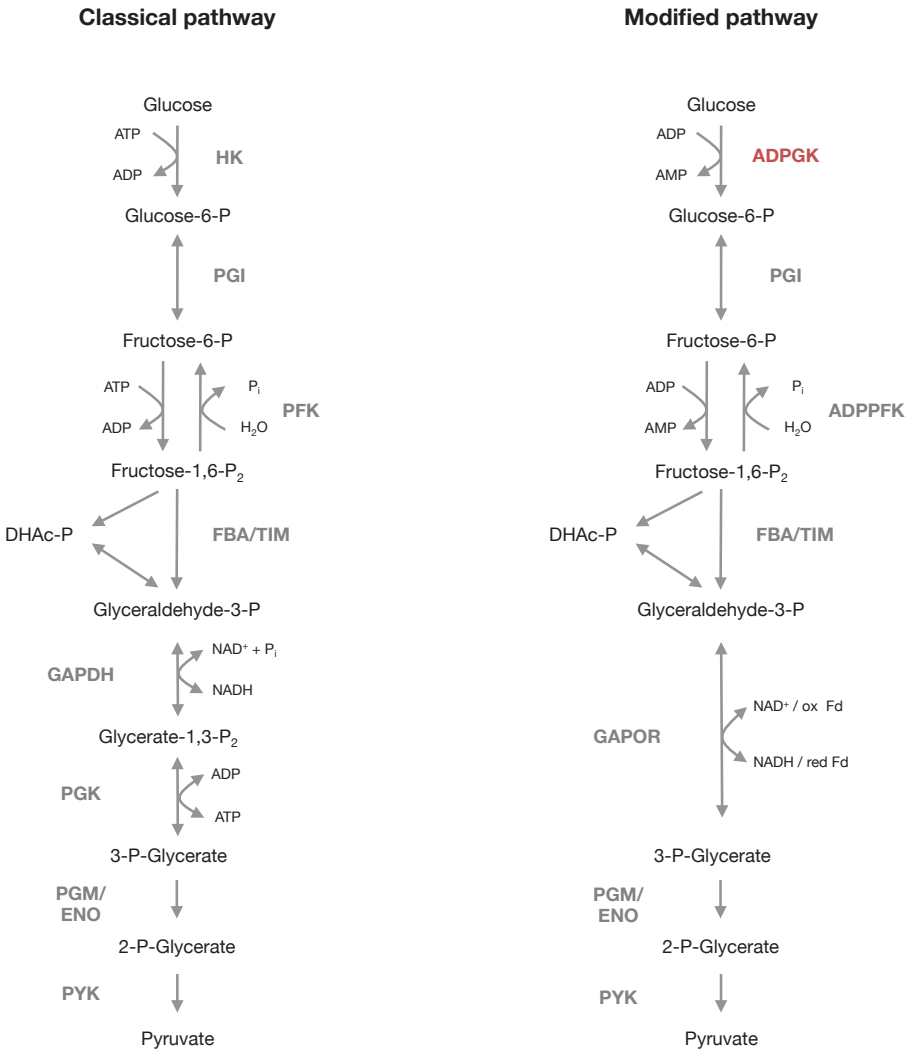

Figure 1: Overview of glycolysis.

Comparison of the classical (left) versus the modified (right) Embden-Meyerhof pathway in archaea, more specifically of *P. furiosus* as example (Verhees et al. 2003). ADPGK is located at the start of the glycolytic pathway, catalysing the phosphorylation of glucose. For details see text.

1.1.3 Regulation of glycolysis

Glycolytic flux in eukaryotes is regulated on the gene expression level and on the protein level by allosteric regulation of the enzymes involved. The main control sites for glycolysis on a protein level are the irreversible reactions catalysed by hexokinase, phosphofructokinase and pyruvate kinase, which are all controlled allosterically. Hexokinase types I, II and III are inhibited by their product, glucose-6-phosphate with K_i values of 0.21, 0.16 and 0.92 mM, respectively, as well as by inorganic phosphate (Grossbard and Schimke 1966; Bianchi et al. 1998). While type II and III are inhibited

by P_i at all concentrations, P_i at low concentrations has an antagonizing effect to the inhibition of glucose-6-phosphate on the type I enzyme. The K_m for glucose of hexokinase types I, II and III are 0.03, 0.3 and 0.003 mM, respectively. Hexokinase I is expressed in most cell types, but it is often referred to as the brain-hexokinase as it is highly expressed in this particular organ. Mammalian hexokinase II is highly expressed in tissues such as skeletal muscles, heart and adipose tissue, and with impaired function of Hexokinase II has been linked to diabetes (Heikkinen et al. 2000). The function of Hexokinase III is not fully understood (Wilson 2003).

The monomeric enzyme glucokinase or hexokinase IV is allosterically regulated by its substrate glucose (Storer and Cornish-Bowden 1976). The proposed mechanism for the allosteric regulation of rat liver glucokinase (Storer and Cornish-Bowden 1977) has been confirmed by structural data of the enzyme in its closed and open forms (Kamata et al. 2004). The mechanism involves a thermodynamically favourable super-open conformation apoenzyme. Upon glucose binding, the enzyme undergoes a conformational change into the open, activated form. This transition happens slowly and is called the slow cycle. When ATP binds to the glucose-bound open conformation, the enzyme transitions into the closed conformation and the phosphoryl transfer reaction is triggered. After releasing the substrates the enzyme is found again in the open conformation. This is called the fast cycle. The open conformation will only slowly revert to the super-open conformation. If glucose binds before the reverse slow cycle is complete, the enzyme can enter the fast cycle again immediately (Kamata et al. 2004). Newer research has also shown that the situation could be even more complex, involving multiple transition states, which are not fully understood (Larion and Miller 2010, 2012). Glucokinase shows no end product inhibition by glucose-6-phosphate below concentrations of 60 mM (Pollard-Knight and Cornish-Bowden 1982). Two tissue-specific promoters, the hepatic and non-hepatic promoter, regulate the mammalian glucokinase gene. The hepatic promoter is responsible for the expression of glucokinase in cells of the liver. The non-hepatic promoter has been shown to regulate expression of glucokinase in cells of brain and gut of mice (Jetton et al. 1994). The K_m of glucokinase is much higher than the K_m of hexokinase types I, II and III, around 6 mM for the β-cell glucokinase (Xu et al. 1995). This is close to the concentration of glucose in blood, which is around 5 mM, so small changes in glucose concentration in blood can have a large effect on the enzyme's activity. This

lack of inhibition by glucose-6-phosphate and the high K_m for glucose make sense for the role of glucokinase in the liver, where it is responsible for the regulation of blood glucose levels by directly influencing the release of insulin (Pollard-Knight and Cornish-Bowden 1982; Postic et al. 2001; Iynedjian 2009).

Phosphofructokinase is a highly regulated enzyme and possibly the most important regulation point of glycolysis as it is one of the rate determining reactions of glycolysis (Voet and Voet 2011). It can be inhibited or activated by various metabolites in a complex manner, but only some of these aspects will be discussed here. Phosphofructokinases are mostly found in a tetrameric oligomerisation state, which is an important feature for their regulation. Several structures of eukaryotic and bacterial enzymes are known (Evans et al. 1981; Schirmer and Evans 1990; Banaszak et al. 2011). Phosphofructokinase is inhibited by high concentrations of ATP and the inhibitory effect can be enhanced by a high citrate concentration. High levels of AMP in the cell, however, alleviate the inhibitory effect of ATP. Another important (and highly potent) activator of phosphofructokinase is fructose-2,6-bisphosphate (Furuya and Uyeda 1980; Van Schaftingen et al. 1980*a*, *b*, 1981; Pilkis et al. 1981; Uyeda et al. 1981; Cséke et al. 1982), which is synthesized by phosphofructokinase-2, a enzyme with dual kinase/phosphatase function. In archaeal glycolysis, no allosteric control mechanism for ADPGK and ADPPFK has been discovered thus far. Analysis of expression of *P. furiosus* glycolytic enzymes has shown that the genes involved are regulated on a transcription level, depending on the nutrient source (Schut et al. 2003). Also, the Thermococcales-Glycolytic-Motif (TGM) was identified from analysis of the genomes of sequenced Thermococcales species and studied in detail in *P. furiosus* and *Thermococcus kodakarensis*. The TGM is a promoter motif, resembling an inverted repeat with the consensus sequence TATCAC-N5-GTGATA. This motif has been identified in all nine genes coding for the glycolytic proteins. The complexity in transcriptional regulators targeting these regulatory sequences varies from species to species among archaea and is not yet fully understood (van de Werken et al. 2006). In contrast to the hexokinases, archaeal ADPGK seems to be inhibited by its product AMP, but not by glucose-6-phosphate (Verhees et al. 2002; Rivas-Pardo et al. 2013). This has also been reported for the enzyme from *Mus musculus* (Ronimus and Morgan 2004).

1.2 Structural and functional features of the ribokinase superfamily

1.2.1 ADPGK and the ribokinase family of proteins

The structures of ADPGK enzymes from three thermophilic archaea, *Pyrococcus horikoshii, Thermococcus litoralis* and *P. furiosus,* have been solved (Ito et al. 2001, 2003; Tsuge et al. 2002; Rivas-Pardo et al. 2013). The first structures of ADPGK that were solved by X-ray crystallography had the ADPGKs from *P. furiosus*, *T. litoralis* and *P. horikoshii* crystallised in a so-called closed, half-open and open conformation, respectively (Ito et al. 2001, 2003; Tsuge et al. 2002). The archaeal ADPGKs consist of two structural domains, a small or lid domain and a large domain, connected by a β-sheet linker region. The substrate-binding site is located in a pocket, formed between the large and the small domains. Upon substrate binding the small and large domain undergo a conformational change that is described as a 'lid-closing' (Ito et al. 2001). The structure of ADPGK from *T. litoralis* is shown in Figure 2 as an example, to illustrate the overall tertiary structure of the ADPGK enzymes.

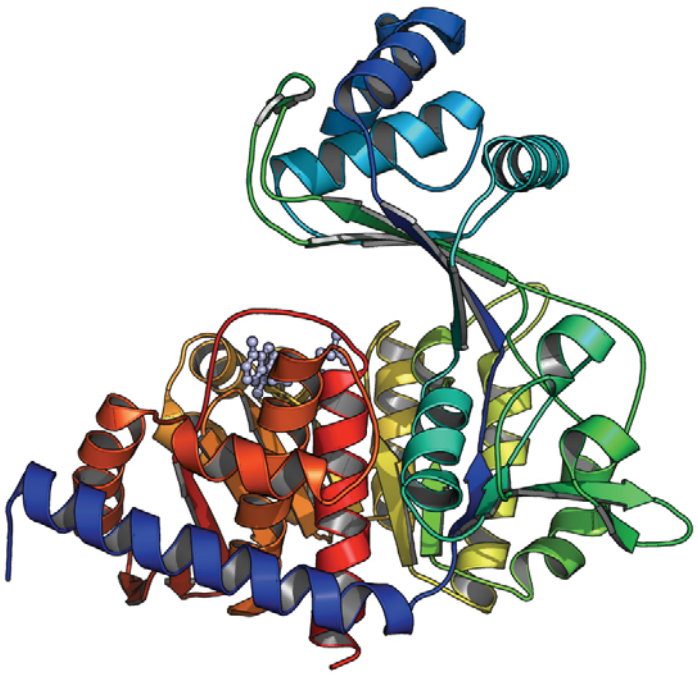

Figure 2: ADPGK from *T. litoralis*.

Cartoon representation of the main chain of ADPGK from *T. litoralis* (PDB 1GC5). The model is coloured from blue to red highlighting the progression of the main chain from the N- to C-terminus. The ligand-binding site is located in the pocket formed by the small and large domain (ligand ADP shown in pale blue, ball and stick representation). The image was generated in PyMOL (DeLano 2002).

Strictly speaking, these are the only structures available of ADPGK enzymes. However, ADPGK does belong to the ribokinase superfamily of proteins, which has been reviewed in detail recently and classified into three branches of different protein folds: Vitamin kinase like, PfkB-like and PfkC-like (Merino and Guixé 2011). The ribokinase-like proteins have a central α/β fold in common, where a central β-sheet is flanked by α-helices on both sides, similar to the Rossman fold of NAD-binding proteins (Rossmann et al. 1974; Sigrell et al. 1998). The most striking difference between members of different branches of the superfamily is found in the small (or lid) domain, which varies in size and secondary structure, or can be completely missing. In Figure 3 the protein structures of selected members of the ribokinase superfamily are visualised for comparison.

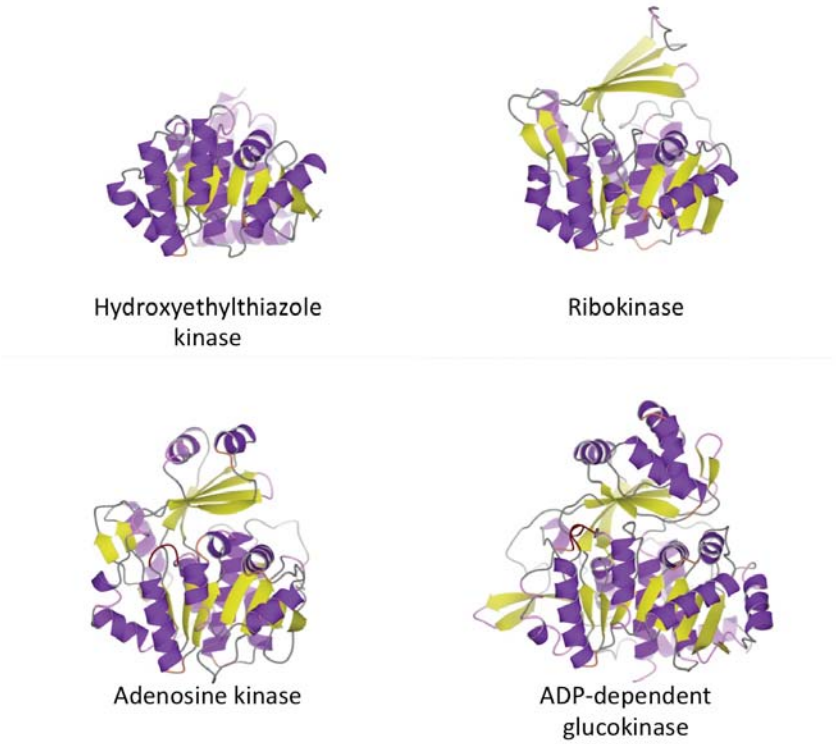

Figure 3: Selected members of the ribokinase superfamily of proteins.
From top left clockwise: Hydroxyethylthiazole kinase from *Bacillus subtilis* (PDB 1EKQ), ribokinase from *E. coli* (PDB 1RKD), ADP-dependent glucokinase from *P. furiosus* (PDB 1UA4), adenosine kinase from *Homo sapiens* (PDB 1BX4). The large domains of the structures were superimposed with the secondary structure match function in CCP4mg (McNicholas et al. 2011). Colour coding is purple for α-helices and 3_{10}-helices, yellow for β-strands and grey for loop regions.

Members of the vitamin kinase-like family lack the small domain completely (see Figure 3) and catalyse the transfer of phosphate to small molecules, such as pyrimidine or pyridoxal. Examples for members of the PfkB-like branch are adenosine kinase, ribokinase, phosphofructokinases and nucleoside kinases. These enzymes commonly catalyse the phosphorylation of various sugar containing-molecules using ATP in the process. A well-studied example is phosphofructokinase-2 from *E. coli* (Cabrera et al. 2010). Phosphofructokinase-1 from *E. coli* on the other hand, does not belong to the PfkB family, but to the PfkA family instead. The PfkA and PfkB families are not phylogenetically related (Ronimus and Morgan 2001). The PfkC branch of the ribokinase superfamily contains the ADP-dependent kinases. These enzymes transfer the terminal phosphate group of ADP to a saccharide substrate.

1.2.2 Structure and function of archaeal ADPGK

There are certain primary structure elements that are conserved throughout the ADP-dependent kinase family and ribokinase superfamily. The sequence identity between the archaeal and the eukaryotic homologs of ADPGKs is generally very low with less than around 20% identity. An alignment of members of the ribokinase family, including ADPGK, ADPPFK, adenosine kinase and ribokinase as examples, is shown in Figure 4. The sequence motif for the active site is GXGD, positioned at the end of a helix, where the aspartate residue (Figure 4, alignment position 567) acts as a catalytic base. This sequence motif is highly conserved in other members of the ribokinase family. It has been shown that mutation of this aspartate to a different amino acid results in the loss of activity (Ito et al. 2001). However, the aspartate residue appears to be not strictly mandatory, as a cysteine residue as been observed in this position in other ribokinase-like enzymes (Ito et al. 2003). In the 4-methyl-5-β-hydroxyethylthiazole kinase from *B. subtilis*, mutation of this cysteine residue to aspartate results in a nine-fold gain of activity (Campobasso et al. 2000). Upon closing of the two domains the terminal phosphate group is stabilized by a conserved arginine residue. This stabilises the phosphate at the catalytic centre and the glucose molecule. The interaction of the catalytic aspartate and the hydroxyl group in 6′-position of the bound glucose triggers the transfer reaction (Ito et al. 2003).

Figure 4: Sequence alignment of members of the ribokinase family.

Sequence alignment of members of the ribokinase family of proteins. Alignment was made with Clustal Omega (Sievers et al. 2011). Pink shading indicates conserved sequence features (see text). hsADPGK = *H. sapiens* ADPGK, mmADPGK = *M. musculus* ADPGK, dr = *Danio rerio* ADPGK, tlADPGK = *T. litoralis* ADPGK, pfADPGK = *P. furiosus* ADPGK, pfADPPFK = *P. furiosus* ADPPFK, tgAK = *Toxoplasma gondii* adenosine kinase, hsAK = *H. sapiens* adenosine kinase, hsRK = *H. sapiens* ribokinase, ec = *E. coli* ribokinase

Another sequence feature essential for the ADP-dependent kinases is an arginine residue (Figure 4, alignment position 270) on the small domain, facing the catalytic site. This arginine is essential for the enzymatic activity and is believed to help stabilise the terminal phosphate during the reaction (Ito et al. 2003). Figure 5 gives an overview of the catalytic site of *P. furiosus* ADPGK, illustrating some of the residues important for catalysis.

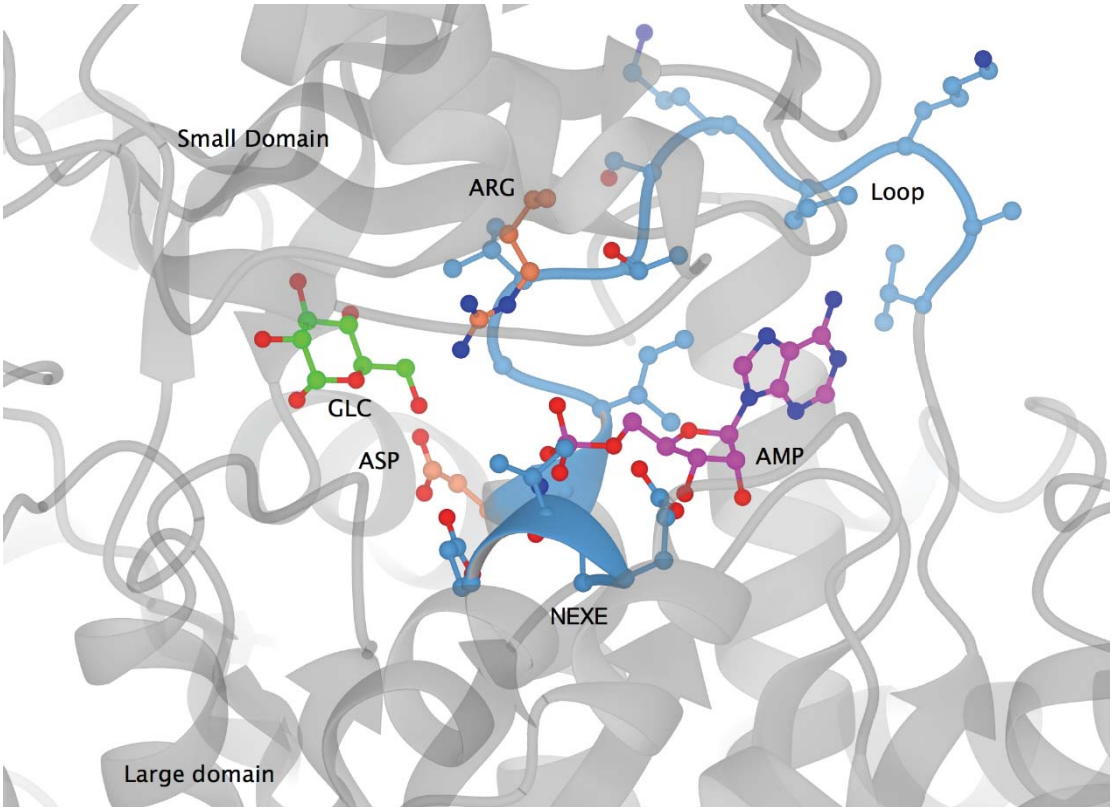

Figure 5: Catalytic site of archaeal ADPGK.

Active site of *P. furiosus* ADPGK (PDB 1UA4). The active site is occupied by glucose (GLC, green) and AMP (AMP, magenta). The key catalytic residues Asp 440 and Arg 197 are orange. The NEXE motif and the nucleotide-binding loop are blue.

The nucleotide-binding site of archaeal ADP-dependent kinases is structurally similar to that of the ATP-dependent ribokinase and adenosine kinases. Their nucleotide-binding elements are located in a similar position and the nucleotide substrate, be it di- or tri-phosphate, is bound in a similar fashion regarding the terminal phosphate groups. However, it appears that the ATP-dependent enzymes allow the nucleoside moiety to enter the active site deeper, thus enabling the enzyme to accommodate three

phosphates instead of two. For archaeal ADPGKs, the active site cavity is lined by residues with large side chains like tyrosine but adenosine kinase or ribokinases have smaller side chains such as threonine or valine. The discrimination between ADP and ATP seems to be more complex than by direct amino acid substitutions, as the binding of ATP also involves a rotation of the adenine moiety. Simply changing these large residues in ADPGK to amino acids with smaller side chains does not immediately generate an enzyme capable of using ATP, which has been shown for the ADPPFK of *P. horikoshii* (Ito et al. 2001; Currie et al. 2009). The base moiety of the nucleotide is recognised by the large nucleotide-binding loop. A phosphate-binding site can be found in members of the ribokinase superfamily with the consensus sequence NEXE (Figure 4, position 379 – 382), which is a defining motif for the ribokinase superfamily (Maj et al. 2002) and has been described as being involved in the nucleotide-binding in the crystal structure of ADP-bound ADPGK from *T. litoralis* (Ito et al. 2001). The first structures of ADPGK solved in the so-called closed, half-closed and open conformation led to theories about the reaction mechanism and movement of the small and large subdomains during catalysis of the ADP-dependent kinases (Ito et al. 2001, 2003; Tsuge et al. 2002). This was supported by mutagenesis studies and crystal structures of a related enzyme in the ribokinase superfamily, ADPPFK from *P. horikoshii* (Currie et al. 2009). A recent study could show the steps of the lid closing by ADPGK from *T. litoralis* by analysis of crystal structures and small angle X-ray scattering, which was supported by kinetic data (Rivas-Pardo et al. 2013). This revealed a sequential ordered enzymatic mechanism, involving a ternary complex, which is shown in the schematic diagram in Figure 6. In inhibition studies it was revealed that the product AMP acts as a competitive inhibitor to ADP and as a mixed type inhibitor to D-glucose. On the other hand D-glucose-6-phosphate, however was a mixed type inhibitor for both substrates, ADP and D-glucose (Rivas-Pardo et al. 2013). This type of product inhibition is indicative for a sequential ordered reaction mechanism, where substrates bind to, and subsequently the formed products leave, the active site for the enzymatic reaction to take place. In the case of *T. litoralis* ADPGK, the inhibition behaviour supports a mechanism where Mg-ADP enters the active site, and Mg-AMP leaves last (shown schematically in Figure 6).

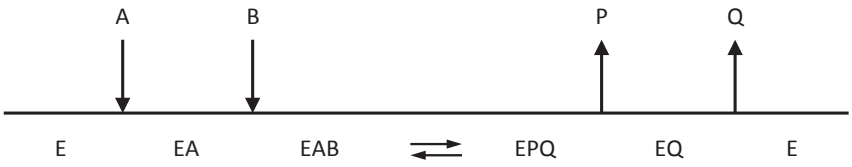

Figure 6: Reaction catalysed by *T. litoralis* ADPGK.

Representation of the sequential ordered bi-bi mechanism of *T. litoralis* ADPGK after Cleland (1963). A is first substrate (Mg-ADP) to enter the active site, B is second substrate (glucose) to enter the active site, P is the first product (glucose-6-phosphate) to leave the active site, Q is the second product (Mg-AMP) to leave the active site, E is the apo-enzyme, EA is the enzyme bound to the first substrate, EAB and EPQ are the ternary complexes with both substrates or products bound, respectively and EQ is the enzyme with second substrate bound.

When substrate binding was observed by small angle X-ray scattering, no, or very little, domain movement was seen when ADPGK was in its apo form or when glucose was bound to the active site. When ADPGK was incubated with ADP, a reduction of the enzyme's radius of gyration was observed, consistent with a proposed domain movement. Finally, the ternary complex with the lid fully closed could be observed when ADPGK was incubated with glucose and a non-hydrolysable ADP-analog. In combination with kinetic studies, this led to the conclusion that Mg-ADP must bind to the enzyme in order to allow glucose to bind. This allows the domain movement and stabilisation of the ternary complex. The product, glucose-6-phosphate is proposed to leave the active site first, followed by Mg^{2+} and AMP. In addition to the enzyme kinetic and small angle X-ray scattering data, crystal structures of the apo- and holo-forms of ADPGK were determined. The structure of the apoenzyme had been published before (Ito et al. 2001) and the holoenzyme was crystallised in complex with D-glucose and Mg-ADPβS. In the structure of the holoenzyme, six hydrogen bonds and one charge-charge interaction were observed between the small and large domains that stabilise the closed form. Some of the residues responsible for these stabilising interactions are highly conserved in different species of archaea, leading to the conclusion that the enzymatic mechanism could be largely conserved as well (Rivas-Pardo et al. 2013).

1.3 ADPGK in eukaryotes

1.3.1 ADPGK in eukaryotes

Eukaryotic ADPGK was first described from *M. musculus* (Ronimus and Morgan 2004), followed by the enzyme from *H. sapiens* (Richter et al. 2012). Eukaryotic ADPGKs were identified by sequence analysis as homologous to the archaeal ADPGKs. In biochemical studies on the ADPGK from mouse, an apparent molecular mass of 60.3 kDa was determined by size-exclusion chromatography, which indicates that the recombinant enzyme is monomeric in solution. In activity assays, the enzyme showed inhibition by high concentrations of glucose (maximum activity around 0.35 mM). The apparent K_m values for ADP and glucose were 0.28 mM and 96 μM, respectively. The recombinant mouse ADPGK was described as highly specific for D-glucose, but could use CDP and GDP as phosphoryl group donors. Although it was found that the enzyme showed highest activity at around pH 6 and pH 9, with specific activities of approximately 16 and 22 U/mg respectively, a role for bimodal pH optima of mouse ADPGK in metabolism has not been established. End product inhibition has been shown for AMP but not glucose-6-phosphate (Ronimus and Morgan 2004).

1.3.2 ADPGK and hypoxia in eukaryotes

Hypoxia occurs during several diseases such as ischemia, strokes, or cancer. The median oxygen partial pressure in healthy tissue ranges between 40 and 60 mmHg, but in hypoxic tumour tissues this value decreases drastically to be less than 10 mmHg. If the oxygen partial pressure in the malignant tissue drops below the threshold of 10 mmHg, a tumour starts to develop resistance to radiation therapies (Brown 1999). Tumour hypoxia has become a very active field of research and attempts are being made to design drugs that are activated under hypoxia (Brown and Wilson 2004). This field is also of particular research interest since some of the most aggressive and fatal forms of cancer form a hypoxic core where cells have very low oxygen supply and it is believed that their ATP generation relies mainly on glycolysis rather than oxidative phosporylation in the mitochondria. This was first observed by Otto Warburg and thus is called the Warburg Effect (Warburg 1947, 1956). It was proposed that an ADP-dependent glucokinase could have positive effects on cell

fitness and survival under hypoxic conditions or act as a regulator for AMP-activated kinase (AMPK) by converting ADP to AMP and serving as an amplifier for an energy depletion signal (Ronimus and Morgan, 2004). Hypoxia leads to ATP depletion, similar to exercise in muscles, and to expression of a large number of target genes. The transcription regulator HIF-1 plays a central role in the regulation of these genes, which acts as a master switch to maintain oxygen homeostasis and regulate the cell's response to changes of oxygen supply (Semenza 2007). Hypoxia-caused ATP shortage in the cell results in increased levels of AMP, which is an indicator of a cell's general energy charge level and can activate the AMP-activated kinase cascade. The function of AMPK is to act as a sensor for the ATP to AMP ratio in a living cell. The signalling cascade will regulate the cells metabolism by shutting down non-vital processes or up-regulating catabolic pathways (glycolysis) to restore the appropriate energy level in the cell. This is called the energy charge hypothesis and was first formulated by Daniel Atkinson in the 1960's. Due to adenylate kinase, the ratios of ATP, ADP and AMP are kept at a certain equilibrium and the ratio of ATP to AMP varies as the square of the ADP to ATP ratio, which makes AMP a very sensitive indicator of the cell energy state (Hardie and Hawley 2001).

Several studies have reported on the role of eukaryotic ADPGK during hypoxic conditions. A systematic RNAi-mediated knock down experiment in *Caenorhabditis elegans* showed that ADPGK increased sensitivity to hypoxia (Mabon et al. 2009). In contrast, two studies of human tumour cell lines H460 and HCT116 could not establish a conclusive role for ADPGK in glycolysis during hypoxia in human cells (Richter et al. 2012, 2013). However, in these studies hexokinase II (HK II) activity was not completely knocked down, which could have masked the contribution of ADPGK to glycolysis in these experiments. In human cells, ADPGK is expressed 4.3 times higher at the RNA level than the average human gene (Richter et al. 2012). The ADPGK transcript was not found to be up-regulated by hypoxia-inducible factor 1 (HIF-1) nor was expression increased under anoxia, which could be consistent with a potential role of cellular energy metabolism. When testing for phenotypes, only a modest additional sensitivity to anoxia was observed when ADPGK and HK2 were suppressed by siRNA silencing. No effect on anaerobic glycolysis was observed, when only ADPGK was suppressed. Overexpression of ADPGK in human tumour cell lines could not rescue the cell during hypoxia or anoxia, with or without co-

suppression of HK II. However, knockdown of ADPGK in H460 cells resulted in lower plating efficiency in aerobic conditions which could indicate increased AMPK signalling (Richter et al. 2012). A knock out of ADPGK with zinc finger nucleases resulted in lower cell survival of H460 under anoxia when HK II and ADPGK were knocked out together (Richter et al. 2013). In a different study, ADPGK, among other candidate genes, was shown to be involved in human peripheral blood monocytes differentiation to immature dendritic cells. The effect was completely lost upon methylation, suggesting that their enhancing activity is critically dependent on their CpG methylation status (Klug et al. 2010). However, the functional implications for ADPGK in this case were not investigated.

1.3.3 ADPGK and T-cell activation in eukaryotes

Recent research has found a link between reactive oxygen species (ROS) induced T-cell activation and ADPGK in mammalian immune cells (Kamiński et al. 2012, 2013). A brief overview of ADPGK-dependent T-cell activation is shown in Figure 7. Upon stimulation of the T-cell receptor (TCR), it was observed that the cells shifted from mitochondrial respiration to aerobic glycolysis, accompanied by uptake of glucose. The metabolic shift was accompanied by major changes in the ultrastructure of the mitochondria. The signal generation was shown to be dependent on increased glycolytic flux and the redirection of the glycolytic flux towards the ubiquinone shuttle in mitochondria. Also, it was shown that the glycolytic flux was ADP dependent and uncoupled from mitochondrial ATP production. The glucose-induced hyper-reduction of the ubiquinone pool and mitochondrial ROS release were dependent on an up-regulated glycolytic flux but enzymatic activities of the major glycolytic regulatory steps, hexokinase (HK), phosphofructokinase (PFK), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), and pyruvate kinase (PK) were found unchanged after TCR triggering. Only the activity of enolase was found to be increased (fat grey arrow in Figure 7), reverting the metabolic flow from 2-phospho-glycerate. The consequent increase of glyercol-3-phosphate leads to a hyper-reduction of the ubiquinone pool in mitochondria via the glycerol-3-phosphate dehydrogenase shuttle. This leads to a reverse electron transfer starting at complex III and generation of the ROS signal at complex I. Upon activation the genes for the

proliferation of T-cells, such as IL-2 and IL-8, were activated by ROS signalling via the transcription factor NF-κB.

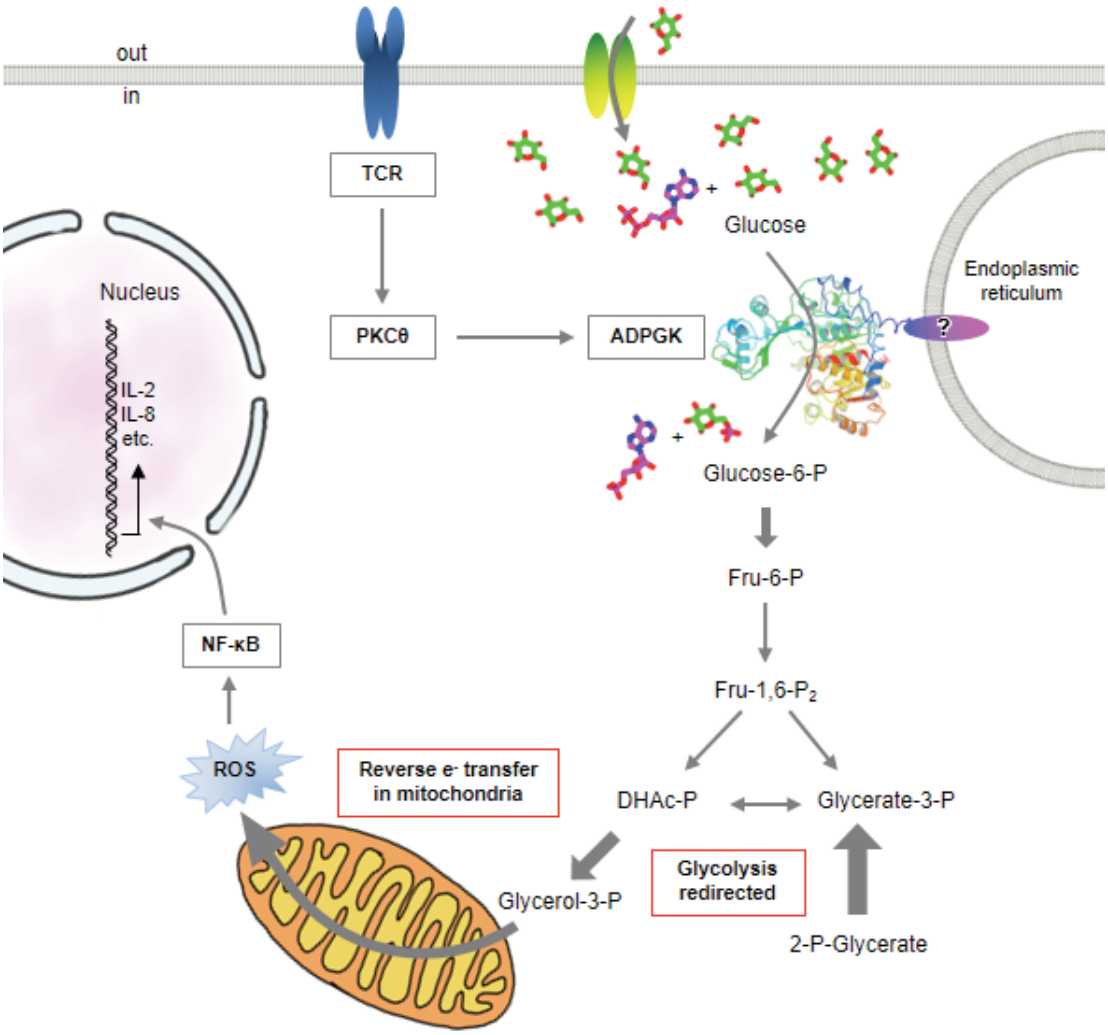

Figure 7: T-cell activation and ADPGK.

Adapted from Kamiński *et al*. (2012) with modifications. ADPGK is part of the redirection of glycolysis and oxidative signal generation in the mitochondria. For further explanation see text.

The increased glycolytic flux was dependent on increased glucose uptake and activation of ADPGK, which was shown to be dependent on protein kinase Cθ (PKCθ in Figure 7). Thus, ADPGK appears to play a critical role in glycolysis in the case of T-cell activation. It was also suggested, that ADPGK is localised in the endoplasmic reticulum (see Figure 7), with the active site protruding into the cytoplasm (Kamiński et al. 2012). A recent screening study has established the presence of a cholesterol

binding site in eukaryotic ADPGK (Hulce et al. 2013). The exact nature of this interaction has not yet been established but a detailed investigation would obviously have important implications for the role of ADPGK in eukaryotic metabolism.

In addition to the above study, two copies of ADPGK genes were identified in *D. rerio*, called *drADPGK1* and *drADPGK2*. Both genes showed about 5-fold lower expression levels when compared to expression levels of the hexokinases I – IV in tissues from liver, muscle, brain and heart. Despite the low expression levels, it was possible to confirm that ADPGK is indeed expressed in zebrafish, and that *Dr*ADPGK1 and *Dr*ADPGK2 are not pseudogenes (González-Alvarez et al. 2009). Interestingly, microarray studies in *Sebastes schlegelii*, the black rockfish, have shown, that ADPGK could be involved in the immune response in fish as well (Kwon et al. 2012). This species of fish is of great commercial interest in Asia, but is very sensitive to stress stimuli and infections, making its cultivation challenging. When *Sebastes schlegelii* leukocytes were treated with a combination of concavalin A and phorbol 12-myristate 13-acetate, the expression level of *ADPGK* was found to be up-regulated more than two-fold after twelve hours. Phorbol 12-myristate 13-acetate has been shown to increase ADPGK expression and/or activity in two separate studies, in fish and human T-cells (Kamiński et al. 2012; Kwon et al. 2012), which could indicate a conserved function for, and regulation/activation of, ADPGK in the immune system of eukaryotes.

1.4 Aims of this work

1.4.1 Structural and functional characterisation of eukaryotic ADPGK

The role of ADPGK in archaea seems to be established as a general replacement for ATP-dependent hexokinases of the glycolytic pathway. The high level of conservation of distinctive sequence features throughout the whole ribokinase family suggests that the structure of the eukaryotic ADPGK is likely similar to that of the archaeal homologs. Recent studies, however, indicate that ADPGK was repurposed during evolution to fulfil a different role in metabolism in eukaryotes. The aims of this thesis were the functional and structural investigation of eukaryotic ADP-dependent glucokinase. The plan was to develop a recombinant expression system for the production of eukaryotic ADPGK in *E. coli* to facilitate analysis of ADPGK. This was especially important for the structural investigations as for protein crystallisation large amounts of protein are required. The constructs were derived from cDNA clones of the human, mouse and zebrafish ADPGK genes, which are commercially available. The design of suitable expression constructs was guided by the archaeal ADPGK enzymes from *T. litoralis*, *P. furiosus* and *P. horikoshii*, for which the three dimensional structures have been solved by X-ray crystallography. For the produced constructs, purification protocols were developed to obtain the target enzyme in the required purity and quantity suitable for the structural and enzymatic studies. Following this, the purified ADPGK was to be used for biochemical characterisation to determine the basic kinetic parameters of the enzyme and to investigate the substrate specificity and potential regulation mechanisms on a enzymatic level as this information is required to understand the role of ADPGK in the metabolism of eukaryotes. The substrate specificity for phosphoryl donors and acceptors other than ADP and glucose was tested. Site-directed mutagenesis was used to identify key residues for enzymatic activity and substrate specificity. The three-dimensional structure of *M. musculus* ADPGK was solved by X-ray crystallography.

2 Materials and Methods

2.1 Equipment and reagents

2.1.1 Water

Water used for all experimental work was purified in a Barnstead NANOpure II system (Thermo Fisher Scientific, USA).

2.1.2 Chemicals

Chemicals used for experimental work were of analytical quality grade or higher, unless indicated otherwise.

2.1.3 cDNA source material

Source vectors are listed in Table 2.1. The clone pBAD-rmADPGK for expression of the *M. musculus* ADPGK (Ronimus and Morgan 2004) was a kind gift from Dr. Ron Ronimus. The cDNA clone for *D. rerio* ADPGK wild type was obtained from Open Biosystems (Thermo Fisher Scientific, USA). The cDNA for *H. sapiens* ADPGK wild type was obtained from Origene (USA). A synthetic cDNA clone for *H. sapiens* ADPGK equivalent to the amino acid sequence of isoform 1 was obtained from GeneArt® (Life Technologies) which was codon optimised for recombinant expression in *E. coli* (Fath et al. 2011).

Table 2.1: List of ADPGK genetic source material.

Organism	Accession	Vector	Description	Selection
H. sapiens	NM_031284	pOTB7	Isoform 1	Cam^R
H. sapiens	Synthetic	pMA-RQ	Equivalent to isoform 1, optimised for expression in *E. coli*.	Amp^R
D. rerio	BC125924	pExpress-1	-	Amp^R
M. musculus	BC021526	pBAD	Isoform 3	Amp^R

The difference between of *H. sapiens* ADPGK isoform 1 and *M. musculus* ADPGK isoform 3, is an inserted glutamine residue following glutamine residue Q313 for isoform 1, resulting in a QQ motif in the expressed protein and amino acid sequence that is longer by one residue. Only the equivalent of isoform 1 is known for the *D. rerio* ADPGK. The cDNA sequences are listed in section 6.2 in the Appendix.

2.1.4 *E. coli* strains for cloning and protein expression

All strains of *E. coli* used in this work are listed in Table 2.2. Strains TOP10 and LMG194 were obtained from Invitrogen (USA). All other strains were obtained from Novagen (Merck Millipore, Germany).

Table 2.2: *E. coli* strains used.

Strain	Genotype and selection marker
TOP10	F^- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL endA1 nupG (Str^R)
BL21 (DE3)	F^- ompT $hsdS_B$ (r_B^- m_B^-) gal dcm (DE3)
Origami (DE3)	Δ(ara leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL F′[lac^+ $lacI^q$ pro] (DE3) gor522::Tn10 trxB (Kan^R, Str^R, Tet^R)
Origami B (DE3)	F^- ompT $hsdS_B$(r_B^- m_B^-) gal dcm lacY1 aphC (DE3) gor522::Tn10 trxB (Kan^R, Tet^R)
Rosetta 2 (DE3)	F^- ompT $hsdS_B$(r_B^- m_B^-) gal dcm (DE3) pRARE2 (Cam^R)
Rosetta-gami (DE3)	Δ(ara–leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL +q 2RRRR4 (DE3) F'[lac lacI pro] gor522::Tn10 trxB pRARE (Kan^R, Str^R, Tet^R, Cam^R)
RosettaBlue (DE3)	endA1 hsdR17(rK12 mK12) supE44 thi-1 recA1 gyrA96 relA1 lac (DE3) [F' $proA^+B^+$ $lacI^q$ ZΔM15 ::Tn10] pRARE (Tet^R, Cam^R)
LMG194	F^- ΔlacX74 ΔaraBADC galE thi rpsL ΔphoA PvuII Δara714 leu::Tn10 (Str^R, Tet^R)

2.1.5 Primers

All primers (see Appendix Table 6.1) used for sequencing, cloning and site-directed mutagenesis of ADPGK were purchased from Integrated DNA Technologies

(Coralville, USA). The purity grade of all primers was standard desalted. Primers for site-directed mutagenesis were obtained as complementary pairs.

2.1.6 Expression vectors

Vectors for expression of recombinant proteins in *E. coli* used for this work are listed in Table 2.3.

Table 2.3: Expression vectors used.

Vector	Description	Selection	Supplier
pBAD TOPO TA	TOPO cloning vector with C-terminal 6×His-tag, *ara* promoter	Amp^R	Invitrogen, USA
pET21-b(+)	C-terminal 6×His-tag, no residues added to N-terminus, *T7* promoter	Amp^R	Merck, Germany
pET-32a(+)	N- and C-terminal 6×His-tag, N-terminal Trx-tag, *T7* promoter	Amp^R	Merck, Germany
pET101/D-TOPO	TOPO cloning vector with C-terminal 6×His-tag, *T7* promoter	Amp^R	Invitrogen, USA
pET151/D-TOPO	TOPO cloning vector with N-terminal 6×His-tag, *T7* promoter	Amp^R	Invitrogen, USA
pGEX-4T-3	N-terminal GST-tag, *trp* promoter	Amp^R	GE Healthcare, UK
pMAL-c2c	N-terminal 6xHis- and MBP-tag, *T7* promoter	Amp^R	New England Biolabs, USA
pMAL-p2c	N-terminal 6xHis- and MBP-tag, periplasmic export sequence, *T7* promoter	Amp^R	New England Biolabs, USA
pProEX HTb	N-terminal 6×His-tag, *trx* promoter	Amp^R	Invitrogen, USA

2.1.7 Crystallisation supplies and screens

Automated setup of crystallisation screens was performed with a mosquito® Crystal liquid handling robot (TTP Labtech, Cambridge, UK). Intelli-plate 96 for high-throughput screening was obtained from Art Robbins Instruments (Sunnyvale, USA). Micro-bridges, 24-well VDX plates, 22 mm siliconised cover slides, Crystal Clear sealing tape, Izit crystal dye, crystal mounts, cryo loops, cryo tools and 96-well microbatch plates were obtained from Hampton Research (USA). Micro Loops™ were

obtained from MiTeGen (USA). Crystallisation screens MIDAS™ HT-96, PACT premier™ , JCSG-plus™ , Morpheus®, Structure Screen 1 and 2 were obtained from Molecular Dimensions (UK). Crystallisation screens PCT™ , Pre-Crystallization Test, Index, Crystal Screen HT, PEG/Ion, Additive Screen, and Silver Bullets Bio were obtained from Hampton Research (USA).

2.1.8 Chromatography columns and system

Chromatography columns HisTrap HP, HisTrap FF Superdex 75 GL, Superdex 200 GL, Superdex 120 PG, HiTrap DEAE Sepharose FF, HiTrap Q Sepharose FF, and GSTrap HP were obtained from GE Healthcare (Little Chlafont, UK). Ion-exchange columns UNO® Q-1 and UNO® Q-6 were obtained from Bio-Rad Laboratories (USA).

2.2 *In silico* analysis of proteins

2.2.1 Alignments

Sequence alignments were performed with Clustal W, Clustal Omega (Larkin et al. 2007; Sievers et al. 2011) or T-COFFEE (Notredame et al. 2000). Sequences and alignments were analysed with CLC Main Workbench software (CLC bio, Aarhus, Denmark). When a sequence logo is displayed with an alignment, residues are coloured by polarity (where green is uncharged polar, red is negatively charged, blue is positively charged and black is all others).

2.2.2 Homology modelling and structure predictions

For homology modelling of the open and closed conformation of ADPGK the protein structure prediction servers Phyre2 (Kelley and Sternberg 2009) and I-TASSER (Roy et al. 2010) were used. Prediction of signal peptides and transmembrane domains was performed with TargetP, SignalP (Emanuelsson et al. 2007; Petersen et al. 2011), TMHMM2.0 (Sonnhammer et al. 1998), Phobius (Käll et al. 2007), iPSORT and PSORT II (Bannai et al. 2002) and Predotar (Small et al. 2004).

2.3 Recombinant protein expression in *E. coli*

2.3.1 Growth media for *E. coli*

The Luria-Bertani broth (LB) growth medium (Invitrogen, USA) was prepared according to the manufacturer's instructions and sterilised by autoclaving at 121 ˚C for 15 minutes. LB-agar plates in Petri dishes were prepared by adding 1% (w/v) of bacto-agar (Invitrogen) prior to autoclaving. The richer 2×YT-medium (16 g tryptone, 10 g yeast extract and 5 g NaCl per litre) used for protein production was prepared to 1.8 – 2 l in 5 l Erlenmeyer flasks and sterilised by autoclaving at 121 ˚C for 15 minutes. Auto-inducing medium ZYM-5052 (10 g tryptone, 5 g yeast extract, 5 g glycerol, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 2 mM $MgSO_4$ and 2.8 mM glucose per litre at neutral pH) was prepared as described by Studier (2005) with the appropriate modifications for lactose-inducing (5.6 mM α-lactose) or arabinose-inducing (0.05 % w/v L-arabinose) expression systems to 2 l in 5 l Erlenmeyer flasks and sterilised by autoclaving at 121 ˚C for 15 minutes. The antibiotics used as selection markers were added after autoclaving, depending on the *E. coli* strain used. All antibiotics were sterile-filtered through a 0.2 μm cellulose acetate filter (Merck Millipore, Germany).

2.3.2 Cryo-stocks of *E. coli* cells

Stocks of *E. coli* cultures for storage at -80 °C were prepared by adding 0.5 ml of culture to 0.5 ml of sterile 80% glycerol in cryo-tubes and flash-frozen in liquid nitrogen before transfer to a -80 ˚C freezer.

2.3.3 Preparation of chemically competent *E. coli* cells

The desired *E. coli* strain was streaked on to a LB agar plate and grown at 37 ˚C. A single colony was used to inoculate 5 ml LB broth media as a pre-culture. From the pre-culture 2 ml was used to inoculate 200 ml LB broth media and the cells grown at 37 ˚C until an OD_{600} of 0.5 – 0.8 was reached. Next, the cells were spun down in a centrifuge at 4,000 × *g* at 4 ˚C for 10 minutes and resuspended in 80 ml ice-cold Buffer 1 (30 mM CH_3CO_2K, 100 mM $RbCl_2$, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15%

(w/v) glycerol, pH 8.5, filtered through 0.2 μm filter) by vortexing. The cells were chilled on ice for 5 minutes and spun down at 4,000 × *g* for 10 minutes at 4 ˚C. The pellet was carefully resuspended in 8 ml ice-cold Buffer 2 (10 mM MOPS/KOH, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% (v/v) glycerol, pH 6.5, filtered through a 0.2 μm filter). Aliquots of 100 μl were prepared, flash-frozen in liquid nitrogen and stored at -80 ˚C.

2.3.4 Transformation of chemically competent *E. coli* cells

An aliquot of competent cells was mixed with DNA (typically between 1 and 100 ng) and incubated on ice for 10 minutes. The uptake of the DNA into the cells was triggered when the transformation mixture had been incubated for 60 seconds at 42 ˚C. After the heat shock the cells were incubated on ice for 2 minutes followed by adding 350 μl LB broth and shaken at 37 ˚C for 15 to 30 minutes. An amount of 10 – 100 μl of the transformed cells was then plated out on LB-agar and incubated at 37 ˚C for 12 hours or until single colonies were visible.

2.3.5 Large scale protein expression in *E. coli*

For the expression of recombinant protein in *E. coli*, a suitable expression strain (see Table 2.2) was selected and transformed as described in section 2.3.4 with an expression vector containing the desired construct of ADPGK. A single colony was picked and used to inoculate a pre-culture in 5 ml of LB broth and grown at 37 ˚C with shaking at 250 rpm for 6 hours. This pre-culture was used to inoculate a second pre-culture in 200 ml LB broth, which was grown at 37 ˚C with shaking at 250 rpm for 12 hours. This second pre-culture was used to inoculate 1.8 - 2 l of growth media (2×YT or LB) in 5 l Erlenmeyer flasks to an initial OD_{600} of 0.08. The cultures were grown until they reached an OD_{600} between 0.5 and 0.8 and inducer was added to start production of recombinant protein. Final inducer concentrations were typically 0.1 mM IPTG or 0.2% (w/v) L-arabinose, depending on the vector used. The cultures were harvested by centrifugation at 4,000 × *g* for 20 minutes at 4 ˚C. The cell pellets were washed once by resuspending in cold TBS buffer and spun down again at 4,000

× *g* for 20 minutes at 4 ˚C. The washed cell pellets were transferred to sterile 50 ml plastic tubes and stored at -20 ˚C for up to two weeks.

2.3.6 Lysis of bacterial cells for soluble protein extraction

Frozen cell pellets were resuspended in lysis buffer at a ratio of 1 to 2 volumes per volume of wet cell pellet, depending on the resulting viscosity. Resuspended cells were lysed by a combination of French press, sonication and lysozyme treatment. Not all three steps were necessarily performed for each purified construct. Lysozyme (Sigma-Aldrich, USA) was added to the resuspended cells to a concentration of no more than 0.1 mg/ml. The resuspended cells were then passed through a French press 2 to 3 times at 4,000 psi. Sonication was performed at 30 – 50% amplitude with a Misonix S-4000 sonicator (Misonix, USA) for approximately 1 minute per 30 ml lysate solution. After cell lysis, 1 U DNase I (Sigma-Aldrich, USA) was added per 10 ml of lysate. Soluble and insoluble components of the lysate were separated by centrifugation at 46,000 × *g* in a Sorvall Evolution™ RC centrifuge (Thermo Fisher Scientific, USA) equipped with a SS-34 rotor (Thermo Fisher Scientific, USA). The lysate was kept on ice during lysis.

2.3.7 Isolation of ADPGK inclusion bodies for refolding

For the isolation of inclusion bodies, competent *E. coli* cells were transformed as described in section 2.3.4 with expression vector containing the desired construct of ADPGK. Cell pellets were prepared as described in section 2.3.5 and lysed by sonication. For this, the cell pellet was resuspended in lysis buffer (50 mM Tris/HCl, 150 mM NaCl, 4 mM DTT, pH 8.5) at a ratio of 10 ml per 1 g wet pellet. Lysozyme (Sigma-Aldrich, USA) was added to the resuspended cells to a concentration of no more than 0.1 mg/ml. Sonication was performed at 30 – 50% amplitude with a Misonix S-4000 sonicator (Misonix, USA) for approximately 1 minute per 30 ml lysate solution. The lysate was kept at 4 ˚C during sonication. Following lysis, 1 U DNase I (Sigma-Aldrich, USA) was added per 10 ml of lysate to remove DNA and the lysate was spun down at 8,000 × *g* in a Sorvall Evolution™ RC centrifuge (Thermo Fisher Scientific, USA) equipped with a SS-34 rotor (Thermo Fisher Scientific, USA)

at 10 ˚C for 15 minutes. The centrifugation was repeated three times with wash buffer 1 (50 mM Tris/HCl, 150 mM NaCl, 2% (v/v) Triton X-100, pH 8.5), wash buffer 2 (50 mM Tris/HCl, 150 mM NaCl, 2 mM CHAPS, pH 8.5) and wash buffer 3 (50 mM Tris/HCl, 2 M, pH 8.5). The washed inclusion bodies were stored at -80 ˚C.

2.4 Molecular cloning of nucleic acids

2.4.1 DNA sequencing

DNA sequencing was performed by the Massey Genome Service using a capillary ABI3730 Genetic Analyzer (Applied Biosystems, USA). Primers used for sequencing are listed in Table 6.1.

2.4.2 Polymerase chain reaction

The DNA-fragments used for cloning the *ADPGK* gene into vector pBAD TOPO® TA (Invitrogen, USA) were amplified using *Taq* polymerase (Roche, Germany). All other DNA fragments, including for directional TOPO® cloning were amplified with *Pwo* polymerase (Roche, Germany). Reaction mixtures contained 1 mM dNTP mix (Roche, Germany), reaction buffer, 2.5 U polymerase, 25 – 50 ng template DNA and 0.5 μM forward and reverse primer (see Table 6.1 in Appendix). The PCR mixture was incubated at 94 ˚C for 30 seconds followed by 22 - 35 cycles of 30 seconds denaturation at 94 ˚C, 30 seconds annealing at 52 – 58 ˚C, 3 – 6 minutes extension at 68 or 72 ˚C and final extension for up to 8 minutes.

2.4.3 Agarose gel electrophoresis

Gels were made of 1% (w/v) agarose in TAE buffer (200 mM Tris/acetate, 50 mM EDTA, pH 8.0). The mixture was heated until the agarose dissolved. Gels were cast and subsequently run in an Owl™ EasyCast™ electrophoresis system (Thermo Fisher Scientific, USA). The DNA samples were mixed with 6× DNA loading dye (40% (w/v) sucrose, bromophenol blue dye) prior to loading the gel and the gels were run at 120 Volts. The 1 kb plus ladder (Invitrogen, USA) was used as a size standard for all gels. The gels were stained in ethidium bromide solution (0.5 μg/ml) for 30 minutes and destained in water for 60 minutes. The stained DNA was visualized with UV light and imaged using GelDoc equipment (Bio-Rad Laboratories, USA).

2.4.4 Restriction digest of DNA for cloning

Restriction digestion of purified plasmid vectors and PCR fragments contained 1 μg DNA, 1 U of the required restriction enzyme, BamHI, EcoRI, EcoRV, HindIII, NcoI, NdeI or XhoI (Roche, Germany), reaction buffer according to the manufacturer's instructions, and 1 mg/mL BSA (New England Biolabs, USA). The total reaction volume was 30 μL. The mixture was incubated for 90 minutes at 37 ˚C (HindIII for 3 to 6 hours) and the enzyme subsequently heat-inactivated at 65 ˚C for at least 15 minutes. Successful digestion was assessed by agarose gel electrophoresis as described in section 2.4.3 and the digested fragments were then purified using the PCR purification kit (Roche, Germany) in preparation for ligation (section 2.4.5).

2.4.5 Ligation using T4 DNA Ligase

Ligation was performed using 1 U T4 DNA Ligase (Invitrogen, USA) in 10 μl total volume of ligase buffer, using a vector to insert molar ratio of 1:5, as estimated from agarose gel DNA band intensities. The mixture was incubated at room temperature for 60 minutes and then transferred to 4 ˚C and incubated for 12 to 16 hours. The whole mixture was transformed into *E. coli* TOP10 cells as described in section 2.3.4.

2.4.6 Ligase-free TOPO® cloning

Cloning in vectors pBAD TOPO® TA, pET101/D-TOPO®, pET151/D-TOPO® (Invitrogen, USA) was carried with a molar ratio of vector to insert of 1:1 according to the manufacturer's instructions. Fresh PCR product was mixed gently with 1 μL salt solution (provided in kit) and 1 μL TOPO® vector in a total volume of 6 μL and incubated for 30 minutes at room temperature. The whole reaction mix was then transformed into chemically competent *E. coli* TOP10 cells as described in section 2.3.4.

2.4.7 Colony PCR

The colony PCR method was used to quickly identify clones with the correct insert. A PCR master mix for *Taq* DNA polymerase (Roche, Germany) was prepared according

to the manufacturer's instructions with the primers used for cloning. When available one of the primers was substituted for a vector-specific primer (see Table 6.1 in Appendix). The transformed *E. coli* TOP10 cells were grown on a LB-agar plate for 16 hours or until colonies were visible and single colonies were picked with a sterile toothpick. The toothpick was dipped briefly in 30 µl PCR mix, and then touched on a replica plate for later use. PCR reactions were analysed by agarose gel electrophoresis as described in section 2.4.3.

2.4.8 Procedure for *in vitro* site-directed mutagenesis

Mutants of human ADPGK were generated using a modified version of the QuikChange® protocol (Stratagene, La Jolla, USA). The whole plasmid carrying the appropriate recombinant gene for ADPGK was amplified with complementary primer pairs carrying the desired mutation using *Pfu* Turbo II HS DNA polymerase (Agilent Technologies, Mulgrave, Australia). The reaction mixture contained 2.5 U polymerase, 500 nM dNTP mix (Roche, Germany), 12 pmoles of each complimentary primer, 10 ng template DNA, and *Pfu* Turbo II HS reaction buffer (Agilent, Santa Clara, USA) in a total volume of 50 µl. The reaction was incubated at 94 ˚C for 30 seconds followed by 16 cycles of 30 seconds denaturation at 94 ˚C, 1 minute annealing at 55 ˚C and 8 minutes elongation at 62 ˚C. After the amplification 10 U DpnI nuclease (New England Biolabs, USA) was added directly to the PCR mix and incubated at 37 ˚C for 90 minutes to remove any template (methylated) DNA. The non-digested non-methylated PCR product was then transformed into chemically competent *E. coli* TOP10 competent cells as described in section 2.3.4. A single colony was selected, grown in 5 ml LB/Amp broth at 37 ˚C overnight and the plasmid purified with a plasmid purification kit (Roche, Germany) and sent for sequencing as described in section 2.4.1 using suitable primers flanking the mutation site.

2.5 Protein purification and analysis

2.5.1 Protein quantitation

The protein concentration was determined based on the absorption of light at a wavelength of 280 nm by aromatic residues tryptophan and tyrosine. The extinction coefficients of proteins were determined with the program ProtParam, available on the ExPASy server (Wilkins et al. 1999; Gasteiger 2003). The protein concentration in mg/ml can be calculated after the Beer-Lambert law using the following equation under consideration of the molecular weight of the protein:

$$A = \varepsilon \times b \times c$$

A = Absorbance of the protein solution at 280 nm

ε = Wavelength-dependent molar absorptivity coefficient at 280 nm (l mol^{-1} cm^{-1})

b = Path length of the sample (cm)

c = Concentration (mol l^{-1})

The determination of protein concentration based on the extinction of UV light does not work reliably with solution of denatured protein with high concentrations of urea or guanidine hydrochloride, which are typically used during protein refolding. For this purpose, the concentration of protein was determined with the Bradford protein assay (Bradford 1976). A series of standards was set up using solution of known concentrations of bovine serum albumin (Sigma-Aldrich, USA). The test protein and the standards were mixed with 3 ml of Bradford reagent (0.01% (w/v) Coomassie Brilliant Blue G-250, 4.7% (w/v) ethanol, 8.5% (w/v) phosphoric acid) and incubated for 5 minutes at room temperature. After the incubation, the absorbance at 595 nm was determined in an UV/Vis spectrophotometer. The protein concentration of the unknown sample was determined from a standard curve (see Figure 83 in Appendix), Samples and standards were measured as triplicates.

2.5.2 SDS-PAGE

SDS-PAGE was performed according to Laemmli's method (Laemmli 1970). Gels used for analysis of ADPGK were prepared as mini-gels with a 10% separation and a 5% stacking gel from an acrylamide solution with a molar ratio of

acrylamide:bisacrylamide of 29:1 (Bio-Rad Laboratories, USA). Protein samples were mixed with 2× sample buffer (125 mM Tris/HCl pH 6.8, 2.5% (w/v) SDS, 10 mM DTT, 50% (w/v) glycerol, 0.05% (w/v) bromophenol blue) and heated to 90 ˚C for 5 minutes before loading on to the gel. Gels were run at 200 Volts for 40 – 50 minutes. As reference for molecular weight, the Precision Plus Protein™ (Bio-Rad Laboratories, USA) was used. The gel lane with standard was always labelled kDa in figures.

2.5.3 Native PAGE

Non-dentauring or native PAGE was performed as described in section 2.5.2 but omitting detergent in the polyacrylamide gel, running buffer and sample buffer. The sample buffer also contained no reducing agent and the protein samples were not heated before loading on the gel. Non-denaturing gels were run at 100 Volts for approximately 90 – 120 minutes.

2.5.4 Coomassie Blue staining of protein gels

For the detection of proteins in polyacrylamide gels the gels were incubated with Coomassie staining solution (2.5 g/L Coomassie Brilliant Blue R-250, 10% (v/v) acetic acid and 50% (v/v) methanol) and destained in destaining solution (10% (v/v) acetic acid and 10% (v/v) methanol). A digital image was taken with a GelDoc system (Bio-Rad Laboratories, USA), photocopier/scanner or a digital camera.

2.5.5 Western blotting

Western blotting was performed as described by Towbin et al. (1979). Samples were transferred from polyacrylamide gels to a nitrocellulose membrane (Hybond™-C extra supported nitrocellulose membrane, Amersham) with a Tank-blot system (Bio-Rad Laboratories, USA) at 100 Volts for 60 minutes at 4 ˚C in transfer buffer (2 M glycine, 50 mM Tris, 20% (v/v) methanol). Successful transfer was assessed by Ponceau S staining (0.1% Ponceau S dye in 5% (v/v) acetic acid), or by the use of pre-stained protein standards (Bio-Rad Laboratories, USA). The membrane was

washed in water and incubated in blocking solution (5% (w/v) skim milk powder in TBST) for 90 minutes. After washing two times in TBST the blot was incubated in primary antibody solution (1% (w/v) skim milk powder in TBST, monoclonal mouse anti-ADPGK antibody, diluted 1:2500) for 120 minutes. The primary monoclonal antibody was obtained from Sapphire Bioscience (Australia). The membrane was washed three times in TBST before incubation in secondary antibody solution (1% (w/v) skim milk in TBST, secondary HRP-conjugate anti-mouse antibody (Abcam, UK), diluted 1:5000) for 45 minutes. The membrane was washed four times before developing in TMB staining solution (Thermo Fisher Scientific, USA).

2.5.6 Procedure for expression trials

The expression strain of *E. coli* was transformed as described in section 2.3.4 with an expression vector containing the desired construct of ADPGK. A single colony was picked and used to inoculate 5 ml LB broth as pre-culture and grown at 37 ˚C with shaking at 250 rpm until the OD_{600} reached 5. This pre-culture was used to inoculate 50 mL growth media in 225 ml Erlenmeyer flasks or 5 ml growth media in glass culture tubes to an initial OD_{600} of 0.08. Inducer was added at an OD_{600} between 0.5 – 0.8. At certain time intervals, 1 ml of cells was harvested in 1.5 mL Eppendorf tubes usin a tabletop centrifuge at 14,000 × *g* for 3 minutes. The supernatant was discarded and the cell pellet was frozen at –20 ˚C. For analysis of the samples, the pellet was resuspended in 200 µl lysis buffer (100 mM Tris/HCl pH 8.0, 100 mM NaCl, 1 mM DTT) and kept at 4 ˚C. The cells were lysed by sonication with a Misonix S-4000 sonicator (Misonix, USA) for 15 seconds at 30% amplitude and put on ice. A sample of whole cell lysate was mixed with SDS sample buffer in preparation for electrophoresis (see section 2.5.2), heated to 95 ˚C for 5 minutes and stored at -20 ˚C. The remaining lysate was spun down in a table-top centrifuge at 21,000 × *g* for 15 minutes at 4 ˚C. A sample of the supernatant for the soluble protein fraction was mixed with SDS sample buffer and stored at -20 ˚C. The supernatant was discarded and the remaining pellet resuspended in a standardized amount of SDS sample buffer heated to 95 ˚C for 5 minutes and stored at -20 ˚C for analysis of insoluble proteins in inclusion bodies. For analysis of protein expression, 5 µl of each sample was loaded

on a 10% gel for SDS-PAGE as described in section 2.5.2. The sample for insoluble protein was usually not analysed by SDS-PAGE when screening for soluble protein.

2.5.7 Immobilised metal ion affinity chromatography

All buffers used for immobilised metal ion affinity chromatography (IMAC) were filtered through a 0.2 μm cellulose acetate filter (Sartorius, Germany) and degassed. Affinity chromatography was performed on a Äkta Explorer chromatography system (GE Healthcare, UK). All steps were performed at 4 °C. Either a Ni^{2+}-NTA HisTrap FF or HP 5 ml column (GE Healthcare, UK) was pre-equilibrated with binding buffer (buffer A). Samples were clarified by centrifugation and filtered through 0.45 μm and 0.22 μm cellulose acetate filters (Sartorius, Germany) before loading on to the column with a P-690 sample pump at a flow rate of 1 ml/min. Protein was eluted with a continuous or stepped gradient of elution buffer containing imidazole (buffer B). The elution fractions were analysed by SDS-PAGE as described in section 2.5.2. The column was afterwards washed, stripped of Ni^{2+} and regenerated according to the manufacturer's instructions. While not in use the columns were stored in 20% (v/v) ethanol.

2.5.8 IMAC purification of denatured protein for refolding

All buffers used for IMAC purification were filtered through a 0.2 μm cellulose acetate filter (Sartorius, Germany) and degassed. IMAC purification of denatured ADPGK in preparation for refolding was performed in buffers containing 8 M urea. The inclusion bodies were prepared as described in section 2.3.7 and resuspended in inclusion body solubilisation buffer (50 mM Tris/HCl, 150 mM NaCl, 8 M Urea, pH 9). The solubilized inclusion bodies were loaded on to a Ni^{2+}-NTA HisTrap FF 5 ml column (GE Healthcare, UK) pre-equilibrated in binding buffer (50 mM Tris/HCl, 250 mM NaCl, 4 mM DTT, 8 M urea, pH 8.0) with a peristaltic pump. The bound protein was washed with binding buffer and eluted stepwise with 2 cv of 10 mM, 100 mM, 225 mM, 350 mM and 500 mM imidazole in binding buffer at a flow rate of 1 ml/min. The elution fractions were analysed by SDS-PAGE as described in section 2.5.2.

2.5.9 Ion-exchange chromatography

All buffers used for ion-exchange chromatography were filtered through a 0.2 μm cellulose acetate filter (Sartorius, Germany) and degassed. Ion-exchange chromatography was performed using a Äkta Explorer chromatography system (GE Healthcare, UK). All steps were performed at 4 ˚C. Samples were dialysed into low ionic strength buffer (≤40 mM NaCl) and filtered through 0.22 μm cellulose acetate filters before loading on to a UnoQ1 or UnoQ6 column (Bio-Rad Laboratories, USA) with a P-960 sample pump at a flow rate of 1 – 5 ml/min, depending on the column used. Protein elution was performed with either a continuous gradient or stepped gradient to 1 M NaCl in binding buffer. The elution fractions were analysed by SDS-PAGE as described in section 2.5.2.

2.5.10 Size-exclusion chromatography

All buffers used for size-exclusion chromatography were filtered through a 0.2 μm cellulose acetate filter (Sartorius, Germany) and degassed. Size-exclusion chromatography was performed on a Äkta Explorer chromatography system (GE Healthcare, UK). All steps were performed at 4 ˚C. Before use, the column was washed in water and pre-equilibrated in running buffer. Samples were concentrated and centrifuged at 21,000 × g for 15 minutes at 4 ˚C or filtered through 0.1 μm PVDF spin filters (Merck Millipore, Germany) to remove aggregated proteins before loading on to the pre-equilibrated column. The sample volume loaded was depended on the column used (250 μl or less for Superdex S200 10/300 GL columns and 5 ml or less for Superdex S200 16/60 columns, both GE Healthcare, UK). The elution fractions were analysed by SDS-PAGE as described in section 2.5.2. To determine the molecular mass and oligomerisation state of ADPGK, molecular weight standards (Sigma-Aldrich, USA) were run and a calibration curve determined (see Figure 84 in Appendix). The calibration curve was plotted as partition coefficient K_{av} against the molecular weight (logMw). The partition coefficient was determined by the following equation:

$$K_{av} = \frac{V_e - V_0}{V_c - V_0}$$

K_{av} = Partition coefficient

V_e = Elution volume in ml

V_0 = Void volume in ml

V_c = Geometric column volume in ml

2.5.11 Enzymatic assays and biochemical characterisation of ADPGK

Assays were performed in duplicate or triplicate in 100 µl per reaction in microcuvettes (Starna, Atascadero, USA) using a Cary 300 UV-Vis spectrometer (Agilent Technologies, Mulgrave, Australia). ADPGK activity was monitored in a coupled assay with glucose-6-phosphate dehydrogenase (EC 1.1.1.49). Assays contained 50 mM buffer Bis-tris propane (pH 7.0 at 37 ˚C), 100 mM KCl, 1 mM $NADP^+$, a molar 1:1 ratio of ADP to $MgCl_2$ and at least 2 U glucose-6-phosphate dehydrogenase from *L. mesenteroides* (Sigma-Aldrich, USA) (Olive et al. 1971). The purified ADPGK enzyme was added last in an amount so that the change in absorbance at a wavelength of 340 nm was in the range of 0.1 to 0.2 per minute. Care was taken to keep glucose-6-phosphate dehydrogenase in excess at all times. For measurement of pH sensitivity, the pH of the Bis-tris propane buffer was varied from pH 6.0 to pH 9.25 in steps of 0.25. The values for the maximum velocity V_{max} and Michaelis constant K_m were determined by non-linear regression with Prism 6.0 (GraphPad Software) using the appropriate model for Michaelis Menten kinetics (Copeland 2000):

$$v = \frac{V_{max}\,[S]}{K_m + [S]}$$

v = Enzyme velocity

V_{max} = Maximum enzyme velocity

$[S]$ = Substrate concentration

K_m = Michaelis-Menten constant

The values V_{max} and K_m were determined by varying one susbstrate while keeping the concentration of the other substrate constant at 1 mM and 0.7 mM for Mg-ADP or

glucose, respectively, as these maximum velocity was reached at these values and the enzyme inhibited by higher concentrations. Therefor the apparent values values K_m^{app} and V_{max}^{app} were reported. The values of K_m^{app} and V_{max}^{app} in this case were determined by non-linear regression with Prism 6.0 (GraphPad Software) using a model considering substrate inhibition (Copeland 2000):

$$v = \frac{V_{max}}{1 + \frac{K_m}{[S]} + \frac{[S]}{K_i}}$$

v = Enzyme velocity

V_{max} = Maximum enzyme velocity

$[S]$ = Substrate concentration

K_m = Michaelis-Menten constant

K_i = Inhibition constant

For inhibition studies, the concentration of Mg-ADP was varied against the concentration of AMP at a constant glucose concentration of 0.7 mM. The kinetic parameters determined by non-linear regression with Prism 6.0 (GraphPad Software) using the appropriate model for competitive inhibition (Copeland 2005):

$$v = \frac{V_{max}[S]}{[S] + K_m\left(1 + \frac{[I]}{K_i}\right)}$$

v = Enzyme velocity

V_{max} = Maximum enzyme velocity

$[S]$ = Substrate concentration

K_m = Michaelis-Menten constant

$[I]$ = Inhibitor concentration

K_i = Inhibition constant

2.5.12 Dye assay for ADPGK activity

The dye assay for activity of ADPGK was basically performed as described for hexokinases (Gabriel and Gerstent 1992), but interchanging ATP required for hexokinase for ADP required for ADPGK. The concentrations for the substrates were

increased, compared to typical assays, for the kinetic characterisation to achieve sufficient staining that could be observed *in situ*. The assay is based on the reduction of thiazolyl blue tetrazolium bromide (Sigma-Aldrich, USA) to a blue formazan precipitate via phenazine methosulfate (Sigma-Aldrich, USA) from the NAD(P)H produced by glucose-6-phosphate dehydrogenase in the coupled assay. The reactions contained 10 mM $NADP^+$, 10 mM $MgCl_2$, 10 mM ADP, 10 mM glucose, 16 μg/ml PMS, 0.1 mg/ml thiazolyl blue, 2 U/ml glucose-6-phosphate dehydrogenase, 50 mM Bis-tris propane at pH 7.0 and were typically performed in PCR-tubes or 96-well plates in a total volume of 100 μl. After adding a solution containing the enzyme, the mixture was incubated for 30 minutes at 37 ˚C and, absorbance readings were measured in plate reader at 490 and 595 nm.

2.5.13 NMR-based ADPGK enzyme assays

The ability of ADPGK to utilise sugar substrates other than glucose was examined by nuclear magnetic resonance (NMR) spectroscopy by monitoring the conversion of ADP into AMP. The reaction mixture contained 50 mM buffer Bis-tris propane (pH 7.0 at 37 ˚C), 100 mM KCl, 1 mM Mg-ADP and 1 mM of variable sugar substrate. The reaction was started by adding 0.0025 mg of purified rhADPGK and incubated at 37 ˚C for 30 minutes. The reaction was stopped by adding 10 mM EDTA and heating to 80 ˚C for 5 minutes. The following protocol for the collection of NMR spectra was developed by Dr. Alexander Goroncy. For NMR measurements, 5% D_2O was added for deuterium frequency lock to compensate for slight magnetic drifts of the NMR magnet. A Bruker Avance 400 MHz NMR spectrometer, operating at 400.13 MHz proton frequency, was used with a 5 mm QNP 1H/13C/31P/19F probehead (Bruker-Biospin GmbH, Rheinstetten, Germany). One-dimensional proton (^1H) spectra were acquired with presaturation or excitation sculpting to suppress the abundant water signal and 32768 data points, 32 scans, 16.0204 ppm spectral width (6410.256 Hz), 2.0 s pulse delay time. One-dimensional phosphorus (^{31}P) spectra were acquired with and without power-gated decoupling with 65536 data points, 2048 scans, 400.9142 ppm spectral width (64935.066 Hz), 4.0 s delay time. Two-dimensional ^{31}P-^1H HMBC spectra were recorded with a pulse sequence that is optimised for long range couplings, no decoupling during acquisition and using shaped pulses for off-

resonance presaturation. Parameters included: 4096 × 64 data points, 16.0204 ppm × 20.0013 ppm sweep width, 256 scans, 3.0 s relaxation delay time. Spectra were processed with Topspin versions 1.3 and 2.1 (Bruker-Biospin GmbH, Rheinstetten, Germany). Control reactions without enzyme or ADP, and with 1 mM glucose were performed. The NMR assignments were confirmed by adding known amounts of ADP, AMP and ATP.

2.5.14 Circular dichroism spectroscopy

All buffers used for circular dichroism spectroscopy were filtered through a 0.2 µm cellulose acetate filter (Sartorius, Germany) and degassed. Circular dichroism spectroscopy measurements were performed with a Chirascan™ CD spectromter (Applied Photophysics, Leatherhead, UK) using a 250 µl quartz glass cuvette with 1 mm path length (Hellma®, Müllheim, Germany). Protein samples were dissolved in water to a concentration of 0.05 to 0.2 mg/ml. Spectra were collected over a range from 190 – 260 nm in 1 nm steps with a measuring time of 0.5 seconds. The temperature was kept constant at 20 ˚C during data collection. Spectra were averaged from 10 repeat measurements using the Pro Data Viewer software and presented as mean residue ellipticity. The mean residue ellipticity ($[\Theta]_{mrw}$) and mean residue weight (MRW) were calculated after the following formulas (Kelly et al. 2005):

$$[\Theta]_{mrw,\lambda} = \frac{MRW \times \Theta_\lambda}{10 \times d \times c}$$

$$MRW = \frac{M}{(N-1)}$$

$[\Theta]_{mrw,\lambda}$ = Mean residue ellipticity at wavelength λ

MRW = Mean residue weight

Θ_λ = Ellipticity at wavelength λ

d = Pathlength of the sample in cm

c = Protein concentration in mg/ml

M = Molecular mass of the protein in Da

N = Number of residues in the protein

2.6 Protein crystallisation and structure determination

2.6.1 Protein crystallisation

Protein samples were judged by SDS-PAGE to ensure at least 95% purity before crystallisation. The samples were transferred into suitable buffer by dialysis or size-exclusion chromatography and directly prior to crystallisation filtered through a spin filter (500 µl Durapore 0.1 µm PVDF membrane filters, Merck Millipore, Germany) or by centrifugation at 21,000 × *g* for 15 minutes at 4 ˚C. High-throughput screening was performed with a Mosquito robot in 96-well Intelli plates (Greiner) for sitting drop vapour diffusion crystallisation. Under-oil crystallisation experiments were set up manually in 96-well microbatch plates. Optimisation screens were set up either automated or manually in 24-well VDX plates as sitting drops in micro-bridges or as hanging drops on siliconised glass cover slides and sealed with grease. Screens were transferred to 21 ˚C or 4 ˚C incubators and examined under a microscope directly after set up, daily for one week and then once a week.

2.6.2 Limited proteolysis

Limited proteolysis (Dong et al. 2007; Wernimont and Edwards 2009) was performed using the Floppy Choppy kit (Jena Bioscience) according to the manufacturer's instructions. Suitable conditions were pre-screened and stable fragments identified by SDS-PAGE. Protease was added to the protein sample directly before setting up the crystallisation screen.

2.6.3 Reductive methylation

Reductive methylation was performed essentially as described by Shaw et al. (2007). Purified mouse ADPGK was exchanged into phosphate buffer (50 mM Na_2HPO_4, 150 mM NaCl, pH 7.5) by size-exclusion chromatography. The elution fractions containing ADPGK were pooled and 500 µl of 2 mg/ml ADPGK solution were mixed with 5 µl 2 M dimethylamine borane complex solution (Sigma-Aldrich, USA) and 20 µl 1 M formaldehyde, incubated at 4 ˚C shaking at 100 rpm for 2 hours. The step of adding 5 µl 2 M dimethylamine borane complex solution and 20 µl 1 M formaldehyde

solution, followed by 2 hours incubation was repeated 3. In a final incubation step, another 5 µl 1 M dimethylamine borane complex solution was added and the mixture was incubated for 12 hours. This was followed by addition of 50 µl 1.5 M Tris pH 8.0 to quench the reaction. No chemicals were added to the control. The methylated protein was purified by size-exclusion chromatography in TBS buffer amines (20 mM Tris, 150 mM NaCl, pH 8.0), which further contributed to quench the reaction, and the eluted protein was analysed by native PAGE as described in section 2.5.3.

2.6.4 Data collection of ADPGK crystals

Crystals of ADPGK were harvested in a Cryo Loop (Hampton Research, USA) or Micro Loop (MiTeGen, USA) of suitable size and briefly incubated in a drop of perfluoropolyether cryoprotectant solution (Hampton Research, USA) before flash-freezing in liquid nitrogen. For data collection the loop holding the crystal was mounted on to the goniometer head in a stream of nitrogen gas at 120 K. Data were collected in-house or at the Australian Synchrotron. In-house data collection was performed with a R-Axis IV^{++} image plate area detector and MicroMax™-007 X-ray source with copper anode (Rigaku, USA). Data collection at the Australian Synchrotron was performed at beamline MX2.

2.6.5 Phasing, refinement and molecular graphics software

Software packages that were used for phasing and refinement were XDS (Kabsch 2010), CCP4 (Winn et al. 2011) and Phenix (Adams et al. 2010). Model re-building and visual interpretation of electron density maps was performed using Coot version 0.7.1 (Emsley et al. 2010). For display and rendered images of protein structures, PyMOL version 0.99 (DeLano 2002), Chimera version 1.8 (Pettersen et al. 2004) and CCP4 molecular graphics version 2.7.3 (McNicholas et al. 2011).

3 Results

3.1 *In silico* analysis, construct design, molecular cloning and expression analysis of eukaryotic ADPGK

3.1.1 Introduction

The *in silico* analysis of a protein is a valuable tool to make informed decisions in the further course of a protein structure and function project. Analysis of sequence motifs and three dimensional structures of eukaryotic ADPGK homology generated by homology modelling, all in comparison with the known experimental structures of archaeal ADPGKs, was used to guide the design of constructs for recombinant expression in *E. coli* and the selection of meaningful sites for functional mutations. For practical reasons, this analysis was performed on the ADPGK sequences of *D. rerio*, *H. sapiens* and *M. musculus* where cDNA was available.

To facilitate biochemical and structural studies of ADPGK, variants of the enzyme were cloned into vectors suitable for recombinant expression in *E. coli*. For this study, variants of human, zebrafish and mouse ADPGK were created with different affinity tags for purification and sequence truncations with varying lengths. The expression constructs were named after the scheme (vector)-(protein)(truncation)-(mutation), e.g. pProEX-rhADPGKΔ50-D481A would be recombinant human ADPGK, with a truncation (Δ) of residues 1 – 50 (inclusive) and point mutation D481A, cloned in to expression vector pProEX HTb. The truncations and mutations were designed based on the results of homology modelling and database searches to identify domain regions of ADPGK suitable for overexpression and crystallisation. As the focus of this work was to analyse recombinant ADPGK produced in *E. coli*, a large number of expression trials were conducted to identify suitable expression vectors, truncations and expression conditions. The ADPGK constructs listed in Table 3.1 were tested with the expression strains listed in Table 2.2, with variable expression temperature, growth media, expression inducer concentration and length of expression before cell harvest. For practical reasons, not all possible combinations could be tested. Since mouse ADPGK could be expressed and purified from pBAD-TOPO-TA expression

vector as reported (Ronimus and Morgan 2004), no large scale expression screen was undertaken, only smaller control experiments.

3.1.2 Sequence similarity and the prediction of sequence motifs of eukaryotic ADPGKs

Signal peptides are amino acid sequences found on the N-termini of proteins, destined for transport to specific cellular compartments, translocation through, or insertion into a membrane and secretory pathways. Signal peptides are often processed by a protease and cleaved off. Predicted signal peptides are annotated in the Uniprot database for *H. sapiens* and *M. musculus* ADPGK, but not for *D. rerio* ADPGK. Prediction of signal peptides was performed with different prediction algorithms: TargetP, TMHMM, SignalP (Emanuelsson et al. 2007; Petersen et al. 2011), Phobius (Käll et al. 2007), iPSORT and PSORT II (Bannai et al. 2002) and Predotar (Small et al. 2004). Most algorithms resulted in consistent prediction of a signal peptide for the ADPGK of *D. rerio, H. sapiens* and *M. musculus*. TargetP predicted all three proteins to contain a signal peptide for a secretory pathway with the highest reliability class. Predotar predicted a subcellular localisation in the ER for all three proteins with a probability higher than 98%. Phobius predicted signal sequence at position 5 – 13 with cleavage site between position 18 and 19 for *D. rerio* ADPGK, at position 12 – 23 with cleavage site between position 33 and 34 for *H. sapiens* ADPGK and at position 6 – 17 with cleavage site between position 22 and 23 for *M. musculus* ADPGK. A sequence alignment indicating the positions of the signal peptides and cleavage sites as predicted by Phobius in Figure 8 shows the difference in predicted position, which is surprising when considering that the sequences of *H. sapiens* and *M. musculus* ADPGK are nearly identical.

Figure 8: Phobius predicted positions of signal peptides.

Sequence alignment of three eukaryotic ADPGKs, dr = *D. rerio*, hs = *H. sapiens*, mm = *M. musculus*. The predicted position of signal peptides is indicated with blue arrow, predicted cleavage site is indicated by red arrowhead.

SignalP predicted one signal peptide for all three of the candidate proteins with cleavage sites between position 20 and 21 for *D. rerio* ADPGK, position 22 and 23 for *H. sapiens* ADPGK and position 18 and 19 for *M. musculus* ADPGK. Predictions for signal peptides and N-terminal transmembrane domains with PSORT (Bannai et al. 2002) were inconsistent between the eukaryotic ADPGKs. The PSORT algorithms (iPSORT and PSORT II) predicted no signal peptide or transmembrane region for *D. rerio* ADPGK. For *H. sapiens* ADPGK, an N-terminal signal peptide but no transmembrane region was predicted. For *M. musculus* ADPGK, two N-terminal transmembrane helices but no signal peptide was predicted. When the protein sequences of *D. rerio, H. sapiens* and *M. musculus* ADPGK were submitted to the TMHMM2.0 webserver (Sonnhammer et al. 1998), no transmembrane helices were predicted (see Figure 88, Figure 89 and Figure 90 in Appendix). However, two regions in the N-terminal part of ADPGK with potential for membrane insertion were predicted in all three homologs. The first region was found around position 1 – 25, the second around position 75 – 100. The score value for the prediction was below the threshold and sometimes N-terminal signal peptides, which are present according to predictions from SignalP 4.1, can be misinterpreted as transmembrane helices.

3.1.3 Structural models for eukaryotic ADPGK

Structural models of *H. sapiens*, *M. musculus* and *D. rerio* ADPGK were generated using different modelling tools, homology modelling, *ab initio* modelling and a combination of the two. All modelling was performed with the I-TASSER and Phyre2 programs, the latter one now superseding the original Phyre algorithm, which was

used at the beginning of this work. The template structures are in most cases automatically chosen by the algorithm and the archaeal ADPGKs from *T. litoralis* (PDB 1GC5), *P. horikoshii* (PDB 1L2L) and *P. furiosus* (1UA4) were usually selected. The sequence identities between the archaeal template proteins and the eukaryotic ADPGKs of *D. rerio*, *H. sapiens* and *M. musculus* was between 24 and 27%. Suitable models for guiding the design of constructs and functional mutants were ultimately selected on subjective assessment of the structure and sequence alignemnts, as errors can be expected when using homology modelling with sequence identities as low as 24%. A model for the closed conformation of *M. musculus* ADPGK was generated with I-TASSER. A model for the closed conformation of *H. sapiens* ADPGK was of inferior quality compared to the *M. musculus* model with incomplete chains and was therefore discarded. Because of the high degree of sequence conservation between the two it was decided that cautious conclusions made from the structure of *M. musculus* model would still be valid for the *H. sapiens* protein. The closed conformation was achieved by excluding the experimental structures with PDB codes 1GC5 and 1L2L, the ADPGKs of *T. litoralis* and *P. horikoshii*, which resemble the more open conformations. This leaves the modelling tool with, at the time, the only structure of ADPGK solved in a closed conformation from *P. furiosus* (PDB code 1UA4). The model is shown in Figure 9. The substrates visible in Figure 9 were placed by superposition with *P. furiosus* ADPGK (PDB 1UA4) for glucose and *T. litoralis* ADPGK (PDB 1GC5) for ADP. The model showed a poor geometry for N-terminal positions 1 – 57, as can be seen in Figure 9 (blue part of the protein chain).

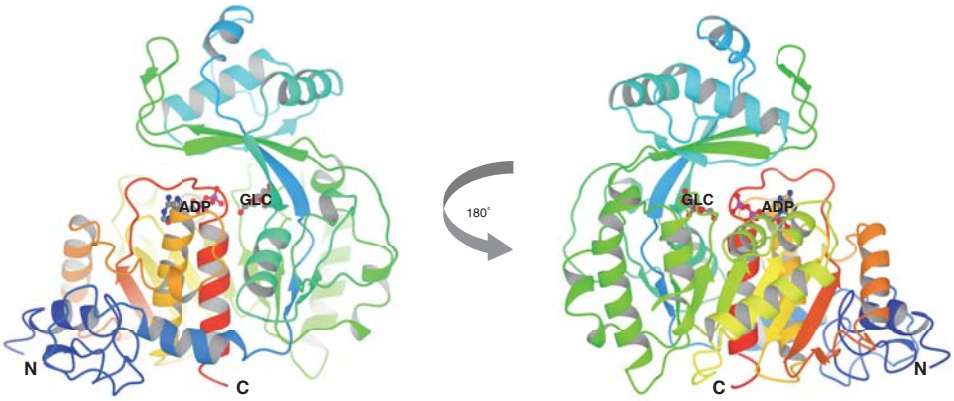

Figure 9: Model of *M. musculus* ADPGK.

Homology model of *M. musculus* ADPGK in the closed conformation, side view and rotated 180°. The N- and C-termini are indicated (N and C, respectively). The model was generated with I-TASSER, based on the structure of *P. furiosus* ADPGK (PDB 1UA4), which was crystallised in the closed conformation. The substrates glucose (labelled GLC) and ADP (labelled ADP) were placed by superimposing the structures PDB 1UA4 and 1GC5, respectively. The N-terminus from position 1 to 57 is of poor or chemically impossible geometry (blue colour).

The poor quality of the N-terminal region structure seen for *M. musculus* ADPGK in Figure 9 was a common problem encountered with most attempts to obtain a structural model of eukaryotic ADPGK, regardless of what algorithm was used or on which organism the prediction was based on. Also, omitting the N-terminal 50 – 60 amino acids from the prediction did not result in improved geometry for the rest of the model. The structural model for *H. sapiens* ADPGK (see Figure 10) generated by Phyre2 showed the best prediction for a model of the N-terminus of a eukaryotic ADPGK. While it is unlikely that the geometry resembles the one found in the actual structure, it was the best prediction available with respect to protein geometry.

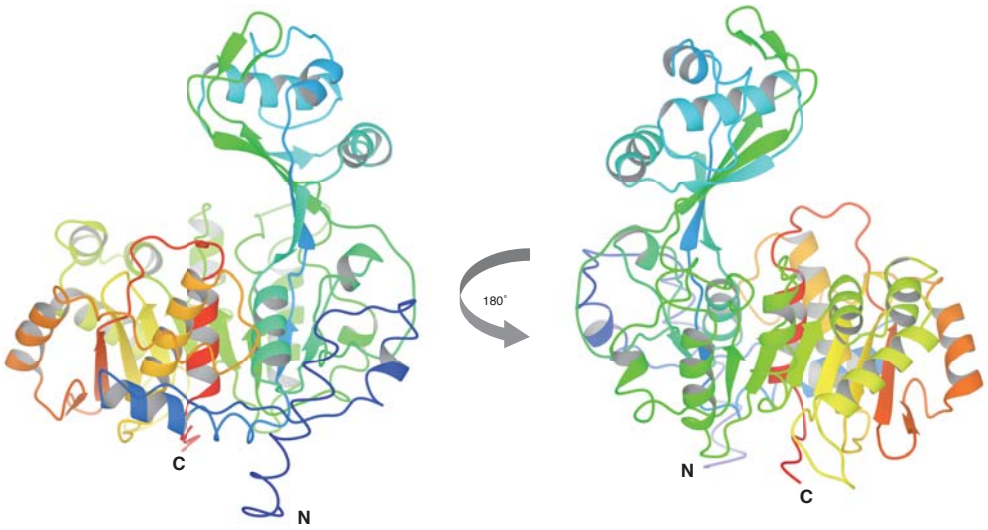

Figure 10: Model of *H. sapiens* ADPGK.

Homology model of *H. sapiens* ADPGK, resembling an open conformation, side view and rotated 180˚. The N- and C-termini are indicated (N and C). The model was generated with Phyre2.

3.1.4 Construct design: Choosing sites for truncations and mutations

Truncations of eukaryotic ADPGK were from the beginning of the project based on the assumption, that the C-terminal part of the enzymes would be structurally very similar. This assumption was based on the fact that several sequence motifs essential for ADPGKs are located in the C-terminal part of the protein, including one of the key catalytic residues almost at the C-terminus of the peptide chain (i.e. residue D481 of 496 total residues in *H. sapiens* ADPGK). A series of truncations starting at the C-terminus would quickly delete the assumed C-terminal helix and the aspartate in question, likely resulting in non-active enzyme. Therefore, all of the efforts in construct design were focused on N-terminal truncations particularly with the hydrophobic region at the N-terminus as discussed previously. In Figure 11, a sequence alignment for ADPGK from *D. rerio*, *H. sapiens*, *M. musculus*, *T. litoralis*, *P. furiosus* and *P. horikoshii* is shown, indicating the positions of the conserved sequence features and intended truncations.

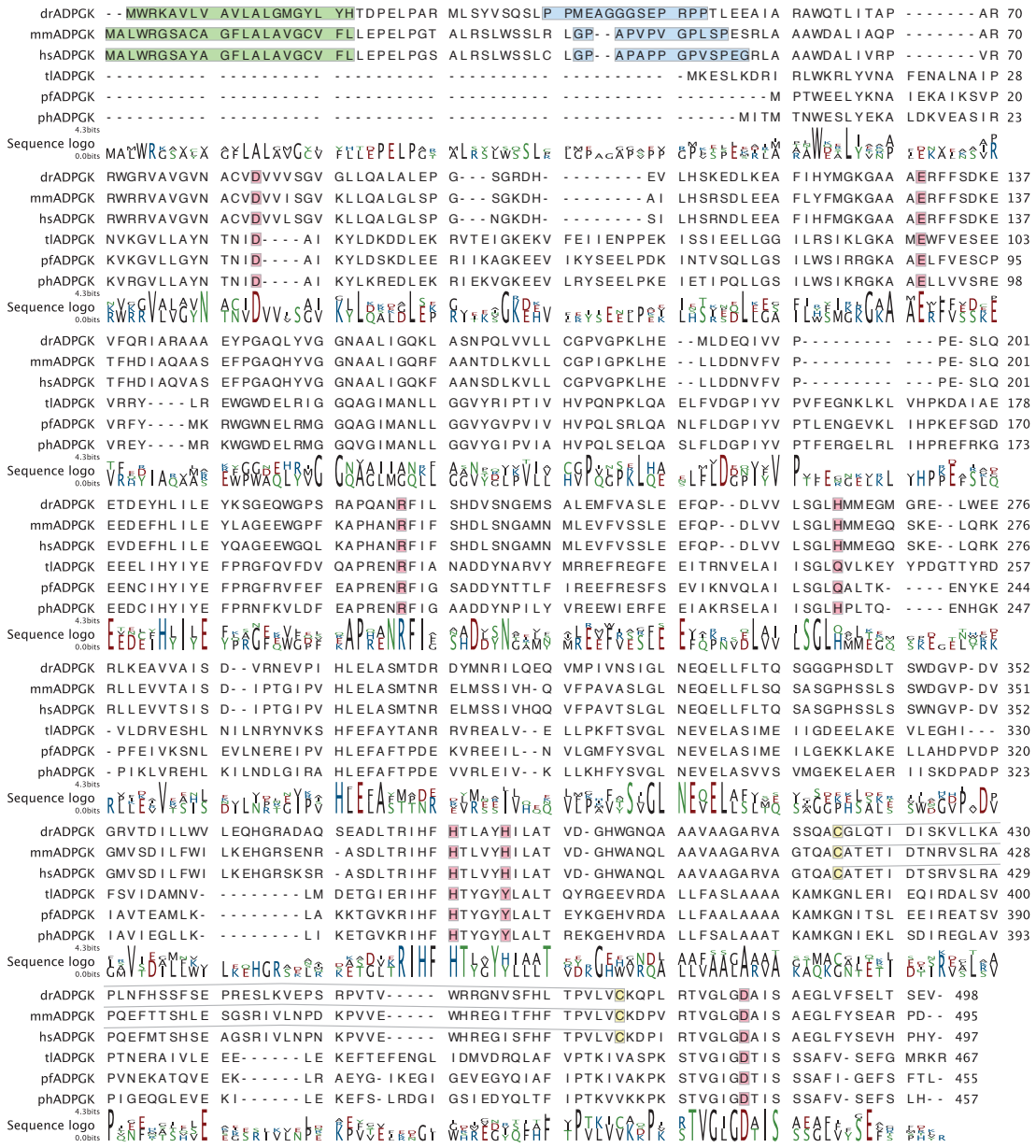

Figure 11: ADPGK sequence alignment for selection of truncation and mutation sites.

Sequence alignment generated with T-COFFEE. dr = *D. rerio*, mm = *M. musculus*, hs = *H. sapiens*, tl = *T. litoralis*, pf = *P. furiosus*, ph = *P. horikoshii*. Sites selected for mutations are red, potential disulphide bond participating cysteines are yellow, predicted signal peptide region is green, proline/glycine rich region is blue.

Signal peptide sequences or membrane anchor sequences, like the putative ones found at the N-terminus of eukaryotic ADPGKs, are stretches of hydrophobic amino acids and only adapt ordered structure when in close proximity to a membrane, signal recognition particle or other binding partner. A disordered N-terminus is potentially problematic for crystal formation. In addition, exposed hydrophobic regions can be

prone to aggregation, which can lead to problems during recombinant expression in *E. coli*, including inclusion body formation. A variety of N-terminally truncated constructs of *H. sapiens* and *M. musculus* ADPGK were designed based on the homology models for the eukaryotic ADPGKs and sequence alignments with archaeal ADPGKs. Initially two truncations for *M. musculus* and *H. sapiens* ADPGK were conceived. The first truncation was around position 25, essentially eliminating the predicted signal peptide (marked green in Figure 11). The second truncation was intended around position 50, for several reasons: The predictions for the exact position of the signal peptide were not always fully consistent, depending on the software used (see Figure 8). Furthermore, the archaeal ADPGKs are approximately 50 amino acids shorter in overall length, when compared to the eukaryotic ADPGKs. This, in combination with a proline and glycine rich sequence motif (marked blue in Figure 11) around position 50 led to the decision to truncate the first 50 amino acids. With the structural model of *H. sapiens* ADPGK, further truncations were designed, which are depicted in Figure 12. Based on the structure, truncations were made at the N-terminus inclusive to position 25, 50, 66, 70, 86, 151 and 239. The truncations intended to successively remove the N-terminus without disrupting the complete structure of ADPGK (see Figure 12). An N-terminal truncation of amino acid 1 – 239 is equivalent to isoform 3 of human ADPGK (Uniprot Q9BRR6-3). The N-terminal fragments 1 – 25, 1 – 50 and 1 – 86 of human ADPGK were cloned as well. This essentially leaves only a small N-terminal fragment comprising the putative signal peptide, while the actual ADPGK domain is deleted. In addition, truncations of mouse ADPGK were made at the N-terminus inclusive to position 24, 49, 51, 74 and 148, with a similar intended truncation pattern shown for the human ADPGK in Figure 12.

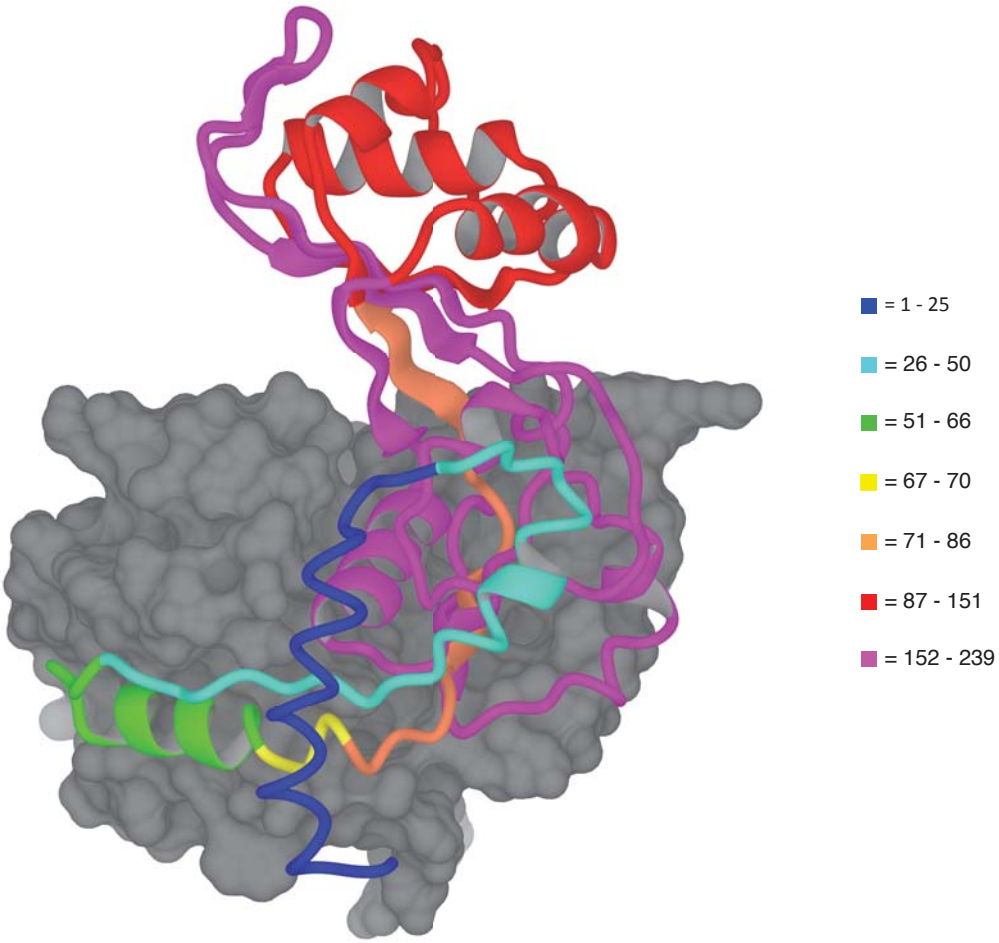

Figure 12: Truncations of *H. sapiens* ADPGK.

Depiction of truncations of *H. sapiens* ADPGK based on a homology model generated by Phyre2. Successive N-terminal truncation regions are displayed as ribbons and coloured according to the legend on the right. Numbers indicate amino acid positions. The remaining part of the large domain of ADPGK is displayed as a grey coloured surface.

The model for the closed conformation in combination with previous results for archaeal ADPGKs was used to select sites for functional mutations. In Figure 13, the active site of the model for *M. musculus* ADPGK is shown, with some of the side chains for amino acid positions selected for mutations displayed. Predicted models for the closed conformation of other eukaryotic ADPGKs were either inferior or could not be modelled (no model was returned by I-TASSER; tested were sequences of *H. sapiens*, *C. elegans* and *D. rerio* ADPGK).

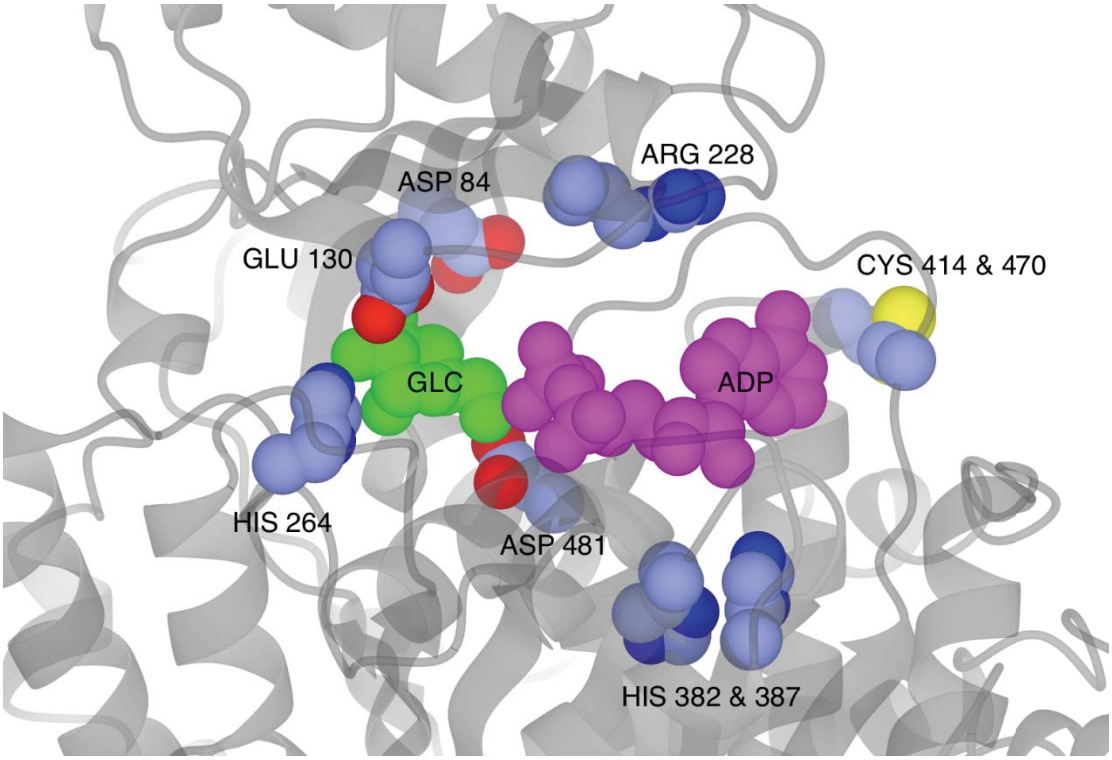

Figure 13: Selection of sites for functional mutants.

Active site view of homology model of *M. musculus* ADPGK generated by I-TASSER. The protein was modelled based on the structure of *P. furiosus* ADPGK to obtain a model of the closed conformation. Colours: glucose (GLC) green, ADP magenta, residues selected for mutations light blue. Residue numbering is according to the *H. sapiens* ADPGK.

Positions D84 and E130 were chosen for mutation due to their apparent proximity to the ligand-binding site. Both residues are conserved in eukarya and archaea. Mutation of these residues could lead to impairment of substrate binding, especially glucose, resulting in decreased activity and an elevated K_m value for glucose. The arginine R228 is conserved throughout the ADPGK and ADPPFK enzymes and believed to be involved in stabilising the terminal phosphate of ADP during the catalysis. Mutation of this position to alanine should result in greatly decreased activity of the mutant enzyme. The histidine residue at position H264 is conserved in eukaryotes but is only found in archaea in the ADPGK *P. horikoshii*. In the homologues of *P. furiosus* and *T. litoralis*, for this position a glutamine residue is found instead. Mutation of this side chain could lead to a change in binding of glucose and possibly alter the K_m of the mutant enzyme. The histidine residues at positions H382 and H387 could potentially be involved in recognition of the nucleotide or the discrimination between di- and tri-phosphates. This mechanism has been first proposed for ADPGK of *T. litoralis* (Ito et

al. 2001), as tri-phosphate utilizing enzymes like ribokinase and adenosine kinase (PDB 1RKD and 1BX4) present residues with smaller side chains at this position. Mutation of the histidine residues to alanine or valine could give valuable insights into the nucleotide binding of eukaryotic ADPGK. The model of *H. sapiens* ADPGK more recently generated by Phyre2 did show C414 and C470 in close proximity, potentially allowing the formation of a disulphide bond. However, this bond was not obvious from previous models calculated with the original Phyre algorithm. The mutation of both cysteine residues to alanine or serine, essentially removing the disulphide bond, could investigate its contribution to stability or activity of eukaryotic ADPGK. A catalytic aspartate residue has been identified as being essential for catalysis throughout the ribokinase superfamily, making this position a natural choice for point mutations. Mutation of the catalytic aspartate to alanine in archaeal ADPGK resulted in reduction of activity (Ito et al. 2001). The corresponding position in *H. sapiens* ADPGK is D481. The mutations planned at this position were D481A, D481E, D481N and D481S. The mutation D481A results in the complete removal of side chain and this mutant would be expected to show greatly reduced activity. The mutation D481S would lead to a smaller, uncharged but yet polar side chain. The mutation to D481E retains the same negatively charged chemical carboxyl group of the residue but should introduce a certain level of steric hindrance due to the increase length of the side chain. The mutation D481N would retain the same size and shape for the side chain, but alter the nature of the interaction.

3.1.5 Subcloning into expression vectors

Vectors pBAD-rmADPGK, pBAD-rmADPGK-NcoI and pProEX-rmADPGK, which were used in this work, were a gift from Ron Ronimus. The PCR fragments for cloning into vectors pBAD TOPO TA, pET101/D-TOPO and pET151/D-TOPO were amplified by PCR as described in section 2.4.2, from template cDNA for human, mouse or zebrafish ADPGK (see Table 2.1) using primers for the desired construct (see Table 6.1 in Appendix). The inserts for directional TOPO cloning used primers introducing the required 5′-CACC site. The success of the PCR reaction was verified by agarose gel electrophoresis as shown in the example in Figure 14. Subsequently,

TOPO cloning was performed as described in section 2.4.6 and the inserts verified by sequencing as described in section 2.4.1.

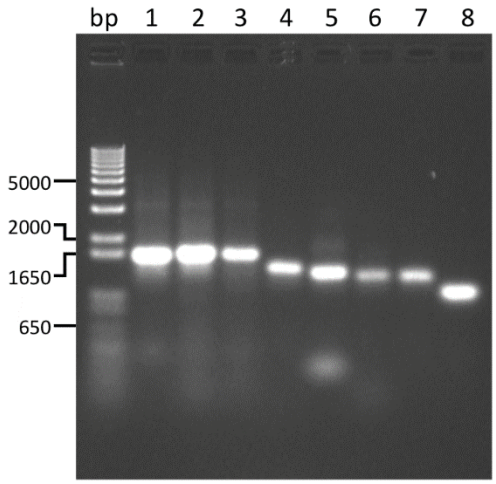

Figure 14: PCR products for TOPO cloning.

Agarose gel showing PCR products for TOPO cloning. 1, human ADPGK full-length; 2, human optimised ADPGK full-length; 3, mouse ADPGK full-length; 4, mouse ADPGK Δ24; 5, mouse ADPGK Δ49; 6 and 7, mouse ADPGK Δ74 (2x); 8, mouse ADPGK Δ86.

In addition to the TOPO-based constructs, the native cDNA variant and the codon-optimised synthetic variant human ADPGK were subcloned into expression vectors pET21(b)+, pET-32a(+), pGEX-4T-3, pMAL-p2c, pMAL-c2c, and pProEX HTb (for pProEX HTb shown in Figure 15). The zebrafish ADPGK was subcloned into expression vectors pET21(b)+, pET-32a(+), pGEX-4T-3, and pProEX HTb Cloning primers were designed with nucleotide overhangs for the PCR products to be efficiently processed by restriction endonucleases (see Table 6.1 in Appendix). Colony PCR was used to identify positive clones containing the desired insert. Figure 16 shows an example of colony PCR screening for cloning human ADPGK in vectors pGEX-4T-3 and pET21-b(+). The final constructs were verified by sequencing of the insert as described in section 2.4.1.

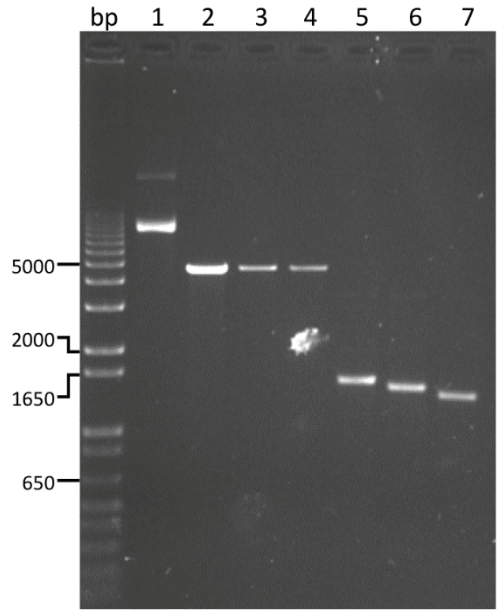

Figure 15: Cloning human ADPGK into pProEX HTb.

Agarose gel showing restriction digest of PCR products and vector DNA for cloning human ADPGK into vector pProEX HTb. 1, vector DNA uncut; 2, vector DNA double-digested (*Bam*HI and *Hin*dIII); 3, vector DNA single digest (*Bam*HI); 4, vector DNA single digest (*Hin*dIII); 5, human ADPGK full-length double-digested (*Bam*HI and *Hin*dIII); 6, human ADPGK Δ25 double-digested (*Bam*HI and *Hin*dIII); 7, human ADPGK Δ50 double-digested (*Bam*HI and *Hin*dIII).

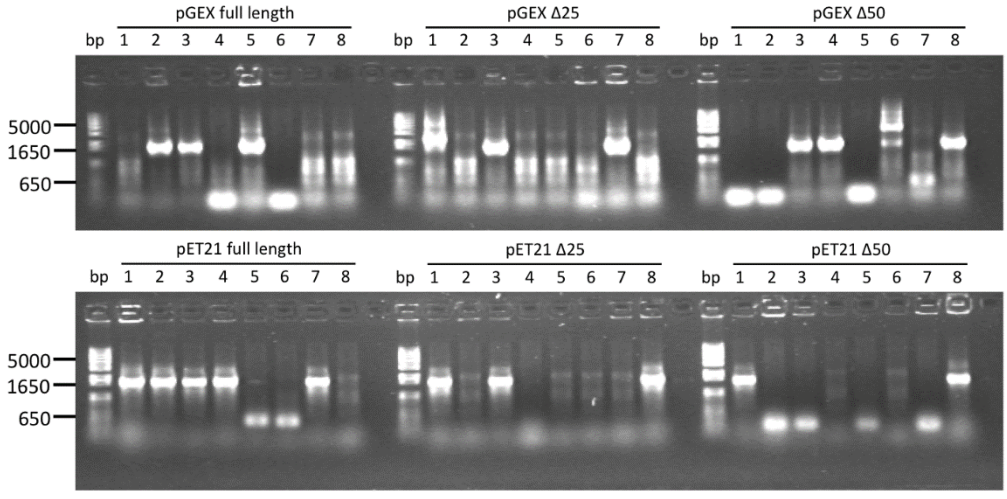

Figure 16: Colony PCR screening for selection of positive clones.

Agarose gel showing colony PCR screening to identify successful cloning into vectors pGEX-4T-3 (top) or pET21-b(+) (bottom). Numbers 1 – 8 represent individual colonies of each construct that were selected for screening. A positive clone should show a single and intense band just below the 1650 bp marker line.

Table 3.1: ADPGK expression constructs.

ADPGK cDNA	Vector	Variants
H. sapiens ADPGK	pBAD TOPO TA	Full-length, +/- NcoI site, Truncation Δ1
	pET21-b(+)	Full-length, Truncations Δ25 and Δ50 +/- C-terminal 6×His-tag
	pET-32a(+)	Full-length, +/- C-terminal 6×His-tag
	pGEX-4T-3	Truncations Δ1, Δ25 and Δ50
	pMAL-p2c/-c2c	Truncation Δ50
	pProEX HTb	Truncations Δ1, Δ25, Δ50, Δ86, Δ27-497, Δ52-497, Δ87-497
H. sapiens ADPGK, sequence codon-optimised for expression in *E. coli*	pBAD TOPO TA	Full-length, Truncations Δ24, Δ50, Δ66, Δ70, Δ151, Δ239
	pET101/D-TOPO	Full-length, Truncation Δ264
	pET151/D-TOPO	Full-length
	pET21-b(+)	Full-length
	pET-32a(+)	Truncation Δ1 +/- C-terminal 6×His-tag
	pGEX-4T-3	Truncation Δ1
	pMAL-p2c/-c2c	Truncation Δ1
	pProEX HTb	Truncation Δ1
M. musculus ADPGK	pBAD TOPO TA	Full-length, +/- V5 epitope tag truncations Δ24, Δ49, Δ51, Δ74, Δ148
	pET101/D-TOPO	Full-length
	pET151/D-TOPO	Full-length
	pProEX HTb	Truncations Δ1, Δ25 and Δ50
D. rerio ADPGK	pBAD TOPO TA	Full-length
	pET101/D-TOPO	Full-length
	pET21-b(+)	Full-length
	pGEX-4T-3	Truncation Δ1
	pProEX HTb	Full-length
	pET-32a(+)	Full-length

3.1.6 Mutants of human ADPGK

The functional mutants generated for this work are listed in Table 3.2 with brief description. Mutants C417A, C470A, C417A:C470A of pBAD-rmADPGK and H207 of pBAD-rhADPGK could not be generated due to time constraints.

Table 3.2: Mutant constructs of ADPGK.

Mutation	Hypothesis
D84A	Residue in proximity to active site, potential interaction with glucose
E130A	Residue in proximity to active site, potential interaction with glucose
H207A	Residue in proximity to active site, potential interaction with glucose
H264A	Residue in proximity to active site, potential interaction with glucose
R228A	Catalytic residue, removing complete side chain
H382A	Possibly involved in nucleotide-binding and discrimination
H387A	Possibly involved in nucleotide-binding and discrimination
H387V:H382V	Possibly involved in nucleotide-binding and discrimination
C414A	Disulphide, removal of one cysteine
C469A	Disulphide, removal of one cysteine
C414A:C469A	Disulphide, complete removal of disulphide
D481A	Catalytic residue, removing functional part of side chain
D481S	Catalytic residue, uncharged polar side chain
D481E	Catalytic residue, steric hindrance for catalysis
D481N	Catalytic residue, removal of carboxyl group

All point mutations of human ADPGK were derived from the pBAD-rhADPGK construct, the full-length protein which was sequence optimised for expression in *E. coli*, cloned into the pBAD TOPO TA expression vector, as this is the most thoroughly characterised variant in this work. In addition, mutation of catalytic aspartate to alanine, D481A, was also derived from construct pProEX-rhADPGKΔ50. The constructs for the point mutations D481A, H387A and R228A were synthesised commercially by Mutagenex (New Brunswick, USA), based on vector pBAD-

rhADPGK as a template. All other mutants were produced using the QuikChange protocol (described in section 2.4.8) with the appropriate primer pair for the desired mutation (see Table 6.1 in Appendix, only one primer of the complementary pair is listed).

3.1.7 Optimisation of expression of human ADPGK from pBAD vector

It was found that the human ADPGK cloned into expression vector pBAD TOPO TA was suitable for protein production. This is similar to the mouse ADPGK, which could be expressed and purified from the same expression system (Ronimus and Morgan 2004). The optimisation for the expression system included screening for inducer concentration, expression time and expression temperature. Expression was only optimised for the synthetic gene, codon optimised for expression in *E. coli*. Due to the unique host strain requirements of the pBAD expression system, the only strain tested was *E. coli* LMG194. As an initial test, the optimum amount of inducer L-arabinose was determined. The final concentration of L-arabinose added to the culture was varied between 0.2 – 0.0000002% (w/v) and samples taken after 2 hours of expression. Protein expression levels were judged by SDS-PAGE and western blotting as shown in Figure 17.

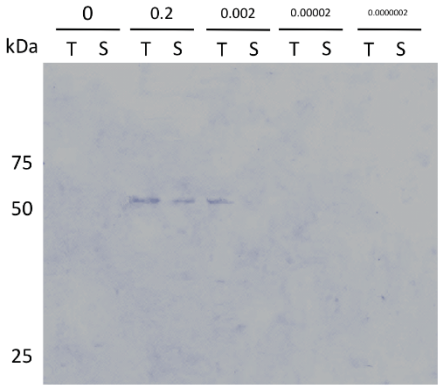

Figure 17: Optimisation of expression: Inducer concentration.

Western blot for determination of optimum L-arabinose concentration for expression of pBAD-rhADPGK in *E. coli* LMG194. Samples for total (T) and soluble (S) protein at concentrations of 0, 0.2, 0.002, 0.00002, and 0.0000002% (w/v) L-arabinose after 2 hours expression at 37 ˚C in LB broth media. ADPGK is visible as single band above the 50 kDa marker line.

The strongest soluble expression levels were achieved only at the highest amount of inducer added. Since adding even larger amounts of L-arabinose would become impractical for large culture volumes of 10 litres or more, a final L-arabinose concentration of 0.2% (w/v) was chosen for induction of human ADPGK. To further optimise expression of human ADPGK in *E. coli* LMG194, protein production was monitored at different time points over several hours and at temperatures of 18 ˚C, 30 ˚C and 37 ˚C. The production of soluble protein was assessed from SDS-PAGE gels and western blots as shown in Figure 18.

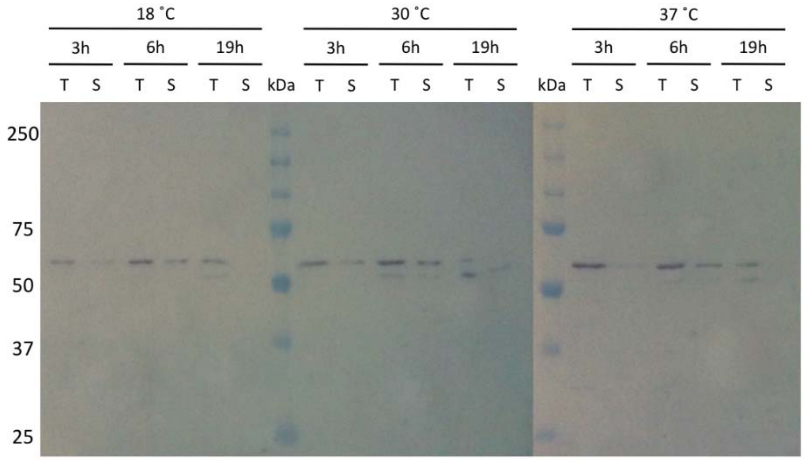

Figure 18: Optimisation of expression: Time and temperature.

Western blots showing the determination of optimum temperature and expression time for expression of human ADPGK from pBAD-rhADPGK in *E. coli* LMG194. The image shown is a composite of two western blots. Cells were grown at temperatures of 18 ˚C, 30 ˚C and 37 ˚C and samples taken for total (T) and soluble (S) protein after 3, 6 and 19 hours. ADPGK is visible above the 50 kDa marker line. After longer expression times, a second band of lower molecular weight would appear in all samples.

As can be seen, protein is produced in all three conditions, but at the higher temperatures of 30 ˚C and 37˚C a second band, possibly due to degradation, starts to appear in the 6 hour and 19 hour samples, and also at 18 ˚C for the 19 hour sample. Also, it appears that at an expression time of 19 hours ADPGK becomes mostly insoluble as no, or only very weak bands, can be seen in these samples. It was therefore decided to produce human ADPGK at temperature of 18 ˚C and not exceed an expression time of 8 hours.

3.1.8 IPTG-inducible expression systems for ADPGK

The IPTG-inducible vectors tested for expression of ADPGK included pProEX HTb, pET21-b(+), pET-32a(+), pET101/D-TOPO, pET151/D-TOPO, pMAL-p2c/c2c and pGEX-4T-3. As most attempts to express ADPGK from these expression vectors were of limited success, only a few representative samples will be presented here. The variable factors in the expression trials were: ADPGK constructs listed in Table 3.1, inducer IPTG concentrations from 0.01 mM – 1 mM, expression temperature between 18 ˚C – 42 ˚C, expression time from 1 hour – 24 hours, host strains (listed in Table 2.2), different growth media, growth in 2.2 l and 10 l fermenter equipment and co-expression of chaperones. Due to practical limitations, not all possible combinations could be tested. Expression of ADPGK from vectors pET21-b(+), pET-32a(+), pGEX-4T-3, pET101/D-TOPO and pET151/D-TOPO resulted generally in low expression levels, especially when compared to pProEX HTb constructs. All constructs based on the pET21-b(+) vector showed no expression of ADPGK. A construct based on the TOPO-cloning vector pET101/D-TOPO did show low expression levels of human ADPGK. The protein was only soluble in a truncated form upon longer expression times of more than 8 hours (see Figure 19). Unfortunately, the degradation did not result in a single product but multiple fragments.

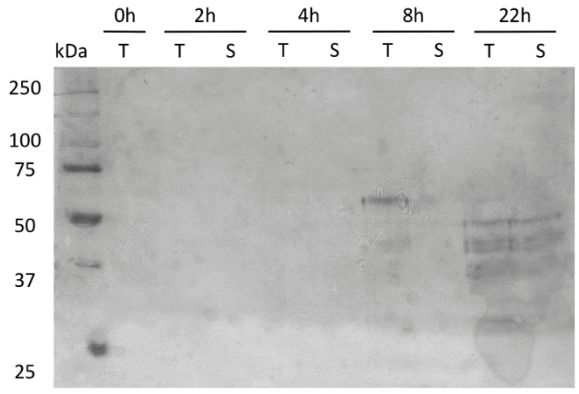

Figure 19: Expression test of pET101-rhADPGK.

Western blot analysis of expression of rhADPGK from pET101-rhADPGK vector in *E. coli* BL21(DE3) expression strain. Cells were grown in LB broth media at 37 ˚C. Samples were taken before induction with 0.1 mM IPTG (0h) and after 2, 4, 8 and 22 hours. T = total protein, S = soluble fraction. ADPGK is detectable after 8 hours expression time, however no soluble protein was visible. After 22 hours, several smaller degradation products of ADPGK became apparent in the soluble fraction.

Of all IPTG-inducible constructs tested, the strongest overexpression of ADPGK was observed from the pProEX HTb vector. Unfortunately, the majority of the protein was in the insoluble fraction, as can be seen in Figure 20, which shows the expression trial for pProEX-rhADPGKΔ50 in the *E. coli* BL21 (DE3) expression host. Upon long expression times (i.e. more than 16 hours) protein would accumulate in the soluble fractions. The ratio of soluble to insoluble protein was unaffected when varying inducer concentration, expression temperature, different media types or expression strains. Lysis buffer additives like reducing agents, detergents or high salt concentrations (data not shown) could not improve the extraction yield.

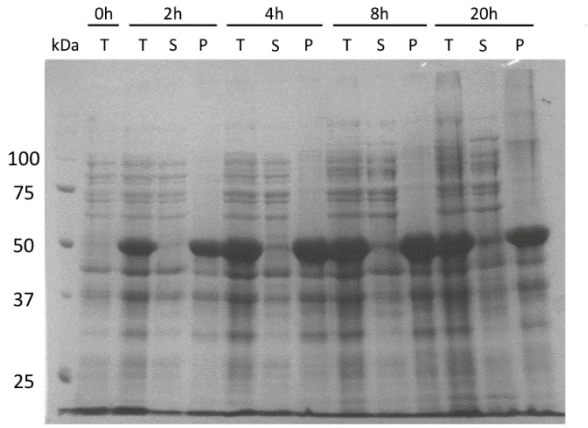

Figure 20: Expression of pProEX-rhADPGKΔ50.

Expression trial of pProEX-rhADPGKΔ50 in *E. coli* BL21 (DE3) at 37 °C. Samples were taken before induction with 0.1 mM IPTG (0 hour) and 2, 4, 8 and 20 hours after induction. T = total protein, S = soluble fraction, P = insoluble pellet fraction. The protein expresses very strongly but only a very small percentage is found in the soluble fractions after long expression times (8 hours and longer).

3.1.9 Expression in Origami hosts

Expression of ADPGK as fusion protein with the purification and expression tags Glutathione-S transferase (GST) and thioredoxin (TRX) were tested for expression from vectors pGEX-4T-3 and pET-32a(+) respectively in the *E. coli* BL21(DE3) expression strain. Unfortunately, this led to low stability of the expressed fusion protein in most cases. In Figure 21, the expression test of pGEX-rhADPGK in *E. coli* BL21(DE3) is shown. Truncated protein from vectors pGEX-rhADPGKΔ25 and pGEX-rhADPGKΔ50 showed comparable results (not shown). As can be seen, the

protein is, to a small proportion, visible in the soluble fraction after an expression time of 2 hours. However, at longer expression times the band is no longer visible, leading to the conclusion that the expression product was not stable and had degraded.

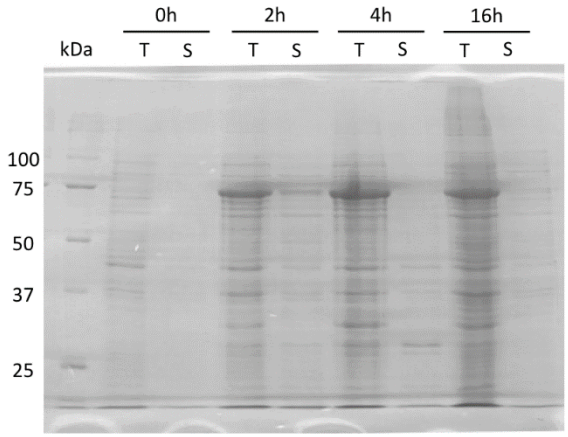

Figure 21: Expression test of pGEX-rhADPGK in *E. coli* BL21(DE3).

Expression of GST-rhADPGK fusion protein from pGEX-rhADPGK vector in *E. coli* BL21(DE3) expression strain. Cells were grown in LB broth media at 37 ˚C. Samples were taken before induction with IPTG and after 2, 4 and 16 hours (0 h, 2 h, 4 h and 16 h labels, respectively) expression time. T = total protein, S = soluble fraction. Due to the large GST-tag, the fusion protein is visible as a band at approximately the same height as the 75 kDa marker. A small portion of recombinant protein is visible in the soluble fraction from the 2 hour sample. The band disappears at longer incubation times.

The instability of the GST-rhADPGK fusion protein could be partly alleviated by expression in *E. coli* Origami (DE3) host. In Figure 22, an expression test for pGEX-rhADPGK is shown. In the Origami expression strain, the protein appears to be stable for a longer time. Despite the initially promising expression results, the work with GST-ADPGK constructs was dismissed, as it was not possible to purify GST-rhADPGK successfully when grown in a larger culture volume. Unfortunately, attempts to reproduce the improved stability of the expressed in Origami cells for other constructs were not successful. Figure 23 shows the expression test for pProEX-rhADPGKΔ50 in Origami B (DE3) host, which did not lead to a higher proportion of protein in the soluble fraction. Expression of pProEX-rhADPGKΔ50 in an Origami (DE3) or Rosetta-gami (DE3) expression host produced comparable results (not shown).

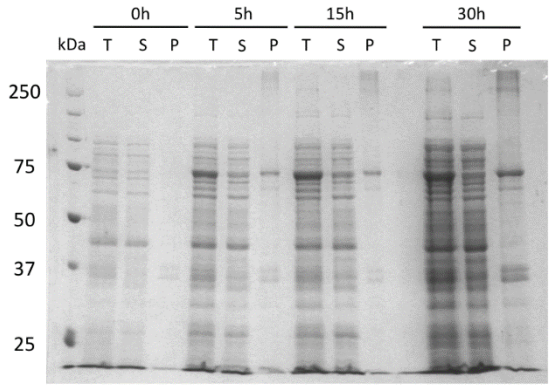

Figure 22: Expression test of pGEX-rhADPGK in *E. coli* Origami (DE3).

Expression of GST-rhADPGK fusion protein from pGEX-rhADPGK vector in the *E. coli* Origami (DE3) expression strain. Cells were grown in LB broth media at 22 ˚C. Samples were taken before induction and after 5, 15 and 30 hours (0 h, 5 h, 15 h and 30 h, respectively) expression time. T = total protein, S = soluble fraction, P = insoluble pellet fraction. Due to the large GST-tag, the fusion protein is visible as a band at approximately the same height as the 75 kDa marker.

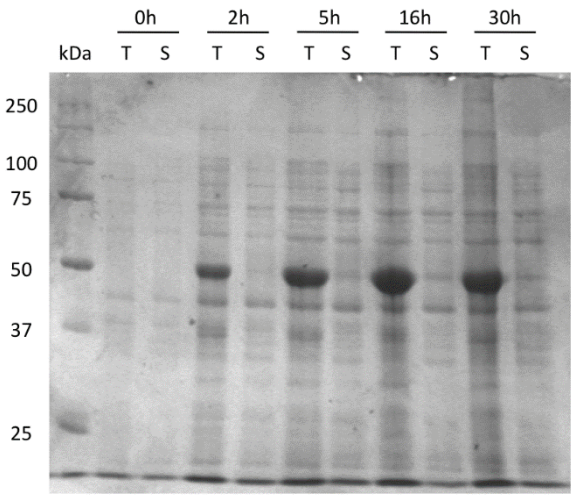

Figure 23: Expression test of pProEX-rhADPGKΔ50 in *E. coli* Origami B (DE3).

Expression of truncated rhADPGKΔ50 protein from pPro-rhADPGKΔ50 vector in *E. coli* Origami B (DE3) expression strain. The cells were grown in LB broth media at 37 ˚C. Samples were taken before induction with 0.1 mM IPTG and after 2, 5, 16 and 30 hours. ADPGK is visible as a very strong band just below the 50 kDa marker line in the total protein fractions of the post-induction samples.

3.1.10 The effect of auto-inducing media on soluble expression of ADPGK

The use of auto-inducing media can have several advantages for protein production in *E. coli*. Auto-inducing media are very convenient in practical use, since after inoculation with the desired expression system usually no further intervention is necessary until the culture is harvested. It is also reported that cultures can grow to higher density and the system is very adaptable to different expression vector systems (in this case arabinose and lactose-inducible host strains and vectors) and expression temperatures (Studier 2005). Protein expression in auto-inducing media was performed as described in section 2.3.1. Auto-inducing media was tested for vectors pProEX-rhADPGKΔ50 and pET32-rhADPGK (with and without the C-terminal 6×His-tag) in *E. coli* BL21(DE3) and vectors pBAD-rmADPGK and pBAD-rhADPGK in *E. coli* LMG194. Human ADPGK was expressed from pET32-rhADPGK (with and without the C-terminal 6×His-tag) in *E. coli* BL21(DE3) in ZYP-5052 auto-inducing media. The cultures were incubated with 25 ml pre-culture grown in LB broth for 12 to 16 hours. Subsequently, the cultures were incubated at 16 °C and samples for SDS-PAGE analysis taken after 2, 4, 8 and 22 hours. Saturation of the cultures was reached after approximately 20-22 hours with a maximum OD_{600} of ~9. In Figure 24, only samples after 22 hours incubation time are shown, since at shorter incubation time no protein was detectable. Interestingly, the variant of ADPGK with both N- and C-terminal 6×His-tags showed minute amounts of protein in the soluble fraction. In comparison, the variant of ADPGK with only N-terminal 6×His-tag did not show a visible band in the soluble fraction and generally less ADPGK was produced.

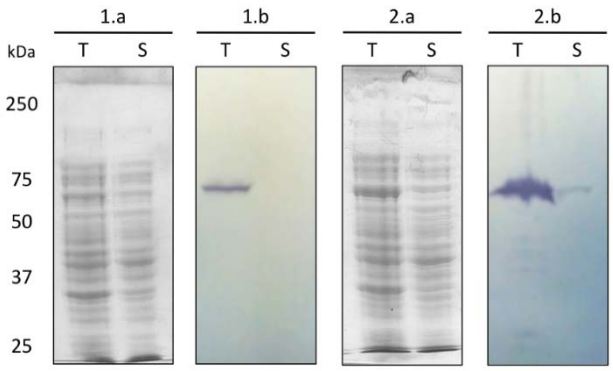

Figure 24: Expression test of pET-32a(+) constructs in auto-induction media.

Expression of human ADPGK from pET-32a(+) in auto-induction media. Expression test of human ADPGK without C-terminal tag (gel in 1.a, western blot in 1.b) and C-terminal 6×His-tag (gel in 2.a, western blot in 2.b) in *E. coli* BL21 (DE3) and ZYP-5250 auto-inducing media at 22 ˚C after 22 hours incubation time. T = total protein, S = soluble fraction.

For the expression of ADPGK from pBAD TOPO TA vector, the system was modified from lactose-inducing to arabinose-inducing (see section 2.3.1). As can be seen in Figure 25, both mouse and human ADPGK showed expression when *E. coli* LMG194 was grown in auto-induction media. However, degradation of both proteins occurred. Mouse ADPGK (Figure 25, left side of western blot) shows a double-banded pattern. Human ADPGK (Figure 25, right side of western blot) shows even more severe degradation, with two major bands below 50 kDa in the soluble fraction. The strain *E. coli* LMG194 is not a protease deficient expression host, which can make degradation during protein expression more prominent. This is also seen for more conventional media as shown in section 3.1.7. However, shorter incubation times (e.g. harvesting after 12 hours or earlier) did not result in any detectable expression of either the mouse or the human ADPGK.

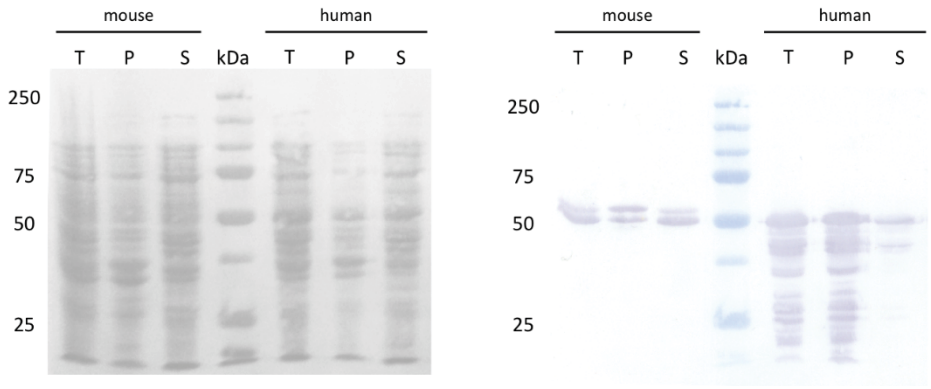

Figure 25: Expression test of pBAD constructs in auto-inducing media.

Western blot showing the expression tests of pBAD-rmADPGK and pBAD-rhADPGK in auto-induction media. Left side shows the blot after Ponceau S staining, right side shows the western blot. *E. coli* LMG194 cells were grown in in ZYP-5250 auto-induction media at 20 ˚C for 22 hours before the cells were harvested. T = total protein, S = soluble fraction, P = insoluble pellet fraction.

3.1.11 Expression trials for zebrafish ADPGK in pBAD and pET vectors

Expression trials of zebrafish ADPGK in vectors pBAD TOPO TA, pET101/D-TOPO and pET151/D-TOPO were unsuccessful. The expression construct pBAD-rzADPGK was tested in *E. coli* LMG194 and expression constructs pET101-rzADPGK and pET151-rzADPGK were screened for expression in *E. coli* BL21 (DE3) and *E. coli* Origami at 37 ˚C and 22 ˚C. The monoclonal antibody raised against the C-terminus of mouse ADPGK was able to detect only human and mouse ADPGK variants. All anti-His6 antibodies that were available could not detect any 6×His-tagged ADPGK protein. No obvious increase in expression of ADPGK after induction with arabinose (pBAD vector) or IPTG (pET vector) could be observed on Coomassie-stained SDS-PAGE gels. The combination of these difficulties led to the decision to abandon the optimisation of zebrafish ADPGK expression.

3.1.12 Summary

In silico analysis was used to identify potentially interesting sites for functional mutants and promising constructs for expression and crystallisation of eukaryotic ADPGK. Truncations of ADPGK were designed that aimed to improve crystallisation and expression behaviour while leaving the remainder of the protein structurally intact and catalytically active. Functional mutants were based on current understanding of the archaeal ADPGK and ADPPFK enzymes. Over 50 constructs, including site-directed mutants and truncations in a variety of expression vectors were created for this work and tested for expression in *E. coli*. During the optimisation of expression, pBAD TOPO TA had been shown to be the overall best choice of the vectors tested for production of the human ADPGK, albeit at still modest levels. For human ADPGK, no obvious difference in expression of soluble protein was found for the codon optimised ADPGK over native non-optimised cDNA derived sequence in all cases. Vector pProEX HTb (in the form of pProEX-rhADPGKΔ50) was only used initially, until the construct pBAD-rhADPGK was available. The most striking advantage of the pBAD vectors is the apparently completely soluble expression of ADPGK. Most other tested vector-strain combination resulted in at least partly insoluble protein. This has led to the decision to derive all functional mutants from the pBAD-rhADPGK construct. Unfortunately, the arabinose-induced protein expression system greatly limited the number of compatible *E. coli* strains and such specialised expression hosts such as Origami were not available for further optimisation. It was not possible in this study to optimise the expression of zebrafish ADPGK since the protein could not be detected in western blots. Furthermore, no form of ADPGK was detectable by anti-6×His-tag antibodies used in western blotting. In addition to the protein expression experiments described in this chapter, expression trials in insect cell lines of *H. sapiens* ADPGK derived from wild type cDNA cloned in vector pFBDM were conducted in collaboration with Dr. Matthew Bennett. Unfortunately, the experiments did not lead to soluble protein expression and the approach was dismissed.

3.2 Purification and refolding of eukaryotic ADPGK

3.2.1 Introduction

To enable biochemical and structural studies of ADPGK, the protein needs to be purified to near homogeneity. Crystallisation experiments require large amounts of protein, and much effort was put into producing ADPGK of sufficient purity to set up crystallisation trials. Most of the crystallisation experiments were conducted with the mouse full-length ADPGK, due to the better behaviour during expression and purification. For the study of the enzyme activity of ADPGK, the human full-length wild-type protein was purified, as well as the functional mutants. The purification of these different constructs and variants of ADPGK will be described in the following sections. Despite the many variants of ADPGK tested for soluble expression in *E. coli*, overall expression levels were either very low or a high proportion of the protein remained in the insoluble fraction. A different approach to this problem is the chemical refolding of ADPGK from inclusion bodies. In this chapter different approaches to refolding ADPGK will be described. The refolding experiments were conducted with a truncated variant of human ADPGK, rhADPGKΔ1-50, as the main intention was to obtain large amounts of protein for structural studies.

3.2.2 Purification of soluble rhADPGKΔ50

For expression, *E. coli* BL21 (DE3) was transformed with pProEX-rhADPGKΔ50 and grown in LB broth medium including 100 μg/ml ampicillin as described in section 2.3.5. Based on results of expression trials the cells were grown in LB broth media at 37 ˚C until OD_{600} of 0.8 before the inducer IPTG was added to 0.1 mM final concentration. Growth was continued at 37 ˚C and cells were harvested by centrifugation 16 hours after induction and frozen at -80 ˚C as described in section 2.3.5. The harvested cells were resuspended in lysis buffer (100 mM Tris pH 8.0, 100 mM NaCl, 4 mM DTT, Roche complete protease inhibitor EDTA free added according to manufacturer's instructions), lysed by sonication and soluble protein separated from insoluble debris as described in section 2.3.6. Imidazole was added to the supernatant to a final concentration of 10 mM and the solution filtered through 0.45 μm and 0.2 μm syringe filters (Sartorius, Germany) before loading on to a

HisTrap HP 5 ml column (GE Healthcare, UK). IMAC purification was performed as described in section 2.5.7. Protein elution was performed with a linear gradient over 20 cv from 10 mM to 500 mM imidazole in phosphate buffer (50 mM Na_2HPO_4, pH 8.0, 500 mM NaCl, 1 mM DTT). The protein eluted at an imidazole concentration of approximately 150 mM. A chromatogram and gel are shown in Figure 26. The theoretical molecular weight of rhADPGKΔ50 was 52.7 kDa, as determined by ProtParam.

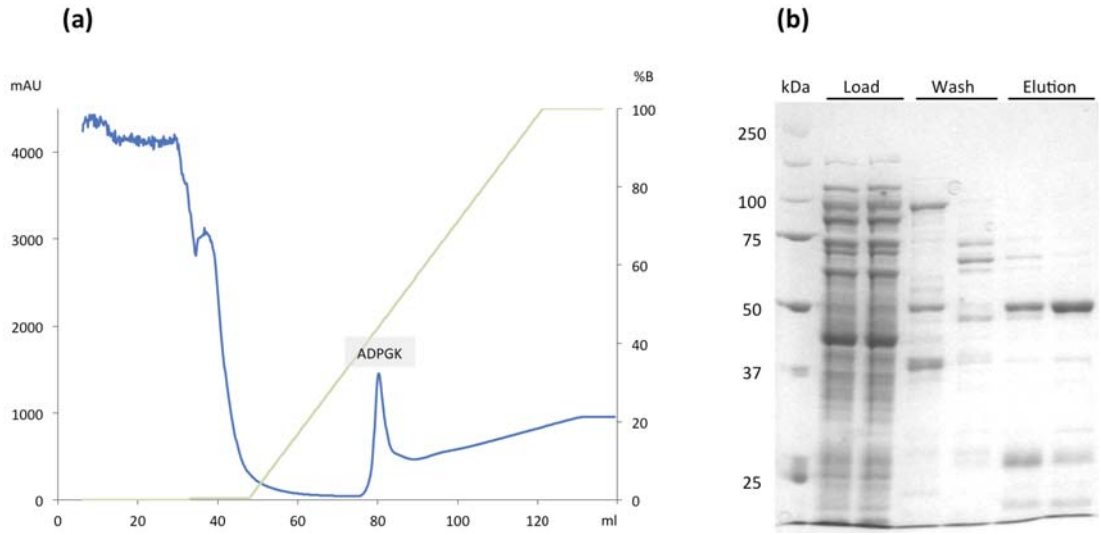

Figure 26: rhADPGKΔ50 IMAC purification.

Overview of the IMAC purification step for rhADPGKΔ50. **(a)** Chromatogram. The protein eluted at a concentration of 150 mM imidazole from the HisTrap HP column. The elution peak containing rhADPGKΔ50 is labelled. **(b)** SDS-PAGE. The gel shows the protein before and after filtering before loading on to the column (Load), the flow-through fractions (Wash) and the elution fractions containing rhADPGKΔ50 (Elution).

The elution fractions containing rhADPGKΔ50 were pooled and the volume reduced by gentle concentration at 4 ˚C in a Vivaspin 20 concentrator (GE Healthcare, UK). Size-exclusion chromatography was performed as described in section 2.5.10. The concentrated protein was centrifuged and filtered before being loaded on to a Superdex S200 10/300 GL Tricorn size-exclusion chromatography column (GE Healthcare, UK) pre-equilibrated with buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1 mM DTT). A chromatogram and corresponding gel are shown in Figure 27. ADPGK eluted as a single peak at an elution volume of 15.46 ml. The molecular weight of

rhADPGKΔ50 as determined by analytical size-exclusion chromatography as described in section 2.5.10 was 67.8 kDa, indicating a monomeric state in solution.

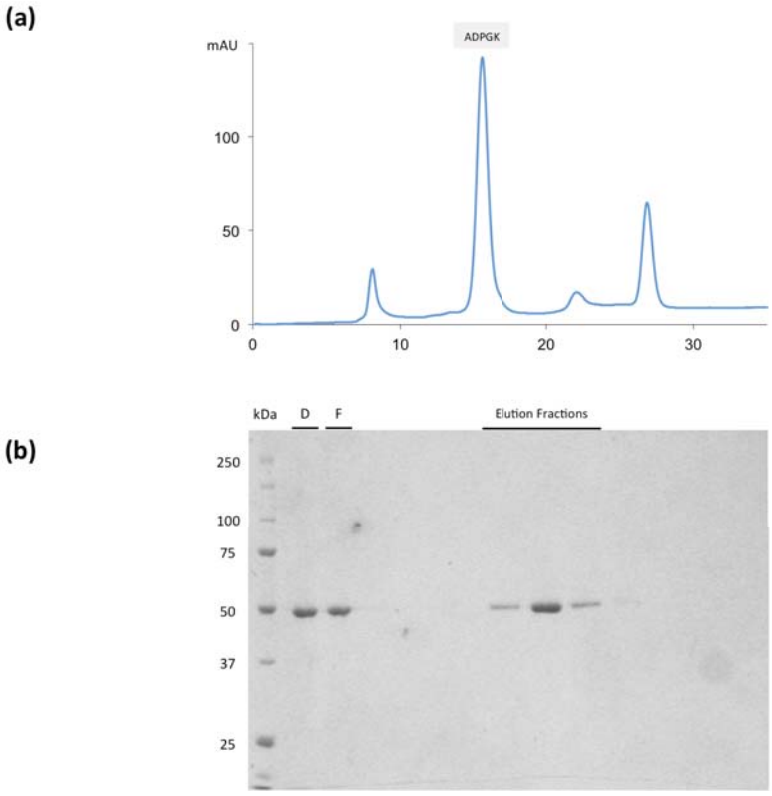

Figure 27: rhADPGKΔ50 size-exclusion chromatography.

Final purification step for rhADPGKΔ50. (a) Chromatogram. The protein eluted as a single peak at 15.46 ml elution volume from a Superdex 200 column. (b) SDS-PAGE. Lanes labelled D and F are sample prepared for loading before and after filtration. Elution fractions containing rhADPGKΔ50 are labelled accordingly. No other contaminants are evident from the Coomassie-stained gel at this protein concentration.

The elution fractions containing purified ADPGK were pooled, concentrated at 4 ˚C in a Vivaspin 500 concentrator (GE Healthcare, UK) and stored at -20 ˚C until used. The typical yield from an initial culture volume of 10 l was about 500 μg purified protein. A summary of the purification and protein prepared for crystallisation is shown in Figure 28.

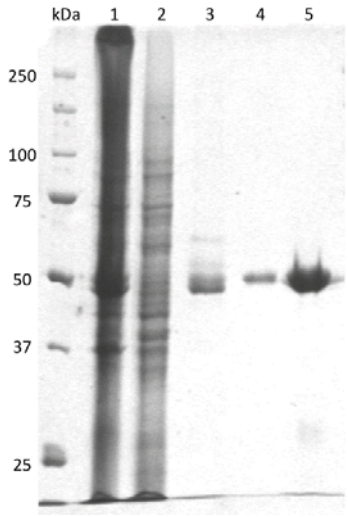

Figure 28: Purified rhADPGKΔ50.

SDS-PAGE of the purification steps of rhADPGKΔ50. **(1)** Crude lysate. **(2)** Soluble fraction after centrifugation. **(3)** IMAC purification elution fraction. **(4)** Size-exclusion chromatography elution fraction. **(5)** 1.5 μg purified rhADPGKΔ50 for crystallisation.

3.2.3 Refolding rhADPGKΔ50 from inclusion bodies

Inclusion bodies were obtained as described in section 2.3.7 from *E. coli* BL21(DE3) expressing rhADPGKΔ50 from construct pProEX-rhADPGKΔ50. This construct was chosen due to the large amount of rhADPGKΔ50 accumulating in the insoluble fraction (see expression trials in section 3.1.8). The procedure for isolation of inclusion bodies and refolding was adapted from a protocol for a full-length construct of ADPGK (Hole 2009) with modifications for the purification. The solubilized, unfolded rhADPGKΔ50 was purified using an IMAC column in the presence of urea and reducing agent as described in section 2.5.8. The majority of denatured rhADPGKΔ50 eluted at 100 mM imidazole (see Figure 29). At this step, the eluted protein had a purity of approximately 80% as assessed from SDS-PAGE.

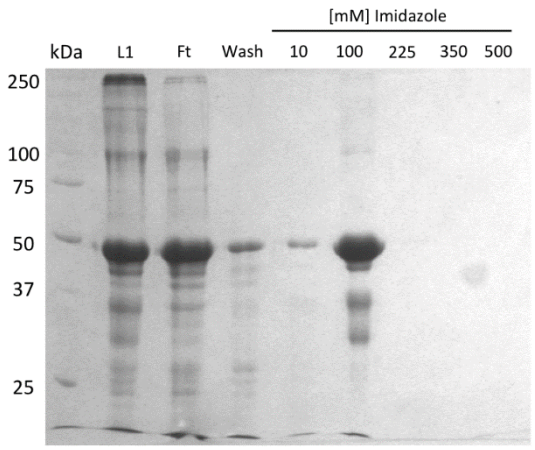

Figure 29: IMAC purification of solubilized inclusion bodies.

Coomassie-stained gel showing the purification of resolubilised and denatured rhADPGKΔ50 over a HisTrap FF column. L = sample prior to loading, Ft = flow-through fraction, Wash = wash fraction, 10, 100, 225, 350 and 500 = elution fraction, where numbers are the concentration of imidazole (mM) in the elution buffer. The main part of rhADPGKΔ50 elutes in the 100 mM imidazole fraction. The load and flow-through show a similar banding pattern as the amount of protein exceeded the binding capacity of the column.

For refolding, the eluted protein was diluted into refolding buffer. The protein was directly diluted 20-fold into 20 mM Tris buffer, pH 8.0, at 4 ˚C and incubated for 12 hours with gentle agitation. The refolded protein was tested for enzymatic activity of ADPGK. Refolded ADPGK was then further purified by ion-exchange chromatography and size-exclusion chromatography. Ion-exchange chromatography was performed as described in section 2.5.9 on a HiTrap Q FF 1 ml column and a HiTrap DEAE FF 1 ml column (GE Healthcare). Refolded rhADPGKΔ50 was eluted with a linear gradient to 1 M NaCl over 25 cv and the elution fractions were analysed by SDS-PAGE and tested for enzymatic activity. Purification using the HiTrap DEAE column, the weaker ion-exchange resin, resulted in a sharper elution peak, when compared to the HiTrap Q column. Attempts were made to purify refolded rhADPGKΔ50 by size-exclusion chromatography, either before or after refolding, and before or after the ion-exchange purification. The protein eluted in multiple peaks directly after the void volume of a Superdex S200 column, indicating that the protein had formed aggregates, or adopts different conformations after refolding. Refolded rhADPGKΔ50 typically only had a specific activity of 0.1 to 0.3 U/mg, regardless the

purification method applied, a reduction of about 10-fold when compared to protein purified from the soluble fraction.

3.2.4 Refolding optimisation of rhADPGKΔ50 with the iFold2 screen

The iFOLD Protein Refolding System 2 (Novagen, Merck Millipore, Germany) was used to identify possible refolding conditions. Inclusion bodies of rhADPGKΔ50 were prepared as described in section 2.3.7. The total protein concentration was determined by the Bradford protein assay (see section 2.5.1) and the concentration then adjusted to 5 mg/ml, and then added to the refolding matrix according to the manufacturer's recommendation. The samples were tested for refolded protein after shaking gently at 21 ˚C for 12 hours. The recommended test for successful refolding from the iFOLD screen is to test the amount of precipitation by measuring the A_{340} blanked against water (Vincentelli et al. 2004). However, soluble protein is not necessarily correctly folded. Therefore, it was decided to find a measurement of enzymatic activity of ADPGK. The standard assay for ADPGK relies on the reduction of $NADP^+$ to NADPH, which can be measured as an increase in A_{340}. To avoid interference between the two processes monitored at the same wavelength, it was decided to couple enzymatic activity to a different wavelength. For the hexokinase enzymes, a simple assay was described by Gabriel and Gerstent (1992), which couples the production of glucose-6-phosphate to reduction of a tetrazolium to formazan dye, resulting in a deep-blue precipitate. The precipitate is visible by eye *in situ*, since it was originally developed to stain polyacrylamide gels for enzymatic activity after electrophoresis. The assay is described in section 2.5.12. For this test, 20 µl aliquots of the refolded protein solution were added to the reaction mixture, which was prepared in a 96-well plate, and incubated for 30 minutes at 37 ˚C. The plate was then scanned in a plate reader. The iFOLD screen does not contain a control reaction for each condition. Therefore, control reactions could not be performed other than testing the reaction mixture of the dye assay with recombinant mouse or human ADPGK from soluble expression in *E. coli*. The highest A_{595} values were observed for conditions A8, B12, C6 and F3, and in these reactions blue staining was also visible *in situ*. For almost all other conditions the A_{595} values ranged between 0.04 and 0.05, which were not investigated further. The chemical composition of these conditions is

detailed in Table 3.3. It has to be noted that the sample refolded in condition B12 lost its activity after less than one day when stored at 4 ˚C. This condition was the only one containing no cyclodextrin but contained L-arginine instead. Conditions containing cyclodextrin retained their biological activity for more than one week when stored at 4 ˚C. Also, all four successful conditions contained a reducing agent, either glutathione in reduced (GSH) and oxidised (GSSG) form or TCEP.

Table 3.3: Successful refolding conditions for ADPGK.

Well	A_{595}	Refolding solution composition
A8	0.099	50 mM EPPS pH 8.0, 10 mM cyclodextrin, 1 mM TCEP, 240 mM NaCl, 10 mM KCl
B12	0.099	50 mM CHES pH 9.0, 500 mM L-arginine, 1 mM EDTA, 6 mM GSH, 4 mM GSSG, 240 mM NaCl, 10 mM KCl
C6	0.078	50 mM EPPS pH 8.0, 10 mM cyclodextrin, 9 mM GSH, 1 mM GSSG, 240 mM NaCl, 10 mM KCl
F3	0.111	50 mM HEPES pH 7.5, 10 mM cyclodextrin, 1 mM TCEP

Unfortunately, it was not possible to reproduce conditions with β-cyclodextrin (Sigma-Aldrich). The cyclodextrin variety used in the iFOLD screens was more specifically methyl-ß-cyclodextrin, which is sold under the brand name CAVASOL™ by Wacker Cemie AG. Methyl-ß-cyclodextrin of suitable quality was not immediately available and the experiment was put on hold.

3.2.5 Purification of soluble rhADPGK

For expression, *E. coli* LMG194 was transformed with pBAD-rhADPGK (cloned from cDNA optimised for expression in *E. coli* as described in section 2.1.3) and grown in 2× YT medium containing 100 μg/ml ampicillin as described in section 2.3.5. Cells were grown at 37 ˚C and expression was induced with a final concentration of 0.2% (w/v) L-arabinose at an OD_{600} of 0.8. Growth was continued for 8 hours at 18 ˚C and the cultures were harvested, washed and stored frozen at -80 ˚C as described in section 2.3.6. Typically 12 l of culture would yield 35 g of wet cell pellet. The cell pellet was resuspended in lysis buffer (50 mM Tris pH 8.0, 100 mM NaCl, 4 mM DTT, Roche complete protease inhibitor EDTA-free) and lysed by sonication, French press and lysozyme treatment and the soluble fraction separated by centrifugation as described in section 2.3.6. Imidazole was added to the supernatant to a final concentration of 25 mM and the solution filtered through 0.45 μm and 0.2 μm syringe filters (Sartorius, Germany) before loading on to a HisTrap FF 5 ml column (GE Healthcare, UK). IMAC purification was performed as described in section 2.5.7. Protein elution was performed with a linear gradient over 40 cv from 25 mM to 500 mM imidazole in phosphate buffer (40 mM Na_2HPO_4, pH 7.8, 500 mM NaCl, 500 μM DTT). The protein eluted at an imidazole concentration of approximately 100 mM in an overlapping peak (see Figure 30). The theoretical molecular weight of rhADPGK was 58.5 kDa, as determined by ProtParam.

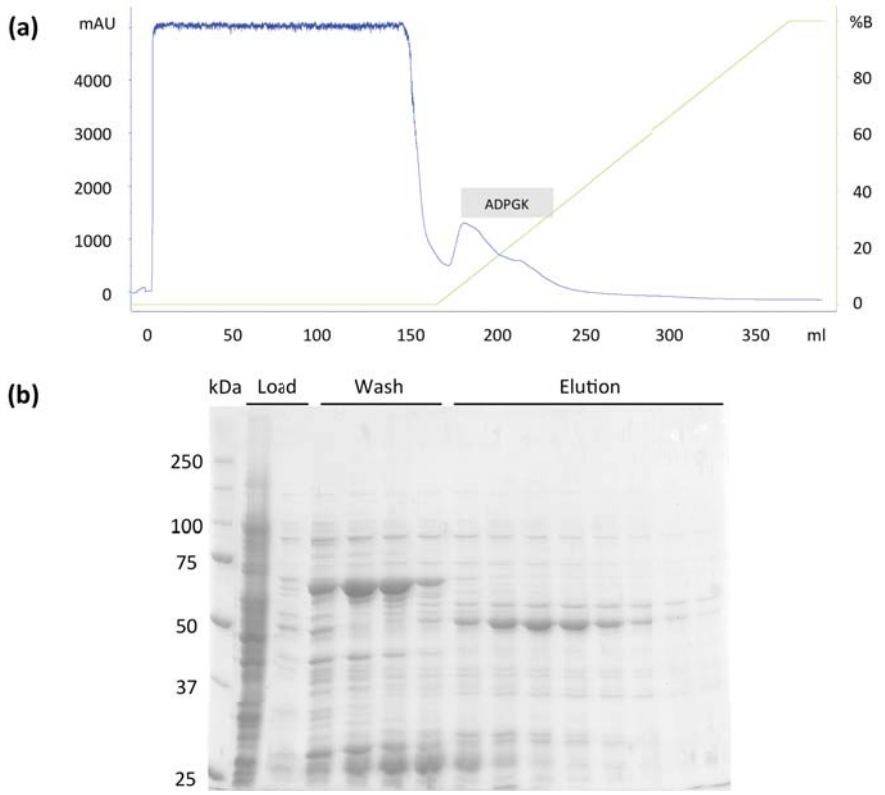

Figure 30: rhADPGK IMAC purification.

IMAC purification of rhADPGK using a HisTrap FF column. **(a)** Chromatogram. The protein eluted in the tail segment of a double peak between 200 and 250 ml elution volume at an imidazole concentration of approximately 100 mM. **(b)** SDS-PAGE. The gel lanes with elution fractions containing rhADPGK are labelled 'Elution'. The first two lanes labelled 'Load' are, from left to right, proteins after filtering before loading on to the column and the first flow-through fraction containing unbound protein. Fractions labelled 'Wash' contain the wash fractions between 150 and 175 ml elution volume.

The elution fractions containing rhADPGK were pooled and dialysed against low ionic strength buffer (40 mM Tris/HCl pH 8.0, 20 mM NaCl 1 mM DTT). The subsequent ion-exchange chromatography step was performed as described in section 2.5.9. The protein solution was filtered before being loaded on to a UnoQ6 column (Bio-Rad Laboratories, USA) pre-equilibrated with binding buffer (40 mM Tris/HCl pH 8.0, 40 mM NaCl, 1 mM DTT). Elution was performed with a segmented gradient over 65 cv to 1 M NaCl and rhADPGK eluted in the first segment at a salt concentration of 150 mM. A chromatogram and corresponding gel are shown in Figure 31.

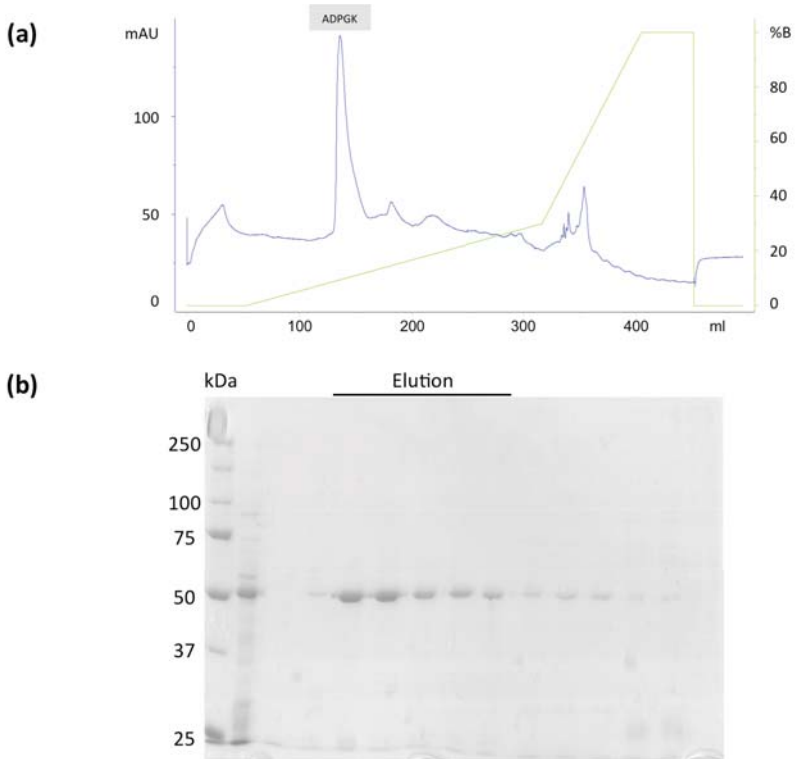

Figure 31: rhADPGK ion-exchange purification.

Overview of the purification of rhADPGK with a UnoQ6 ion-exchange column. **(a)** Chromatogram. The protein eluted in a major peak at a concentration of 150 mM NaCl. **(b)** SDS-PAGE: The gel lanes with elution fractions corresponding to the major peak at 150 ml elution volume containing rhADPGK are labelled 'Elution'. The lane directly right of the molecular weight marker is protein sample before loading on to the column.

Elution fractions containing ADPGK were pooled and the volume reduced by gentle concentration at 4 ˚C using a Vivaspin 20 concentrator (GE Healthcare, UK). Protein purity was judged by SDS-PAGE and rhADPGK was either directly concentrated, glycerol added to 20% (w/v) and stored at -20 °C for use in activity assays or subjected to further purification by size-exclusion chromatography for all other purposes. Size-exclusion chromatography was performed as described in section 2.5.10. The concentrated protein was centrifuged and filtered before being loaded on a Tricorn Superdex 200 10/300 GL or a HiLoad 16/600 Superdex 200 pg size-exclusion chromatography column (GE Healthcare, UK) pre-equilibrated with buffer (50 mM Tris pH 7.6, 150 mM NaCl, 1 mM DTT). A chromatogram and corresponding gel when using a Tricorn Superdex 200 10/300 GL column are shown in Figure 32.

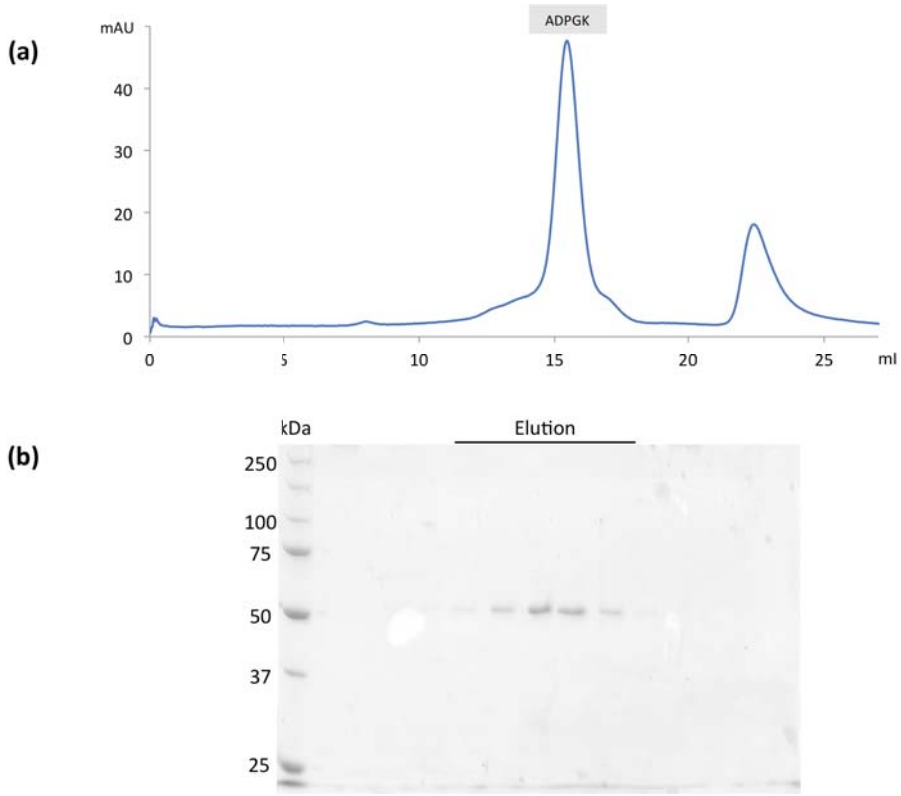

Figure 32: rhADPGK size-exclusion chromtaography.

Overview of the final purification step for rhADPGK using a Tricorn Superdex 200 10/300 GL column. **(a)** Chromatogram. The protein eluted as a single peak at 15.5 ml elution volume from the Superdex S200 column. **(b)** SDS-PAGE: The gel lanes labelled elution contain purified rhADPGK, visible as intense band just above the 50 kDa marker.

rhADPGK eluted as a single peak at a elution volume of 15.5 ml. The molecular weight of rhADPGK as determined by analytical size-exclusion chromatography as described in section 2.5.10 was 68.1 kDa, indicating a monomeric state in solution. The elution fractions containing rhADPGK were pooled, concentrated at 4 ˚C in a Vivaspin 500 concentrator (GE Healthcare, UK) and stored at -80 ˚C.

3.2.6 Purification of soluble rhADPGKΔ151

For expression, *E. coli* LMG194 were transformed with pBAD-rhADPGKΔ1-151 (cloned from cDNA optimised for expression in *E. coli* as described in section 2.1.3) and grown in 2× YT medium containing 100 μg/ml ampicillin as described in section 2.3.5. All other steps were performed as described in section 0 for the full-length ADPGK cloned into the same vector. The protein was purified by IMAC and ion-

exchange chromatography. A chromatogram for the ion-exchange chromatography purification step is shown in Figure 33.

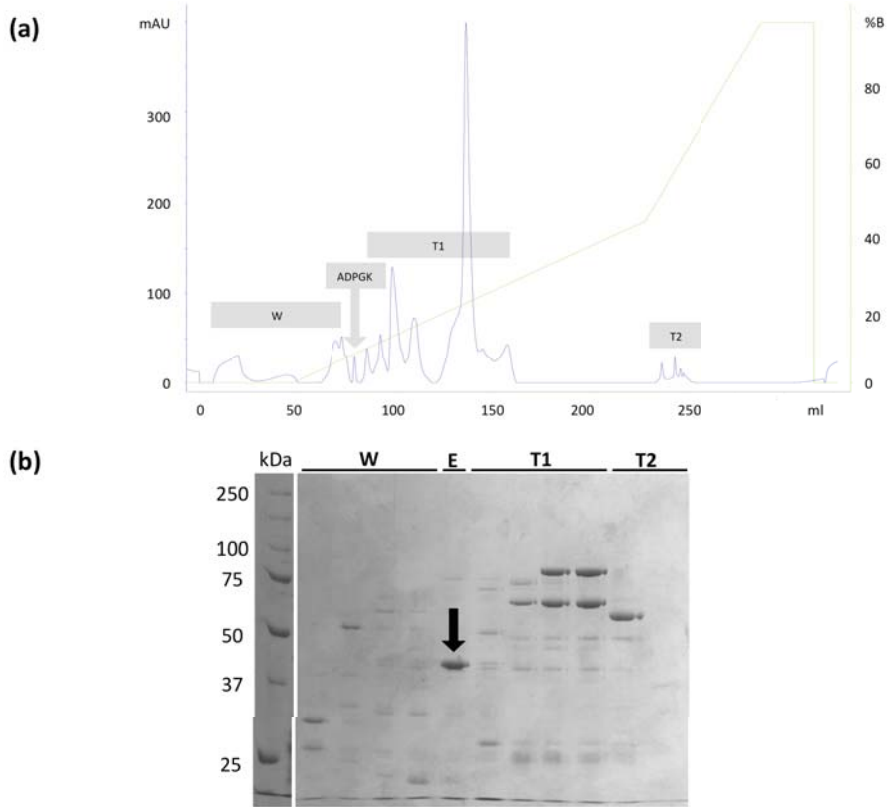

Figure 33: rhADPGKΔ151 ion-exchange chromatography.

Overview of the ion-exchange purification of rhADPGKΔ151 using a UnoQ6 column. (a) Chromatogram. The protein eluted as a minor peak at a NaCl concentration of 120 mM that could be easily dismissed as a contaminant. The peak is indicated by a grey arrow. (b) SDS-PAGE. The elution fraction containing rhADPGKΔ151 is indicated by a black arrow. Wash fractions (W) and the peaks tailing the elution fraction containing ADPGK (T1 and T2) are marked on the chromatogram and gel accordingly.

rhADPGKΔ151 eluted as a minor peak at a concentration of 120 mM NaCl from the UnoQ6 column. The theoretical molecular weight of rhADPGKΔ151 was 42.8 kDa, as determined by ProtParam. The protein was only available in very small amounts and no enzymatic activity was detected for this variant when tested as described in section 2.5.11.

3.2.7 Purification of rhADPGK functional mutants

The following section will describe the purification of functional mutants of ADPGK. The purification was always performed as described for the corresponding soluble wild type construct, either rhADPGK (cloned from cDNA optimised for expression in *E. coli* as described in section 2.1.3) or rhADPGKΔ50. Not all generated mutants could be tested within the time frame of this work. In general, the protein yield after purification was found to be much lower for the mutant constructs. To avoid cross-contamination, separate HisTrap FF affinity columns were used. The UnoQ6 ion-exchange column was washed thoroughly in 20 cv high salt buffer and 20 cv water after each use. For the expression of rhADPGKΔ50-D481A, *E. coli* BL21(DE3) were transformed with pProEX-rhADPGKΔ50-D481A and grown in LB broth medium containing 100 μg/ml ampicillin as described in section 2.3.5. All other steps were performed as described in section 3.2.2 for the truncated ADPGK cloned into the same vector. Mutant rhADPGKΔ50-D481A was purified by IMAC and size-exclusion chromatography. The protein could only be purified in very small quantities and no enzymatic activity was detected for this variant when tested as described in section 2.5.11. For the expression of rhADPGK with the point mutations D84A, H264A, R228A, H382A, H387A, H387V:H382V, D481A, D481S, D481N and D481E, *E. coli* LMG194 cells were transformed with the corresponding variant of vector pBAD-rhADPGK (cloned from cDNA optimised for expression in *E. coli* as described in section 2.1.3) and grown in 2× YT medium containing 100 μg/ml ampicillin as described in section 2.3.5. All other steps were performed as described in section 0 for the full-length ADPGK. The point mutation variants of rhADPGK were purified by IMAC and ion-exchange chromatography. A representative chromatogram and gel for the purification of rhADPGK-D481E are shown in Figure 34.

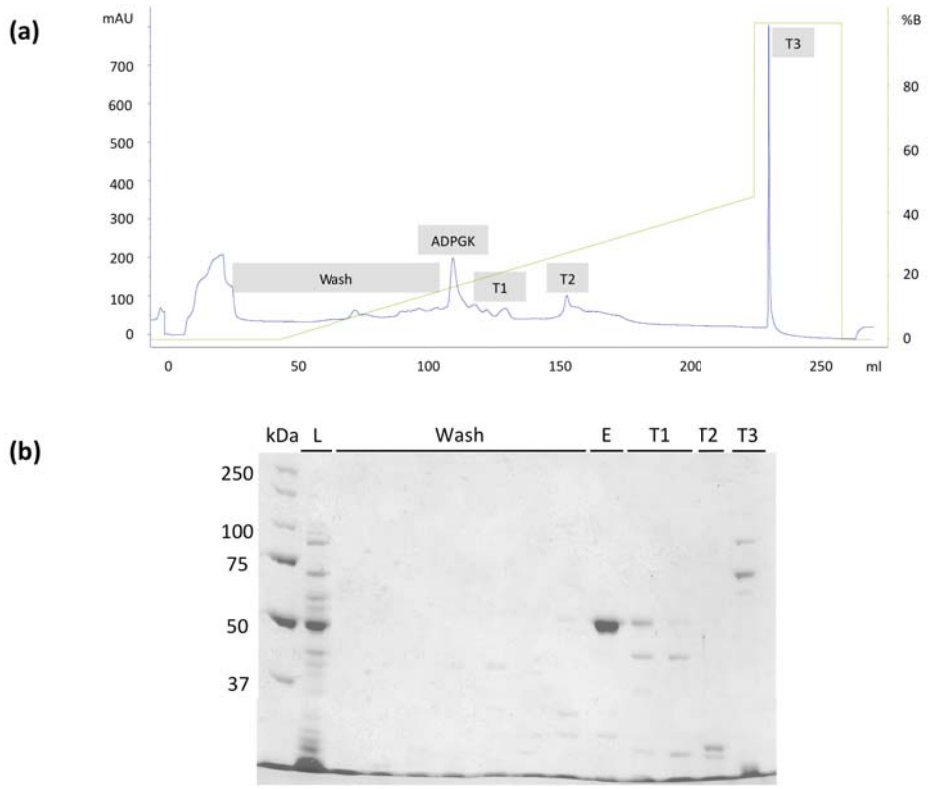

Figure 34: rhADPGK-D481E ion-exchange chromatography.

Overview of the ion-exchange chromatography step of the purification of rhADPGK-D481E. **(a)** Chromatogram. The protein eluted at a salt concentration of around 150 NaCl from the UnoQ6 column and the respective elution fraction peak is indicated. **(b)** SDS-PAGE. The protein sample before loading on to the column is labelled L, the elution fraction containing rhADPGK-D481E is labelled E. Selected fractions of the wash step and peaks tailing the elution fraction (Wash, T1, T2 and T3) are marked on the gel and chromatogram accordingly.

Unfortunately, in some cases two bands were apparent for rhADPGK on a western blot with the monoclonal antibody. The pattern was visible in the elution fractions of the affinity and ion-exchange purification steps. The second band was not always visible with Coomassie staining. For rhADPGK-R228A, the two bands were visible after IMAC purification (shown in Figure 35), but disappeared after ion-exchange chromatography (shown in Figure 36). A gel and chromatogram showing the double-banded pattern after ion-exchange chromatography are shown in Figure 37 for the purification of rhADPGK-H264A. The additional band is likely a degradation product of ADPGK. This has been observed in expression tests for rhADPGK, shown in Figure 18 (section 3.1.7.), depending on the expression time and temperature. The double-banded pattern is therefore likely caused by expression times being too long or the protein not being processed quickly enough after cell lysis. The additional band

was observed during the purification of rhADPGK-D84A, rhADPGK-R228A, rhADPGK-H264A and rhADPGK-D481A.

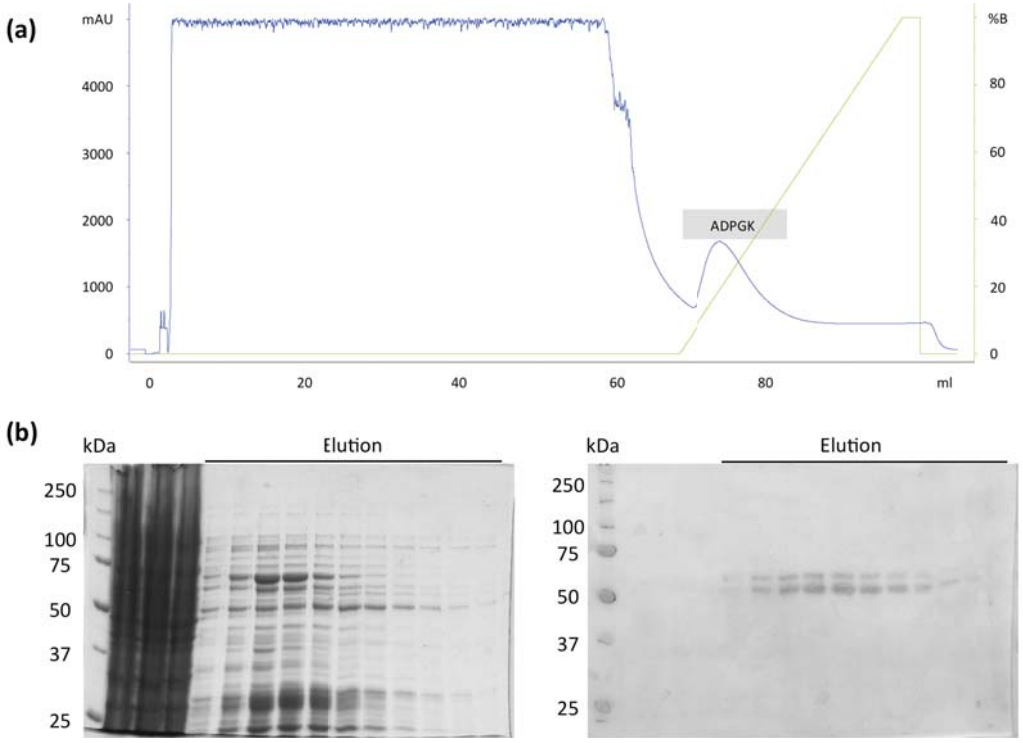

Figure 35: rhADPGK-R228A IMAC purification.

IMAC purification of rhADPGK-R228A using a HisTrap FF column. **(a)** Chromatogram. The protein eluted in a single, but broad peak at an imidazole concentration of approximately 200 mM. **(b)** SDS-PAGE (left) and western blot (right). The gel lanes with elution fractions containing rhADPGK-R228A are labelled Elution. On the western blot, two bands are visible for rhADPGK-R228A.

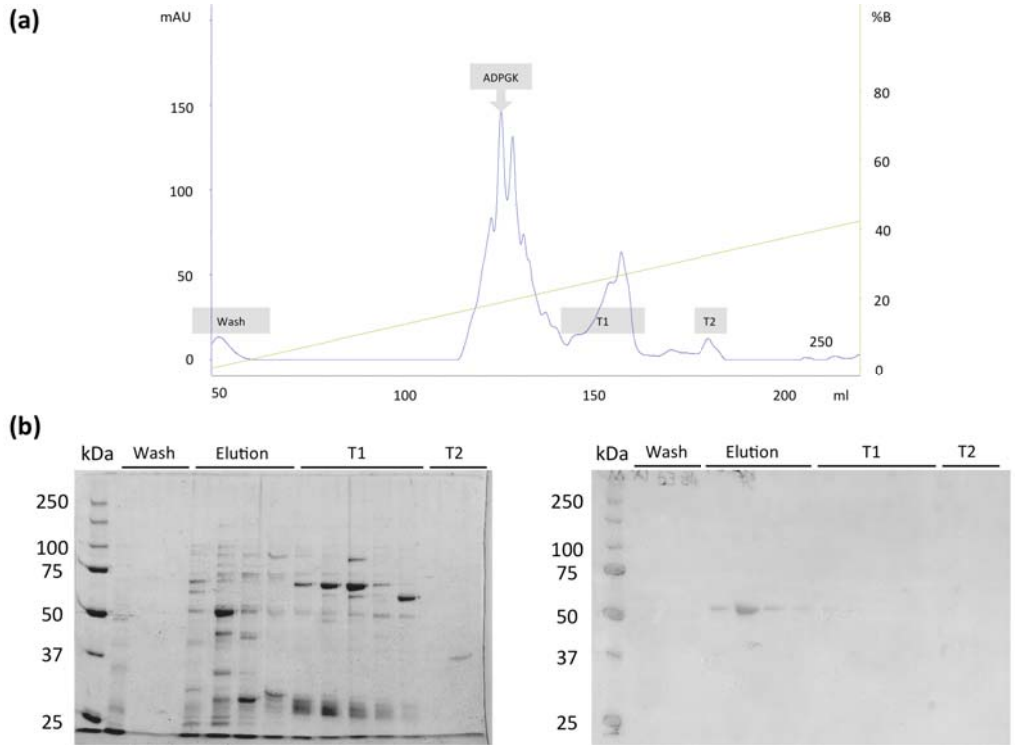

Figure 36: rhADPGK-R228A ion-exchange chromtapgraphy.

Overview of the ion-exchange chromatography step of the purification of rhADPGK-R228A. **(a)** Chromatogram. The protein eluted in a group of overlapping peaks (ADPGK marked with arrow) at a NaCl concentration of around 200 mM from the UnoQ6 column. **(b)** SDS-PAGE (left) and western blot (right). The elution fractions containing rhADPGK-R228A are labelled 'Elution'. The protein is visible as a single band on the western blot. On the Coomassie-stained gel impurities are apparent. Selected fractions from the wash steps and peaks tailing the elution (T1 and T2) are marked accordingly on the gel and chromatogram.

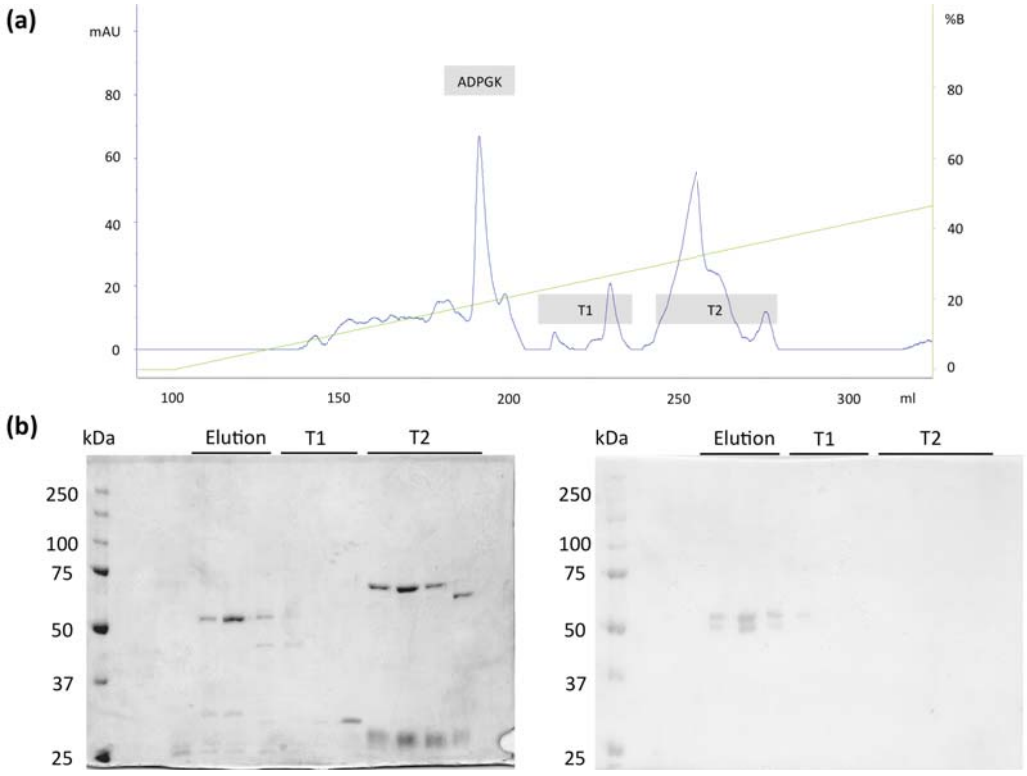

Figure 37: rhADPGK-H264A ion-exchange chromatography.

Overview of the ion-exchange chromatography step of the purification of rhADPGK-H264A. **(a)** Chromatogram. The protein eluted at a salt concentration of around 200 mM NaCl from the UnoQ6 column. **(b)** SDS-PAGE (left) and western blot (right). The protein sample before loading on to the column is labelled L, the elution fraction containing rhADPGK-H264A is labelled 'Elution'. On the western blot, two bands are visible for rhADPGK-H264A.

It was not possible to achieve the same degree of purity as the wild type for all ADPGK mutants. In Figure 36 and Figure 38, the chromatogram and gel for the ion-exchange purification and size-exclusion chromatography purification steps of rhADPGK-R228A are shown, respectively. The protein elutes in a group of overlapping peaks from the UnoQ6 column. The elution fractions contain several impurities after the ion-exchange step, which could not be removed by size-exclusion chromatography. Therefore, it was decided to purify all mutant constructs of rhADPGK only by affinity and ion-exchange chromatography.

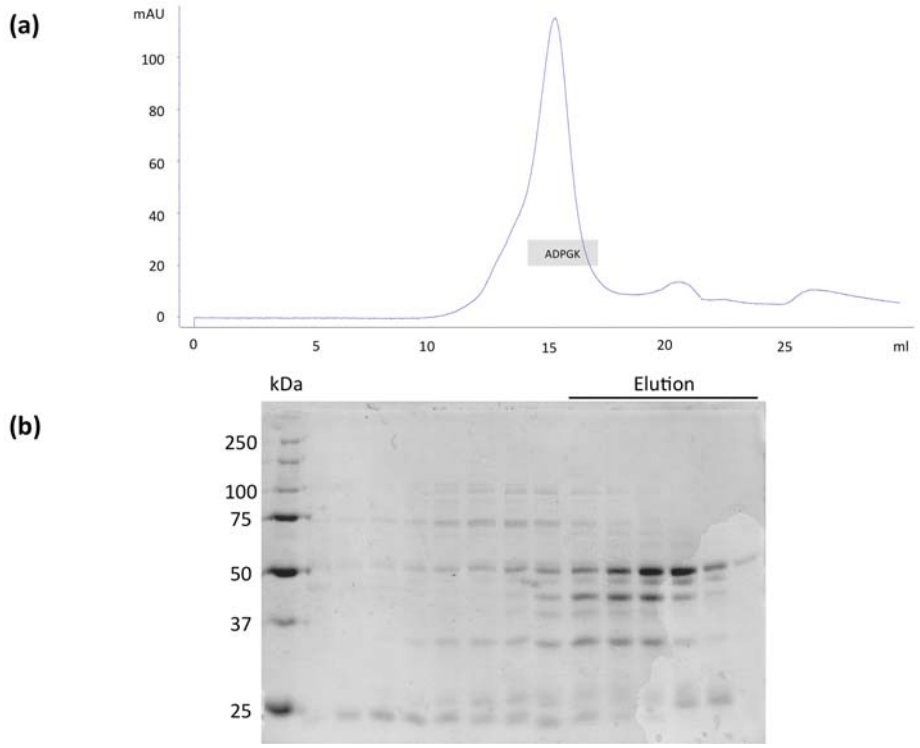

Figure 38: rhADPGK-R228A size-exclusion chromatography.

Overview of the size-exclusion chromatography purification step for rhADPGK-R228A. **(a)** Chromatogram. The protein eluted as a relatively broad peak at around 15 ml elution volume from the Superdex S200 column. **(b)** SDS-PAGE: The gel shows that the impurities could not be efficiently separated from ADPGK in this case by size-exclusion chromatography.

3.2.8 Purification of soluble rmADPGK

Production of mouse ADPGK was performed as described (Ronimus and Morgan 2004) with modifications. For expression, *E. coli* LMG194 was transformed with pBAD-rmADPGK and grown in 2× YT medium containing 100 µg/ml ampicillin as described in section 2.3.5. Cultures were grown at 37 ˚C to an OD_{600} of 0.8. The cells were then brought to a temperature of 22 ˚C and the inducer L-arabinose was added to a final concentration of 0.2% (w/v). Growth was continued and cells were harvested 12 hours after induction by centrifugation, washed in cold TBS buffer and frozen at -80 ˚C as described in section 2.3.5. The weight of the washed, wet cell pellets obtained from 12 l medium was typically around 40 g. The cell pellets were resuspended in lysis buffer (50 mM Tris pH 8.0, 500 mM NaCl, 1 mM DTT, Roche complete protease inhibitor tablet EDTA-free), lysed by French press and the soluble fraction separated by centrifugation as described in section 2.3.6. Imidazole was added to the supernatant to a final concentration of 20 mM and the solution filtered

through 0.45 μm and 0.2 μm syringe filters (Sartorius, Germany) before loading on to a HisTrap FF 5 ml column (GE Healthcare, UK). IMAC purification was performed as described in section 2.5.7. Protein was eluted using a linear gradient from 20 mM to 500 mM imidazole in phosphate buffer (25 mM Na_2HPO_4, pH 8.0, 500 mM NaCl, 500 μM DTT) and rmADPGK eluted at a imidazole concentration of 120 mM. The subsequent ion-exchange chromatography step was performed as described in section 2.5.9. The fractions containing rmADPGK were pooled, dialyzed against low ionic strength buffer (40 mM Tris/HCl pH 8.0, 20 mM NaCl) and filtered through 0.2 μm syringe filter before being loaded on to a UnoQ6 column (Bio-Rad Laboratories, USA) pre-equilibrated with binding buffer (40 mM Tris/HCl pH 8.0, 20 mM NaCl). Elution was performed with a step gradient over 35 cv to 1 M NaCl and rmADPGK eluted in the first segment at a salt concentration of 150 mM. A chromatogram and corresponding gel are shown in Figure 39. The theoretical molecular weight of rmADPGK was 58.4 kDa, as determined by ProtParam. Protein purity was judged by SDS-PAGE and ADPGK was either directly concentrated and flash-frozen in liquid nitrogen for storage at -80 ˚C or subjected to further purification by size-exclusion chromatography. The elution fractions containing rmADPGK from ion-exchange chromatography were pooled and the volume reduced by concentration at 4 ˚C in a Vivaspin 20 concentrator (GE Healthcare, UK).

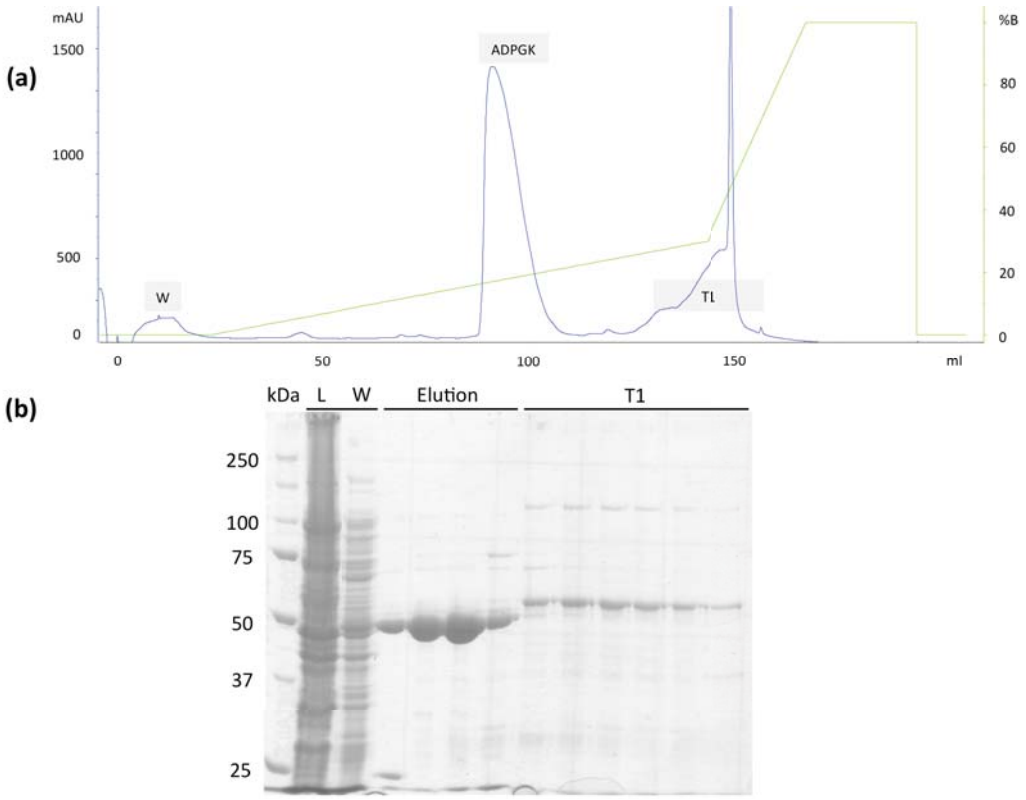

Figure 39: rmADPGK ion-exchange chromatography.

Overview of the purification of rmADPGK with a UnoQ6 ion-exchange column. **(a)** Chromatogram. The protein eluted in the first major peak at a concentration of 150 mM NaCl. **(b)** SDS-PAGE: The gel lanes with elution fractions containing rmADPGK are labelled 'Elution'. The first two lanes labelled 'Load' are, from left to right, protein after filtering before loading on to the column and the first flow-through fraction containing unbound protein.

Size-exclusion chromatography was performed as described in section 2.5.10. The concentrated protein was centrifuged and filtered before being loaded on a Tricorn Superdex S200 10/300 GL size-exclusion chromatography column (GE Healthcare, UK) pre-equilibrated with buffer (50 mM Tris, pH 8.0, 150 mM NaCl). A chromatogram and corresponding gel are shown in Figure 40, where eluted rmADPGK can be seen as a single peak at an elution volume of 15.3 ml.

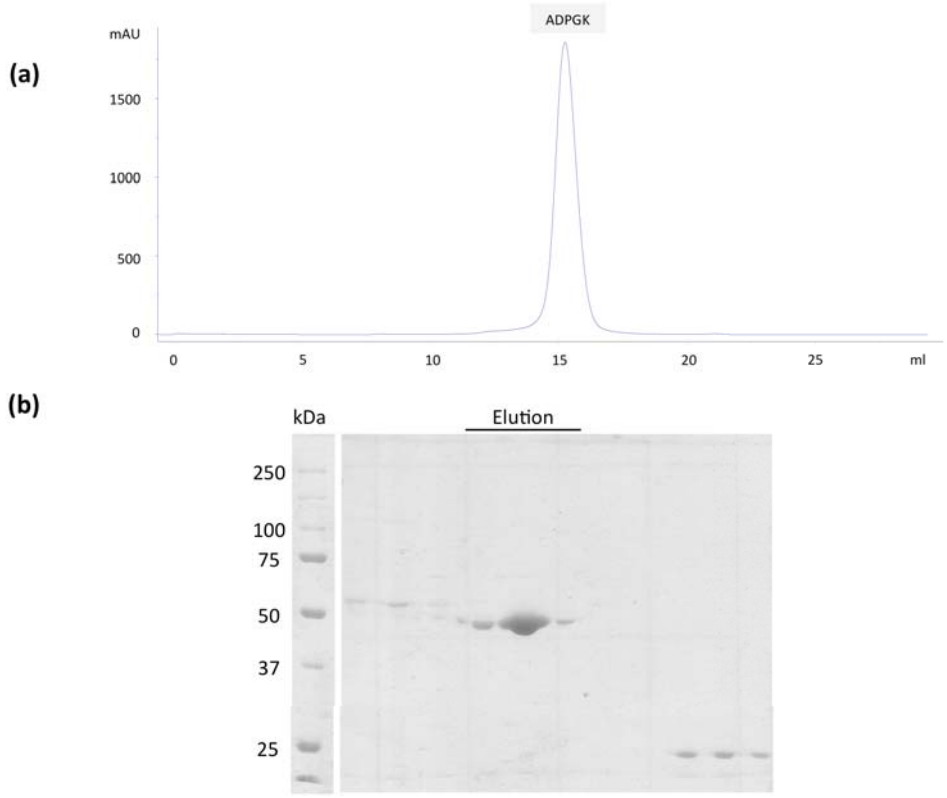

Figure 40: rmADPGK size-exclusion chromatography.

Overview of the final purification step for rmADPGK. **(a)** Chromatogram. The protein eluted as a single peak at 15.3 ml elution volume from the Tricorn Superdex S200 10/300 GL size-exclusion chromatography column. **(b)** SDS-PAGE: The gel lanes labelled 'Elution' contain purified ADPGK, visible as intense band below the 50 kDa marker. Contaminants of larger and small molecular weight are visible left and right of the elution fraction containing rmADPGK and were successfully removed.

The elution fractions containing rmADPGK were pooled, concentrated, flash frozen in liquid nitrogen and stored at -80 ˚C. Purified rmADPGK after size-exclusion chromatography had over 95% purity as assessed by SDS-PAGE. A gel is shown in Figure 41. Typical yield of an initial culture volume of 12 l was around 10 mg of purified protein.

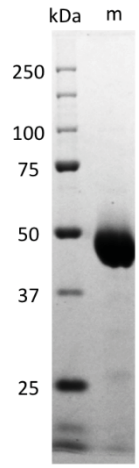

Figure 41: Purified rmADPGK for crystallisation.

Purified mouse ADPGK (m) after Coomassie staining. ADPGK is visible as a single and intense band just below the 50 kDa marker line.

3.2.9 Purification of soluble rmADPGKΔ51

For expression, *E. coli* LMG194 were transformed with pBAD-rmADPGKΔ51 and grown in 2× YT medium containing 100 μg/ml ampicillin as described in section 2.3.5. All other steps were performed as described in section 3.2.8 for the full-length ADPGK cloned into the same vector. The protein was purified by IMAC and ion-exchange chromatography. A size-exclusion chromatography purification step was not performed. After purification, rmADPGKΔ51 was flash frozen in liquid nitrogen and stored at -80 ˚C until required for crystallisation experiments.

3.2.10 Purification of soluble rzADPGK

The expression conditions of ADPGK from *D. rerio* could not be optimised during expression tests for mostly unknown reasons (see section 3.1.11). Despite the problems during these initial tests, it was decided to attempt to purify rzADPGK using the same protocol that was used for rhADPGK described in section 0, based on the assumption that the protein might express in *E. coli*, despite not being detectable on SDS-PAGE by coomassie staining or western blot (as was explained in section 3.1.11). For expression, *E. coli* LMG194 were transformed with pBAD-rzADPGK and grown in 2× YT medium containing 100 μg/ml ampicillin as described in section 2.3.5. The cells were grown at 37 ˚C and expression was induced with a final

concentration of 0.2% (w/v) L-arabinose at an OD_{600} of 0.8. Growth was continued for 12 hours at 18 ˚C and the cultures were harvested, washed and stored frozen at -20 ˚C as described in section 2.3.6. IMAC purification was carried out exactly as described for rhADPGK in section 0. The resulting protein fractions were tested for enzymatic activity and the active fractions were pooled. The fraction contained a large number of *E. coli* proteins. Unfortunately, it was not possible to separate rzADPGK from these contaminants by ion-exchange chromatography and size-exclusion chromatography purification when treated as described for rhADPGK. Also, the enzyme lost its activity in the purification steps after the IMAC purification step. The elution fractions from IMAC purification containing rzADPGK were pooled and stored at 4 ˚C for one day before being used for activity assays.

3.2.11 Summary

Different variants of recombinant ADPGK were produced in *E. coli* and purified. For the purpose of crystallisation, ADPGK was purified as full-length protein rhADPGK and rmADPGK, and in a truncated form, rmADPGKΔ51 and rhADPGKΔ50. The protein from *M. musculus* was more stable during purification, making it the preferred choice for the set up of large crystallisation trials, compared to the protein from *H. sapiens*. In Figure 42, comparison of CD spectra for full-length and truncated ADPGK variants are shown. The CD spectra for rmADPGK, rhADPGK, and rhADPGKΔ50, showed that the overall structure and composition of secondary structure elements of the truncated forms of mouse and human ADPGK is likely to be very similar, and that the truncation of the N-terminus likely does not lead to a major structural rearrangement, compared to the full-length protein.

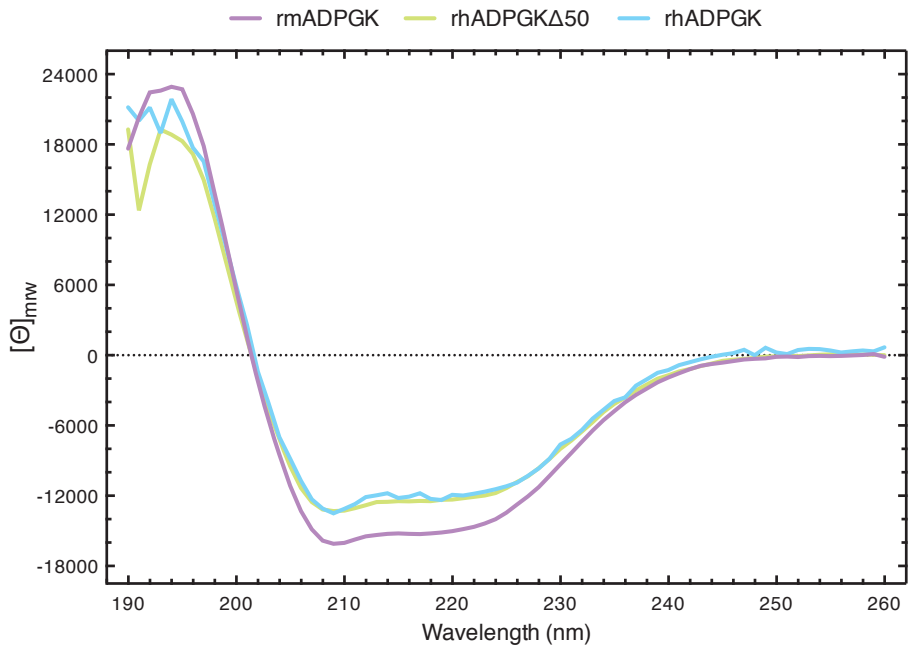

Figure 42: Comparison of CD spectra of purified eukaryotic ADPGK.

The circular dichroism spectra of the purified full-length proteins rmADPGK rhADPGK and the truncated variant rhADPGKΔ50 are shown as mean residue ellipticity $[\Theta]_{mrw}$. The spectra show an overall similar shape, indicating a similar composition of secondary structure elements and therefore likely similar overall fold.

For biochemical studies, a variety of functional mutants of rhADPGK were produced and purified. The ADPGK from *D. rerio* could not be purified to homogeneity and only the crude elution fractions after IMAC purification could be obtained. The refolding of rhADPGKΔ50 had limited success. When the protein was refolded after an established protocol for a full-length construct of ADPGK (Hole 2009), the refolded protein had a reduced specific activity when compared to protein purified from soluble expression in *E. coli*. In an attempt to optimise the refolding protocol for ADPGK with the iFOLD screen, several promising conditions were identified. Unfortunately, the chemicals required to reproduce the experiments were not immediately available. The experiments have confirmed that it is possible to refold human ADPGK, although more optimisation may be required to make this approach more desirable than the purification of ADPGK from a source optimised for soluble expression.

3.3 Activity of ADPGK

3.3.1 Introduction

In this chapter the enzymatic characterisation of ADPGK is presented. Kinetic parameters were determined for rhADPGK, rhADPGKΔ50, rzADPGK and functional mutants of rhADPGK and rhADPGKΔ50. The determination of kinetic constants and measurement of activity was performed as described in section 2.5.11. The substrate specificity of rhADPGK was investigated by NMR spectroscopy methods as described in section 2.5.13. For determination of the apparent values for the maximum velocity V_{max} and Michaelis constant K_m , the concentration of one substrate was varied while either Mg-ADP or glucose concentration was kept constant. Since high concentrations of either substrate were inhibitory, kinetics experiments were performed at substrate concentrations where the inhibitory effect would still be negligible as possible (see section 2.5.11). Therefore apparent values K_m^{app} and V_{max}^{app} are reported. The scope of analysis is limited for rhADPGKΔ50, rzADPGK and certain functional mutants of rhADPGK, due to limited availability of the respective proteins, as the analysis of bisubstrate reaction would usually require testing out the reaction mechanism and the use of a general reate equation (Alberty 1953) to determine the kinetic parameters.

3.3.2 Determination of kinetic parameters for rhADPGK

Purified rhADPGK was transferred in to storage buffer (20 mM Tris/acetate, 150 mM NaCl, pH 8.0) by dialysis after ion-exchange chromatography. Glycerol was added to a final concentration of 20% (v/v) and the enzyme stored at -80 ˚C. The enzyme did retain its activity upon thawing after storage for up to one year. To determine the pH sensitivity of the catalytic rate of rhADPGK, the enzyme was found to have the highest activity at a pH of 7.5, with 18%, 91%, 87% and 40% of activity remaining at pH 6.0, 7.0, 8.0 and 9.0, respectively (shown in Figure 43). The pH was tested in steps of 0.25 from pH 6.0 to 9.25 in Bis-tris propane buffer, as this buffer had previously been proven to be suitable for studying the pH sensitivity of *M. musculus* ADPGK (Ronimus and Morgan 2004).

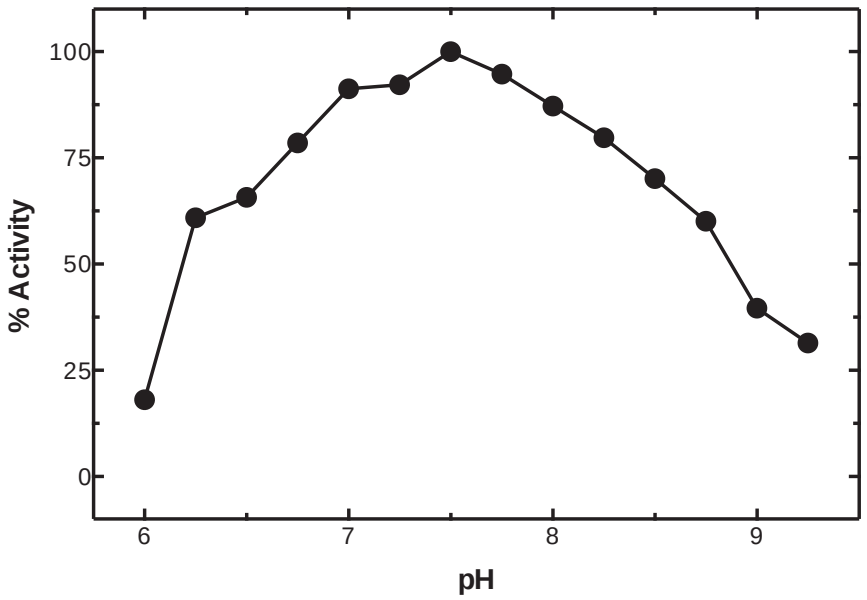

Figure 43: Determination of dependence of enzymatic activity on pH for rhADPGK.

The plot shows the dependence of enzymatic activity of rhADPGK on the pH. The enzyme shows maximum activity at pH 7.5. The pH was varied in 0.25 steps in 50 mM Bis-tris propane buffer and activity measured with 0.7 mM D-glucose and 1 mM ADP.

When DTT was added in amounts of 10 or 40 mM to the storage buffer of rhADPGK , no change in activity was observed when compared to an untreated control sample. The enzyme was found to have a V_{max}^{app} of 30 U/mg and followed Michaelis-Menten kinetics up to a certain concentration of substrates, around 1 mM for glucose and around 2 mM for Mg-ADP. ADPGK was inhibited by higher concentrations of the substrate. Kinetic parameters for rhADPGK were determined as described in section 2.5.11 with a model considering substrate inhibition. The K_m^{app} and K_i values for glucose were 0.48 mM (+/- 0.09) and 2.9 mM (+/- 0.6), respectively. The K_m^{app} and K_i values for Mg-ADP were 0.56 mM (+/- 0.094) and 9 mM (+/- 1.5), respectively (shown in Figure 44 and Figure 45 for glucose and Mg-ADP, respectively).

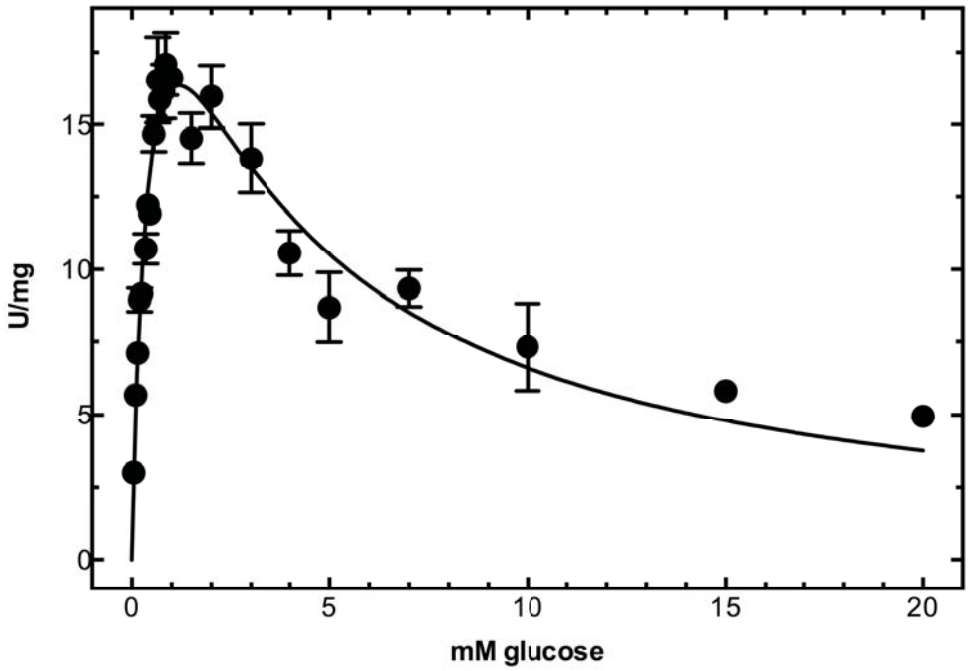

Figure 44: Enzyme kinetics of rhADPGK, K_m^{app} for glucose.

The plot shows the rate of enzymatic activity of rhADPGK as a function of increasing concentration of D-glucose. The enzyme is inhibited by glucose concentrations above 1 mM. The K_m^{app} and K_i for D-glucose were 0.48 mM and 2.9. $R^2 = 0.95$.

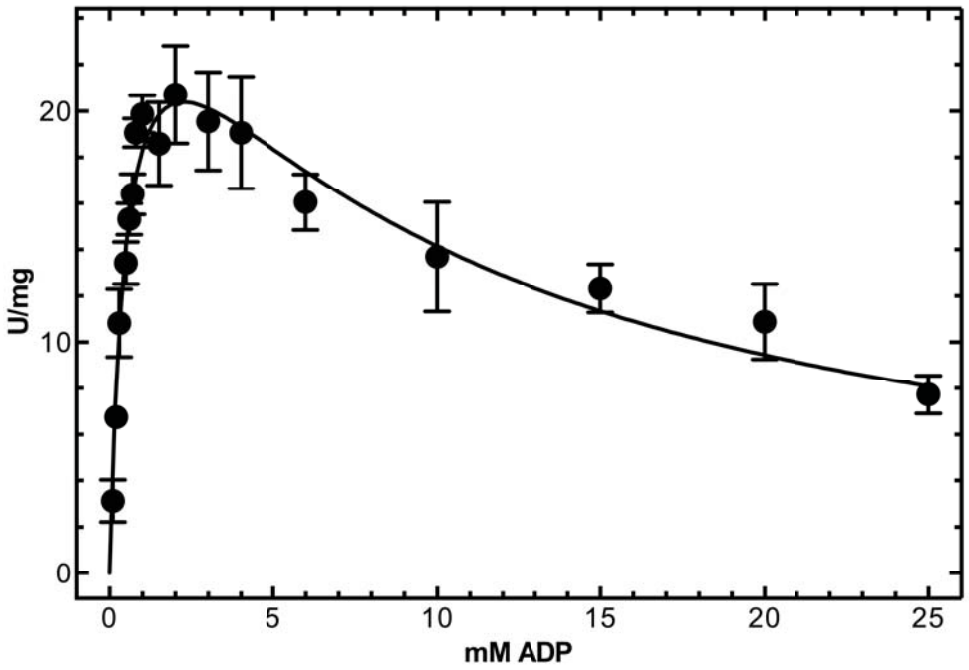

Figure 45: Enzyme kinetics of rhADPGK, K_m^{app} for ADP.

The plot shows the rate of enzymatic activity of rhADPGK as a function of increasing concentration of Mg-ADP. The enzyme was inhibited by ADP concentrations above 2 mM. The K_m^{app} and K_i were 0.56 mM and 9 mM, respectively, at a constant glucose concentration of 0.7 mM. $R^2 = 0.96$.

Product inhibition was detected for AMP. The inhibitory effect of AMP is shown in Figure 46. AMP was found to be a competitive inhibitor to Mg-ADP for ADPGK, with a K_i of 0.50 mM (+/- 0.07). For the inhibition experiment, AMP concentrations of 1, 2, 5, and 10 mM were tested. No extra Mg^{2+} ion was added to the reaction in addition to AMP as adding excess $MgCl_2$ was found to be inhibitory by itself. Concentrations of 5 mM, 10 mM and 20 mM were tested, with 100%, 48% and 51% of activity remaining, respectively. Glucose-6-phosphate was shown not to be inhibitory to mouse ADPGK activity in previous studies (Ronimus and Morgan 2004).

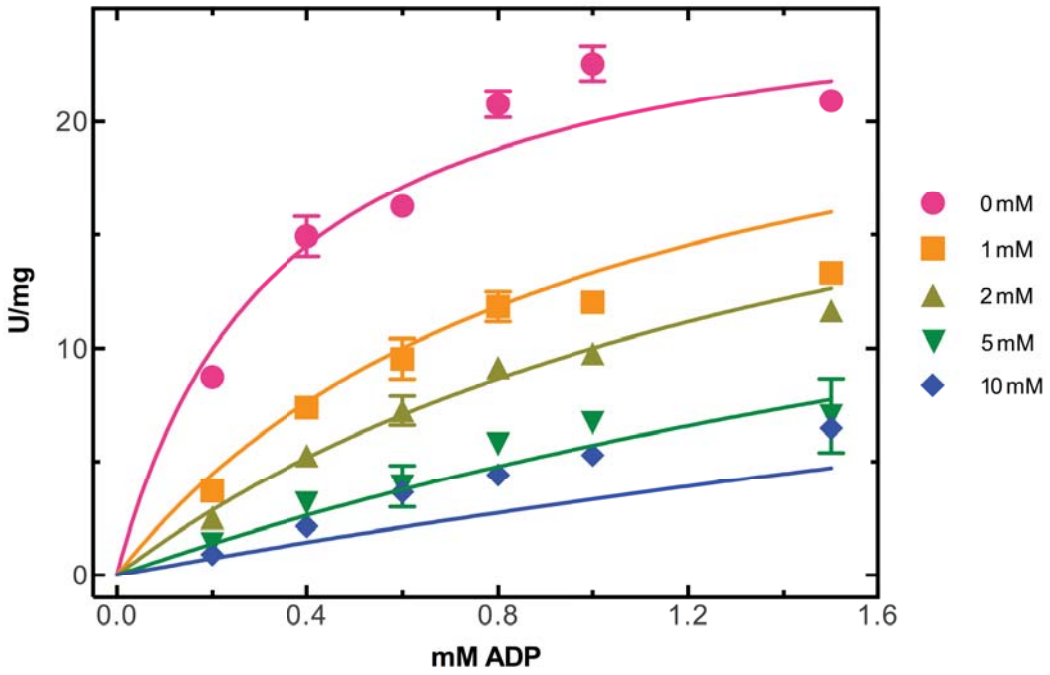

Figure 46: Enzyme kinetics of rhADPGK, product inhibition by AMP.

The plot shows the effect of the product AMP on the enzymatic activity of rhADPGK as a function of increasing concentration of Mg-ADP. AMP is a competitive inhibitor for Mg-ADP binding to ADPGK with a K_i of 0.50 mM. The legend on the right shows the concentration of AMP. The kinetic parameters were determined by non-linear regression with Prism 6.0. Global $R^2 = 0.96$.

Additionally, ADPGK was tested for inhibition by Mg-ATP. For this experiment it was decided to add Mg-ATP, to avoid quenching of Mg^{2+} when ATP is added in large

amounts. The resulting data showed that ATP is likely binding to ADPGK. The maximum catalytic activity was reduced to 21% at a concentration of 9 mM Mg-ATP. However, since Mg^{2+} by itself also has an inhibitory effect, the effect of ATP was not separable from Mg^{2+} inhibition. The reduction of the maximum catalytic activity was, however, still stronger with Mg^{2+} and ATP combined, than Mg^{2+} on its own.

3.3.3 Determination of kinetic parameters for functional mutants

Mutants and variants of ADPGK are listed in Table 3.2. It has to be noted that rhADPGKΔ50 and rhADPGK have the same amino acid sequence, apart from the N-terminal truncation for the former construct and different purification tags, since the proteins were cloned from different cDNA and expressed and purified from different vectors (see Table 3.1). The truncated version of ADPGK missing the first 50 amino acid residues, rhADPGKΔ50, showed comparable activity to the wild type enzyme. Mutant rhADPGK-H382A did not express in *E. coli* and therefore could not be purified. Mutants rhADPGKΔ50-D481A, rhADPGK-H387A, rhADPGK-D481E and rhADPGK-H382V:H387V had no detectable activity although these proteins were detectable by western blotting upon purification. Reduced activity was observed for mutant rhADPGK-D84A (0.00043 U/mg), rhADPGK-R228A (0.31 U/mg), rhADPGK-H264A (0.0051 U/mg), rhADPGK-D481A (0.15 U/mg) and rhADPGK-D481N (0.017 U/mg). A variant of human ADPGK with a truncation comprising the first 151 amino acids, rhADPGKΔ151 showed no activity, although the protein could be purified and was detectable on a western blot. For mutant rhADPGK-H264A, the K_m^{app} for ADP was 0.25 mM (+/- 0.06).

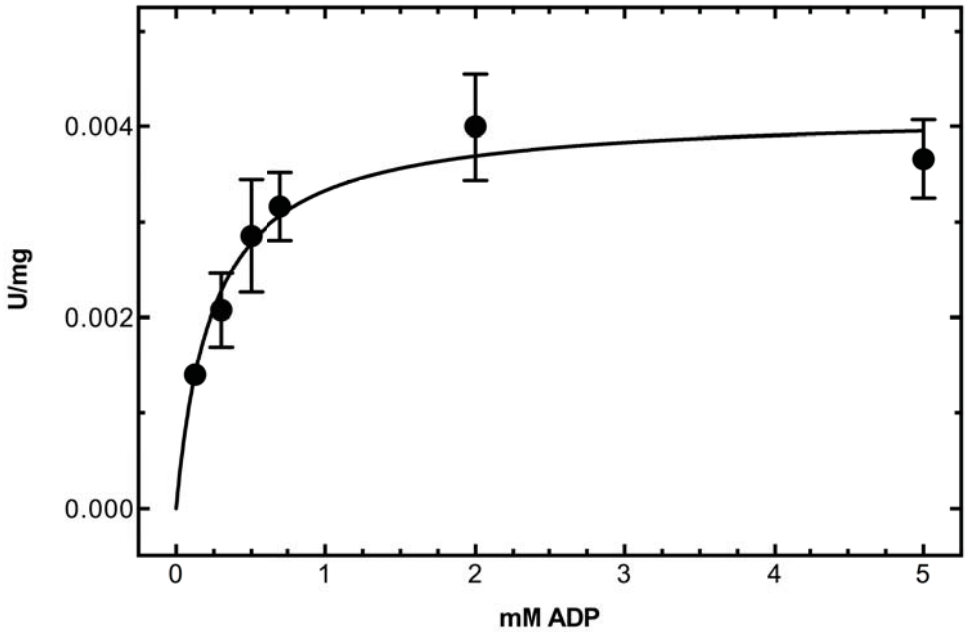

Figure 47: Enzyme kinetics of rhADPGK-H264A, K_m^{app} for ADP.

The plot shows the rate of enzymatic activity of rhADPGK-H264A as a function of increasing concentration of Mg-ADP at a constant glucose concentration of 0.7 mM. The K_m^{app} for ADP was 0.25 mM. $R^2 = 0.85$.

The K_m^{app} for glucose of rhADPGK-H264A was 5 mM (+/- 0.6), which is considerably higher than the wild type enzyme. The V_{max} was 0.0051 U/mg. This mutant did not show inhibition by glucose with data collected for concentration up to 55 mM as can be seen in Figure 48.

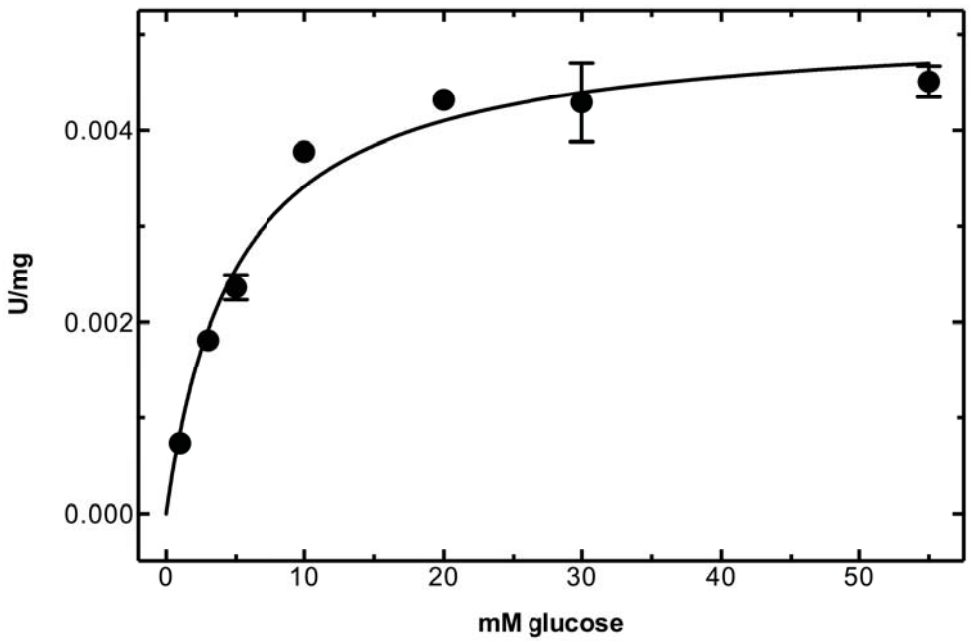

Figure 48: Enzyme kinetics of rhADPGK-H264A, K_m^{app} for glucose.

The plot shows the rate of enzymatic activity of mutant rhADPGK-H264A as a function of increasing concentration of D-glucose. The enzyme showed no inhibition by glucose at concentrations of up to 55 mM. The K_m^{app} for glucose was 5 mM at a constant concentration of 1 mM ADP. $R^2 = 0.97$.

For mutant rhADPGK-R228A, the K_m^{app} for Mg-ADP was 0.27 mM (+/- 0.06) and the K_m^{app} for glucose was 0.23 mM (+/- 0.06) as shown in Figure 49 and Figure 50, respectively. The V_{max}^{app} was 0.31 U/mg. Only a very limited amount of sample was available for this mutant.

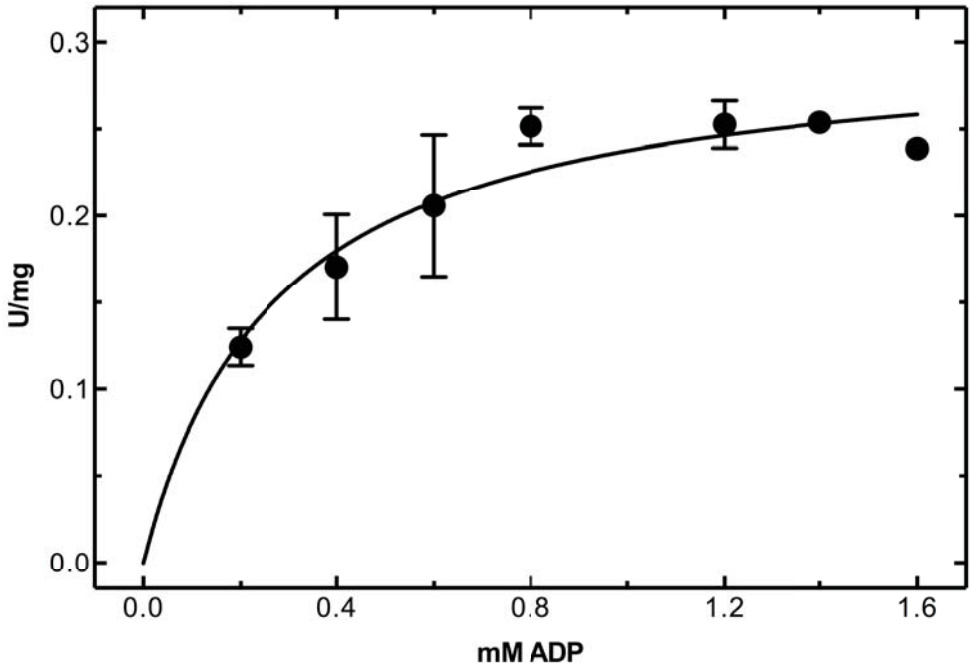

Figure 49: Enzyme kinetics of rhADPGK-R228A, K_m^{app} for ADP.

The plot shows the rate of enzymatic activity of rhADPGK-H264A as a function of increasing concentration of Mg-ADP. The K_m^{app} for ADP was 0.27 mM. $R^2 = 0.84$.

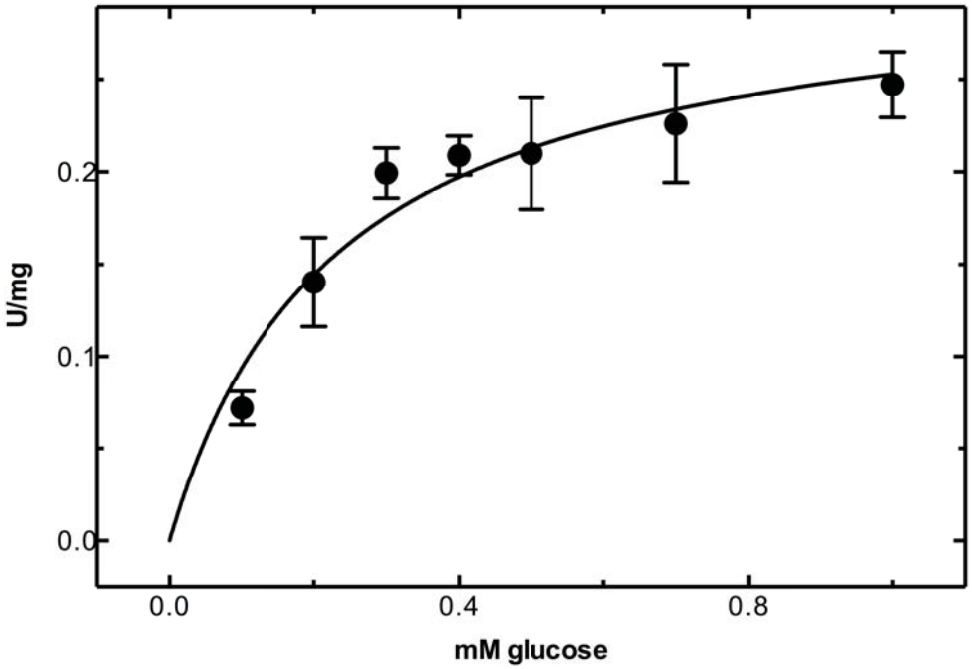

Figure 50: Enzyme kinetics of rhADPGK-R228A, K_m^{app} for glucose.

The plot shows the rate of enzymatic activity of rhADPGK-H264A as a function of increasing concentration of D-glucose. The K_m^{app} for glucose was 0.23 mM at a constant concentration of 1 mM Mg-ADP. $R^2 = 0.88$.

3.3.4 Alternative substrates for rhADPGK

The ability of rhADPGK to use alternative nucleotide substrates ATP, GDP, CDP, TDP and UDP was tested, in equimolar ratios with the counter ion Mg^{2+} (added as $MgCl_2$). These tests were conducted with the standard coupled assay using glucose-6-phosphate dehydrogenase as linker enzyme, since phosphorylation should still produce the compatible glucose-6-phosphate. When rhADPGK was assayed with 1 mM GDP it had a V_{max}^{app} of 1.46 U/mg (+/- 0.16) and a K_m^{app} of 2.9 mM (+/- 0.8) as shown in Figure 51. When tested with 1 mM CDP rhADPGK had a V_{max} of 0.049 U/mg. The K_m^{app} value for CDP was not determined due to the low activity. No activity was detectable with ATP (tested at 0.1, 0.2, 2.0 and 6.0 mM), UDP or TDP (tested at 1 and 10 mM).

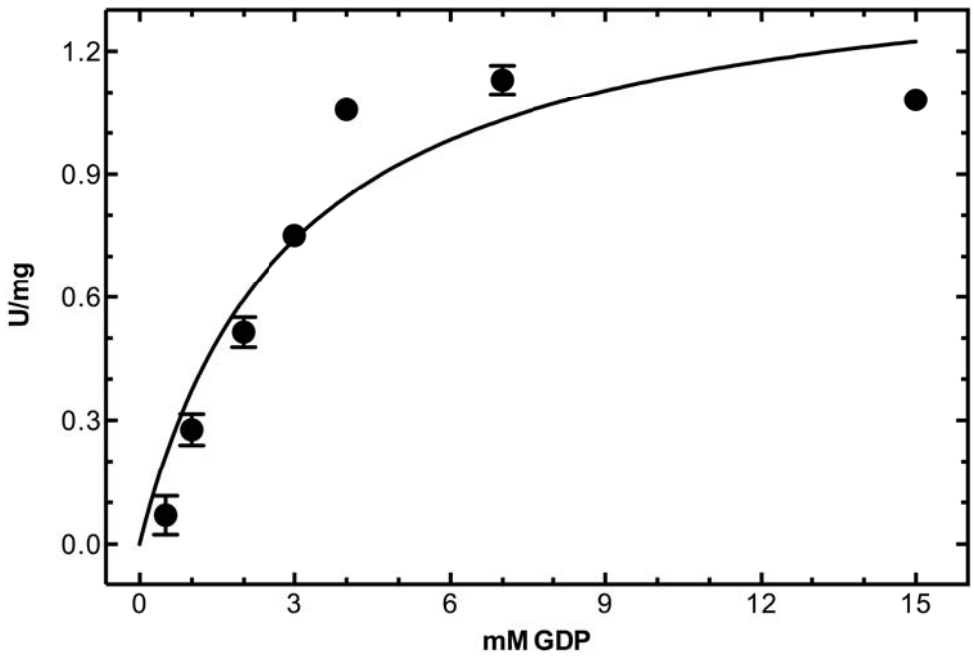

Figure 51: Enzyme kinetics of rhADPGK, K_m^{app} for GDP.

The plot shows the rate of enzymatic activity of rhADPGK as a function of increasing concentration of Mg-GDP at a constant concentration of 0.7 mM glucose. The K_m^{app} for GDP was 2.9 mM. $R^2 = 0.89$.

Alternative sugar substrates were screened for wild type rhADPGK and functional mutant rhADPGK-H264A by using direct measurement of ADP consumption by ^{31}P-NMR spectroscopy as described in section 2.5.13. All NMR measurements were conducted in collaboration with Dr. Alexander Goroncy. NMR spectroscopy is an

excellent tool for this particular question due to its high sensitivity and the possibility to measure the product directly, thus eliminating the need for linking enzymes usually required in coupled enzymatic kinetic activity assays for ADPGK. D-glucose and Mg-ADP were used as controls and the disappearance of the ^{31}P-signals representing ADP and appearance of signals representing AMP and glucose-6-phosphate could be observed (see Figure 52).

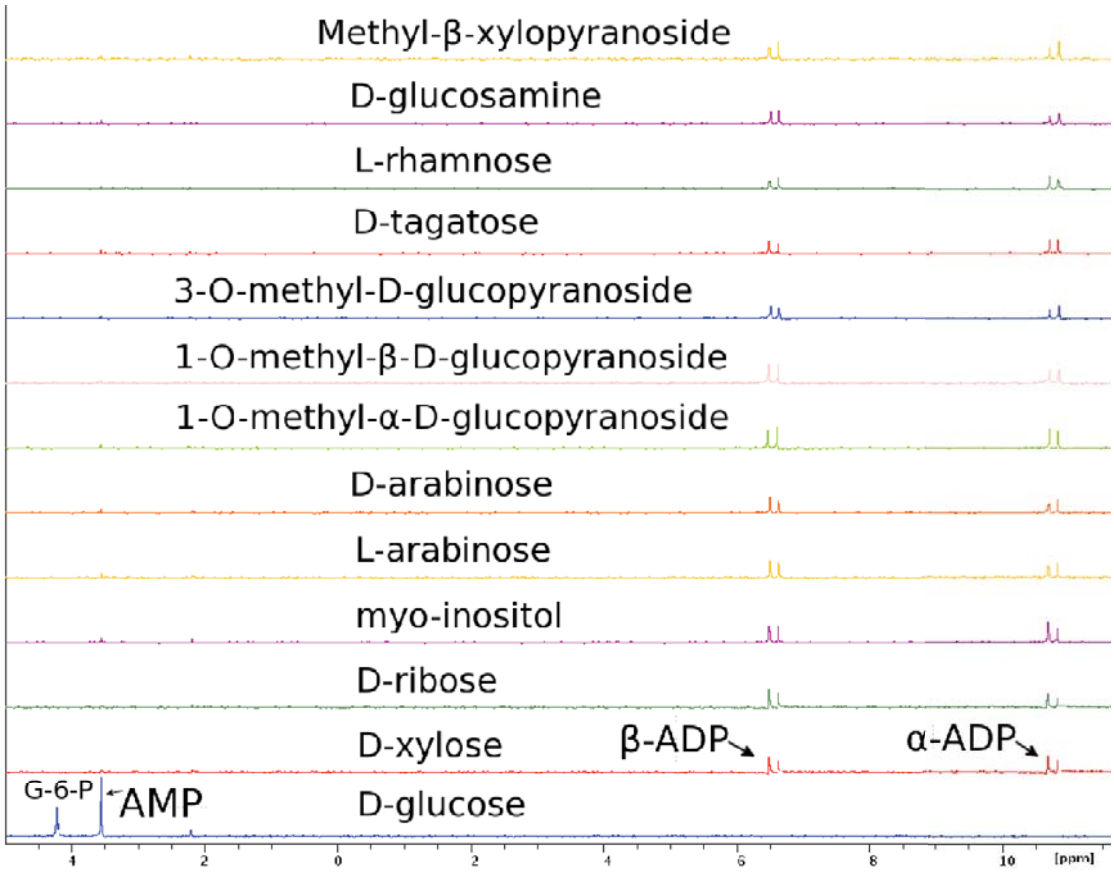

Figure 52: Alternative substrates tested for *H. sapiens* ADPGK.

Representative ^{31}P -NMR spectra for screening substrate specificity of rhADPGK. A control reaction with Mg-ADP and D-glucose is shown at the bottom, where the signals for the products AMP and D-glucose-6-phosphate are indicated. No product was detectable for all other spectra and the signal intensity for the α- and β-anomers of ADP remained unchanged. Concentrations of substrates were 1 mM for sugars and Mg-ADP. The figure was prepared by Dr. Alexander Goroncy.

Assignments were confirmed by adding known amounts of ADP, AMP, and glucose-6-phosphate. None of the tested substrates served as phosphoryl acceptor. The compounds tested were L-glucose, 2-deoxy-D-glucose, D-galactose, D-fructose, D-

mannose, D-tagatose, L-rhamnose, D-glucosamine, D-fructose-6-phosphate, α-D-glucose-1-phosphate, 1-*O*-methyl-α-D-glucopyranoside, 1-*O*-methyl-β-D-glucopyranoside, 3-*O*-methyl-D-glucopyranoside, methyl-β-D-xylopyranoside, D-arabinose, L-arabinose, D-ribose, D-xylose, D-glucose-6-phosphate, myo-inositol, and D-fructose-2,6-bisphosphate. No phosphoryl-acceptor other than D-glucose was accepted as a substrate for ADPGK. Alternative nucleotide substrates were tested for the double-mutant rhADPGK-H382V:H387V with D-glucose as a phosphoryl group acceptor by ^{31}P-NMR. This double mutant was designed with the aim to potentially alter the nucleotide specificity, together with the mutants rhADPGK-H382A and rhADPGK-H387A. Tested nucleotides were ADP, ATP, CDP, GDP, TDP and UDP at a concentration of 1 mM with Mg^{2+} as counter ion. No activity was found for any nucleotide for this mutant.

3.3.5 Determination of kinetic parameters for rhADPGKΔ50

ADPGK variant rhADPGKΔ50 was overexpressed in *E. coli* and purified as described in section 3.2.2. When the protein was prepared for enzymatic assays, the concentration was adjusted to 0.5 mg/ml in storage buffer (10 mM Tris/HCl, 150 mM NaCl, 0.5 mM DTT, pH 7.5). To determine the pH sensitivity of the catalytic rate of rhADPGKΔ50, the enzyme was found to have the highest activity at a pH of 8.0, with 16%, 79% and 47% of activity remaining at pH 6.0, 7.0 and 9.0, respectively. The pH was tested in Bis-tris propane buffer at a constant concentration of 0.7 mM glucose and 1 mM Mg-ADP, as this buffer has proven to be most suited for studying the pH sensitivity of *M. musculus* ADPGK due to its broad buffering range (Ronimus and Morgan 2004). The test was limited to the pH values listed. To determine the sensitivity to salt concentration of the reaction buffer, rhADPGKΔ50 was tested for activity at a KCl concentration of 20, 200 and 400 mM at a constant concentration of 0.7 mM for glucose and 1 mM for Mg-ADP. The enzyme had the highest activity at 100 mM KCl, with 98, 90% and 89% of activity remaining at 20, 200 and 400 mM KCl respectively (see Figure 53).

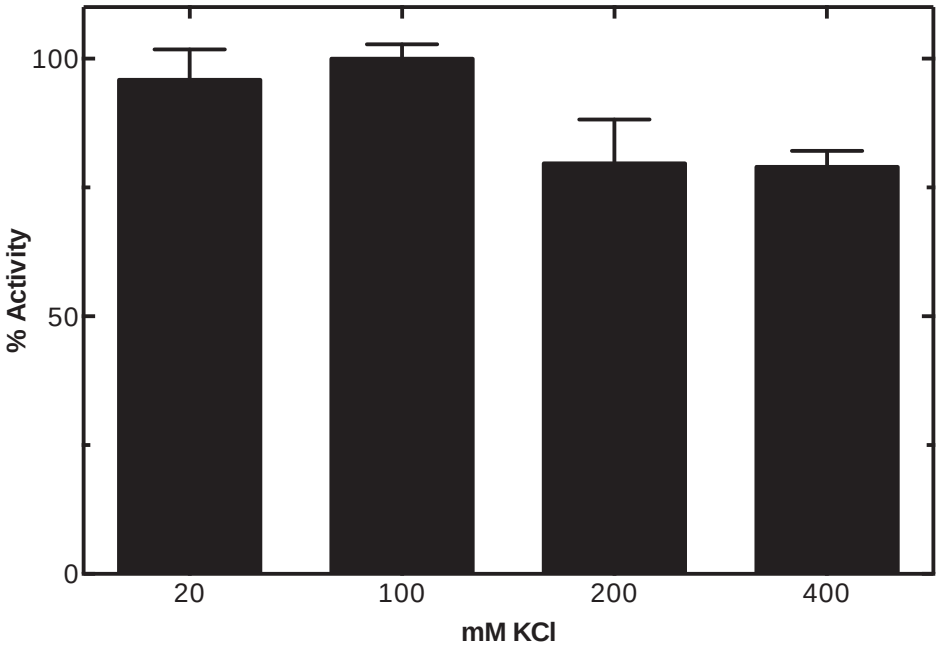

Figure 53: Enzyme kinetics of rhADPGKΔ50, effect of salt concentration.

The bar plot shows the effect of 20, 100, 200 and 400 mM KCl on the catalytic activity of rhADPGKΔ50. Maximum catalytic activity was observed at 100 mM. Effect of salt concentrations was tested at 0.7 mM glucose and 1 mM Mg-ADP.

For rhADPGKΔ50 the K_m^{app} for glucose was 0.37 mM (+/- 0.07) and the V_{max}^{app} was 32 U/mg (+/- 2) (see Figure 54). The K_m^{app} for ADP could not be tested convincingly due to insufficient data points. With the limited data available the K_m^{app} for ADP would be 0.19 mM (+/- 0.08). However, this result can only be interpreted with caution, due to the lack of data available. No catalytic activity of rhADPGKΔ50 was observed with ATP (tested at 1 mM).

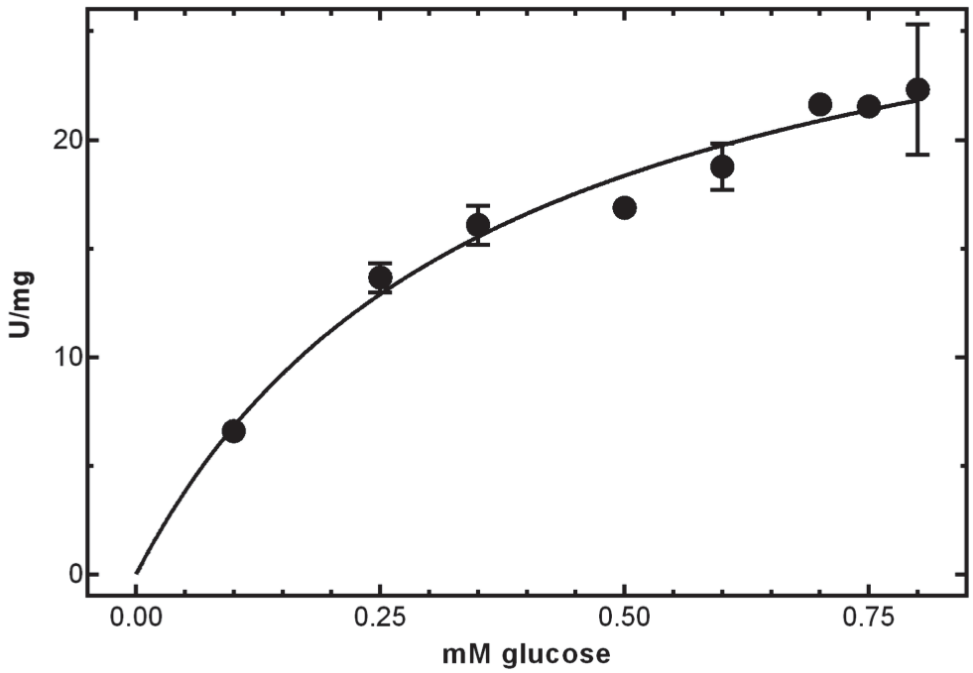

Figure 54: Enzyme kinetics of rhADPGKΔ50, K_m^{app} for glucose.

The plot shows the rate of enzymatic activity of rhADPGKΔ50 as a function of increasing concentration of D-glucose at a constant concanetration of 1 mM Mg-ADP. The K_m^{app} for glucose was 0.37 mM. $R^2 = 0.95$.

3.3.6 Determination of kinetic parameters for rzADPGK

The samples of rzADPGK could only be obtained as crude elution from IMAC purification as described in section 3.2.10. The elution fractions showing ADPGK activity were pooled and used in activity assays without further purification or transfer into a storage buffer. Enzymatic assays with rzADPGK were performed at 37 ˚C for comparison to ADPGK from mouse and human. However, zebrafish have a lower body temperature, depending on the water temperature which is typically maintained at around 28 ˚C (Gronquist and Berges 2013). The V_{max}^{app} of rzADPGK was 0.005 U/mg. To determine the pH sensitivity of the catalytic rate of rzADPGK, the enzyme was found to have the highest activity at a pH of 6.75, with 78%, 80%, 37% and 7% of activity remaining at pH 6.0, 7.0, 8.0 and 9.0, respectively. The pH was tested in steps of 0.25 or 0.5 from pH 6.0 to 9.0 in Bis-tris propane buffer (shown in Figure 55). Other kinetic parameters were not determined due to the limited availability of rzADPGK enzyme.

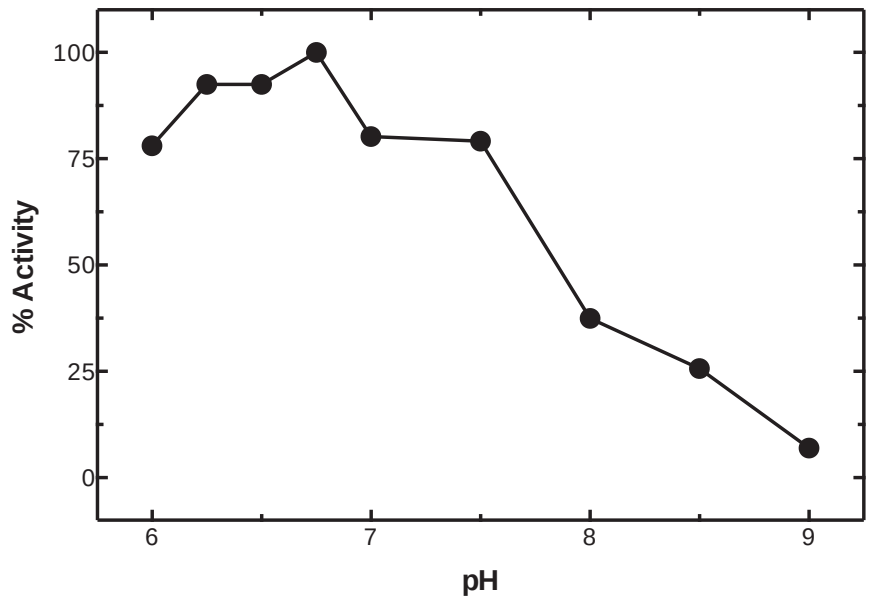

Figure 55: Determination of dependence of enzymatic activity on pH for rzADPGK.

The plot shows the dependence of enzymatic activity of rzADPGK of the pH. The enzyme shows maximum activity at pH 6.75. The pH was varied in 0.25 or 0.5 steps in 50 mM Bis-tris propane buffer and activity measured with 0.5 mM D-glucose and 1 mM ADP.

3.3.7 Summary

The kinetic parameters for variants of ADPGK determined in this work, including functional mutants, are summarised in Table 3.4. In a few cases, activity with ATP has been reported in the literature for ADP-dependent enzymes, e.g. for the ADPPFK from *P. furiosus* (Guixé and Merino 2009). However, this is likely due to ADP contamination in the ATP solutions used. No such activity with ATP was observed for rhADPGK that would suggest contamination of the ATP solutions. Overall, the enzymatic characterisation carried out in this work could only determine the apparent values for the Michaelis constant and velocity. The limited availability of the human ADPGK, especially the case for most functional mutants, was prohibitive to a more thorough kinetic analysis. In Table 3.4 the values for K_m^{app}, V_{max}^{app}, k_{cat} and ${k_{cat}}/{K_m}$ are reported. However, mostly the values V_{max}^{app} and K_m^{app} will be used for comparison to other ADPGKs, as most results were reported as these.

Table 3.4: Overview of the kinetic parameters of ADPGK and functional mutants.

Variant	V_{max}^{app}	k_{cat}	Mg-ADP		D-glucose	
			K_m^{app}	k_{cat}/K_m	K_m^{app}	k_{cat}/K_m
rhADPGKΔ50	32	28.1	0.19	1.5×10^8	0.37	7.6×10^7
ATP instead of ADP	No activity detectable.					
rhADPGKΔ50-D481A	No activity detectable.					
rhADPGK	30	29.5	0.48	6.1×10^7	0.56	5.2×10^7
CDP instead of ADP	0.049	-	-	-	-	-
GDP instead of ADP	1.46	1.4	2.93	4.9×10^5	-	-
rhADPGK Δ151	No activity detectable.					
rhADPGK D84A	0.00043	4.2×10^{-4}	-	-	-	-
rhADPGK E130A	Not tested or purified.					
rhADPGK R228A	0.31	0.3	0.27	1.1×10^6	0.23	1.3×10^6
rhADPGK H264A	0.0051	0.005	0.61	8.2×10^3	5.04	9.9×10^2
rhADPGK H382A	Does not express in *E. coli*.-					
rhADPGK H382V:H387V	No activity detectable. No altered nucleotide specificity.					
rhADPGK H387A	No activity detectable-					
rhADPGK D481A	0.15	0.15	-	-	-	-
rhADPGK D481E	No activity detectable.					
rhADPGK D481N	0.017	0.017	-	-	-	-
rhADPGK D481S	Purified but not tested for activity.					

3.4 Crystal structure determination of mouse ADPGK

3.4.1 Introduction

Prior to this work, a total of five structures of ADPGK enzymes had been reported (Ito et al. 2001, 2003; Tsuge et al. 2002; Rivas-Pardo et al. 2013), all of which are of archaeal origin. The ADPGK enzymes of *P. furiosus*, *T. litoralis* and *P. horikoshii* were crystallised in different ligand bound states and conformations, having their small and large domains at different angles towards each other, often referred to as the open, half-open or closed conformation. Sometimes binding of a protein and ligand can alter its conformation and thus improve, or diminish, the chances of crystallising a given protein sample. Given that the sequence identity of these archaeal proteins to their eukaryotic counterparts is less than 20%, the crystallisation conditions for the archaeal proteins were not necessarily applicable here. In this chapter, the crystallisation and structure determination of *M. musculus* ADPGK will be described.

3.4.2 Crystallisation of ADPGK

Protein samples for crystallisation were prepared as described in section 2.6.1, ideally ensuring that ADPGK was of a purity of over 90% as assessed by SDS-PAGE. Potential aggregates were removed by filtration or centrifugation directly prior to crystallisation. A large number of crystallisation trials were set up for ADPGK. The strategy used for this work was mostly facilitated by access to a Mosquito pipetting robot (TTP Labtech) enabling the quick set-up of large throughput crystallisation screens in 96-well plates. The screening strategy chosen was based on the commercial sparse matrix screens (Jancarick and Kim 1991). Four variants of ADPGK could be purified in large enough quantities to allow crystallisation screening, rhADPGK, rhADPGKΔ50, rmADPGK and rmADPGKΔ51. In addition to different protein constructs, attempts were made to improve the chance of successful crystallisation by varying incubation temperature (4 ˚C and 21 ˚C), different protein concentrations (between 2 and 20 mg/ml), co-crystallisation of a ligand or substrate, reductive methylation (Walter et al. 2006; Shaw et al. 2007), limited proteolysis (Wernimont and Edwards 2009) and using different commercially available crystallisation screens (see section 2.1.7). Limited proteolysis can improve the crystallisation behaviour of a protein by producing smaller, less flexible, fragments through proteolytic digestion.

The reagents for limited proteolysis were purchased from Jena Bioscience (Germany) as the Floppy Choppy kit and samples prepared as described in section 2.6.2. Pre-screening of suitable conditions was performed prior to setting up the crystallisation trials. Based on outcomes of earlier crystallisation screens the concentration of rmADPGK was adjusted to 9 mg/ml and incubated with α-chymotrypsin, trypsin, subtilisin and papain, each in 1/10, 1/100 and 1/1000 dilutions of the protease. Samples were analysed by SDS-PAGE after 30 minutes incubation. As shown in Figure 56, at a dilution factor of 1/10, subtilisin and papain digested the target protein completely within 30 minutes, and were therefore deemed not suitable for this experiment. Trypsin and α-chymotrypsin were chosen as more suitable proteases since they seemed to have only limited access to cleavage sites of ADPGK, leading to fewer degradation products. Based on the results of the pre-screening these two proteases were used at a 1/200 dilution for the crystallisation screens.

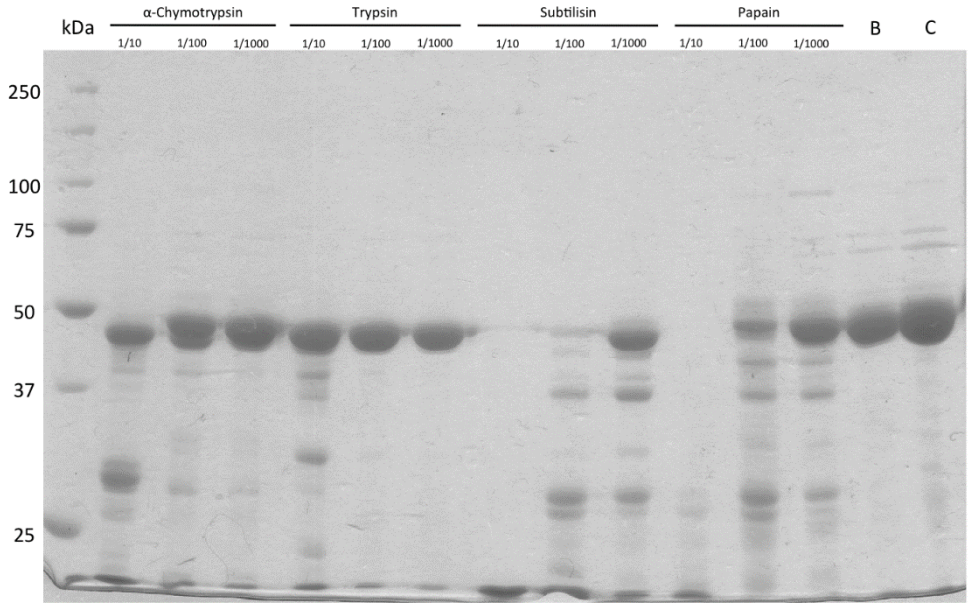

Figure 56: Limited proteolysis of mouse ADPGK.

Proteolysis fragment screening of mouse ADPGK using the Floppy Choppy kit (Jena Bioscience). 1 μg of protein was loaded per lane. The gels show samples of ADPGK incubated with α-chymotrypsin, trypsin, subtilisin and papain, each in 1/10, 1/100 and 1/1000 dilutions of the protease (labelled accordingly). Lane B is the control reaction of ADPGK incubated with proteolysis buffer only. Lane C is untreated protein.

Another form of protein modification to improve crystallisation chances is reductive methylation. With this method, solvent accessible lysine residues and the N-terminus will be methylated (Walter et al. 2006; Shaw et al. 2007). This will affect protein solubility by altering the isoelectric point of the protein slightly. Also, the protein will become slightly more hydrophobic, which can allow for previously unavailable hydrophobic contacts to be formed upon crystallisation. Mouse ADPGK was purified by size-exclusion chromatography as described in section 2.5.10 and reductive methylation was performed as described in section 2.6.3. Native blue PAGE confirmed the methylation of the product after further purification by size-exclusion chromatography as can be seen in Figure 57.

Figure 57: Reductive methylation of mouse ADPGK.
The native blue PAGE gel shows mouse ADPGK non-methylated **(1)** and methylated **(2)**. Due to the change in the proteins pI caused by the methylation of lysines, the methylated protein migrates faster in the gel during electrophoresis.

Methylation of the surface lysine residues increases the overall negative charge of the protein at the given pH for electrophoresis. This leads to a faster migration of the methylated sample when compared to the non-methylated control sample during electrophoresis (see Figure 57). The added molecular weight of 28 Da per methylated lysine appears to be negligible in this case in decreasing migration in the gel. The methylated as well as the non-methylated mouse ADPGK proteins were used to set up

crystallization screens in 96-well plate format. While confirmation of the successful methylation by mass spectrometry is recommended (Walter et al. 2006; Shaw et al. 2007), this was not possible due to technical difficulties at the time of the experiment.

Co-crystallisation of ADPGK and its substrates and products was attempted with D-glucose (10 mM), D-glucose-6-phosphate (10 mM), AMP (10 mM), $MgCl_2$ and ADP together (6 mM each), non-hydrolysable ADP-analog AMPCP (5 mM) and D-glucose, $MgCl_2$ and ADPGK together (5 mM each). In the case of co-crystallisation with glucose small, round crystals were obtained in many conditions, including the control experiment without protein, which indicates that the small crystals were likely to be glucose. Also X-ray diffraction patterns collected from these crystals showed the lowest resolution spots at around 5 Å and a large spot separation, indicating crystals with unit cell dimensions too small to contain protein. None of the above conditions tested led to diffracting protein crystals.

Initial crystals of ADPGK could only be obtained for construct rmADPGKΔ51, a truncated version of mouse ADPGK, missing the first 51 N-terminal amino acids. The purification of rmADPGKΔ51 is described in section 3.2.9 above. The initial hits of needle-type crystals could be seen in a 96-well plate set up with the Mosquito robot as sitting drops with 400 nl total drop volume and a screen set up manually in 24-well plates as hanging drops with 2 µl drop volume from a protein solution at a concentration of 9 mg/ml after an incubation time of 2 weeks at 21 ˚C in condition A9 of the JCSG+ HT-96 screen (Molecular Dimensions) and condition D10 of the Structure Screen I + II HT-96 (Molecular Dimensions). Condition A9 of the JCSG+ screen contains 0.2 M NH_4Cl, 20% (w/v) PEG 3350. Condition D10 of the Structure Screen contains 0.05 M potassium dihydrogen phosphate and 20% (w/v) PEG 8000. Improved crystals of rmADPGKΔ51 were obtained from optimisation when combining the initially successful condition with the Silver Bullet Bio screen (Hampton Research) as an additive screen. Improved crystals were found in conditions A1 and H2. Crystals from both screens are shown in Figure 58. Condition A1 of the Silver Bullets Bio screen contained 0.16% (w/v) L-citrulline, 0.16% (w/v) L-ornithine hydrochloride, 0.16% (w/v) urea, 0.16% (w/v) oxalic acid, 0.16% (w/v) kanamycin monosulfate, 0.16% (w/v) L-arginine and 0.02 M HEPES sodium pH 6.8. Condition H2 contained 0.16% (w/v) adenosine 5′-triphosphate disodium salt hydrate,

0.16% (w/v) pyridoxal 5-phosphate monohydrate, 0.16% (w/v) creatine monohydrate, 0.16% (w/v) thymine, 0.16% (w/v) L-malic acid sodium salt, 0.16% (w/v) spermine and 0.02 M HEPES sodium pH 6.8. The optimised crystal grown in condition A1, as shown in Figure 58, was suitable for X-ray diffraction and taken to the Australian synchrotron for data collection.

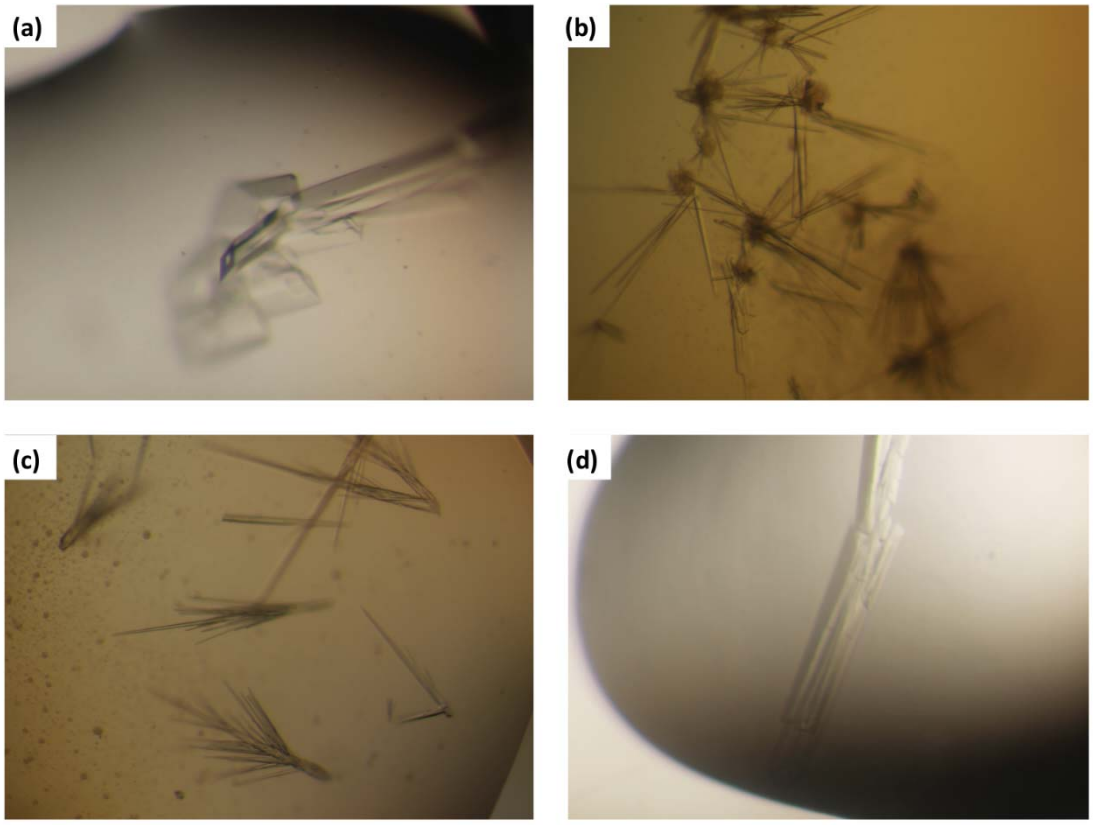

Figure 58: Crystals of ADPGK.

Crystals of rmADPGKΔ51. **(a)** Crystals of rmADPGKΔ51 grown in condition A1 of the Silver Bullets Bio screen. **(b)** Non-optimised crystals of rmADPGKΔ51 grown in condition A9 of the JCSG+ screen. **(c)** Non-optimised crystals of rmADPGKΔ51 grown in condition D10 of the JCSG+ screen. **(d)** Crystals of rmADPGKΔ51 grown in condition H2 of the Silver Bullets Bio screen.

3.4.3 Data collection and indexing

Data were collected at the Australian Synchrotron from a single crystal grown from rmADPGKΔ51 in condition A1 of the Silver Bullets Bio for optimising condition A9 of the JCSG+ screen (described in section 3.4.2 above). The crystals were scooped in a nylon loop, briefly incubated in perfluoropolyether cryo-oil as cryoprotectant and

flash-frozen in liquid nitrogen. Data collection and indexing was performed by Dr. Ron Ronimus and Dr. Vince Carbone at the Australian Synchrotron. In total, 240 images were collected at the MX-2 beamline at 100 K with an oscillation angle of 0.5˚ over 120˚ oscillation range at a wavelength of 0.9184 Å. Two separate datasets were collected with varied exposure times for optimal collection the low- and high-resolution data. During the course of the data collection, cracks in the crystal became more and more evident over time. While the perfluoropolyether oil used was apparently effective as a cryoprotectant, future experiments could include optimising the cryoprotectant used for ADPGK crystals (personal communication, Dr. Vince Carbone and Dr. Ron Ronimus).

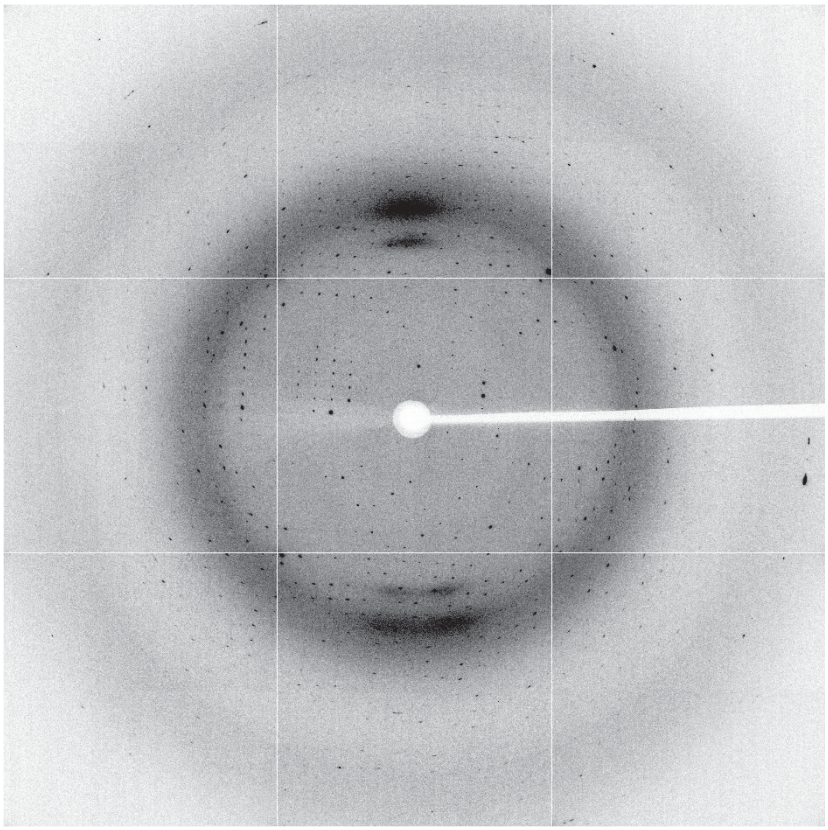

Figure 59: X-ray diffraction image of the ADPGK crystal.

X-ray diffraction image of a rmADPGKΔ51 crystal grown in condition A1 of the Silver Bullet Bio screen (Hampton Research) from a data set collected at the Australian synchrotron. The space group was orthorhombic $P\,2_12_12_1$.

The images were indexed with the program XDS (Kabsch 2010) and scaled and averaged with the program SCALA (Evans 2006). The dataset statistics are shown in Table 3.5. The protein had crystallised in the orthorhombic space group $P\ 2_12_12_1$, which was derived from the occurrence of the systematic absences typical for this space group. The space group was then later confirmed by the molecular replacement solution. The unit cell dimensions were a = 45.92 Å; b = 58.68 Å; c = 160.94 Å; and $\alpha = \beta = \gamma = 90^\circ$.

Table 3.5: Dataset statistics for the rmADPGKΔ51 dataset.

Parameter	rmADPGKΔ51 dataset
Wavelength	0.9184 Å
Distance to detector	315 mm
Resolution range	50 – 2.1 Å
Space group	$P\ 2_12_12_1$
Unit cell dimensions	a = 45.92 Å b = 58.68 Å c = 160.94 Å $\alpha = \beta = \gamma = 90^\circ$
Total reflections	126165
Unique reflections	26240
Multiplicity	4.8
Completeness	99.4% (98.7% highest resolution shell)
R_{merge}	11.6% (67.3%)
Mean I/σ	12.1 (2.2)

3.4.4 Phasing and refinement of the *M. musculus* ADPGK structure

The numbering of the residues in the structure of *M. musculus* ADPGK, derived from the rmADPGKΔ51 construct, in the following sections will be relative to the positions in the native full-length protein to avoid confusion in sequence alignments or structural superpositions. Molecular replacement was used to obtain the phase information required to solve the structure of the *Mus musclus* ADPGK. This was done using the program Phaser (McCoy et al. 2007) as described in section 2.6.5. The solvent content of the crystal was 40.8%, with one molecule in the asymmetric unit.

The Matthews' coefficient was V_M of 2.3 Å^3/Da (Matthews 1968). No structure of a eukaryotic ADPGK was solved prior to this work. Therefore, the structures of archaeal ADPGK were chosen as search models for the molecular replacement, as these were the most closely related based on sequence alignments. These models included the crystal structures of the ADPGKs from *T. litoralis* (PDB 1GC5), *P. horikoshii* (PDB 1L2L) and *P. furiosus* (PDB 1UA4) and, due to its structural similarity to the archaeal ADPGK, the ADPPFK from *P. horikoshii* OT3 (PDB 1U2X and 2DRW). Also, homology models for *M. musculus* ADPGK, generated with Phyre2 (Kelley and Sternberg 2009) and I-TASSER (Roy et al. 2010), were tested as search models. The same models were used for design of the functional mutants and truncations as described in section 3.1. Structures of ribokinase and adenosine kinase were not considered suitable candidates. This is mainly because the small domain is mostly missing and the connectivity of secondary structure elements in the core-fold is very different in parts of the C-terminal region. The search models were modified with Ensembler and Chainsaw from Phenix (Adams et al. 2010) and CCP4 (Winn et al. 2011). In this process, side chains of non-conserved amino acids and inserted/deleted loop regions were removed. Alternatively, the whole search model was also reduced to a poly-alanine model. Because a definitive solution for the molecular replacement with the complete structure was not immediately found, the small and large domains were used separately as search models. No solutions could be found with search models, truncated or not, which were based on the homology models of *M. musculus* or *H. sapiens* ADPGK. A solution was found first with a search model based on the ADPGK from *T. litoralis* (PDB 1GC5). Attempts at using Phenix Autobuild with the initial molecular replacement solutions were not successful and the rebuilding had to be performed manually. Representative steps of the progression of the molecular replacement solution are shown in Figure 60.

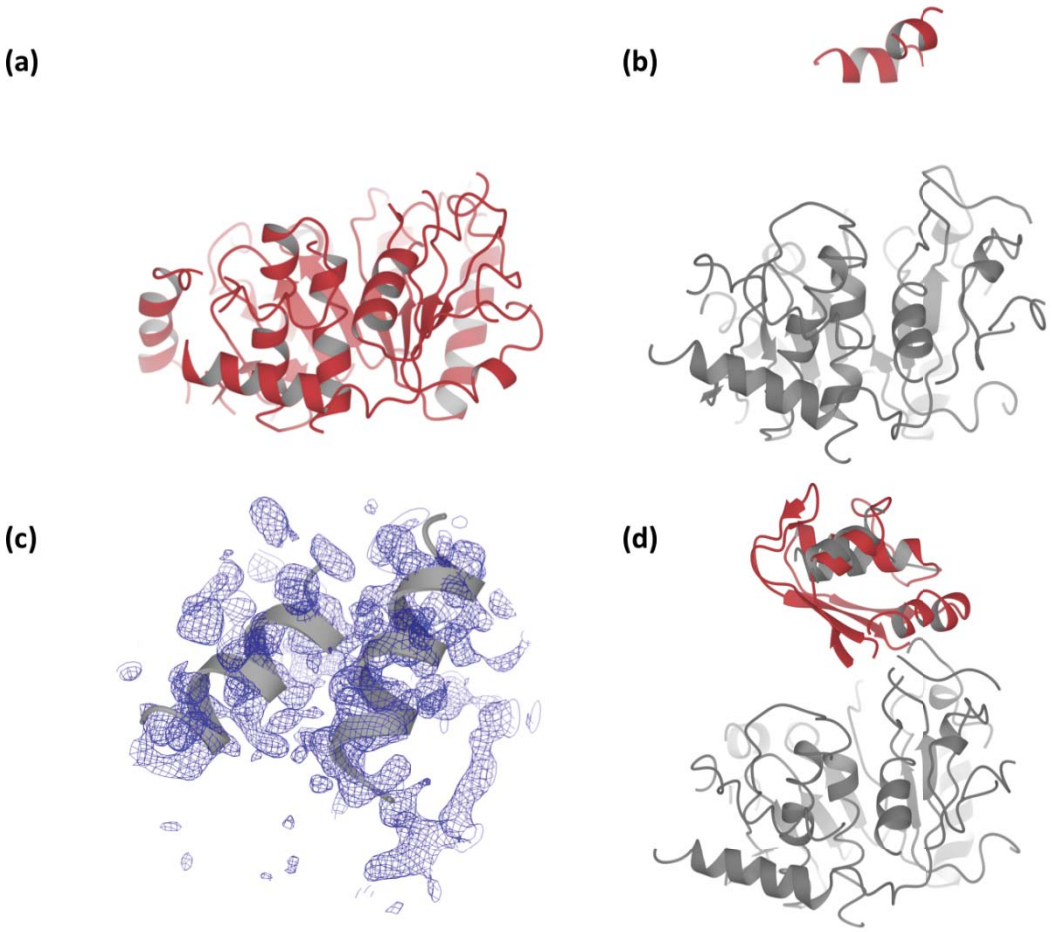

Figure 60: Molecular replacement with Phaser.

Stepwise building a molecular replacement solution with Phaser and Coot. The protein backbone is shown as ribbon representation, the side chains are not displayed for clarity. **(a)** The truncated large domain (shown in red) of *T. litoralis* ADPGK (PDB 1GC5) was used as the initial search model. **(b)** The search model was modified in Coot and two helices (shown in red) of the small domain could be built into the electron density. **(c)** The two helices, building the 'core' of the small domain, were verified in Phaser with additional electron density (shown in blue) becoming visible. **(d)** The truncated lid domain (shown in red) of *T. litoralis* ADPGK (PDB 1GC5) could be superimposed on the two helices in Coot and used for the subsequent refinement steps. The images were generated with CCP4 molecular graphics. Electron density is displayed at 1.0 σ.

The large subdomain fragment of 1GC5 had a Z-score of 7.4 for the translation search and after refinement a Z-score of 13.3 for the translation search and a log-likelihood-gain of 336. The model is shown in Figure 60, panel (a). A Z-score above 8 indicates a certain solution for the molecular replacement (McCoy et al. 2007). However, the initial molecular replacement solutions did not refine in REFMAC or Phenix, with R_{free} higher than 0.52. The model was then manually rebuilt and refined in Coot

(Emsley et al. 2010) using real space refinement, which was done mostly by deleting parts of the model which were not explained by the electron density map. The model was run back through Phaser to validate the improvements in the model by assessing map quality. With the improved electron density maps obtained from the repeated Phaser runs, it was then possible to build two helices in the model, which are part of the small domain. The molecular replacement solution and the two additional helices are shown in Figure 60, panel (b) and (c). It was then possible to superimpose a poly-alanine model of the small domain of *T. litoralis* ADPGK (PDB 1GC5), generated with Ensembler with the two helices in the small domain, using Coot, which is shown in Figure 60, panel (d). The model containing the poly-alanine small domain fragment could then be used to successively rebuild the structure in Coot and running the improved model back through Phaser. The quality of the molecular replacement solutions was assessed on the basis of the quality of the electron density maps and the improvement of the Z-score for the translation search in Phaser. However, during this rebuilding process, emphasis was on the improved electron density maps. Large amino acid side chains were used to locate sequence regions in the electron density maps as sequence alignments proved unreliable for the small domain. The unexpected observation of the juxtaposition of two cysteines forming a disulphide bridge, the interpretable density for small region of the vector tag at the N-terminus and other mouse ADPGK sequence insertions, not in any search model, confirmed the correct structure solution. The structure of rmADPGKΔ51 was then refined using Phenix and REFMAC5 (see section 2.6.5) and the R_{free} for the whole model decreased from 0.47 to 0.34 in less than ten cycles of manual rebuilding in Coot and refinement. For the refinement process, 5% of the reflections were assigned to a test set to calculate R_{free} (Brunger 1992). The refinement statistics for the final model of mouse ADPGK are summarised in Table 3.6. The model was validated using the Molprobity (Davis et al. 2007; Chen et al. 2010) validation tools of Phenix (Adams et al. 2010).

Table 3.6: Crystallographic data.

Parameter	rmADPGKΔ51
R_{work}	0.21
R_{free}	0.26
Resolution	47.42 – 2.10 Å
$RMSD_{bonds}$	0.0153 Å
$RMSD_{angles}$	1.62°
Average B-factor	32.1 $Å^2$
Ramachandran favored	95.8%
Ramachandran allowed	4.2%
Ramachandran outliers	0%
Clashscore	3.67

The structure of mouse ADPGK was refined to an R_{work} and R_{free} of 0.21 and 0.26, respectively. The geometry of the model was validated by analysis of the φ and ψ angles of the protein backbone (Ramachandran et al. 1963) and the χ angles of the residue side chains. The Ramachandran plot showed most residues in the favoured (95.8%) and the rest in generally allowed (4.2%) regions. No outliers were seen in the Ramachandran plot (see Figure 61).

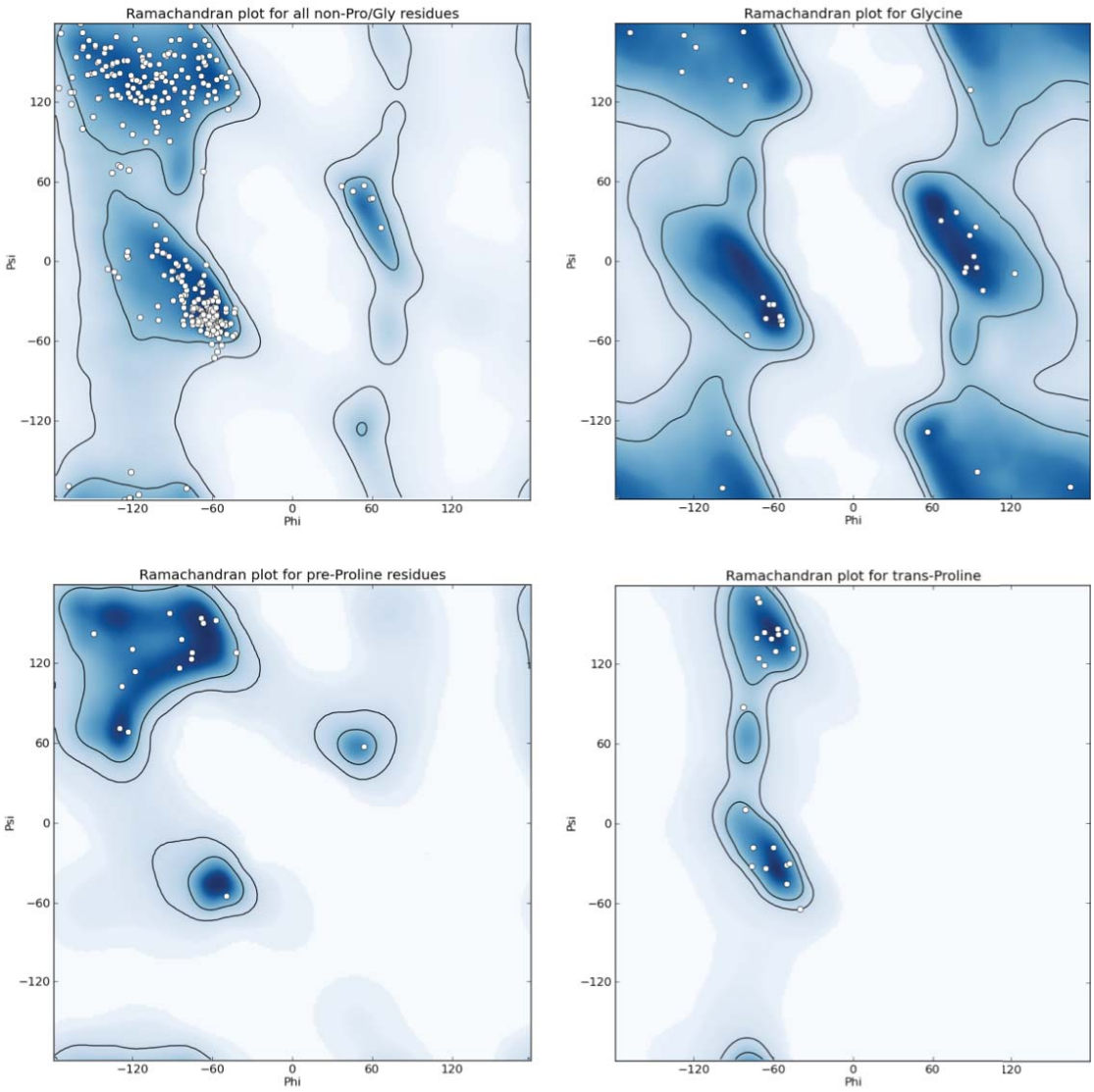

Figure 61 Ramachandran plots for rmADPGKΔ51.

The Ramachandran plots for the rmADPGKΔ51 structure were generated with the Molprobity tools in Phenix. The Ramachandran favoured and allowed regions are coloured blue and light blue, respectively. Residues are represented as white circles. Ramachandran plots from top left clockwise are: plot for all residues except glycine and proline, plot for all glycine residues, plot for all trans-proline residues, plot for all pre-proline residues. Of all residues, 95.8% are in the favoured regions. No outliers were found.

During the refinement process waters were automatically placed with Coot with F_O-F_C map peaks greater than 2.8 σ, $2F_O$-F_C map peaks greater than 1.8 σ and a distance between 2.4 Å and 3.2 Å to suitable protein atoms for hydrogen bonding. In subsequent refinement steps, the waters were manually validated. The finished structure contained 296 water molecules. No other ligands or ions could be identified

unambiguously in the electron density maps. The overall B-factor calculated by Molprobity was 32.1 Å^2. The content of the unit cell for the solved crystal structure is shown in Figure 62.

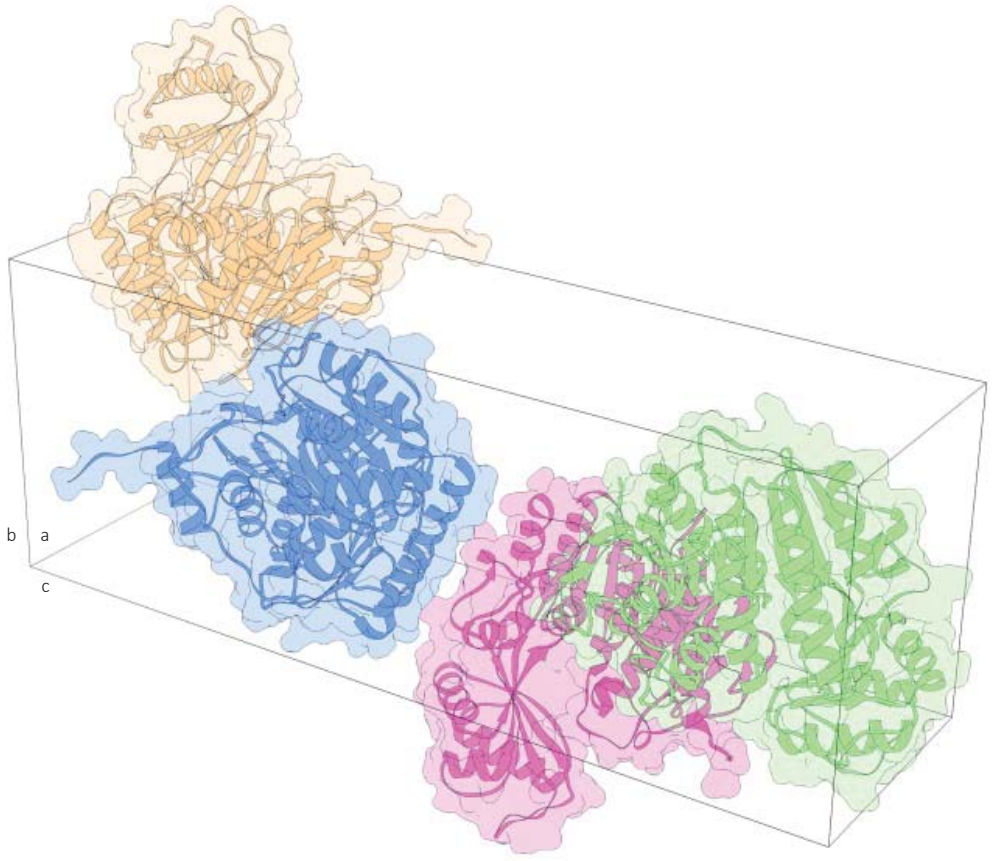

Figure 62: Contents of one unit cell.

The figure shows the content of the unit cell for the rmADPGKΔ51 crystal. The space group was *P* $2_12_12_1$.The unit cell dimensions were a = 45.92 Å, b = 58.68 and Å c = 160.94 Å, with all angles 90°.

The lid domain and the central part of the large domain have lower overall B-factors (see Figure 63). The highest B-factors are found in the C-terminal part of the large domain. The closing of the lid domain is essential for the catalytic mechanism of the enzyme and one could expect an intrinsic flexibility in the hinge region. However, this is clearly not the case. The rigidity of the hinge region is probably required for stable crystals to form under these conditions as the lid domain does form crystal contacts in the lattice.

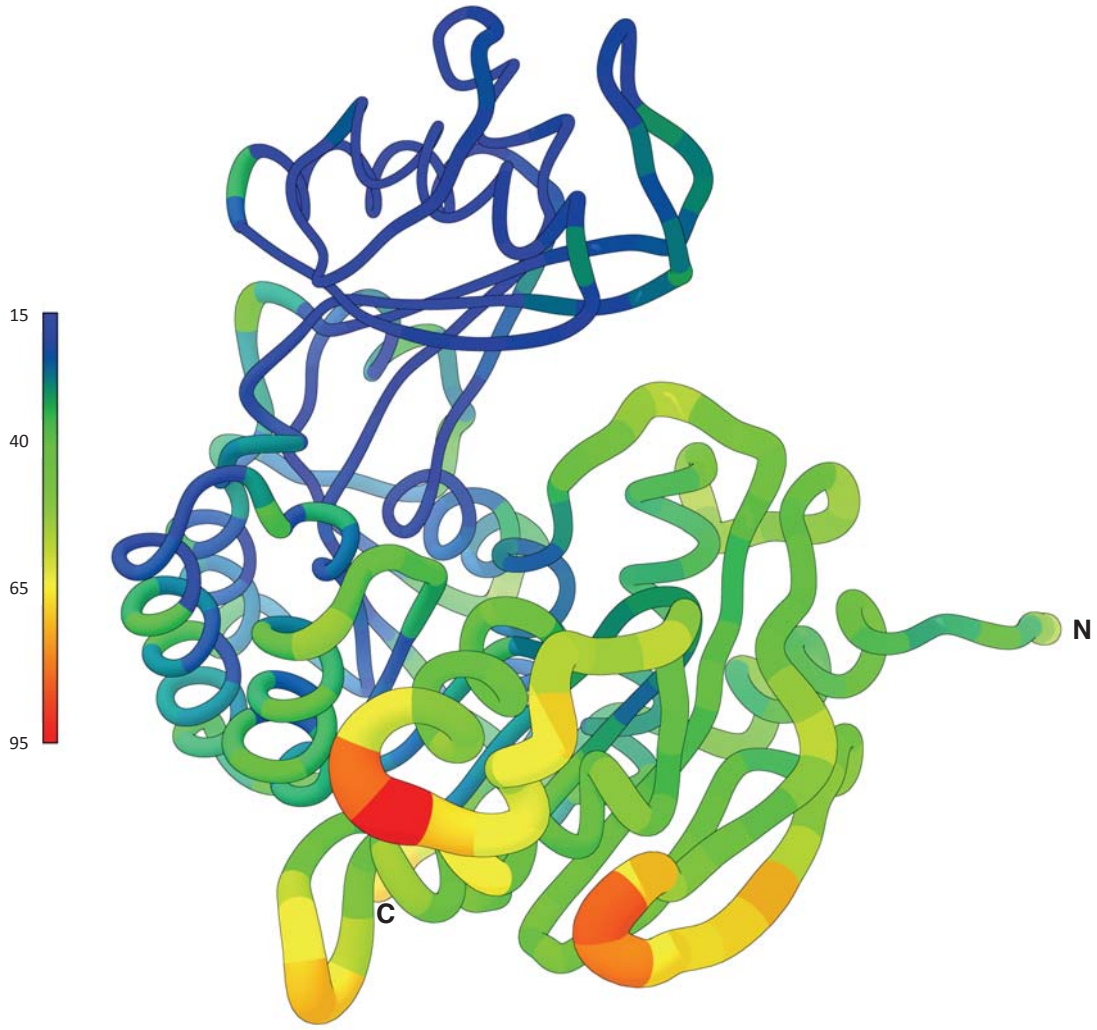

Figure 63: B-factor analysis of the *M. musculus* ADPGK structure.

Ribbon representation of the rmADPGKΔ51 backbone with colour key representing the average temperature factors for each residue in $Å^2$ (to the left). The N- and C-termini of the protein are indicated. The lid domain and N-terminal core part of the large domain have overall lower B-factors. The highest B-factors are seen on the surface of large domain towards the C-terminus. The image was generated using Chimera (Pettersen et al. 2004).

No density was observed for the side chain hydroxyl group of serine residue S308 and therefore the side chain of this residue was cut off at the β-carbon position, deleting the hydroxyl group. Of the 486 possible residues, 449 were explained in the electron density map, while 9 residues were missing at the N-terminus and 28 were missing at the C-terminus. The missing residues were vector encoded, including the N-terminal start codon with following enterokinase site and the C-terminal region with V5 epitope, 6×His-tag and linker regions. Also, a disulphide bond has been found in the structure of mouse ADPGK, which is formed between C469, located at the N-terminal

end of the nucleotide-binding loop, and C414. The cysteine bond is well defined by the experimental data, as can be seen in Figure 64.

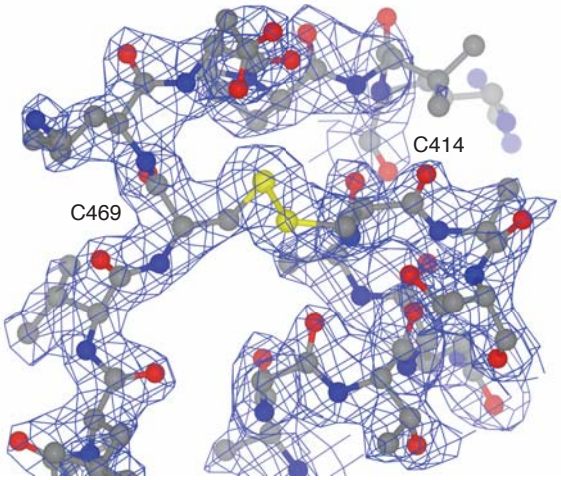

Figure 64: Electron density around the disulphide bond.

Image depicting the electron density around the disulphide bond found in mouse ADPGK. The electron density is displayed at a contour level of 1.0 σ. The bond is formed between cysteine residues number 469 (on the left) and 414 (on the right). The electron density around the disulphide bond is well defined.

The truncated variant rmADPGKΔ51 was found to be monomeric in the crystal structure, with contacts between molecules likely being crystallisation contacts, and not a biological assembly. To validate this observation, the final model for rmADPGKΔ51 was analysed with the PISA service at the European Bioinformatics Institute (Krissinel and Henrick 2007). The protein was classified as not forming complexes in solution by PISA. For rmADPGKΔ51, five crystal contact interfaces were identified, which are detailed in Table 3.7.

Table 3.7: Crystal contact analysis with PISA.

Interface	Contact area in $Å^2$	ΔG in kcal M^{-1}	Hydrogen bonds	Salt bridges	Disulphide
1	738.9	-4.1	14	4	No
2	447.3	-1.2	7	0	No
3	318.9	1.3	4	1	No
4	197.2	2.5	3	2	No
5	166.7	1.9	0	0	No

An overview over the crystal contacts as determined by PISA in the rmADPGKΔ51 crystal is shown in Figure 65. The surface area of rmADPGKΔ51 in total was 19,349.3 $Å^2$ as determined by PISA. The largest contact area of 738.9 $Å^2$ is just below the threshold of 856 $Å^2$ used by PISA for scoring the interactions (Ponstingl et al. 2000). This is supported by the fact that *M. musculus* ADPGK has been shown to be a monomer in solution, with a molecular mass of 60.3 kDa when determined by analytical size-exclusion chromatography (Ronimus and Morgan 2004).

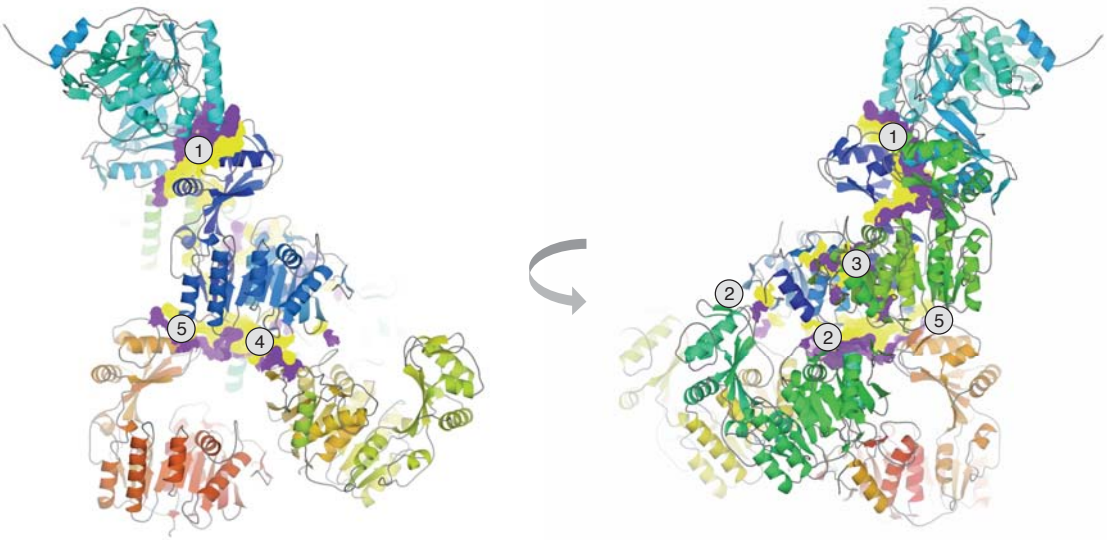

Figure 65: Crystal contacts in the rmADPGKΔ51 crystal.

The figure shows the crystal contacts determined by PISA. Crystal contacts are displayed as yellow (for blue centre molecule) and purple (other ADPGK molecules that contact the blue molecule) surfaces. The contacts are numbered according to Table 3.7.

3.4.5 Structure of mouse ADPGK

An overview of the structure of the mouse ADPGK is shown in Figure 66. The small domain is comprised of a five-stranded β-sheet, flanked by three α-helices on the outward facing side of the domain. The uncovered, inward-facing side of this β-sheet forms the lid part of the active site. The large domain is composed of a central ten-stranded β-sheet flanked by 5 α-helices on each side. The structure of *M. musculus* ADPGK solved in this work was the apoenzyme, with no ligand found in the active site apart from water molecules. The only other ADPGK to date solved in the apo-form is the *P. horikoshii* ADPGK (PDB 1L2L). In the crystal structure of this archaeal ADPGK, several segments of the protein chain were disordered and are therefore missing in the final structure, including the large nucleotide-binding loop. The ADPGKs of *T. litoralis* (PDB 1GC5, 4B8R, 4B8S) and *P. furiosus* (PDB 1UA4) are ligand bound with either AMP or ADP found in the nucleotide binding pocket. In contrast, the large nucleotide-binding loop is well defined in the electron density for *M. musculus* ADPGK. One explanation for this could be the presence of the disulphide bond in the eukaryotic ADPGK, which could help stabilise the otherwise possibly flexible loop in the absence of a bound ligand.

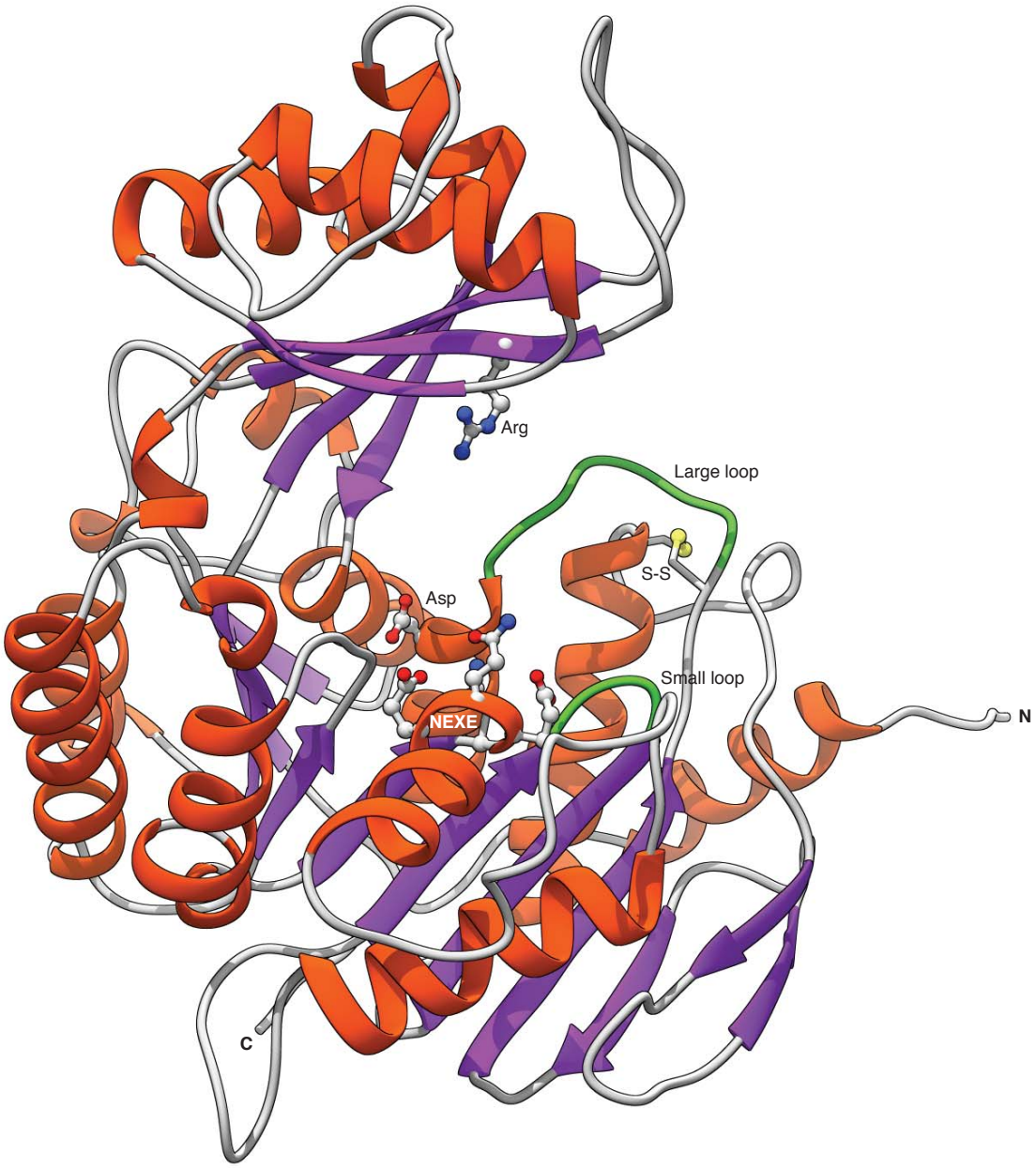

Figure 66: Ribokinase-like fold of *M. musculus* ADPGK.

Ribbon representation of the main chain of *M. musculus* ADPGKΔ51. The N- and C-termini are indicated. Loop and coiled regions are coloured grey, α-helices are coloured orange and β-sheets are coloured purple. The disulphide bond is displayed as spheres with sulphurs in yellow (S-S). The catalytic residues D480 (located on the large domain) and R228 (located on the lid domain), which are facing the active site, are displayed as sticks, coloured by atom type. The large and the small nucleotide-binding loops are coloured green. The residues forming the NEXE motif are displayed as sticks, coloured by atom type (where carbons are grey).

The structure of the *M. musculus* ADPGK has an overall topology very similar to that of other members of the PfkC (or ADP-dependent kinase like) branch of the

ribokinase superfamily of protein folds, with the small lid domain and the large domain. A superposition of mouse ADPGK and a representative archaeal ADPGK (*T. litoralis* ADPGK PDB 4B8R, which was also crystallised in open conformation) is shown in Figure 67. An overview of selected members of the ribokinase-superfamily is shown in Figure 3 in the introduction chapter.

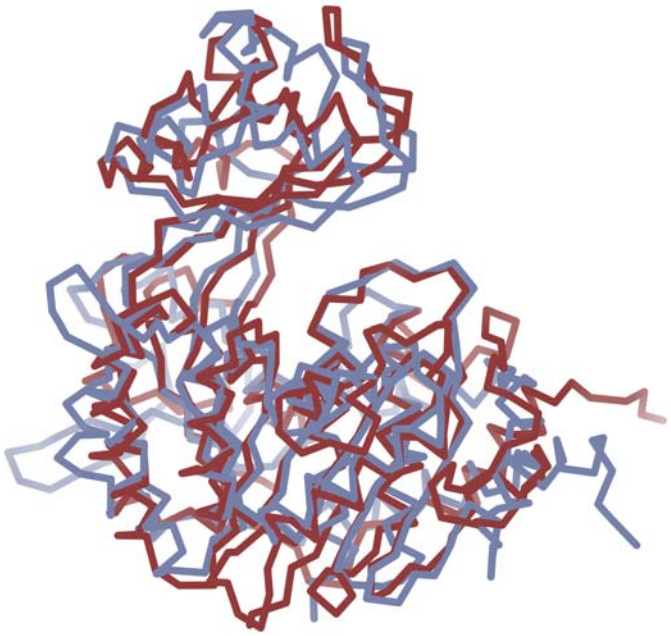

Figure 67: Structural superposition of archaeal and eukaryotic ADPGK.
The figure shows the superposition of rmADPGKΔ51 (red) and *T. litoralis* ADPGK (blue, PDB 4B8R). The core of the large domain is structurally conserved and the overall fold of *M. musculus* ADPGK is very similar to the archaeal ADPGK enzymes.

The crystallised rmADPGKΔ51 was based on the isoform 3 sequence of *M. musculus* ADPGK, as was briefly mentioned in section 2.1.3. The difference between isoform 1 and 3 is an inserted glutamine residue following another glutamine residue at position 313 for isoform 1, resulting in a QQ sequence. The same is true for *H. sapiens* ADPGK, where isoform 1 has the additional glutamine when compared to isoform 2. No additional isoforms are known for *D. rerio* ADPGK, where an EQ sequence is found instead. In Figure 68, a sequence alignment for *D. rerio* ADPGK, *H. sapiens* ADPGK isoform 1 and *M. musculus* ADPGK isoform 3 is shown.

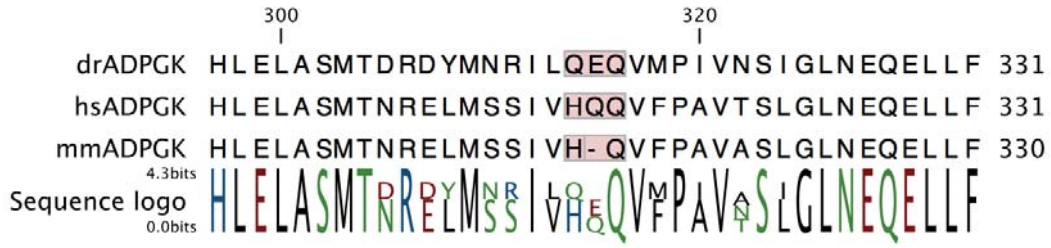

Figure 68: Sequence alignment of the region differing between the isoforms of ADPGK.

The alignment shows a sequence fragment *D. rerio* (dr) ADPGK, *H. sapiens* (hs) ADPGK isoform 1 and *M. musculus* (mm) ADPGK isoform 3. The sequences were aligned with Clustal W (Sievers et al. 2011).

The difference between isoform 1 and 3 of mouse ADPGK, and by homology the isoform 1 and 2 of human ADPGK, affects an area on the surface of the protein, which is shown in Figure 69. The hydrophobic residues preceding (IV) and following (VF) pack against the hydrophobic interface presented by the central β-sheet (e.g. V311 and L298 in Figure 69). An additional residue with a large and polar side chain, like glutamine, is unlikely to be able to replace the valine residue V314 and shift the following sequence by one residue without consequences. This additional residue could result in producing a disordered region, which could affect the proteins stability, catalytic activity or crystallisation behaviour.

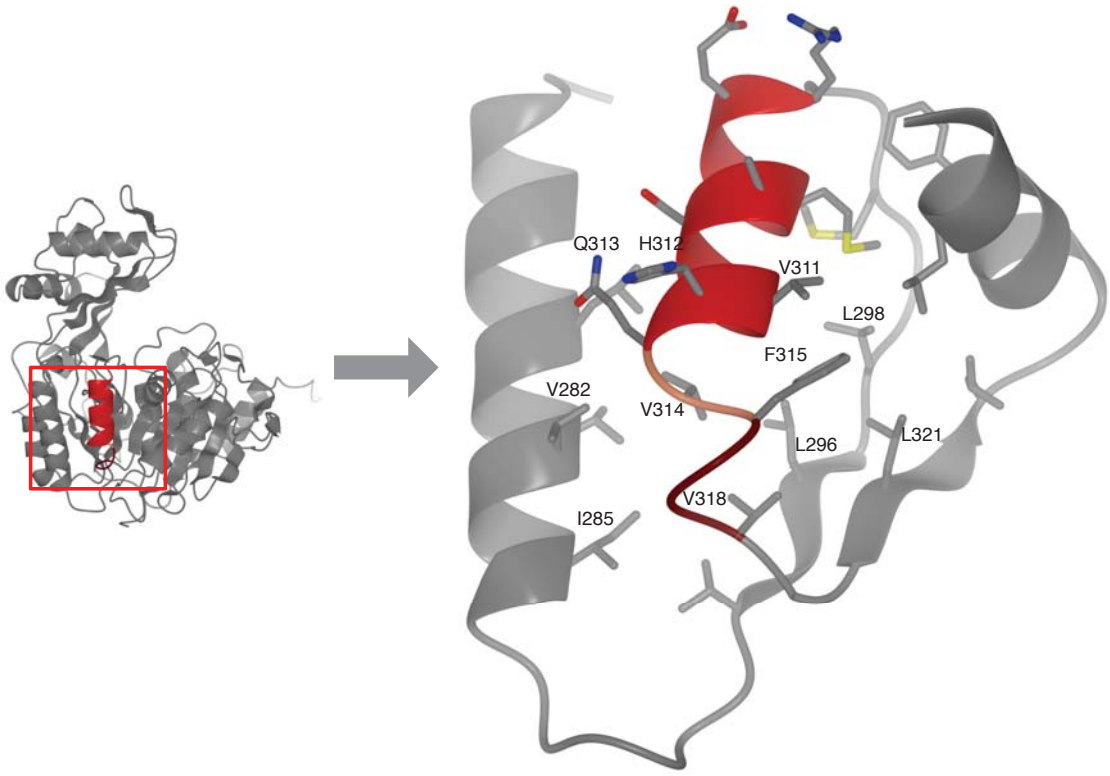

Figure 69: Difference between isoform 1 and 3 of *M. musculus* ADPGK.

The truncated variant rmADPGKΔ51 was derived from isoform 3 of *M. musculus* ADPGK (figure on the left). The additional glutamine residue in isoform 1, Q314 would follow Q313 (located at the bottom of the red helix), falling in the place of V314, possibly interfering with the hydrophobic packing of the inward facing side of the helix against the hydrophobic interior of the central β-sheet (enlarged figure on the right).

3.4.6 The active site of the *M. musculus* ADPGK

Several important functional motifs and features are conserved between the *M. musculus* and archaeal ADPGKs. A structure-based sequence alignment of archaeal and *M. musculus* ADPGK is shown in Figure 70. The conserved motifs shown in the alignment will be discussed in the following sections.

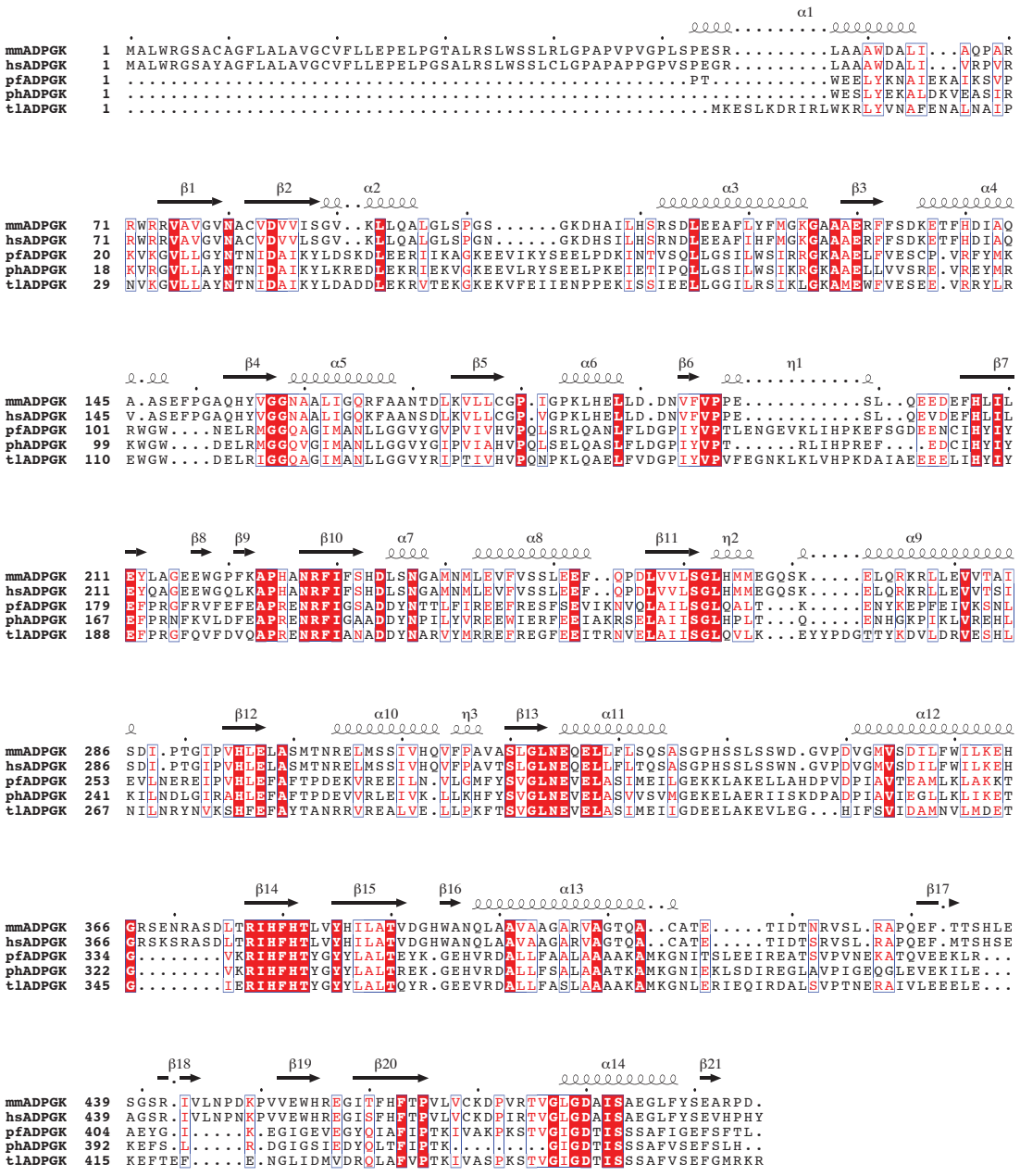

Figure 70: Structure-based sequence alignment

Structure-based sequence alignment of archaeal and *M. musculus* ADPGK. Numbers on residues refer to positions in mouse ADPGK. Secondary structure elements are shown for *M. musculus* ADPGK The alignment was generated using PROMALS3D and the graphical output was generated using ESPript 3.0. The alignment shows the full sequences for the mouse and human proteins.

A distinctive feature of the active site of mouse ADPGK, compared to the rest of the protein's surface, is its negative net charge (see Figure 71). This would be consistent with the requirement of binding of the cation requirement for catalysis. The archaeal ADPGKs of *P. horikoshii* and *T. litoralis*, which have been crystallised in the open conformation, also show a negatively charged ligand-binding site (PDB 1L2L and

4B8R, respectively). However this feature is less pronounced in other members of the ribokinase family, e.g. adenosine kinase or ribokinase (not shown).

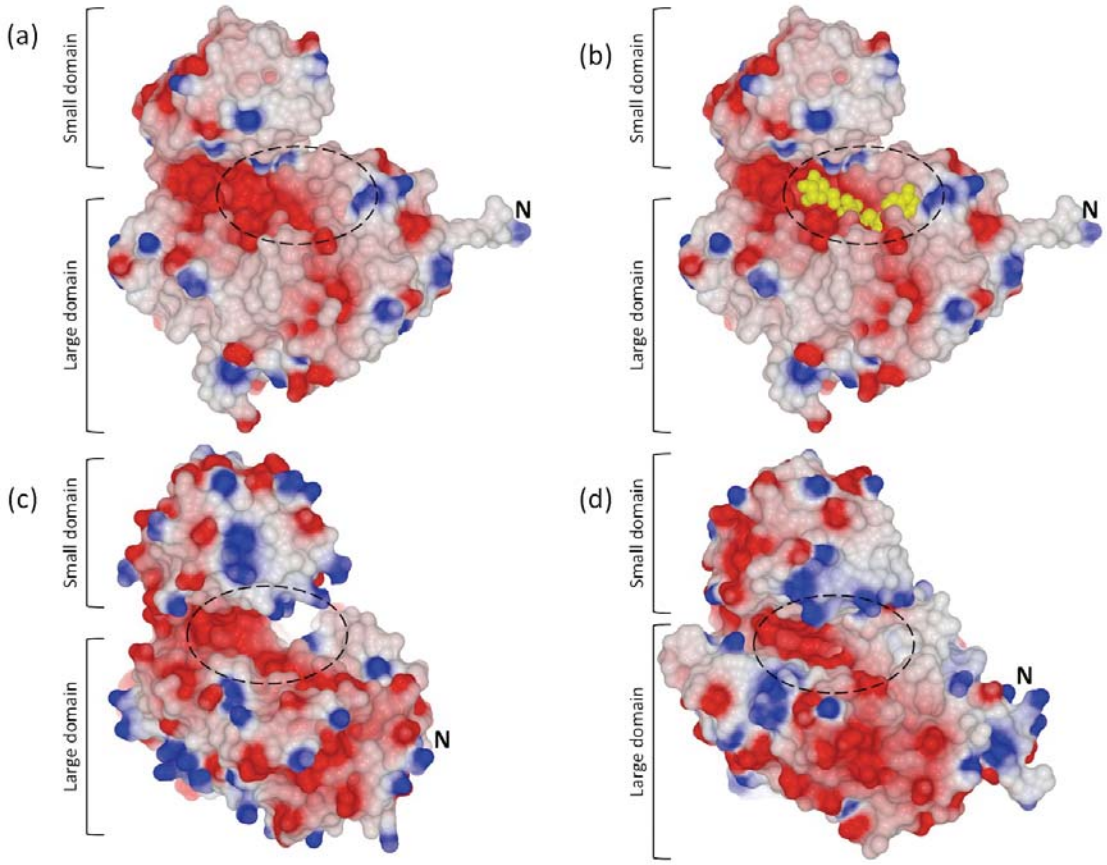

Figure 71: Electrostatic surface of mouse ADPGK

Surface representation with colour gradient from negatively charged (red = -0.5 V) to positively charged (blue = 0.5 V) of ADPGK enzymes. Ligands are coloured yellow and were not included in the electrostatic calculation. The calculation of the electrostatic potential and images were generated with CCP4 molecular graphics software (McNicholas et al. 2011). **(a)** *M. musculus* ADPGK without ligands. **(b)** *M. musculus* ADPGK with ligands. **(c)** *P. horikoshii* ADPGK (PDB 1L2L). **(d)** *T. litoralis* ADPGK (PDB 4B8R). The molecules were superimposed to give comparable viewing angles. The nucleotide-binding pocket is marked with a dashed oval and the approximate position of the N-terminus is marked with '**N**'. The left image shows the apo-form of mouse ADPGK. The right image shows a possible situation for the holoenzyme with glucose and ADP (both yellow) bound to the active site. The ligand molecules were placed from superimposing of *T. litoralis* ADPGK (PDB 1GC5 for ADP position, 4B8S for glucose position) on the structure of *Mus muculus* ADPGK. The nucleotide-binding site appears to have a distinct negative charge compared to the rest of the protein, similar to the archaeal ADPGKs shown.

In Figure 72, the glucose binding pockets of *M. musculus* ADPGK and *T. litoralis* ADPGK (PDB 4B8S) are shown. In the structure of mouse ADPGK, a ring-like arrangement of ordered water molecules was found in the active site. When the structure of glucose bound archaeal ADPGK was superimposed on the mouse ADPGK, the hydroxyl groups of α-D-glucose in positions 1′, 2′, 3′, 4′ and 6′ aligned with waters 156, 159, 86, 163 and 273, respectively. The glucose-bound structure of *T. litoralis* ADPGK (PDB 4B8S) resulted in a closer match of the glucose molecule on the ordered waters than the structure of *P. furiosus* ADPGK (PDB 1UA4), which is why the former is displayed in Figure 72 and further used for comparison. The superposition can be used to infer several potential conserved interactions between certain protein residues and D-glucose. Only α-D-glucose is shown in Figure 72, as the alignment with the ordered water molecules was subjectively better than for β-D-glucose. Asparagine residue N80 interacts with water 156, which aligns with the 1′-hydroxyl group of glucose. This asparagine residue is conserved in *T. litoralis* ADPGK (N38) and the interaction is formed in both cases via the δ-nitrogen of the amino carbonyl group. This interaction will likely only be possible with α-D-glucose, since the 1′-hydroxyl group will be positioned away from the asparagine residue for β-D-glucose. Water 125 in *M. musculus* ADPGK aligns with the carboxyl group of E96 of *T. litoralis* ADPGK. The corresponding residue in *M. musculus* ADPGK E130 is positioned at a greater distance from the binding pocket, as this protein was crystallised in the open conformation. This residue likely interacts with the 2′-hydroxyl group of glucose, which aligns with water 159. A histidine residue H264 forms an interaction with water 173. The histidine is replaced with glutamine Q243 in *T. litoralis* ADPGK. However, both residues form an interaction with a water molecule via the ε-nitrogen of the imidazole or amide carbonyl group, respectively. The 3′-hydroxyl group of glucose overlaps with water 86, which forms interactions with a histidine residue H207 and an aspartic acid residue D84. Both residues are conserved in the *T. litoralis* ADPGK. A cysteine residue C82 is located in the hinge region, with its side chain pointed towards the active site. The way the side chain SH is directed towards the glucose-binding site, it is likely to interact with the 3′-hydroxyl group of the glucose molecule (see Figure 72). This residue is aspartate in the *T. litoralis* ADPGK. The 4′-hydroxyl group of the glucose molecule aligns with water

163. Possible interactions are formed here with aspartic acid residue D84, similar to water 86 and the 3′-hydroxyl group, and the peptide backbone of the GG motif at positions 157 and 158. Water 273 aligns with the 6′-hydroxyl group of glucose, the phosphoryl acceptor site. An interaction is formed with the carboxyl group of aspartic acid D480, which is one of the main catalytic residues identified by sequence alignments in ADPGK enzymes. Another interaction formed by water 273 is a hydrogen bond with water 256.

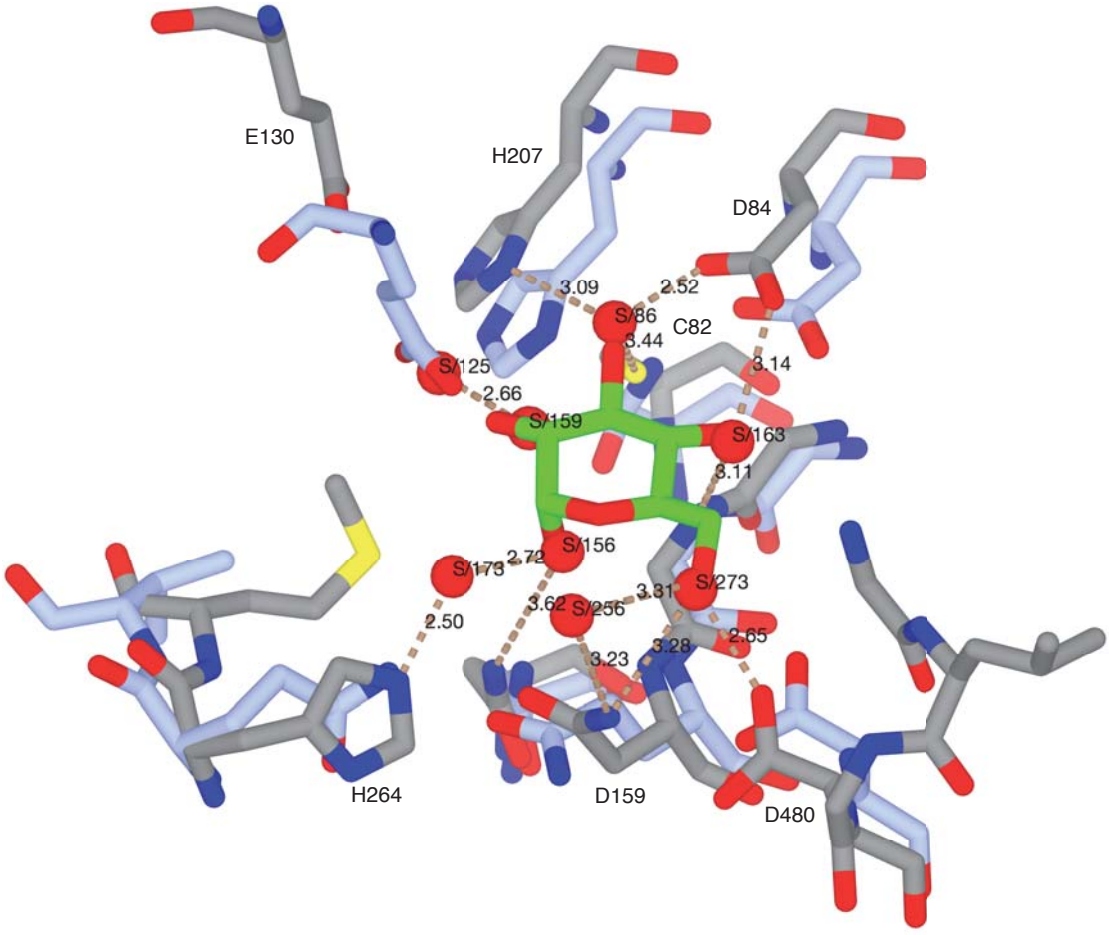

Figure 72: Inferring the binding of α-D-glucose from positions of ordered water molecules.

The image shows a superposition of rmADPGKΔ51 (coloured by atom type with carbon atoms in grey) and the closed, glucose/AMP bound form of *T. litoralis* ADPGK (PDB 4B8S, coloured by atom type with carbon atoms in light blue). Waters are displayed as spheres (red) with labels in black (S/#). The position of α-D-glucose (coloured by atom type with carbon atoms in green) from 4B8S was manually adjusted in Coot after superimposing the structures, to better match the water molecules. Distances are drawn with brown dashed lines with the number indicating distance in Å. All labels refer to positions and actual distances between atoms in rmADPGKΔ51.

The nucleotide-binding site of *M. musculus* ADPGK is shown in Figure 73 as a superposition with the *T. litoralis* ADPGK (PDB 1GC5), the latter being the only structure of an ADP-bound ADPGK available as of yet (Ito et al. 2001). The ADPGK of *M. musculus* was crystallised as apoenzyme, with no ligands visible in the nucleotide-binding site. Distances between atoms shown in Figure 73 are measured between *M. musclus* ADPGK and the ADP molecule of *T. litoralis* ADPGK (PDB 1GC5). Ordered water molecules in the nucleotide-binding pocket were not as obviously suggestive of a relationship to atom positions of the ADP molecule, as was observed for α-D-glucose. The nucleotide-binding site of the mouse ADPGK is highly similar to the archaeal proteins, apart from the disulphide bond. The NEXE motif, proposed to be involved in recognition of the Mg^{2+} ion required for catalysis is conserved in the mouse ADPGK (Ito et al. 2001), as well as the large and small nucleotide-binding loops. Interactions with the nucleotide appear to be conserved when compared to the binding site of *T. litoralis* ADPGK. The nitrogen and oxygen atoms of the peptide backbone in residue C469 seem to be capable of forming hydrogen bonds with the amino group in position 6 and the 1 position of the adenine moiety. These interactions are formed by the peptide backbone of residue V440 and the nucleotide in *T. litoralis* ADPGK. Residue T475 is conserved in *T. litoralis* ADPGK (T446), where the γ-hydroxyl group of the side chain forms a hydrogen bond with the 7 position of the adenine moiety. This interaction appears to be possible in a similar way in the mouse ADPGK. The 2′- and 3′-hydroxyl groups of the ribose moiety likely interact with nitrogen and oxygen atoms of the peptide backbone of the small nucleotide-binding loop (residues 382-385). It is hypothesised that the side chains of the histidine residues H381 and H386 (H352 and Y357 in 1GC5, respectively) are responsible for the specificity for ADP by preventing the nucleotide from entering deeper into the nucleotide-binding pocket, which would allow a nucleotide triphosphate to bind in an orientation suitable to act as a phosphoryl donor (Ito et al. 2001). This mode of binding was observed, for example, for the human adenosine kinase (Mathews et al. 1998) or *E. coli* ribokinase (Sigrell et al. 1998). The large side chains are proposed to build up a steric hindrance thus excluding the larger triphosphates (Ito et al. 2001). However, to date it has not been possible to reverse engineer an ADPGK or ADPPFK enzyme to allow either to accept a nucleotide

triphosphate as phosphoryl donor by mutating these positions to residues with smaller side chains (Currie et al. 2009).

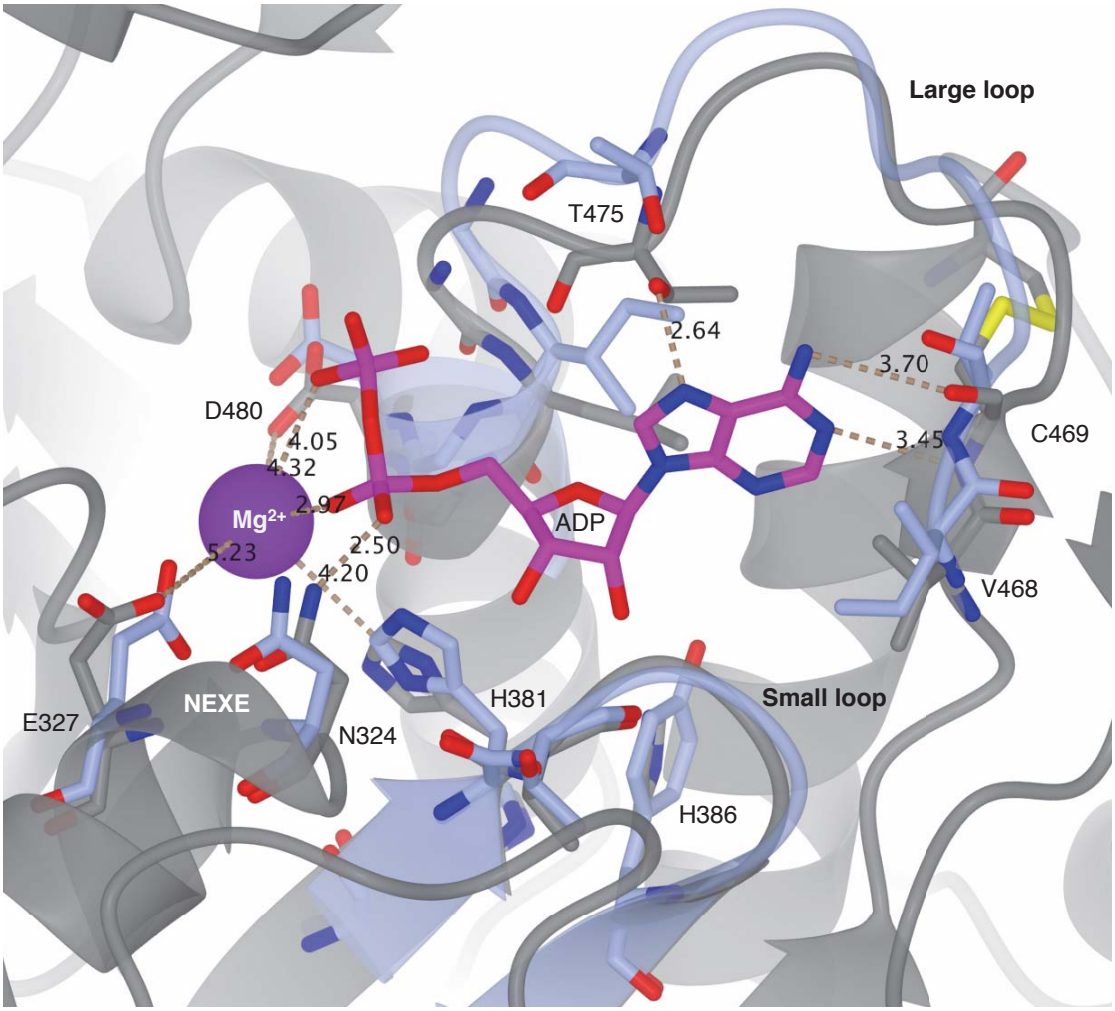

Figure 73: Proposed nucleotide binding of rmADPGKΔ51.

The figure shows a superposition of the nucleotide-binding site of rmADPGKΔ51 (grey) and *T. litoralis* ADPGK (PDB 1GC5) (light blue). For placement of the Mg^{2+} ion (purple), the structure of human adenosine kinase (PDB 2FV7) was superimposed with the mouse ADPGK and the position of the ion slightly adjusted in Coot. The highly conserved NEXE motif is displayed below the Mg^{2+} ion. For the small and large nucleotide-binding loops, the backbone of 1GC5 is displayed in ribbon representations. All residue labels refer to positions in *M. musculus* ADPGK. Measured distances are in Å between residues of *M. musculus* ADPGK and ADP from *T. litoralis* ADPGK (PDB 1GC5).

The only other ADPGK solved in an apo-form is from *P. horikoshii* (PDB 1L2L) (Tsuge et al. 2002). Interestingly, the large nucleotide-binding loop is disordered and not included in the *P. horikoshii* structure (PDB 1L2L). This is in contrast to the *M.*

musculus ADPGK, where the large nucleotide-binding loop is well defined. A possible explanation could be the disulphide bond formed between C414 and C469, attaching the large nucleotide-binding loop to the top of a helix formed by residues 398 – 404. This could possibly be enough to restrict the dynamic movement of the loop in the crystal to be defined in the electron density in the absence of a nucleotide.

The structure of the *T. litoralis* ADPGK has been crystallised in different substrate bound states: as apoenzyme in open conformation (PDB 4B8R), in complex with ADP (PDB 1GC5) and in closed conformation complexed with D-glucose and AMP (PDB 4B8S) (Ito et al. 2001; Rivas-Pardo et al. 2013). These structures in combination with SAXS structural information and enzyme kinetics have identified the five conserved clusters of interacting residues between the large and small domains, that help to induce the open and closed conformations, thus playing an important role in the reaction mechanism of ADPGK. Interestingly, it appears that some of these interactions are conserved in *M. musculus* ADPGK. However, for some of the clusters only the type of interaction is conserved, while the positions of the participating residues can be switched, and therefore will not necessarily show up in a sequence alignment. An overview of the interactions is shown in Figure 74 on page 145, where the interactions are numbered in the same fashion from 1 through 5, according to Rivas-Pardo et al. (2013). Only the open conformation of *T. litoralis* ADPGK (PDB 4B8R) is shown and compared with *M. musculus* ADPGK here, as it more closely resembles the orientation of the two subdomains to each other. The residues involved in the interaction in cluster 1 (see Figure 74) appear to be highly conserved when compared to the *T. litoralis* ADPGK. The residues E211, T475 and V476 are directly conserved in archaeal ADPGK as E188, T446 and V446, respectively. This would enable stabilising hydrogen bounding between the side chain of E211 and the hydroxyl group of T475 and the backbone amide of T475 and V476, akin to the interaction observed for the closed form of *T. litoralis* ADPGK. The interaction in cluster 2 of *T. litoralis* ADPGK is formed by a cation-π interaction between the aromatic side chain of tyrosine residue Y354 on the small nucleotide binding loop of the large subunit and a positive charge on the side chain of arginine residue R202 on the 'lip' of the small lid domain. The corresponding residues in these positions in *M. musculus* ADPGK would be L383 on the large domain and H225 on the small domain, respectively (see Figure 74). The cation-π interaction found in the

archaeal ADPGK is not directly conserved here. This type of interaction is preferably formed between the aromatic side chain of phenylalanine, tyrosine, or tryptophan with the functional group of the side chain of lysine or arginine (Minoux and Chipot 1999). However, histidine does have the potential to form π-stacking and cation-π interactions, although histidine can have the potential to contribute either a π-system or the positive charge, depending on protonation state (McGaughey et al. 1998; Gallivan and Dougherty 1999; Wang et al. 2006). The histidine residue H225 found on the small domain of *M. musculus* ADPGK does however have the potential to act as either partner in cation-π interaction, due to the potential for a positive charge and an aromatic ring system on the side chain. In the proximity of the leucine residue L383, occupying the position of the aromatic residue Y354 in *T. litoralis* ADPGK, the *M. musculus* ADPGK possesses an arginine residue R427, which could contribute the positive charge to this interaction. However, the interaction is certainly not as clear as the residues appear to be switched, or, alternatively, it is possible that this interaction may not be conserved at all. Another problem arises when considering that both residue types, histidine and arginine, do have the potential for a positive charge on their side chains, which could result in a more repulsive interaction, depending on the pH of the surrounding medium. The pKa values of arginine and histidine side chains are 12 and 6, respectively (Voet and Voet 2011). Assuming a localisation of ADPGK in the endoplasmic reticulum or the cytoplasm, the environment would likely have a pH between 6.8 and 7.4 (Wu et al. 2000), but would depend on the cell type and the cell's metabolic state. At this pH, arginine would carry more of a positive charge, contributing the cationic group, and histidine would loose its positive charge and could contribute the π-system. Therefore, a cation-π interaction in cluster 2 could be possible in the *M. musculus* ADPGK although with different positions of the functional groups forming the interaction. Arginine residue R117 plays a central part for the interactions in cluster 3 of *T. lioralis* ADPGK (see Figure 74). The hydrogen bonds are formed in the closed conformation between arginine residue R117 on the small domain and the glycine and serine residues G386 and S445 on the large domain. In addition, glutamic acid residue E115 forms hydrogen bonds with arginine residue R117 and tyrosine Y46, which is located on the small domain. The situation for the ADPGK of *M. musculus* is not entirely clear. However, the central arginine residue is structurally conserved as R474, occupying the position of S445 of *T. litoralis*

ADPGK. The position of the arginine R117 of *T. litoralis* ADPGK is occupied by tyrosine residue T155 in *M. musculus* ADPGK. The glutamic acid residue of *T. litoralis* ADPGK is replaced by glutamate Q153. Finally, tyrosine residue Y46 of *T. litoralis* ADPGK is replaced by serine residue S88 in *M. musculus* ADPGK. This pattern of functional groups leaves the eukaryotic ADPGK with a similar arrangement of hydrogen donor and acceptor groups, to potentially form similar stabilising interactions upon domain closing like the archaeal ADPGK cluster 3. In cluster 4 (see Figure 74) of *T. litoralis* ADPGK, the guanidine group of the side chain of arginine residue R191 located on the 'lip' of the small domain forms hydrogen bonds with the hydroxyl group of the side chain of serine residue S442 and the carbonyl backbone oxygen of alanine residue A441, both located on the large nucleotide binding loop. In the ADPGK of *M. musculus*, this interaction could be inverted as well. The arginine residue on the 'lip' of the small domain is not conserved, but alanine residue A214 is found in its place. However, on the large nucleotide binding loop, lysine residue K470 is found in the same equivalent position as alanine A441 in *T. litoralis* ADPGK. A likely interacting partner on the small domain could be the carbonyl backbone oxygen of leucine residue L213, acting as a hydrogen acceptor. A lysine residue would only contribute one hydrogen bond to the interaction through its side chain amide group, unlike the arginine, making the interaction with the backbone atom with no other obvious hydrogen bond acceptor in direct proximity plausible. The last interaction shown in Figure 74, cluster 5, is likely a repulsive interaction in *T. litoralis* ADPGK. The authors state that the interaction between lysine residues K74 on the small domain and K 246 on the large domain is likely disadvantageous and functions like a spring-loaded mechanism to revert the enzyme into an open conformation once the products are released, to facilitate the binding of new substrates. The lysine residues do not exist in the *M. musculus* ADPGK in the respective locations. On the small domain, this position is occupied by isoleucine residue I108 and on the large domain by glutamate residue Q269. However, since the topology of the main chain is slightly different and more bulging in direction of the small domain in the *M. musculus* ADPGK (main chain between M265 and Q269 in Figure 74, panel 5), it is possible that a slightly repulsive steric interaction is formed upon transition to the closed state between I108, Q269 and main chain atoms.

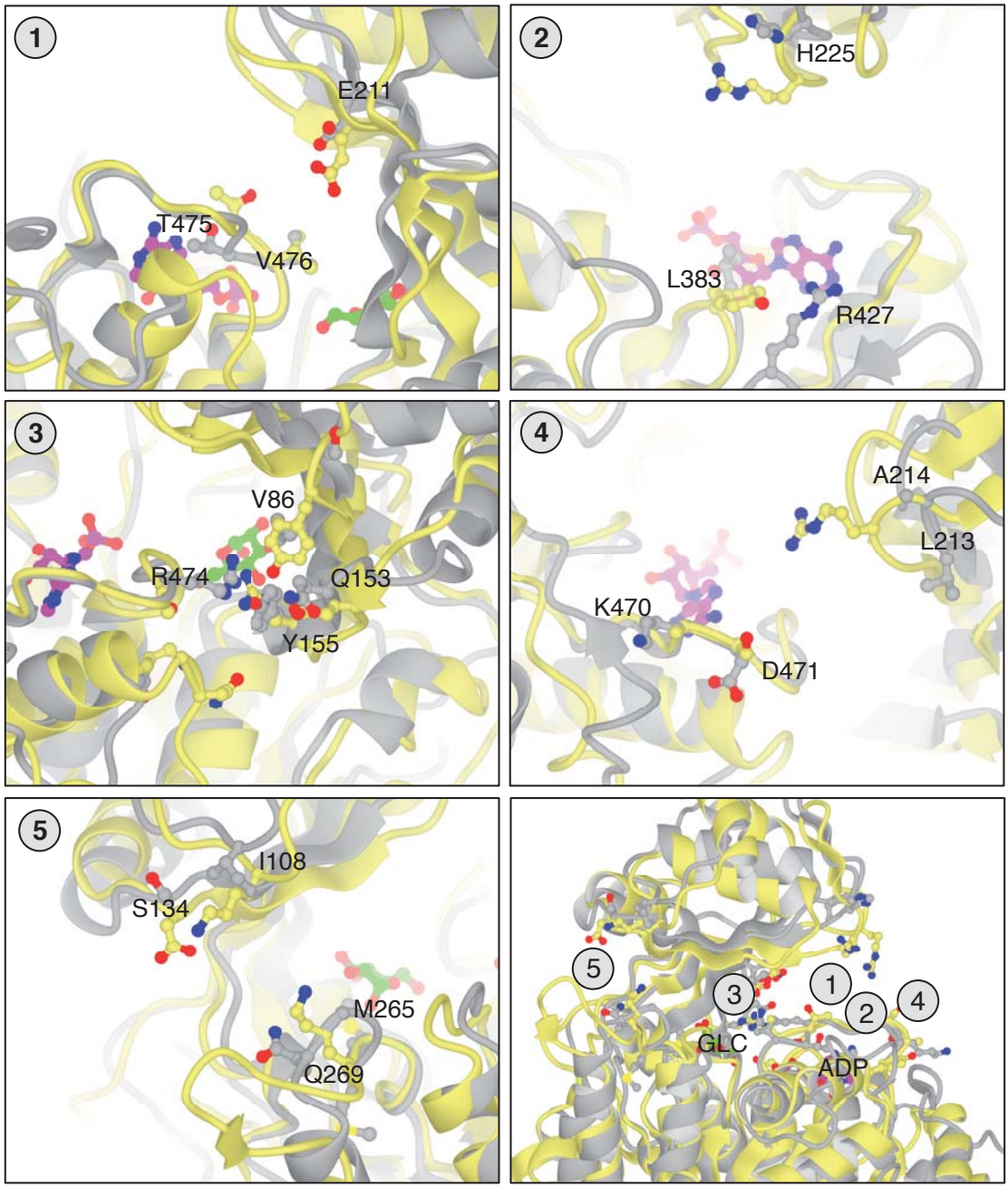

Figure 74: Conserved interactions between small and large domain of eukaryotic and archaeal ADPGKs.

The images in panels 1 through 5 show the five interaction clusters for the domain movement during catalysis described by Rivas-Pardo et al. (2013) in a superposition of *M. musculus* ADPGK (grey, rmADPGKΔ51 structure) and the open conformation *T. litoralis* ADPGK (yellow, PDB 4B8R). Glucose (green) and AMP (magenta) of the closed conformation of *T. litoralis* ADPGK (PDB 4B8S) are shown. The panel on the bottom right shows an overview of the approximate position of the interaction indicated by number in the molecules. Residue positions refer to *M. musculus* ADPGK only. For explanation see text.

3.4.7 Summary

In this work, crystals of a truncated form of *M. musculus* ADPGK, missing the first 51 amino acids, could be grown and diffracting crystals were obtained by optimising conditions with the Silver Bullet Bio screen. The crystal had the space group $P\,2_1\,2_1\,2_1$ and diffracted to 2.1 Å, and the structure of *M. musculus* ADPGK could be refined to an R_{work} and R_{free} of 0.21 and 0.26. The structure of the apoenzyme was solved, with no ligand visible in the substrate binding pockets. Overall, the structure of *M. musculus* ADPGK had high similarity to the known structures of archaeal ADPGKs. This includes the structural conservation of functional residues for binding and recognition of substrates glucose and Mg-ADP and conserved clusters for interaction between the small and large domain responsible for the transition between the open and closed states of the enzyme. This high structural and functional conservation stands in contrast to the low sequence identity. Interestingly, a disulphide bond was found in the structure of the *M. musculus* ADPGK. This finding was unexpected, since ADPGK was shown to be localised in the cytoplasm (Kamiński et al. 2012). The disulphide could possibly help to stabilise the large nucleotide-binding loop in the absence of ligand in the crystal structure. Also, a region was identified concerning the different isoforms of human and mouse ADPGK, where an additional glutamine residue is found in position 314 for the isoform 1, which could have an effect on the crystallisation behaviour on the protein. In the putative glucose-binding pocket of *M. musculus* ADPGK, a cluster of water molecules was found, arranged in a ring-like shape where the water oxygen atoms convincingly matched the positions of the hydroxyl groups of α-D-glucose. This made it possible to infer the position of the glucose substrate and side chains based on actual experimentally determined distances between the side chains and water molecules of *M. musculus* ADPGK.

4 Discussion and future outlook

4.1 Recombinant expression and crystallisation of ADPGK

4.1.1 Expression, refolding and purification of eukaryotic ADPGKs

Since an experimentally determined structure of mouse ADPGK was not available at the beginning of this work, the design of truncation constructs for expression and crystallisation and functional mutants for functional studies was based on sequence alignments and homology modelling with the few achaeal structures (PDB 1UA4, 1L2L and 1GC5) available at the beginning of the project (as described in section 3.1). The availability of an experimentally determined structure of a eukaryotic ADPGK will enable these predictions to be more accurate in the future, assuming that other eukaryotic ADPGKs will exhibit similar folding. The structural prediction for the residues close to the active site of mouse ADPGK were found to be largely correct, some structural elements in the peripheral parts of the protein and especially the N-terminal could not be predicted sufficiently. A comparison of the homology model for human ADPGK (resembles open conformation) and the experimentally determined structure of mouse ADPGK is shown in Figure 75. The figure shows a superposition of the structures with a RMSD of 1.1 Å (for Cα atoms). For this comparison to work, the assumption has to be made that the structure of human ADPGK would be nearly identical to the mouse ADPGK under comparable conditions. The region around the N-terminus up to residue 71 (on right side of the image in Figure 75) is the most obvious difference between the modelled and experimental structure, mostly because the experimental structure is missing residues 1 to 51. The missing N-terminal region is certainly of interest, as it is one of the main differences in the structure to the archaeal ADPGKs. Unfortunately, a consequence of the truncated experimental structure of a eukaryotic ADPGK also means, that homology models generated of other eukaryotic ADPGKs, will not lead to increased confidence in the predicted structure of these N-terminal residues. The differences in the small domain depicted in Figure 75 are mostly due to the different domain angles towards each other. The marked region between residues 99 and 108 resembles a helix in the archaeal ADPGK, but no such structural element is observed in the structure of mouse ADPGK. The region is correctly modelled, however in the worng

location. Three regions on the large domain are differences in the peripheral parts of the protein, spanning residues 244-251, 267-274 and 338-349. Again, the latter of these regions is α-helical in the archaeal ADPGKs, however there is no such secondary structure element found in the eukaryotic ADPGK. Interestingly, the last region marked in Figure 75, spanning residues 415 to 453, covers almost the complete stretch between the cysteine residues forming the disulphide bond. While the positions of the two cysteine residues are predicted correctly, the part of the structure connecting the two residues is not.

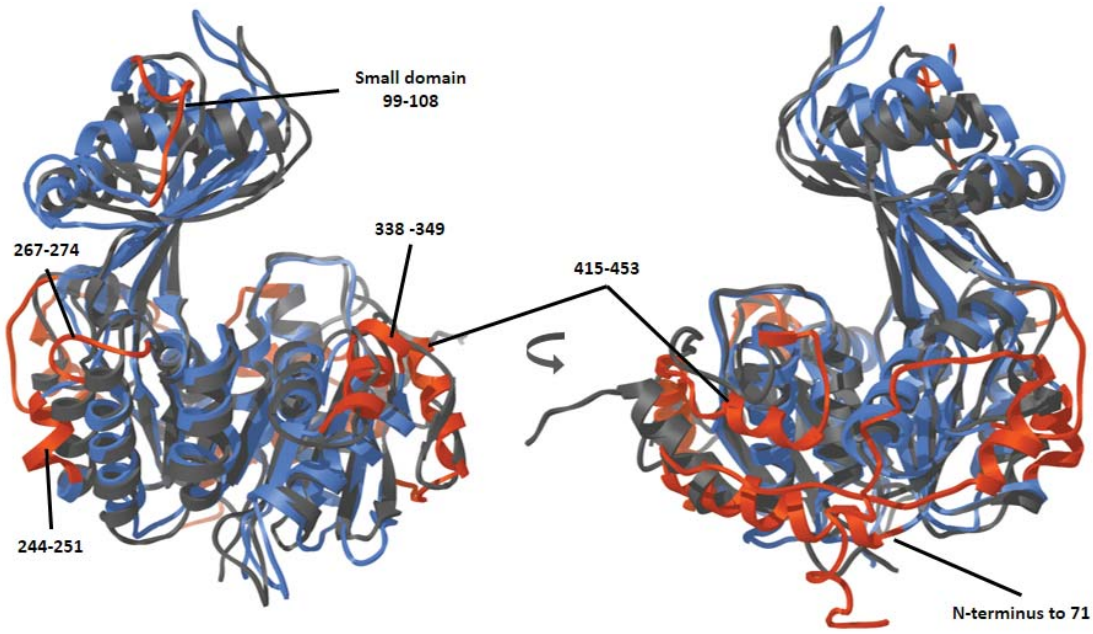

Figure 75: Comparison of homology modelling and experimentally determined structure: Overall structure.

Structural superposition of predicted and experimental structure of eukaryotic ADPGK. The experimental structure of mouse ADPGK is shown in grey. The homology model for human ADPGK is shown in blue. The superposition was generated using Chimera. The RMSD was 1.08 Å for the Cα trace, Areas with large discrepancies (RMSD > 4 Å) are coloured red in the homology model. Regions of interest are marked, for explanation see text.

With the crystal structure of mouse ADPGK at hand, the positions for all selected mutation targets could be verified and the functional residues are indeed found in the anticipated positions. An overview is shown in Figure 76, which shows the Cα-trace of mouse ADPGK with the positions of residues targeted for site directed mutagenesis

indicated in the experimental structure and the homology model of human ADPGK (the latter being chosen for the comparison due to the similar lid-opening angle). It is evident that, while not all rotamers are predicted correctly, all residues are in the correct positions.

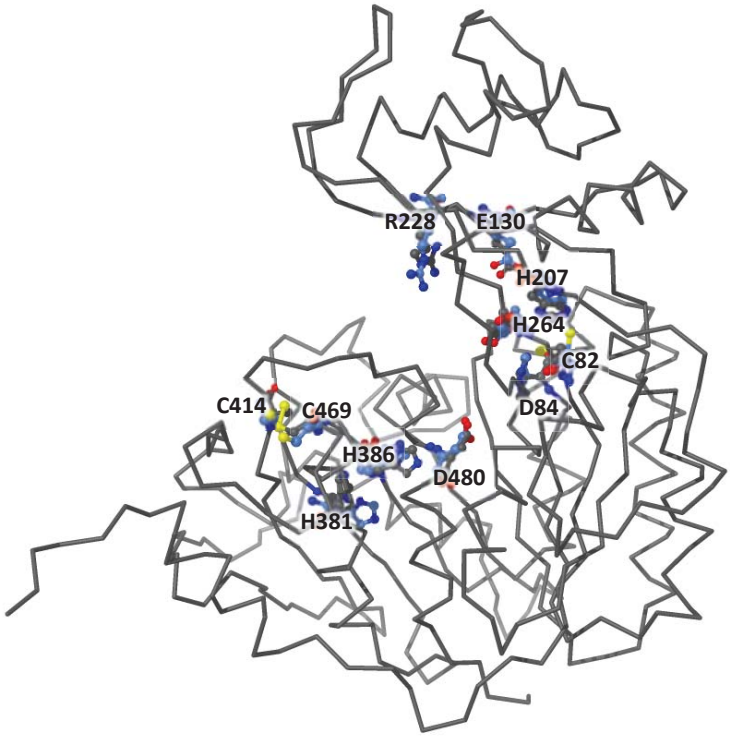

Figure 76: Comparison of homology modelling and experimental structure: Active site residues targeted for mutation.

The figure shows the position of residues targeted for site directed mutagenesis in the homology model of human ADPGK (in blue, only mutaed residues shown) as superposition on the experimentally determined structure of mouse ADPGK (grey, Cα trace and mutades residues shown). Apart from not showing the same rotamers for some residues (H381 and H386 most pronounced) the predictions were correct.

The expression of human ADPGK in *E. coli* was from the beginning hampered by low yields of soluble protein. The reasons for the low soluble expression in the conditions tested are ultimately not known but there are some possibilities that could offer an explanation. As can be seen from the crystal structure of the mouse ADPGK, the fold of this protein is rather complicated, akin to the archaeal ADPGKs, and the protein has to form a disulphide bond upon folding (see Figure 77).

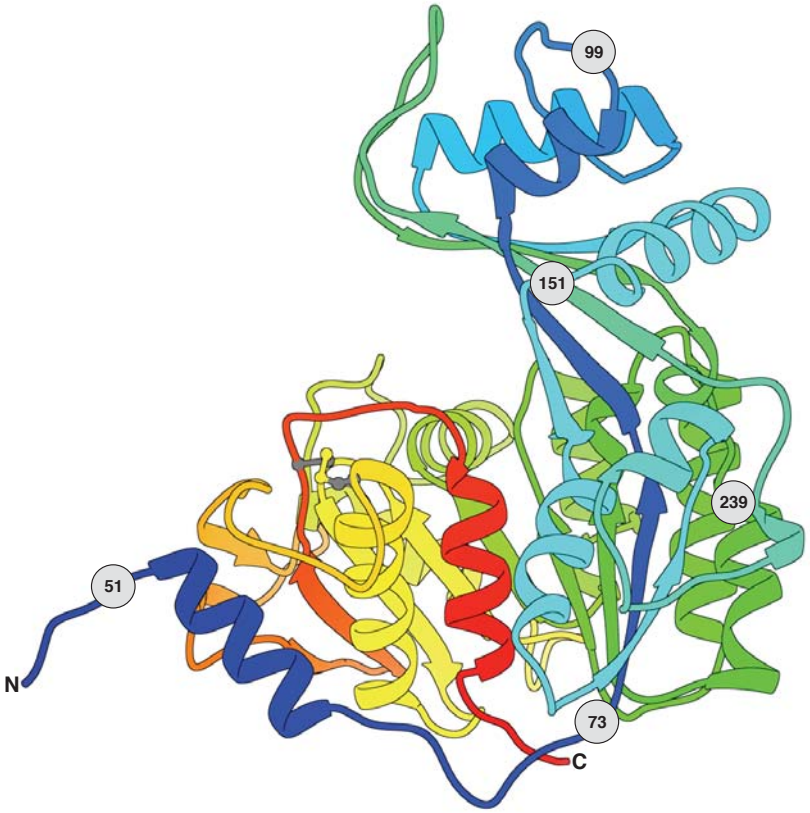

Figure 77: Domain folding of eukaryotic ADPGK.

The structure of rmADPGKΔ51 consists of the large and small domain, as also found in the archaeal ADPGKs. The main chain is displayed as a ribbon and coloured with a rainbow gradient from the N-terminus (blue) to the C-terminus (red). The disulphide bond is displayed as grey (carbon) and yellow (sulfur) ball and stick representation. Numbers in grey circles refer to residue positions (see text for details).

As can be seen from Figure 77, the small and large domains are not simply connected linearly like beads on a string but the peptide chain has to fold back and forth multiple times between the two domains, before the lid is completely formed (represented by the blue, cyan and green section of the structure in Figure 77). This complex fold could lead to problems during the protein expression, and would be very restrictive to the truncations that can be made without disrupting the folding completely. The eukaryotic ADPGKs have a potential N-terminal signal peptide, as has been described in section 3.1. The truncation of residues up to around position 50 has likely removed this N-terminal signal peptide completely in constructs like the rmADPGKΔ51 variant. However, it seems likely that any truncations after residue 73 (the residue before the first β-strand) would potentially be catastrophic for the folding of the

protein, especially since the first strand following residue 73 of the central β-sheet is not a flanking, but an internal strand. For a truncation beyond residue 73, it is likely that all residues up to about residue 239 will need to be truncated as well, essentially removing the whole lid domain and the linking β-sheet hinge region. However, this suggestion is in contrast to the observation that an ADPGK with a truncation of the first 151 amino acids, rhADPGKΔ151, could be purified (see section 3.2.6). The purified rhADPGKΔ151 did not have any catalytic activity.

The disulphide bond found in the structure of *M. musculus* ADPGK (see section 3.4.4) could be problematic during the expression recombinant proteins in *E. coli*, as the bacterial cells do not deal in the same way with the formation of covalent disulphide bonds as eukaryotic cells. In addition, the successful refolding conditions described in section 3.2.4 all contained a reducing agent, although it is likely that this would be required for any protein containing at least one cysteine residue. In mammals, disulphide-containing proteins are often secreted and therefore transported to the endoplasmic reticulum, where an oxidising milieu is maintained and disulphide bonds can form. For the transport into the endoplasmic reticulum a signal sequence on the target protein is recognised by the signal recognition particle, which then forms a complex with the nascent protein chain and the ribosome. This complex then binds to the endoplasmic reticulum and is transported into the endoplasmic reticulum via the Sec61 translocon (Braakman and Bulleid 2011). In prokaryotes like *E. coli*, which do not have an endoplasmic reticulum, secretory proteins are exported into the periplasm, where an oxidising environment is found and disulphide bonds can form (Inaba 2009; Kadokura and Beckwith 2010). The folding environment of the endoplasmic reticulum is different to the periplasm, and very different to the cytoplasm of *E. coli* not only because of the different redox states but also their composition of chaperones (not discussed in detail here). For the expression of eukaryotic proteins in *E. coli*, there are number of tools available to the experimenter, that can help circumvent the problems arising from these differences. One possibility, which has been explored in this work for the eukaryotic ADPGKs (see section 3.1.9), is the use of an *E. coli* expression host like Origami (or one of its commercially available variants), which has an altered genotype (see section 2.1.4 for details) that helps with disulphide formation directly in the *E. coli* cytoplasm. However, the Origami strains tested for this work did not lead to an increase in protein yield (see section 3.1.9). A possible

conclusion is that the disulphide bond is not necessarily required for the stabile folding of ADPGK and that the protein is able to fold in the reducing *E. coli* cytoplasm regardless. Another problem could arise from a disordered stretch around Q313 is that this part of the protein could be more prone to undesirable proteolysis by *E. coli* proteases. Considering that similar degradation issues could be detected for mouse ADPGK isoform 3 (featuring a single Q) and human ADPGK isoform 1 (featuring a double Q), proteolysis specific to this site is probably not the main cause of the different behaviour for the mouse and human ADPGK. It is more likely that this particular difference in the amino acid sequence between the isoform 1 and 3 of *M. musculus* ADPGK could affect the crystallisation behaviour, as will be discussed in section 4.1.2 below.

4.1.2 Crystallisation of eukaryotic ADPGKs

Several constructs with N-terminal truncations of ADPGK from *M. musculus* and *H. sapiens* were generated (see Table 3.1 for an overview). The aim of these truncations was to remove N-terminal regions or the potential signal sequences, with a propensity for disorder (see section 3.1). These regions would likely make unfavourable entropic contributions to the formation of crystals. C-terminal truncations or internal deletions were ruled out from the beginning, based on the assumption that eukaryotic ADPGK would have a fold similar to the archaeal ADPGKs. Therefore, one of the main catalytic sites, the GXGD motif, would be located very close to the C-terminus. Of the tested constructs it was only possible to obtain crystals for rmADPGKΔ51, a construct with a truncation of the first 51 amino acids based on isoform 3 of mouse ADPGK. Crystallisation trials for the human ADPGK or a full-length construct of either mouse or human ADPGK have not been successful thus far. However, the crystal contacts in the solved structure of the mouse ADPGK could shed some light on the underlying problem. Of the contacts in the rmADPGKΔ51 crystal listed in Table 3.7, two sites are of particular interest. The first site of interest is interface number 3 in Table 3.7 (or Figure 65, on the right side), which could be very specific to the isoform 3 of mouse ADPGK. The interaction is depicted in Figure 78.

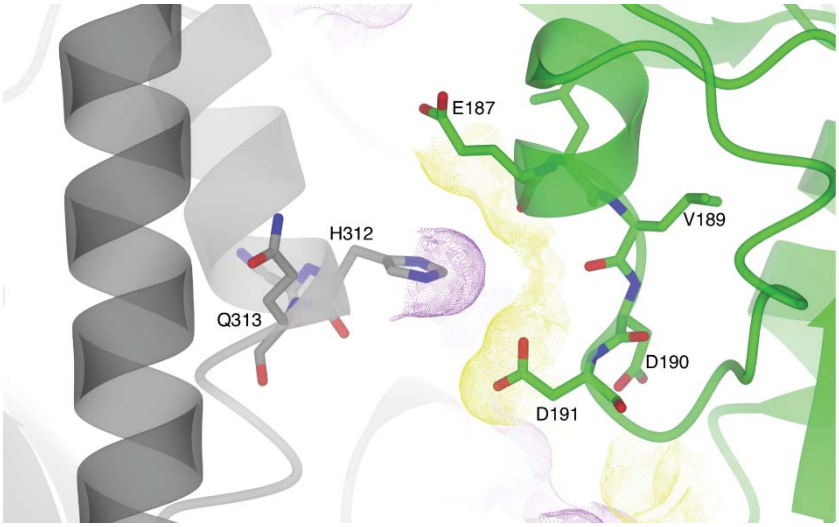

Figure 78: Crystal contact specific for of *M. musculus* ADPGK isoform 3.

The figure shows a crystal contact identified by PISA between two symmetry mates (grey and green molecule), which is potentially very specific for the isoform 3 of mouse ADPGK. The isoform 1 of ADPGK incorporates an additional glutamine residue Q314, which would follow Q313 shown in the grey coloured molecule, which could prevent the formation of this crystal contact. The van der Waals radii of contacting atoms are displayed as yellow and purple surfaces.

Due to the high level of sequence conservation between mouse and human ADPGK, one would expect to find an analogous situation for isoform 1 and 2 of human ADPGK. A histidine residue H312 (grey molecule in Figure 78) makes contact with the side chain of aspartate residue D191 and the backbone of valine residue V189 in a different molecule (green molecule in Figure 78). The potential problem could arise in isoform 1 of mouse ADPGK. Two glutamine residues, Q313 and Q314 follow the histidine. Only one glutamine residue is found in isoform 3 (see Figure 68 for an alignment of this particular region). The additional glutamine could disrupt the hydrophobic packing of the inward facing side of the helix (see Figure 69), introducing a larger bulge in the last helical turn or producing a disordered stretch of sequence. In both cases, this would likely have an unfavourable contribution to crystal formation.

The second crystal contact of interest involves mostly the contact surface number 2 in Table 3.7 (see Figure 65), which affects the N-terminus of ADPGK. The first N-terminal helix of rmADPGKΔ51 is capped by proline residue P53 (yellow molecule in Figure 79), which packs against the side chain of the arginine residue R71 in a symmetry related molecule (green molecule in Figure 79). The side chain of arginine

residue R71 (green molecule in Figure 79) packs between two different symmetry mates (yellow and cyan molecule in Figure 79), alanine residue A50 (yellow molecule in Figure 79) at the N-terminus and proline and histidine residue P224 and H225 (cyan molecule in Figure 79) of the small domain. For the crystal contact between P53 and R71 to form, the position of the proline could be mandatory. In a truncated construct starting after the proline, the α-helix could extend further making the formation of this particular crystal form impossible. In a full-length construct, P53 is preceded by prolines in positions 50 and 47. Dihedral angles for position 50 fall in the allowed regions of the Ramachandran plot for proline in the rmADPGKΔ51 structures. Given the tightly stacked interaction of A50 and R73, a longer construct incorporating these residues could have a positive or negative effect on the crystallisation behaviour. The rigidity due to the restriction in dihedral angles of the proline residues could make a favourable (or at least negligible) entropic contribution to the crystal formation. On the other hand, in a full-length construct the N-terminal amino acids preceding these crystal contacts could form a small sub-domain by themselves or pack against the rest of the protein and in either case would likely interfere with the crystal formation. In the case with the N-terminal region being disordered, and the rest of the protein fold not altered, obtaining a crystal in this form of a full-length ADPGK would possibly not have yielded much additional structural information.

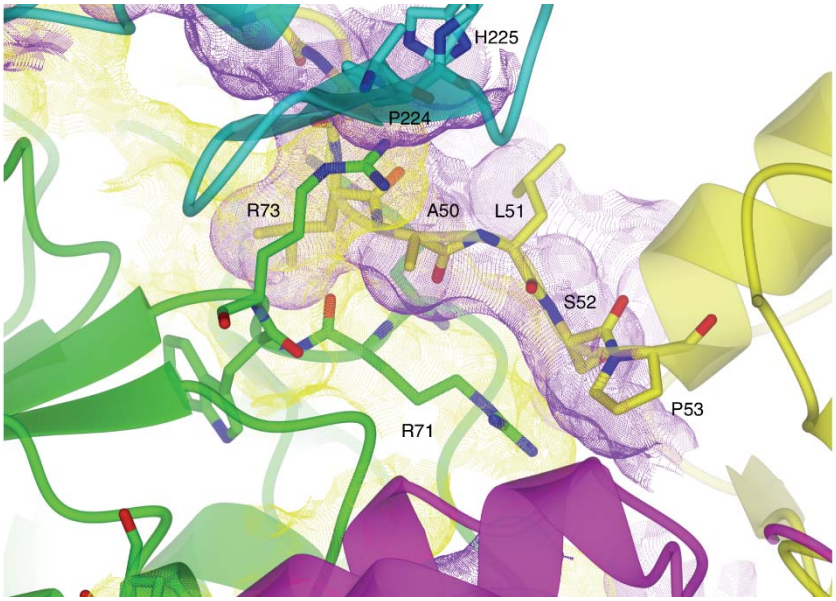

Figure 79: Crystal contacts of the truncated N-terminus of rmADPGKΔ51.

The figure shows the crystal contacts around the N-terminus of rmADPGKΔ51, where three different symmetry mates are involved (yellow, green and cyan molecules). The van der Waals radii of contacting atoms are displayed as yellow and purple surfaces. The proline P53 at the beginning of the N-terminal helix (yellow) packs against the sidechain of arginine R71 (green). The side chain of arginine R73 (green) packs between alanine A50 (yellow) and proline P224 and histidine H225 (cyan).

When a protein is crystallised, it is always desirable to obtain diffraction data of the protein in different functional states. In the case for ADPGK, this would include the different stages of substrate binding, lid closing and possibly a catalytic transition state. Unfortunately, given the B-factors (shown in Figure 63) and the distribution of the crystal contacts (shown in Figure 65), there is a possibility that the particular crystal form found for rmADPGKΔ51 is unable to undergo a domain movement upon ligand binding in soaking experiments that introduce the ligands into the crystals. The apparent rigidity of the 'axis' formed by the lid domain, the linking β-sheet hinge and the more N-terminal side of the large domain (indicated by lower B-factors in these particular regions as shown in Figure 63) indicates that the crystal packing could collapse upon ligand binding, dissolving the crystal. Hence to obtain the structure of a ligand bound ADPGK, eventually a different crystal form needs to be grown.

4.2 Structure and function of eukaryotic ADPGK

4.2.1 The structure of eukaryotic ADPGK

The overall fold of *M. musculus* ADPGK is remarkably similar to the archaeal ADPGKs (see section 3.4.5). However, the solved structure comprises residues 51 to 496, excluding the putative N-terminal signal peptide. The region between arginine residue 74 to serine residue 99 has been described to be an amphipathic, membrane associated helix in a recent study (Kamiński et al. 2012). However, this is not supported by the solved structure of rmADPGKΔ51. The region in question (between residue position markers 73 and 99 in Figure 77, coloured blue) is an internal strand of the central β-sheet, transitioning into a hinge-region β-strand and ultimately a short helix part of the small lid domain (blue α-helix preceding marker 99 in Figure 77). This helix in the small domain packs against the hydrophobic core of the lid. A membrane association of this part of the protein is therefore unlikely without a major structural rearrangement.

Interestingly, a disulphide bond was found in the structure of mouse ADPGK. Both cysteine residues are conserved in ADPGK of other eukaryotes, but not in archaea, as can be seen in the protein sequence alignment in Figure 80. In a BLAST search with the *H. sapiens* ADPGK sequence an uncharacterized protein from *Tetrahymena thermophila* strain SB210 (accession code I7M6W2) was found. A disulphide bond is not directly identified for ttADPGK from the sequence alignment. Other sequence motifs, however, are conserved in this unicellular protozoon, including the catalytic aspartate/arginine pair and the NEXE motif for recognition of Mg^{2+}. ADPGKs of insects (si, bm, dm) and sea urchin (sp) show multiple cysteine residues in the region in question (see Figure 80). The disulphide bond found in the mouse ADPGK still appears to be conserved, with the other cysteine residues being an additional pair, potentially forming another disulphide. The equivalent residues in the structure of mouse ADPGK, identified based on the sequence alignment in Figure 80 are structurally in close proximity to each other, making an additional disulphide at least in theory possible.

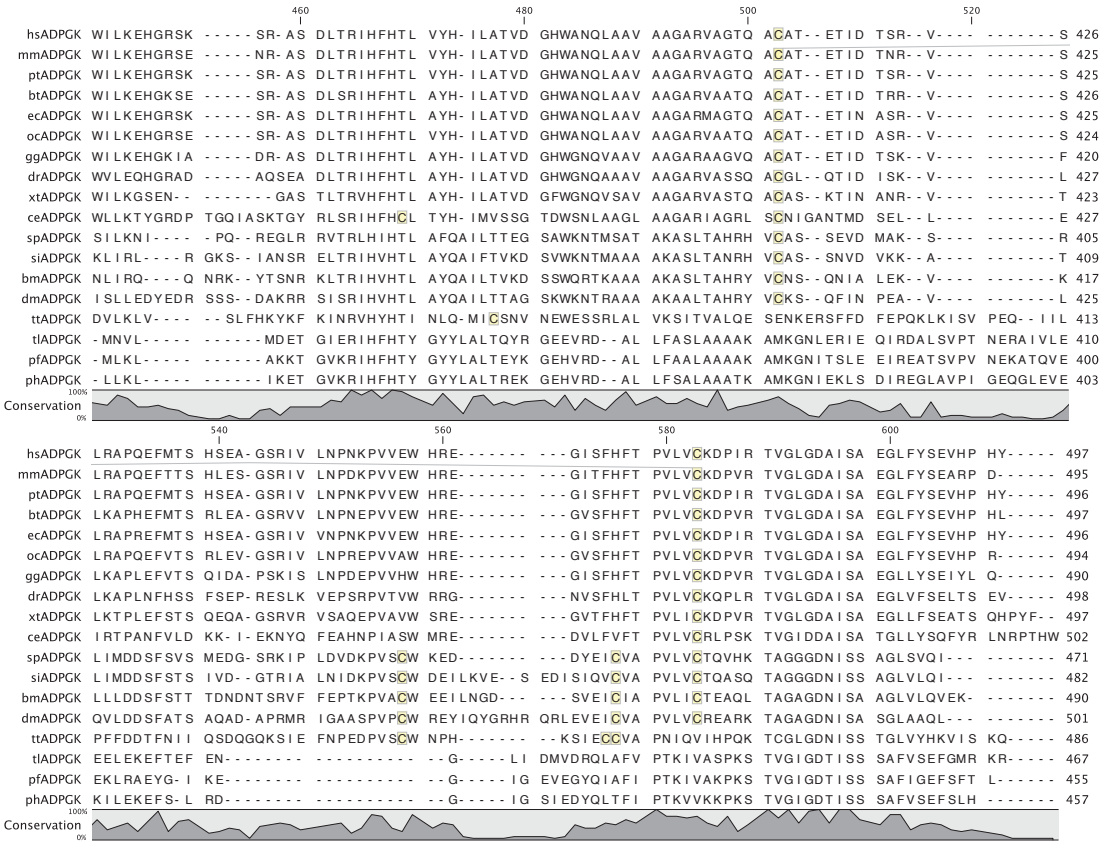

Figure 80: Sequence alignment of C-terminal region of ADPGK.

The sequence alignment shows the C-terminal region of ADPGKs from various eukaryotes and archaea. Species: hs = *H. sapiens*, mm = *M. musculus*, pt = *Pan troglodytes*, bt = *Bos taurus*, ec = *Equus caballus*, oc = *Oryctolagus cuniculus*, gg = *Gallus gallus*, dr = *D. rerio*, xt = *Xenopus tropicalis*, ce = *Caenorhabditis elegans*, sp = *Stronglycentrotus purpuratus*, si = *Solenopsis invicta*, bm = *Bombyx mori*, dm = *Drosophila melanogaster*, tt = *Tetrahymena thermophila*, tl = *T. litoralis*, pf = *P. furiosus*, ph = *P. horikoshii*. Cysteine residues, including the cysteine residues homologous to those involved in the formation of the disulphide bond observed in *M. musculus* ADPGK are highlighted with a yellow box. The disulphide bond is indicated as a grey link between the cysteine residues for mmADPGK only. The disulphide bond is highly conserved in eukarya, but is absent in archaea.

The finding of the conserved disulphide (or multiple) is a bit of a surprise, as it is thought that ADPGK in human cells is anchored in the membrane of the endoplasmic reticulum facing the cytosol with its active site region, near where the disulphide bond is found (Kamiński et al. 2012). A protein with two cysteine residues in such close proximity that they are able to form a disulphide bond would be expected to be located in a compartment with oxidising environment, like the extracellular space or the lumen of the endoplasmic reticulum. However, intracellular disulphides do exist (Mallick et al. 2002), although they are usually of a transient nature and part of a

catalytic mechanism, e.g. in the case of thioredoxin (Prinz et al. 1997), or function as a redox sensor, e.g. in the case of the transcription factor OxyR (Choi et al. 2001). The membrane association of eukaryotic ADPGK is further supported by the predicted N-terminal signal peptide and reports of a cholesterol binding motif in the human ADPGK (Hulce et al. 2013). Based on the structure of the *M. musculus* ADPGK, the deduction could be made that a potential transmembrane domain would be likely located in the part of the protein that was not crystallised – the first 50 amino acids - as the region proposed by Kamiński *et al.* (2012) would require a major rearrangement of the protein. The more commonly found cholesterol binding sites are the cholesterol recognition/interaction amino acid consensus sequence, short CRAC, and a similar, inverted motif called CARC (Li and Papadopoulos 1998; Baier et al. 2011). The consensus sequence for CRAC is (L/V)-X_{1-5}-(Y)-X_{1-5}-(K/R) and the consensus sequence for CARC is (K/R)-X_{1-5}-(Y/F)-X_{1-5}-(L/V). The CARC motif starts with a lysine or arginine residue with a positively charge side chain, which can form interactions with the hydroxyl group of cholesterol. The conserved aromatic phenylalanine or tyrosine residue in the centre of the motif can form a π-stack with cholesterol. The motif concludes with a branched, aliphatic amino acid, which can form van der Waals interactions with cholesterol. Other motifs are known, but they are less clearly defined. Also larger sequence separation between the defining residues has been observed and the aromatic residues can sometimes be substituted by tryptophan residues (Fantini and Barrantes 2013). The N-terminal region under discussion here is shown in Figure 81. The most compact possible CARC domain of human ADPGK would be 5-RGSAYAGFL-13. The tyrosine residue is not conserved in mouse ADPGK, however it is found in other vertebrates. This motif would fulfil all the sequence criteria required for the CARC domain. A second CARC like domain for cholesterol binding could be the 33-RSLWSSL-39 motif. Given the loose definition of the cholesterol binding motifs (Fantini and Barrantes 2013), other combinations with different lengths are certainly contrivable for the eukaryotic ADPGKs. Due to the lack of experimental data however, it is certainly safer to abide by the defined motifs as much as possible. A sequence motif resembling a CRAC-like domain is not directly evident for the human and mouse ADPGKs.

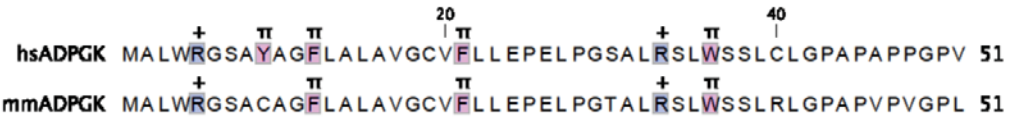

Figure 81: A putative cholesterol-binding site of eukaryotic ADPGK.

The alignment shows the first 51 amino acid residues of *H. sapiens* (hs) and *M. musculus* (mm) ADPGK. Residues with positively charge side chain are marked blue (+), residues with a side chain capable of forming a π-stack with cholesterol, are marked magenta (π).

However, it seems that with the given parameters for the cholesterol binding domains on the N-terminus of eukaryotic ADPGK, that there are two possible CARC-like domains, one starting at position 5, the other at position 33. Given the observation, that the C-terminal domain of eukaryotic ADPGK is located in the cytoplasm (Kamiński et al. 2012), the simplest explanation here could be that the second CARC-like domain (33-RSLWSSL-39) interacts with cholesterol in the membrane. This would mean that the rest of the N-terminus is indeed the predicted signal peptide and will be cleaved off, making eukaryotic ADPGK a type I ER membrane protein (Lodish et al. 2012). In this case, the disulphide bond is likely not required for stability and the two free cysteine residues apparently would not interfere with the protein's function. Should the first CARC domain be responsible for the cholesterol interaction, eukaryotic ADPGK would have a type II or type III insertion mechanism and the C-terminal domain would be localised in the endoplasmic reticulum lumen or cytoplasm, respectively. The direction would be determined by the positively charged residue, either in position 5 (before the transmembrane domain) or in position 33 (after the transmembrane domain) (Lodish et al. 2012). Should the CARC domain starting at position 5 (5-RGSAYAGFL-13) be true, eukaryotic ADPGK would not have its C-terminal domain in the cytoplasm, contradicting the experimental results of Kamiński *et al*. (2012). This would however allow the formation of the disulphide bond observed in the crystal structure of *M. musculus* ADPGK. In contrast, HKI and HKII have been shown to bind to the outer mitochondrial membrane mediated by an N-terminal insertion motif resembling an amphipathic α-helix, which may insert at least partly into the outer mitochondrial membrane (Xie and Wilson 1988; Pastorino and Hoek 2008). Therefore, association to other membranes would be at least possible for glycolytic enzymes. The proposed cholesterol binding domains are speculative and certainly require experimental confirmation.

4.2.2 Enzymatic characterisation of eukaryotic ADPGK

The K_m for glucose of human ADPGK (rhADPGK) was 0.56 mM, more than 10-fold higher than the K_m of 0.096 mM reported for the mouse ADPGK (Ronimus and Morgan 2004). This makes the human ADPGK comparable to HK2 (0.3 mM), and mouse ADPGK comparable to HK1 (0.03 mM) (Wilson 2003), regarding the K_m values. Also, the maximum activity of the mouse ADPGK was observed at a glucose concentration of 0.35 mM, while the maximum activity of human ADPGK was observed at a glucose concentration of 0.75 mM. Both enzymes were inhibitited by high concentrations of glucose, with activity reduced to ~50% at a concentrations of 5 mM or higher observed for the human ADPGK. The effect was even stronger for mouse ADPGK (Ronimus and Morgan 2004). Regardless of the differences, the kinetic characteristics would make both enzymes sensitive to changes in intracellular glucose concentration, typically around 5 mM in erythrocytes (Nelson and Cox 2012). This would especially be true, if the glucose concentration lowers in the cell, as the catalytic rate of ADPGK would then increase, at least to a certain concentration. The K_m values for ADP of human and mouse ADPGK of 0.48 mM and 0.28 mM are within a comparable range (Ronimus and Morgan 2004). Both, the human and mouse ADPGK, were inhibited by high concentrations of Mg-ADP. Inhibition by high concentrations of substrate is normal for most enzymes and often occurs at a concentration that is not physiologically relevant, which is most likely the case for ADP, unlike glucose. The typical intracellular concentration for ADP is 0.25 mM in the cytoplasm (Nelson and Cox 2012), while the K_i for Mg-ADP of human ADPGK was 9.1 mM. Mouse ADPGK did not show product inhibition by glucose-6-phosphate (Ronimus and Morgan 2004), while human ADPGK was not tested for this type of inhibition. Human and mouse ADPGK showed product inhibition by AMP. The K_i for AMP was 0.5 mM, making ADPGK sensible to concentration changes in the intracellular concentration of AMP which is typically 0.02 mM in the cytoplasm (Nelson and Cox 2012). This is around 10-fold higher than the K_i reported for *P. furiosus* ADPGK, which was 0.06 mM (Ito et al. 2003). Human ADPGK showed inhibition by Mg-ATP, however it was not possible to determine an accurate K_i, as the assay was inconsistent and difficult to interpret for unknown reasons. Also, since Mg-ATP was added to the assay and Mg^{2+} has been shown to be inhibitory on its own, the

two effects are impossible to separate in this case. The effect was still stronger than the inhibition by Mg^{2+} alone. The remaining activity at an Mg-ATP concentration of 9 mM was around 21%. This would mean, that changes in ATP concentration are likely to have an effect on the catalytic rate of eukaryotic ADPGK, as the typical intracellular concentration is around 2.25 mM (Nelson and Cox 2012). To summarise, eukaryotic ADPGK is a glycolytic enzyme highly specific for its substrates Mg-ADP and D-glucose. Given the results from inhibition experiments, it is possible, that ATP, ADP, AMP and D-glucose exhibit a regulatory effect on the catalytic rate *in vivo*, as the kinetic properties of eukaryotic ADPGK would make the enzyme highly sensitive to changes in concentration of these metabolites. The ratio of ATP to AMP is the cell's measurement of its energy charge state (Hardie and Hawley 2001), and the apparent sensitivity of ADPGK to the ATP:ADP:AMP ratio in a cell is certainly intriguing.

The role of catalytic key residues was investigated by site-directed mutagenesis. When the catalytic aspartate of human ADPGK was changed to alanine (D481A), asparagine (D481N) or glutamate (D481E), the residual activity was reduced to 0.5%, 0.05% or 0%, respectively. A truncated variant with the D481A mutation, rhADPGKΔ50-D481A, also showed no activity. The catalytic aspartate is apparently critical for catalysis, and a variation in the side chains' functional group by mutation has a dramatic effect. The mutation D481N shows a greater reduction in activity than D481A, where the functional group is completely removed. Asparagine is likely unable to act as a strong enough catalytic base to activate glucose for the phosphoryl transfer. The mutation to glutamate leaves the functional group of the side chain in place, although the steric hindrance due to the larger residue probably makes the positioning of the sugar substrate impossible. The mutation R228A in human ADPGK resulted in catalytic activity as low as 1% of the wild type enzyme. The K_m for ADP was 0.27 mM and the K_m for glucose 0.23 mM for this mutant, which is still comparable the wild type enzyme with a K_m for ADP of 0.48 mM and a K_m for glucose of 0.58 mM. The enzyme is likely unable to stabilise the terminal phosphate sufficiently during the reaction while the substrate binding is only affected to a lesser extent.

The nucleotide specificity of archaeal ADPGKs, ADPPFKs and ADPGK/PFKs has been discussed in a recent review (Guixé and Merino 2009). Most of the archaeal enzymes were shown to be able to utilise other nucleotide diphosphates than ADP. For example, the ADPGK from *P. furiosus* and *T. litoralis* have been shown to use CDP at a similar rate, albeit with a much higher K_m value. Much lower catalytic rates were reported for the nucleotide diphosphates GDP, UDP and IDP. On the other hand, no alternative nucleotide substrate utilisation was observed for the ADPGK from *A. fulgidus*. The ADPGK from *M. musculus* has been shown to be able to utilise other sugar and nucleotide substrates (Ronimus and Morgan 2004). Mouse ADPGK was able to use GDP and CDP as phosphoryl donor with 55% and 12% activity remaining, respectively. When the nucleotide specificity of *H. sapiens* ADPGK was tested (see section 3.3.4), the enzyme showed only activity with the nucleotide diphosphates ADP (100%) and to lower extent GDP (4.7% of control) and CDP (0.2% of control). No activity with the nucleotide diphosphates UDP and TDP or the nucleotide triphosphate ATP was detected. None of the ADP-dependent kinases could use nucleotide triphosphates as phosphoryl donor. This includes the functional mutant rhADPGK-H382V:H387V, which was designed with the aim to generate a mutant that could possibly utilise ATP. An equivalent mutation in *P. furiosus* ADPPFK also did not result in an enzyme with catalytic activity with ATP (Currie et al. 2009). The adenosine moiety of the phosphoryl donor is likely stabilised by hydrogen bond interactions with the side chain hydroxyl group of threonine residue T475 and backbone amide groups of cysteine residue 469 of the large nucleotide-binding loop (see panel (a) in Figure 82). The binding of GDP could be impaired due to the missing hydrogen donor group in the 6-position and hydrogen acceptor in 1-position of the purine base, leaving only the interactions with the side chain hydroxyl group of threonine in place (see panel (b) in Figure 82). The pyrimidine-based phosphoryl donors resulted in much lower (CDP) or no (UDP, TDP) catalytic activity with the human ADPGK. This is probably due to the much smaller pyrimidine base being unable to fill the binding pocket and stabilise the nucleotide-binding loop sufficiently, when compared to the larger purines, which could result in distorted geometry and inefficient positioning of the terminal phosphate (see panels (c) and (d) in Figure 82).

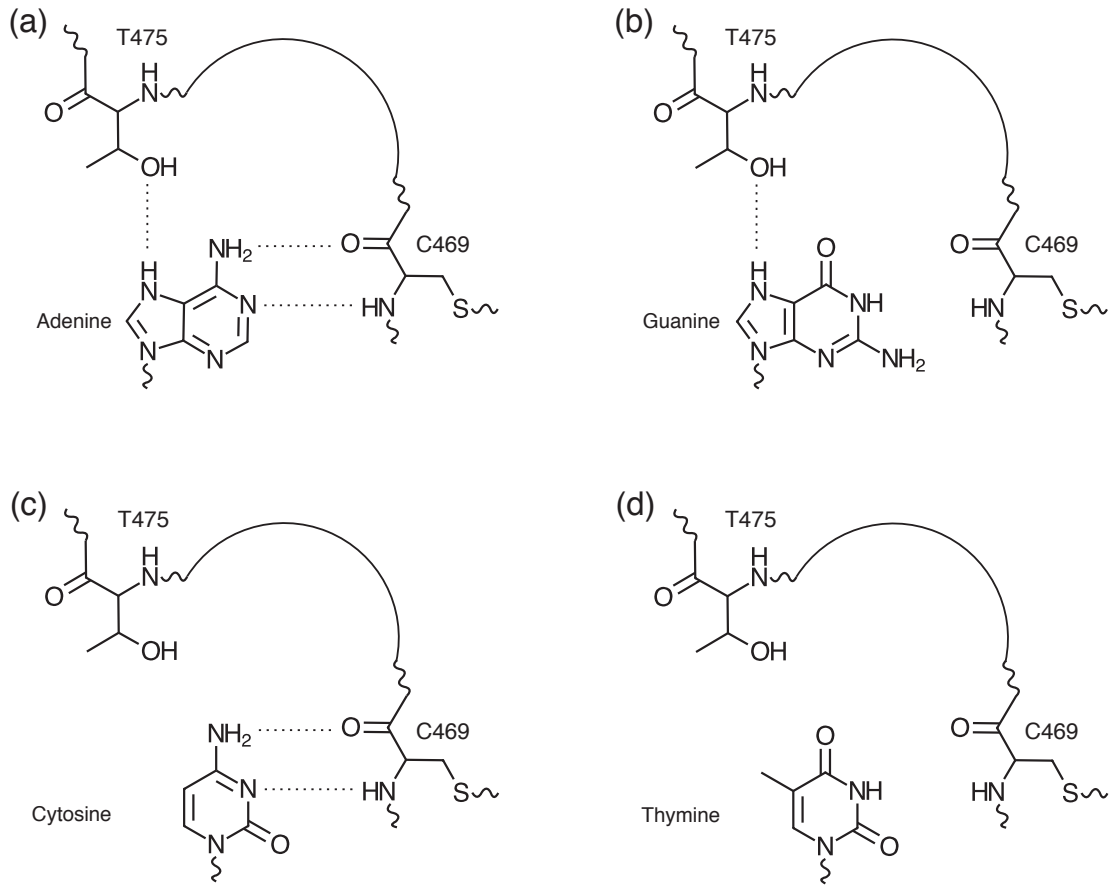

Figure 82: Phosphoryl donor selectivity of eukaryotic ADPGKs.

Schematic representation of the nucleotide-binding site of mouse ADPGK. The nucleotide group of the phosphoryl donor is displayed; a waved line indicates the connected ribose and phosphate moiety. Interactions with amino acids in the large nucleotide-binding loop (threonine residue T475 and cysteine residue C469) are indicated as dotted lines. The large nucleotide-binding loop is displayed as a curved line. (a) Adenosine diphosphate is the preferred nucleotide, and forms a hydrogen bond with the side chain of T475 and two hydrogen bonds with the backbone atoms of C469. (b) Guanosine diphosphate placed in the same orientation could potentially only form the stabilising interaction with the side chain hydroxyl group of T475. (c) Cytidine diphosphate could form the stabilising interaction with the backbone of C469, or, alternatively, a hydrogen bond with the hydroxyl group of T475. (d) Thymidine diphosphate or uridine diphosphate (without the methyl group in 5′ position) may not be able to form any of the stabilising interactions in the way that was shown seen for the other nucleotides, resulting in a distorted geometry of the interaction and incorrect positioning of the terminal phosphate, or no binding at all.

The large nucleotide-binding loop of the eukaryotic ADPGK is stabilised by a disulphide bond between cysteine residues C414 and C469 in the structure of

rmADPGKΔ51 (see section 3.4 and Figure 80). It is interesting to note that the loop of rmADPGKΔ51 is visible in the electron density, despite the absence of a stabilising ligand in the binding pocket. The only other ADPGK structure without a nucleotide bound is the *P. horikoshii* ADPGK (PDB 1L2L), where the large nucleotide-binding loop is apparently too flexible to be visible in the electron density. The disulphide bond could be a stabilising factor for the loop of *M. musculus* ADPGK. It is also possible that a more rigid loop, stabilised by the conserved disulphide bond, could affect the nucleotide selectivity of the eukaryotic enzymes. The ADPGKs from *P. furiosus* and *T. litoralis* have been shown to able to utilise CDP to a much higher extent (111% and 122% of ADP control, respectively) than the eukaryotic ADPGKs and were also able to utilise UDP (2% and 3% of ADP control, respectively) as phosphoryl donor (Koga et al. 2000). Therefore, the additional conformational restraint of the large nucleotide-binding loop through a disulphide bond may explain the more stringent nucleotide selectivity of the eukaryotic enzymes. With regard to experimental work, an increased rigidity could mean that for functional mutants addressing the nucleotide specificity, the effect will be negligible unless the disulphide bond is removed as well – either by a mutation of the cysteine residues or a chemical reduction. Breaking the bond chemically may be hindered by the fact that the reducing agent would have to be able to react with the cysteine residues, which is not guaranteed as the bond is not directly exposed on the surface of the protein. Mutation of the cysteine residues could therefore be the preferred option, assuming that the protein is still able to fold without the disulphide, which is a likely possibility, considering the high similarity of the overall fold between the eukaryotic and archaeal ADPGKs, the latter being stable without a disulphide.

The ADPGK of *H. sapiens* was tested for its ability to accept molecules other than D-glucose as alternative phosphoryl acceptor (see section 3.3.4). It was found that none of the molecules tested could substitute for D-glucose. The utilisation of alternative phosphoryl acceptors was determined by monitoring ADP-disappearance and AMP-appearance by ^{31}P-NMR, thus eliminating possible complications of a linked enzymatic assay, where the produced phosphorylated products may not be compatible with the linker-enzymes. Other studies investigating the phosphoryl acceptor specificity did so by using a coupled enzymatic assay. The ADPGK enzymes from other sources were found to have some residual activity with other hexoses. Mouse

ADPGK had 10 and 20% of catalytic activity with fructose and mannose, respectively. Furthermore, the ADPGK from *T. litoralis* and *P. furiosus* had less than 10% of their activity remaining with the hexoses galactose, fructose or mannose (Koga et al. 2000; Verhees et al. 2002; Guixé and Merino 2009). The archaeal enzymes were also able to utilise 2-deoxyglucose as phosphoryl acceptor to a certain extent. The ADPGK of *A. fulgidus* was not able to use hexoses other than D-glucose, and glucosamine and 2-deoxyglucose only to a very low extent. Mouse (Ronimus and Morgan 2004) and human ADPGK were not able to use 2-deoxyglucose. This was also true for the functional mutant rhADPGK-H264A, which was tested for different phosphoryl group acceptors, because of its elevated K_m for glucose and therefore possible significance for substrate binding.

The screen included a sensible selection of candidate molecules that were available in our lab at the time, and, while it is certainly in no way completely comprehensive in addressing all possible interactions between enzyme and substrate, it nicely supports the apparently tight packing that is anticipated when glucose is modelled in the active site by alignment of the hydroxyl groups on ordered water molecules in the crystallised *M. musculus* ADPGK. The interactions between the protein side chains and the inferred positions of the 2′-, 3′-, 4′- and 6′-hydroxyl groups of D-glucose appear to be in a reasonable position for hydrogen bond interactions. The tight packing in the binding site leaves apparently very little room for variation (see Figure 72 on page 139), explaining why no substrate with variations in these positions could be utilised. Furthermore, 2-deoxy-D-glucose was not accepted as a substrate, meaning that the interaction of the 2′-hydroxyl group of glucose and the side chain of glutamate residue E130 is essential for catalysis. The binding site of *M. musculus* ADPGK appears to provide the most space for the 1′-hydroxyl group of glucose (when compared to the other positions), possibly enabling accommodation of the α- and the β-anomers of D-glucose. A functional mutant of *H. sapiens* ADPGK, rhADPGK-H264A, was also unable to transfer the phosphoryl group to an acceptor other than D-glucose, including variations in the 1′-position (see section 3.3.4). The mutant was selected for screening for different phosphoryl group acceptors, because of its elevated K_m for glucose and therefore possible role for this residue in substrate binding. The fact that eukaryotic ADPGK is apparently unable to accept a phosphoryl

acceptor modified in the 2'-position could have interesting implications for probes used in positron emission tomography (PET), an imaging technique widely in visualization of malignant tumour tissues which exploits the elevated glucose consumption of tumour cells. Much research is conducted towards the development of new improved PET imaging probes, but the most frequently used for PET imaging of tumours is 2-^{18}F-fluoro-2-deoxy-D-glucose which is metabolized after uptake into the cells by the hexokinases to 2-^{18}F-fluoro-2-deoxy-D-glucose-6-phosphate, which in turn can not be further metabolised and can be visualized by the PET scanning technique (Oriuchi et al. 2006; Zhu et al. 2013). A eukaryotic ADPGK unable to convert 2FDG to 2FDG6P could therefore lead to an underestimation of the total conversion of glucose to glucose-6-phosphate, which is an important metric in diagnosis.

4.3 Outlook and future experiments

4.3.1 Outlook and future experiments

While the work presented in this thesis has led to a suite of interesting discoveries about ADPGK, many open questions remain. In particular, the crystal structure of mouse ADPGK, which was solved to a resolution of 2.1 Å in this work, revealed some surprising findings. From a technical point of view, to overcome the difficulties met when attempting to crystallise eukaryotic ADPGK in the same crystal form as the mouse ADPGK crystals presented in this work, the exact length of an N-terminal truncation could be crucial. Analysis of the crystal packing has shown that the proline residue P53 could be required to end the N-terminal alpha helix at the exact right point (see Figure 79 on page 155) to facilitate the formation of the crystal contacts at the N-terminus. The generation of expression constructs with the equivalent truncation of the zebrafish and human ADPGK would therefore be a sensible continuation of the project. In addition, the crystallised protein was derived from a cDNA clone of isoform 3 of mouse ADPGK, which differs from the ADPGK clones derived from human and zebrafish cDNA used in this work insofar, that it is missing a glutamine residue in position 313-314. The residue forms a crystal contact and is situated in such a way on the solvent facing side of an α-helix, that an additional polar residue could disrupt the crystal packing as well as the hydrophobic packing on the protein-core facing side of the helix. The deletion of this residue from the expression constructs for human and zebrafish ADPGK is therefore likely worthwhile.

With all the work that has been done on the enzymatic characterisation of human and mouse ADPGK (Ronimus and Morgan 2004; Richter et al. 2012), it is only logical to try and obtain structures of eukaryotic ADPGK with substrate or enzymatic product bound in the catalytic site. The majority of the archaeal ADPGK structures were crystallised in the presence of ADP or AMP and D-glucose together (Ito et al. 2001, 2003; Rivas-Pardo et al. 2013). Attempts to co-crystallise eukaryotic ADPGK with a suitable ligand, which were undertaken for this work, were not successful. However, these experiments were conducted without prior knowledge, and lack of a successful crystallisation condition, but were added to high throughput screens to promote crystal formation in the first place. Should it not be possible to crystallise eukaryotic ADPGK in the presence of a ligand, SAXS measurements of the enzyme-ligand

complex could reveal more information about the catalytic mechanism, which has been successful for the ADPGK of *T. litoralis* (Rivas-Pardo et al. 2013).

Due to time constraints it was not possible to finish the work on the functional mutants of human ADPGK. In the light of the structure of mouse ADPGK, certainly the glutamate residue E130 in the active site is worth further investigation (see Figure 72 on page 139), not only in regard to the potential impact for PET scanning probes. Investigation of this functional mutant would be facilitated by the fact that an expression construct is already available. Also, the role of cysteine residue C82 for substrate binding could prove to be interesting. One of the surprising findings in the crystal structure was the presence of two cysteine residues in so close proximity, that they were able to form a disulphide bond (see Figure 64 on page 129). Given that the structure of the eukaryotic ADPGK and the archaeal ADPGKs are so similar, it is very possible that the protein is still able to fold in its native state (or something very close) without losing its stability when both cysteine residues are removed by site-directed mutagenesis. The contribution to the protein's stability could be assessed by thermal denaturation of the disulphide-free mutant and wild type, while following the unfolding by circular dichroism spectroscopy. Additionally, it would be very interesting to see what effect the removal of the disulphide bond has on the nucleotide-binding loop of ADPGK. The removal of the disulphide could lead to a more flexible nucleotide-binding loop and altered substrate specificity. This mutant could be used in combination with mutations addressing the ADP/ATP selectivity, which was previously unsuccessful (see sections 3.3.3 and 3.3.4).

A cholesterol-binding domain has been experimentally confirmed for eukaryotic ADPGK (Hulce et al. 2013). However, the nature of this domain is not entirely clear and has been discussed in section 4.2.1. An investigation of this interaction certainly would have important implications for the role of ADPGK in eukaryotic metabolism. To identify the cholesterol binding domain, the fact that the π-interaction formed by the central aromatic amino acid with cholesterol is absolutely essential for the functionality of the motif (Fantini and Barrantes 2013) could be used, to change candidate residues by site-directed mutagenesis. However, these experiments would likely have to be conducted either *in vivo*, or with the protocol used by Kamiński *et al.* (2012) to determine the ER localisation of ADPGK.

5 References

Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* 66 (Pt 2): 213–221.

Alberty, R. A. 1953. The Relationship between Michaelis Constants, Maximum Velocities and the Equilibrium Constant for an Enzyme-catalyzed Reaction. *Journal of the American Chemical Society* 75 (8): 1928–1932.

Baier, C. J., Fantini, J. and Barrantes, F. J. 2011. Disclosure of cholesterol recognition motifs in transmembrane domains of the human nicotinic acetylcholine receptor. *Scientific Reports* 1 (69): 1–7.

Banaszak, K., Mechin, I., Obmolova, G., Oldham, M., Chang, S. H., Ruiz, T., Radermacher, M., Kopperschläger, G. and Rypniewski, W. 2011. The crystal structures of eukaryotic phosphofructokinases from baker's yeast and rabbit skeletal muscle. *Journal of Molecular Biology* 407 (2): 284–297.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. and Miyano, S. 2002. Extensive feature detection of N-terminal protein sorting signals. *Bioinformatics* 18 (2): 298–305.

Bianchi, M., Serafini, G., Bartolucci, E., Giammarini, C. and Magnani, M. 1998. Enzymatic properties of overexpressed human hexokinase fragments. *Molecular and Cellular Biochemistry* 189 (1-2): 185–193.

Braakman, I. and Bulleid, N. J. 2011. Protein folding and modification in the mammalian endoplasmic reticulum. *Annual Review of Biochemistry* 80: 71–99.

Bradford, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* 72: 248–254.

Brown, J. M. 1999. The hypoxic cell: a target for selective cancer therapy--eighteenth Bruce F. Cain Memorial Award lecture. *Cancer Research* 59 (23): 5863–5870.

Brown, J. M. and Wilson, W. R. 2004. Exploiting tumour hypoxia in cancer treatment. *Nature Reviews Cancer* 4 (6): 437–447.

Brunger, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355: 472–475.

Cabrera, R., Babul, J. and Guixé, V. 2010. Ribokinase family evolution and the role of conserved residues at the active site of the PfkB subfamily representative, Pfk-2 from *Escherichia coli*. *Archives of Biochemistry and Biophysics* 502 (1): 23–30.

Campobasso, N., Mathews, I. I., Begley, T. P. and Ealick, S. E. 2000. Crystal structure of 4-methyl-5-beta-hydroxyethylthiazole kinase from *Bacillus subtilis* at 1.5 A resolution. *Biochemistry* 39: 7868–7877.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography* 66 (Pt 1): 12–21.

Choi, H., Kim, S., Mukhopadhyay, P., Cho, S., Woo, J., Storz, G. and Ryu, S. 2001. Structural basis of the redox switch in the OxyR transcription factor. *Cell* 105: 103–113.

Cleland, W. W. 1963. The kinetics of enzyme-catalyzed reactions with two or more substrates or products: I. Nomenclature and rate equations. *Biochimica et Biophysica Acta* 67: 104–137.

Copeland, R. A. 2000. *Enzymes: A Practical Introduction to Structure, Mechanism and Data Analysis*. Wiley.

Copeland, R. A. 2005. *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists*. Wiley.

Cséke, C., Weeden, N. F., Buchanan, B. B. and Uyeda, K. 1982. A special fructose bisphosphate functions as a cytoplasmic regulatory metabolite in green leaves. *Proceedings of the National Academy of Sciences of the United States of America* 79 (14): 4322–4326.

Currie, M. A., Merino, F., Skarina, T., Wong, A. H. Y., Singer, A., Brown, G., Savchenko, A., Caniuguir, A., Guixé, V., Yakunin, A. F. and Jia, Z. 2009. ADP-dependent 6-phosphofructokinase from *Pyrococcus horikoshii* OT3: structure determination and biochemical characterization of PH1645. *The Journal of Biological Chemistry* 284 (34): 22664–22671.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. and Richardson, D. C. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 35 (Web Server issue): W375–83.

DeLano, W. L. 2002. *The PyMOL User's Manual*. DeLano Scientific.

Dong, A., Xu, X. and Edwards, A. M. 2007. *In situ* proteolysis for protein crystallization and structure determination. *Nature Methods* 4 (12): 1019–1021.

Dörr, C., Zaparty, M., Tjaden, B., Brinkmann, H. and Siebers, B. 2003. The hexokinase of the hyperthermophile *Thermoproteus tenax*. *The Journal of Biological Chemistry* 278 (21): 18744–18753.

Emanuelsson, O., Brunak, S., Von Heijne, G. and Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. *Nature Protocols* 2: 953–971.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. 2010. Features and development of *Coot*. *Acta Crystallographica Section D: Biological Crystallography* 66 (Pt 4): 486–501.

Evans, P. R., Farrants, G. W., Hudson, P. J. and Britton, H. G. 1981. Phosphofructokinase: structure and control. *Philosophical Transactions of the Royal Society B: Biological Sciences* 293 (1063): 53–62.

Evans, P. 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D: Biological Crystallography* 62: 72–82.

Fantini, J. and Barrantes, F. J. 2013. How cholesterol interacts with membrane proteins: an exploration of cholesterol-binding sites including CRAC, CARC, and tilted domains. *Frontiers in Physiology* 4 (31): 1–9.

Fath, S., Bauer, A. P., Liss, M., Spriestersbach, A., Maertens, B., Hahn, P., Ludwig, C., Schäfer, F., Graf, M. and Wagner, R. 2011. Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. *PLOS ONE* 6 (3): e17596.

Furuya, E. and Uyeda, K. 1980. An activation factor of liver phosphofructokinase. *Proceedings of the National Academy of Sciences of the United States of America* 77 (10): 5861–5864.

Gabriel, O. and Gerstent, D. M. 1992. Staining for enzymatic activity after gel electrophoresis. *Analytical Biochemistry* 203 (1): 1–180.

Gallivan, J. P. and Dougherty, D. a. 1999. Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America* 96 (17): 9459–9464.

Gasteiger, E. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* 31 (13): 3784–3788.

González-Alvarez, R., Ortega-Cuellar, D., Hernández-Mendoza, A., Moreno-Arriola, E., Villaseñor-Mendoza, K., Gálvez-Mariscal, A., Pérez-Cruz, M. E., Morales-

Salas, I. and Velázquez-Arellano, A. 2009. The hexokinase gene family in the zebrafish: structure, expression, functional and phylogenetic analysis. *Comparative Biochemistry and Physiology - Part B: Biochemistry & Molecular Biology* 152 (2): 189–195.

Gronquist, D. and Berges, J. A. 2013. Effects of aquarium-related stressors on the zebrafish: a comparison of behavioral, physiological, and biochemical indicators. *Journal of Aquatic Animal Health* 25 (1): 53–65.

Grossbard, L. and Schimke, R. T. 1966. Multiple Hexokinases of Rat Tissues. *Journal of Biological Chemistry* 241: 3546–3560.

Guixé, V. and Merino, F. 2009. The ADP-dependent sugar kinase family: kinetic and evolutionary aspects. *IUBMB life* 61 (7): 753–761.

Hansen, T., Reichstein, B., Schmid, R. and Schoenheit, P. 2002. The First Archaeal ATP-Dependent Glucokinase , from the Hyperthermophilic Crenarchaeon *Aeropyrum pernix*, Represents a Monomeric , Extremely Thermophilic ROK Glucokinase with Broad Hexose Specificity. *Journal of Bacteriology* 184 (21): 5955–5965.

Hardie, D. G. and Hawley, S. A. 2001. AMP-activated protein kinase: the energy charge hypothesis revisited. *BioEssays* 23 (12): 1112–1119.

Heikkinen, S., Suppola, S., Malkki, M., Deeb, S. S., Jänne, J. and Laakso, M. 2000. Mouse hexokinase II gene: structure, cDNA, promoter analysis, and expression pattern. *Mammalian Genome* 11 (2): 91–96.

Hole, R. 2009. Mammalian ADP-dependendent glucokinase, Massey University.

Hulce, J. J., Cognetta, A. B., Niphakis, M. J., Tully, S. E. and Cravatt, B. F. 2013. Proteome-wide mapping of cholesterol-interacting proteins in mammalian cells. *Nature Methods* 10 (3): 259–264.

Inaba, K. 2009. Disulfide bond formation system in *Escherichia coli*. *Journal of Biochemistry* 146 (5): 591–597.

Ito, S., Fushinobu, S., Yoshioka, I., Koga, S., Matsuzawa, H. and Wakagi, T. 2001. Structural basis for the ADP-specificity of a novel glucokinase from a hyperthermophilic archaeon. *Structure* 9 (3): 205–214.

Ito, S., Fushinobu, S., Jeong, J.-J., Yoshioka, I., Koga, S., Shoun, H. and Wakagi, T. 2003. Crystal Structure of an ADP-dependent glucokinase from Pyrococcus furiosus: implications for a sugar-induced conformational change in ADP-dependent kinase. *Journal of Molecular Biology* 331 (4): 871–883.

Iynedjian, P. B. 2009. Molecular physiology of mammalian glucokinase. *Cellular and Molecular Life Sciences* 66 (1): 27–42.

Jancarick, J. and Kim, S. 1991. Sparse matrix sampling: a screening method crystallization of proteins. *Journal of Applied Crystallography* 24 (4): 409–411.

Jetton, T. L., Liang, Y., Pettepher, C. C., Zimmerman, E. C., Cox, F. G., Horvath, K., Matschinsky, F. M. and Magnuson, M. A. 1994. Analysis of upstream glucokinase promoter activity in transgenic mice and identification of glucokinase in rare neuroendocrine cells in the brain and gut. *The Journal of Biological Chemistry* 269 (5): 3641–3654.

Kabsch, W. 2010. XDS. *Acta Crystallographica Section D: Biological Crystallography* 66: 125–132.

Kadokura, H. and Beckwith, J. 2010. Mechanisms of oxidative protein folding in the bacterial cell envelope. *Antioxidants & Redox Signaling* 13 (8): 1231–1246.

Käll, L., Krogh, A. and Sonnhammer, E. L. L. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Research* 35 (Web Server issue): W429–32.

Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J.-I. and Nagata, Y. 2004. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* 12 (3): 429–438.

Kamiński, M. M., Sauer, S. W., Kamiński, M., Opp, S., Ruppert, T., Grigaravičius, P., Grudnik, P., Gröne, H.-J., Krammer, P. H. and Gülow, K. 2012. T cell activation is driven by an ADP-dependent glucokinase linking enhanced glycolysis with mitochondrial reactive oxygen species generation. *Cell Reports* 2 (5): 1300–1315.

Kamiński, M. M., Röth, D., Krammer, P. H. and Gülow, K. 2013. Mitochondria as oxidative signaling organelles in T-cell activation: physiological role and pathological implications. *Archivum Immunologiae et Therapiae Experimentalis* 61 (5): 367–384.

Kelley, L. A. and Sternberg, M. J. E. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* 4 (3): 363–371.

Kelly, S. M., Jess, T. J. and Price, N. C. 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta* 1751 (2): 119–139.

Kengen, S. W., Tuininga, J. E., De Bok, F. A., Stams, A. J. and De Vos, W. M. 1995. Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *The Journal of Biological Chemistry* 270 (51): 30453–30457.

Kengen, S. W., De Bok, F. A, Van Loo, N. D., Dijkema, C., Stams, A. J. and De Vos, W. M. 1994. Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. *The Journal of Biological Chemistry* 269 (26): 17537–17541.

Klug, M., Heinz, S., Gebhard, C., Schwarzfischer, L., Krause, S. W., Andreesen, R. and Rehli, M. 2010. Active DNA demethylation in human postmitotic cells correlates with activating histone modifications, but not transcription levels. *Genome biology* 11 (6): R63.

Koga, S., Yoshioka, I., Sakuraba, H., Takahashi, M., Sakasegawa, S., Shimizu, S. and Ohshima, T. 2000. Biochemical characterization, cloning, and sequencing of ADP-dependent (AMP-forming) glucokinase from two hyperthermophilic

archaea, *Pyrococcus furiosus* and *Thermococcus litoralis*. *Journal of Biochemistry* 128 (6): 1079–1085.

Kornberg, A., Rao, N. N. and Ault-Riche, D. 1999. Inorganic polyphosphate: a molecule of many functions. *Annual Review of Biochemistry* 68: 98–125.

Krissinel, E. and Henrick, K. 2007. Inference of macromolecular assemblies from crystaline state. *Journal of Molecular Biology* 327: 774–797.

Kwon, M.-G., Park, M.-A., Hwang, J.-Y., Baeck, G.-W., Jung, J.-M., Kim, J.-W., Park, D.-W. and Park, C. 2012. Expression analysis of lipopolysaccharide-and/or concanavalin A/phorbol myristate acetate-stimulated black rockfish peripheral blood leukocytes using a cDNA microarray to identify up-regulated genes. *Genes & Genomics* 34 (2): 111–123.

Labes, A. and Schönheit, P. 2003. ADP-dependent glucokinase from the hyperthermophilic sulfate-reducing archaeon Archaeoglobus fulgidus strain 7324. *Archives of Microbiology* 180 (1): 69–75.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680–685.

Larion, M. and Miller, B. G. 2010. Global fit analysis of glucose binding curves reveals a minimal model for kinetic cooperativity in human glucokinase. *Biochemistry* 49 (41): 8902–8911.

Larion, M. and Miller, B. G. 2012. Homotropic allosteric regulation in monomeric mammalian glucokinase. *Archives of Biochemistry and Biophysics* 519: 103–111.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A, McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23 (21): 2947–2948.

Li, H. and Papadopoulos, V. 1998. Peripheral-type benzodiazepine receptor function in cholesterol transport. Identification of a putative cholesterol recognition/interaction amino acid sequence and consensus pattern. *Endocrinology* 139 (12): 4991–4997.

Lodish, H., Berk, A., Kaiser, C. A., Krieger, M., Bretscher, A., Ploegh, H., Amon, A. and Scott, M. P. 2012. *Molecular Cell Biology*. Freeman & Company, W. H., .

Mabon, M. E., Mao, X., Jiao, Y., Scott, B. A. and Crowder, C. M. 2009. Systematic identification of gene activities promoting hypoxic death. *Genetics* 181 (2): 483–496.

Maj, M. C., Singh, B. and Gupta, R. S. 2002. Pentavalent ions dependency is a conserved property of adenosine kinase from diverse sources: identification of a novel motif implicated in phosphate and magnesium ion binding and substrate inhibition. *Biochemistry* 41 (12): 4059–4069.

Mallick, P., Boutz, D. R., Eisenberg, D. and Yeates, T. O. 2002. Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proceedings of the National Academy of Sciences of the United States of America* 99 (15): 9679–9684.

Mathews, I. I., Erion, M. D. and Ealick, S. E. 1998. Structure of human adenosine kinase at 1.5 A resolution. *Biochemistry* 37 (45): 15607–15620.

Matthews, B. W. 1968. Solvent Content of Protein. *Journal of Molecular Biology* 33: 491–497.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* 40 (Pt 4): 658–674.

McGaughey, G. B., Gagnes, M. and Rappe, A. K. 1998. π-Stacking Interactions. Alive and well in proteins. *Journal of Biological Chemistry* 273 (25): 15458–15463.

McNicholas, S., Potterton, E., Wilson, K. S. and Noble, M. E. M. 2011. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D: Biological Crystallography* 67 (Pt 4): 386–394.

Merino, F. and Guixé, V. 2011. On the specialization history of the ADP-dependent sugar kinase family. *In*: Friedberg, F. (ed.), *Gene Duplication*, 237 – 256. .

Minoux, H. and Chipot, C. 1999. Cation-π Interactions in proteins : can simple models provide an accurate description? *Journal of the American Chemical Society* 121 (44): 10366–10372.

Mukai, T., Kawai, S., Mori, S., Mikami, B. and Murata, K. 2004. Crystal structure of bacterial inorganic polyphosphate/ATP-glucomannokinase. *The Journal of Biological Chemistry* 279 (48): 50591–50600.

Nelson, D. L. and Cox, M. M. 2012. *Lehninger Principles of Biochemistry*. Freeman & Company, W. H., .

Notredame, C., Higgins, D. G. and Heringa, J. 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* 302 (1): 205–217.

Olive, C., Geroch, M. E. and Levy, H. R. 1971. Glucose dehydrogenase from *Leuconostoc mesenteroides*. *The Journal of Biological Chemistry* 246 (7): 2047–2057.

Oriuchi, N., Higuchi, T., Ishikita, T., Miyakubo, M., Hanaoka, H., Iida, Y. and Endo, K. 2006. Present role and future prospects of positron emission tomography in clinical oncology. *Cancer Science* 97 (12): 1291–1297.

Pastorino, J. G. and Hoek, J. B. 2008. Regulation of hexokinase binding to VDAC. *Journal of Bioenergetics and Biomembranes* 40 (3): 171–182.

Petersen, T. N., Brunak, S., Von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8 (10): 785–786.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25 (13): 1605–1612.

Pilkis, S. J., El-maghrabi, M. R., Pilkis, J., Claus, T. H. and Cumming, D. A. 1981. Fructose-2,6-Bisphosphate. A new activator of phosphofructokinase. *The Journal of Biological Chemistry* 256 (7): 3171–3174.

Pollard-Knight, D. and Cornish-Bowden, A. 1982. Mechanism of liver glucokinase. *Molecular and Cellular Biochemistry* 44 (2): 71–80.

Ponstingl, H., Henrick, K. and Thornton, J. M. 2000. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins* 41 (1): 47–57.

Postic, C., Shiota, M. and Magnuson, M. A. 2001. Cell-specific roles of glucokinase in glucose homeostasis. *Recent Progress in Hormone Research* 56: 195–217.

Prinz, W. A., Åslund, F., Holmgren, A. and Beckwith, J. 1997. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *Journal of Biological Chemistry* 272 (25): 15661–15667.

Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* 7 (1): 95–99.

Richter, S., Morrison, S., Connor, T., Su, J., Print, C. G., Ronimus, R. S., McGee, S. L. and Wilson, W. R. 2013. Zinc finger nuclease mediated knockout of ADP-dependent glucokinase in cancer cell lines: effects on cell survival and mitochondrial oxidative metabolism. *PLOS ONE* 8 (6): e65267.

Richter, S., Richter, J. P., Mehta, S. Y., Gribble, A. M., Sutherland-Smith, A. J., Stowell, K. M., Print, C. G., Ronimus, R. S. and Wilson, W. R. 2012. Expression and role in glycolysis of human ADP-dependent glucokinase. *Molecular and Cellular Biochemistry* 328: 131–145.

Rivas-Pardo, J. A., Herrera-Morande, A., Castro-Fernandez, V., Fernandez, F. J., Vega, M. C. and Guixé, V. 2013. Crystal Structure, SAXS and kinetic mechanism of hyperthermophilic ADP-dependent glucokinase from *Thermococcus litoralis* reveal a conserved mechanism for catalysis. *PLOS ONE* 8 (6): 1–12.

Ronimus, R. S. and Morgan, H. W. 2001. The biochemical properties and phylogenies of phosphofructokinases from extremophiles. *Extremophiles* 5 (6): 357–373.

Ronimus, R. S. and Morgan, H. W. 2003. Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* 1 (3): 199–221.

Ronimus, R. S. and Morgan, H. W. 2004. Cloning and biochemical characterization of a novel mouse ADP-dependent glucokinase. *Biochemical and Biophysical Research Communications* 315 (3): 652–658.

Rossmann, M. G., Moras, D. and Olsen, K. W. 1974. Chemical and biological evolution of nucleotide-binding protein. *Nature* 250: 194–199.

Roy, A., Kucukural, A. and Zhang, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols* 5 (4): 725–738.

Sakuraba, H., Yoshioka, I., Koga, S., Takahashi, M., Kitahama, Y., Satomura, T., Kawakami, R. and Ohshima, T. 2002. ADP-dependent glucokinase/phosphofructokinase, a novel bifunctional enzyme from the hyperthermophilic archaeon *Methanococcus jannaschii*. *The Journal of Biological Chemistry* 277 (15): 12495–12498.

Van Schaftingen, E., Hue, L. and Hers, H.-G. 1980a. Fructose 2, 6-bisphosphate, the probable structure of the glucose-and glucagon-sensitive stimulator of phosphofructokinase. *Biochemical Journal* 192: 897–901.

Van Schaftingen, E., Hue, L. and Hers, H.-G. 1980b. Control of the fructose 6-phosphate/fructose 1,6-bisphosphate cycle in isolated hepatocytes by glucose and glucagon. *Biochemical Journal* 192: 887–895.

Van Schaftingen, E., Jett, M. F., Hue, L. and Hers, H.-G. 1981. Control of liver 6-phosphofructokinase by fructose 2,6-bisphosphate and other effectors. *Proceedings of the National Academy of Sciences of the United States of America* 78 (6): 3483–3486.

Schirmer, T. and Evans, P. R. 1990. Structural basis of the allosteric behaviour of phosphofructokinase. *Nature* 343: 140–145.

Schut, G. J., Brehm, S. D., Datta, S. and Adams, M. W. W. 2003. Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon : *Pyrococcus furiosus* grown on carbohydrates or peptides. *Journal of Bacteriology* 185 (13): 3935–3947.

Semenza, G. L. 2007. Life with oxygen. *Science* 318 (5847): 62–64.

Shaw, N., Cheng, C. and Liu, Z.-J. 2007. Procedure for reductive methylation of protein to improve crystallizability. *Protocol Exchange*.

Siebers, B., Tjaden, B., Michalke, K., Dörr, C., Ahmed, H., Zaparty, M., Gordon, P., Sensen, C. W., Zibat, A., Klenk, H.-P., Schuster, S. C. and Hensel, R. 2004. Reconstruction of the central carbohydrate metabolism of Thermoproteus tenax by use of genomic and biochemical data. *Journal of Bacteriology* 186 (7): 2179–2194.

Siebers, B. and Schönheit, P. 2005. Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Current Opinion in Microbiology* 8 (6): 695–705.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. and Higgins, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7 (539): 539.

Sigrell, J. A., Cameron, A. D., Jones, T. A. and Mowbray, S. L. 1998. Structure of *Escherichia coli* ribokinase in complex with ribose and dinucleotide determined to 1.8 A resolution: insights into a new family of kinase structures. *Structure* 6 (2): 183–193.

Small, I., Peeters, N., Legeai, F. and Lurin, C. 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. *Proteomics* 4 (6): 1581–1590.

Sonnhammer, E. L. L., von Heijne, G. and Krogh, A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings of the 6th International Conference on Intelligent Systems for Molecular Biology* 1: 175–182.

Storer, A. C. and Cornish-Bowden, A. 1976. Kinetics of rat liver glucokinase. *Biochemical Journal* 159: 7–14.

Storer, A. C. and Cornish-Bowden, A. 1977. Kinetic evidence for a "mnemonical" mechanism for rat liver glucokinase. *The Biochemical Journal* 165 (1): 61–69.

Studier, F. W. 2005. Protein Production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* 41: 207–234.

Tanaka, S., Lee, S., Hamaoka, K., Kato, J., Takiguchi, N., Nakamura, K., Ohtake, H. and Kuroda, A. 2003. Strictly polyphosphate-dependent glucokinase in a polyphosphate-accumulating bacterium, *Microlunatus phosphovorus*. *Journal of Bacteriology* 185 (18): 5654–5656.

Towbin, H., Staehelin, T. and Gordon, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets : Procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* 76 (9): 4350–4354.

Tsuge, H., Sakuraba, H., Kobe, T., Kujime, A., Katunuma, N. and Ohshima, T. 2002. Crystal structure of the ADP-dependent glucokinase from *Pyrococcus horikoshii*

at 2.0-Å resolution : A large conformational change in ADP-dependent glucokinase. *Protein Science* (1): 2456–2463.

Tuininga, J. E., Verhees, C. H., Van Der Oost, J., Kengen, S. W. M., Stams, A. J. M. and De Vos, W. M. 1999. Molecular and biochemical characterization of the ADP-dependent phosphofructokinase from the hyperthermophilic archaeon Pyrococcus furiosus. *Journal of Biological Chemistry* 274 (30): 21023–21028.

Uyeda, K., Furuya, E. and Sherry, A. D. 1981. The structure of "activation factor" for phosphofructokinase. *Journal of Biological Chemistry* 256: 8679–8684.

Verhees, C. H., Tuininga, J. E., Kengen, S., Stams, A. J. M., Van Der Oost, J. and De Vos, W. M. 2001. ADP-dependent phosphofructokinases in mesophilic and thermophilic methanogenic archaea. *Journal of Bacteriology* 183 (24): 7145–7153.

Verhees, C. H., Koot, D. G. M., Ettema, T. J. G., Dijkema, C., De Vos, W. M. and Van Der Oost, J. 2002. Biochemical adaptions of two sugar kinases from the hyperthermophilic archaeon Pyrococcus furiosus. *Biochemical Journal* 366: 121–127.

Verhees, C. H., Kengen, S. W. M., Tuininga, J. E., Schut, G. J., Adams, M. W. W., De Vos, W. M. and Van Der Oost, J. 2003. The unique features of glycolytic pathways in Archaea. *The Biochemical Journal* 375 (Pt 2): 231–246.

Vincentelli, R., Canaan, S., Campanacci, V., Valencia, C., Maurin, D., Frassinetti, F., Scappucini-Calvo, L., Bourne, Y., Cambillau, C. and Bignon, C. 2004. High-throughput automated refolding screening of inclusion bodies. *Protein Science* 13: 2782–2792.

Voet, D. and Voet, J. G. 2011. *Biochemistry*. Wiley.

Walter, T. S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J. E., Owens, R. J., Stuart, D. I. and Grimes, J. M. 2006. Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* 14 (11): 1617–1622.

Wang, L., Sun, N., Terzyan, S., Zhang, X. and Benson, D. R. 2006. A histidine/tryptophan pi-stacking interaction stabilizes the heme-independent folding core of microsomal apocytochrome b5 relative to that of mitochondrial apocytochrome b5. *Biochemistry* 45 (46): 13750–13759.

Warburg, O. H. 1947. *Ideen zur Fermentchemie der Tumoren*. Akademie-Verlag.

Warburg, O. H. 1956. On respiratory impairment in cancer cells. *Science* 124: 269–270.

Van de Werken, H. J. G., Verhees, C. H., Akerboom, J., De Vos, W. M. and Van Der Oost, J. 2006. Identification of a glycolytic regulon in the archaea *Pyrococcus* and *Thermococcus*. *FEMS Microbiology Letters* 260 (1): 69–76.

Wernimont, A. and Edwards, A. 2009. In situ proteolysis to generate crystals for structure determination: an update. *PLOS ONE* 4 (4): e5094.

Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D. and Hochstrasser, D. F. 1999. Protein identification and analysis tools in the ExPASy server. *Methods in Molecular Biology* 112: 571–607.

Wilson, J. E. 2003. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *Journal of Experimental Biology* 206 (12): 2049–2057.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. a, Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* 67 (Pt 4): 235–242.

Woese, C. R., Kandler, O. and Wheelis, M. L. 1990. Towards a natural system of organisms : proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences of the United States of America* 87: 4576–4579.

Wu, M. M., Llopis, J., Adams, S., McCaffery, J. M., Kulomaa, M. S., Machen, T. E., Moore, H. P. and Tsien, R. Y. 2000. Organelle pH studies using targeted avidin and fluorescein-biotin. *Chemistry & Biology* 7 (3): 197–209.

Xie, G. C. and Wilson, J. E. 1988. Rat brain hexokinase: the hydrophobic N-terminus of the mitochondrially bound enzyme is inserted in the lipid bilayer. *Archives of Biochemistry and Biophysics* 267 (2): 803–810.

Xu, L. Z., Harrison, R. W., Weber, I. T. and Pilkis, S. J. 1995. Human beta-cell glucokinase: dual role of Ser151 in catalysis and hexose affinity. *The Journal of Biological Chemistry* 270 (17): 9939–9946.

Zhu, L., Ploessl, K. and Kung, H. F. 2013. Expanding the scope of fluorine tags for PET imaging. *Science* 342: 429–430.

6 Appendix

6.1 Primers

Table 6.1: Primers.

Target	Description	Sequence
Vector	M13 reverse sequencing primer (-48)	AGCGGATAACAATTTCACACAGG
Vector	T7 Terminator sequencing primer	GCTAGTTATTGCTCAGCGG
pBAD vector	Amplify any pBAD insert for directional TOPO cloning	CACCATGGCCTCTGGATCCGGTGATG
pBAD vector	Amplify any pBAD insert for TOPO cloning	CATCACCGGATCCAGAGGCCATGGTG
pBAD vector	Amplify any pBAD insert for TOPO cloning reverse primer	ACCGGTCATCATCATCATCATCATTGAGTT
D. rerio ADPGK (wt)	Sequencing primer (765)	GCGAATCGCTTCATTCTCTC
D. rerio ADPGK (wt)	Sequencing primer (-933)	CAGCCTCTTTGAGTCGTTCC
D. rerio ADPGK (wt)	Cloning full length protein into pET21-b(+) (BamHI)	GCGTTACGGATCCGATGTGGAGGAAGGCTGTG CTGGTG
D. rerio ADPGK (wt)	Cloning into pProEX HTb, pET-32a(+), pGEX-4T-3, pET21-b(+), reverse primer	CCGCTCGAGAACCTCAACCTCAGACGTGAGCT CTGA
D. rerio ADPGK (wt)	TOPO coning Δ1-45 protein into pBAD	TCAGAACCCCGCCCACCC
D. rerio ADPGK (wt)	TOPO coning Δ1-51 protein into pBAD	ACACTGGAGGAGGCCATC
D. rerio ADPGK (wt)	TOPO coning into pBAD, pET101/D and pET151/D reverse primer	AACCTCAGACGTGAGCTC
D. rerio ADPGK (wt)	Cloning full-length protein into pET-32a(+) (NcoI, mutation W2G)	CATACGCCATGGGGAGGAAGGCTGTGC
D. rerio ADPGK (wt)	TOPO cloning full-length protein into pBAD	ATGTGGAGGAAGGCTGTGCTG
D. rerio ADPGK (wt)	Cloning full-length protein into pET101/D and pET151/D	CACCATGTGGAGGAAGGCTGTGCTG
D. rerio ADPGK (wt)	Cloning Δ1 protein into pGEX-4T-3 (BamHI)	GCTTACGGATCCTGGAGGAAGGCTGTG
H. sapiens ADPGK (wt)	Sequencing primer (791)	ACATGATGGAGGGACAAAGC
H. sapiens ADPGK (wt)	Sequencing primer (-1008)	AGAGGCTGACTGGGTGAGAA
H. sapiens ADPGK (wt)	Cloning Δ1 protein into pProEX HTb	GACGGAATTCACGCGCTGTGGCGCGGCTCC
H. sapiens ADPGK (wt)	Cloning Δ1-25 protein into pProEX HTb	CGGAATTCACGAGCTGCCAGGCTCGGCG
H. sapiens ADPGK (wt)	Cloning Δ1-50 protein into pProEX HTb	CGGAATTCACGTCTCCCCCGAGGGCCGG
H. sapiens ADPGK (wt)	Cloning Δ1-86 protein into pProEX HTb	CGGAATTCACCTCTCAGGGGTGAAGCT

H. sapiens ADPGK (wt)	Cloning Δ27-496 protein into pProEX HTb	GCGGAAGCTTCTATTATTATGGCTCCAGCAGT AAGAC
H. sapiens ADPGK (wt)	Cloning Δ52-496 protein into pProEX HTb	GCGGAAGCTTCTATTATTAGACGGGTCCCGGG
H. sapiens ADPGK (wt)	Cloning Δ87-496 protein into pProEX HTb	GCGGAAGCTTCTATTATTACACCACATCAACA CATGCATTGACTCC
H. sapiens ADPGK (wt)	Cloning into pProEX HTb, reverse primer	GCGGAAGCTTCTACTAATAGTGAGGGTGTACT TC
H. sapiens ADPGK (wt)	Cloning full-length protein into pET21-b(+) (NdeI)	GCGTTACCATATGGCGCTGTGGCGCGGCTCC
H. sapiens ADPGK (wt)	Cloning Δ1-25 protein into pET21-b(+) (NdeI)	GCGTTACCATATGGAGCTGCCAGGCTCGGCG
H. sapiens ADPGK (wt)	Cloning Δ1-50 protein into pET21-b(+) (NdeI)	GCGTTACCATATGGTCTCCCCCGAGGGCCGG
H. sapiens ADPGK (wt)	Cloning into pET21-b(+) and pGEX-4T-3 reverse primer (XhoI)	CCGCTCGAGATAGTGAGGGTGTACTTC
H. sapiens ADPGK (wt)	Cloning into pET21-b(+) reverse primer with stop codon (XhoI)	CCGCTCGAGCTAATAGTGAGGGTGTACTTC
H. sapiens ADPGK (wt)	Cloning Δ1 protein into pGEX-4T-3 (EcoRI)	CGGAATTCCGCGCTGTGGCGCGGCTCC
H. sapiens ADPGK (wt)	Cloning Δ1-25 protein into pGEX-4T-3 (EcoRI)	CGGAATTCCGAGCTGCCAGGCTCGGCG
H. sapiens ADPGK (wt)	Cloning Δ1-50 protein into pGEX-4T-3 (EcoRI)	CGGAATTCCGTCTCCCCCGAGGGCCGG
H. sapiens ADPGK (wt)	Cloning Δ1-50 protein into pMAL-p2c/-c2c	GACGGAATTCACGCGCTGTGGCGCGGCTCC
H. sapiens ADPGK (wt)	TOPO cloning full-length protein into pBAD including N-terminal NcoI site	ACCATGGCGCTGTGGCGCGGCTCC
H. sapiens ADPGK (wt)	TOPO into pBAD reverse primer	ATAGTGAGGGTGTACTTCCGA
H. sapiens ADPGK (wt)	TOPO cloning into pBAD reverse primer including stop codon	CTAATAGTGAGGGTGTACTTC
H. sapiens ADPGK (wt)	TOPO cloning Δ1 protein into pBAD	GCGCTGTGGCGCGGCTCCGCG
H. sapiens ADPGK (wt)	Cloning Δ1 protein into into pMAL-p2c/-c2c	GACGGAATTCACGCGCTGTGGCGCGGCTCC
H. sapiens ADPGK (wt)	Cloning into pMAL-p2c/-c2c reverse primer	GCGGAAGCTTCTACTAATAGTGAGGGTGTACT TC
H. sapiens ADPGK (wt)	Cloning full-length protein into pET-32a(+) (NcoI)	GCCATACGCCATGGCGCTGTGGCGCGGCTCC
H. sapiens ADPGK (wt)	Cloning full-length protein into pFBDM	TCCCACCATCGGGCGATGGCGCTGTGGCGCGG CTCC
H. sapiens ADPGK (wt)	Cloning into pFBDM, reverse primer	AGCGGTTTCTTTACCATAGTGAGGGTGTACTT C
H. sapiens ADPGK (wt)	Mutation D481A	CGAACTGTAGGCCTTGGA**GCT**GCCATTTCAGC CGAAGGA

H. sapiens ADPGK (opt)	Sequencing primer (600)	GCACCGCATGCAAATCGTTT
H. sapiens ADPGK (opt)	Sequencing primer (-600)	AAACGATTTGCATGCGGTGC
H. sapiens ADPGK (opt)	Sequencing primer (1100)	CATTTTCATACCCTGGTGTA
H. sapiens ADPGK (opt)	Sequencing primer (-1100)	TACACCAGGGTATGAAAATG
H. sapiens ADPGK (opt)	TOPO cloning full-length protein into pBAD	ATGGCACTGTGGCGTGGTAGC
H. sapiens ADPGK (opt)	TOPO cloning Δ1-24 protein into pBAD	CCGGAACTGCCTGGTAGCGCACTG
H. sapiens ADPGK (opt)	TOPO cloning Δ1-50 protein into pBAD	GTTAGTCCGGAAGGTCGTCTGGCAGCA
H. sapiens ADPGK (opt)	TOPO cloning Δ1-51 protein into pBAD	AGTCCGGAAGGTCGTCTGGC
H. sapiens ADPGK (opt)	TOPO cloning Δ1-66 protein into pBAD	CGTCCGGTTCGTCGTTGGCGTCG
H. sapiens ADPGK (opt)	TOPO cloning Δ1-70 protein into pBAD	CGTTGGCGTCGTGTTGCAGTTGGTGTT
H. sapiens ADPGK (opt)	TOPO cloning Δ1-151 protein into pBAD	GCACAGCATTATGTTGGTGGTAATGCAGCACTGA
H. sapiens ADPGK (opt)	TOPO cloning into pBAD reverse primer	ATAATGCGGATGAACTTCGCTATAAAACAG
H. sapiens ADPGK (opt)	Cloning full length protein into pET21-b(+) (BamHI)	CGGGGTCCGATGGCACTGTGGCGTGGT
H. sapiens ADPGK (opt)	Cloning Δ1 protein into pET-32a(+), pMAL-p2c/-c2c, pProEX HTb and pGEX-4T-3 (BamHI)	CGGGGTCCGCACTGTGGCGTGGTAGCG
H. sapiens ADPGK (opt)	Cloning into pMAL-p2c/-c2c reverse primer with stop codons (HindIII)	CGAAGCTTGTTATTATTAATAATGCGGATGAAC
H. sapiens ADPGK (opt)	Cloning into pProEX HTb. pET-32a(+) pGEX-4T-3 pET21-b(+) reverse primer with stop codons (XhoI)	GGTACCTCGAGTTATTATTAATAATGCGGATGAAC
H. sapiens ADPGK (opt)	Cloning full-length protein into pET-32a(+) (NcoI)	GCTCCATGGCACTGTGGCGTG
H. sapiens ADPGK (opt)	Cloning into pProEX HTb. pET-32a(+) pET101/D pET21-b(+) reverse primer (XhoI)	GGTACCTCGAGATAATGCGGA
H. sapiens ADPGK (opt)	TOPO cloning full-length protein into pET101/D and pET151/D	CACCATGGCACTGTGGCGTG
H. sapiens ADPGK (opt)	Mutation D84A	GTTAATGCATGTGTT**GCT**GTTGTTCTGAGC
H. sapiens ADPGK (opt)	Mutation E130A	AAAGGTGCAGCAGCC**GCA**CGTTTTTTTAGC
H. sapiens ADPGK (opt)	Mutation H207A	GAAGTTGACGAATTT**GCT**CTGATTCTGGAA
H. sapiens ADPGK (opt)	Mutation H264A	GTGCTGAGCGGTCTG**GCT**ATGATGGAAGGT
H. sapiens ADPGK (opt)	Mutation D481N	ACCGTTGGTCTGGGT**AAT**GCAATTAGCGCA
H. sapiens ADPGK (opt)	Mutation D481E	ACCGTTGGTCTGGGT**GAA**GCAATTAGCGCA
H. sapiens ADPGK (opt)	Mutation D481S	ACCGTTGGTCTGGGT**TCT**GCAATTAGCGCA
H. sapiens ADPGK (opt)	Mutation R408A	GTTGCAGCCGGTGCA**GCT**GTTGCAGGCACCCAG

H. sapiens ADPGK (opt)	Mutation H382V	ACCCGTATTCATTTT**GTT**ACCCTGGTGTAT
H. sapiens ADPGK (opt)	Mutation H382A	ACCCGTATTCATTTT**GCU** ACCCTGGTGTAT
H. sapiens ADPGK (opt)	Mutation H387V	ACCCTGGTGTAT**GTT**ATTCTGGCAACCGTT
H. sapiens ADPGK (opt)	Mutation D481A (Mutagenex)	TCTGGGTGCTGCAATTAGCGC
H. sapiens ADPGK (opt)	Mutation R228A (Mutagenex)	TGCAAATGCTTTTATTTTT
H. sapiens ADPGK (opt)	Mutation H387A (Mutagenex)	GTGTATGCTATTCTGG
M. musculus ADPGK (wt)	Sequencing primer (932)	TGTGCATCAGCAGGTCTTTC
M. musculus ADPGK (wt)	Sequencing primer (-932)	GAAAGACCTGCTGATGCACA
M. musculus ADPGK (wt)	Cloning full-length protein into pET101/D and pET151/D	CACCATGGCGCTGTGGCGCGGCTCTG
M. musculus ADPGK (wt)	Cloning into pET101/D and pET151/D reverse primer	ATCAGGGCGTGCTTCTGAATAGAAGAGGC
M. musculus ADPGK (wt)	Cloning Δ1 protein into pProEX HTb	CGGAATTCACGCGCTGTGGCGCGGCTCTG
M. musculus ADPGK (wt)	Cloning Δ1-25 protein into pProEX HTb	CGGAATTCACGAGCTGCCGGGCACGGCG
M. musculus ADPGK (wt)	Cloning Δ1-50 protein into pProEX HTb	CGGAATTCACCTGTCCCCGGAGAGCCGGC
M. musculus ADPGK (wt)	Cloning into pProEX HTb reverse primer	GCGGAAGCTTCTACTAATCAGGGCGTGCTTCT G
M. musculus ADPGK (wt)	TOPO cloning Δ1-24 protein into pBAD	CCGGAGCTGCCGGGCAC
M. musculus ADPGK (wt)	TOPO cloning Δ1-49 protein into pBAD	CCTCTGTCCCCGGAGAGCCG
M. musculus ADPGK (wt)	TOPO cloning Δ1-51 protein into pBAD	CTGTCCCCGGAGAGCCGGCT
M. musculus ADPGK (wt)	TOPO cloning Δ1-74 protein into pBAD	GTGGCTGTGGGAGTCAACGCCTG
M. musculus ADPGK (wt)	TOPO cloning Δ1-148 protein into pBAD	TTTCCAGGAGCTCAGCATTATGTAGGAGGA
M. musculus ADPGK (wt)	TOPO into pBAD reverse primer	ATCAGGGCGTGCTTCTGAATAGAAGAGG
M. musculus ADPGK (wt)	Mutation C417A	GTAGCTGGGACACAAGCC**GCT**GCCACAGAAAC CATAG
M. musculus ADPGK (wt)	Mutation C470A	CACACCTGTGTTGGTG**GCT**AAAGACCCCGTTC GAAC

6.2 cDNA sequences

Table 6.2: Nucleotide sequence of optimised *H. sapiens* ADPGK

```
GAGCTCCATGGCACTGTGGCGTGGTAGCGCATATGCAGGTTTTCTGGCACTGGCAGTTGGTTGTGTTTTTCTGCTGGAACCGGAAC
TGCCTGGTAGCGCACTGCGTAGCCTGTGGTCAAGCCTGTGTCTGGGTCCGGCACCGGCACCTCCGGGTCCGGTTAGTCCGGAAGGT
CGTCTGGCAGCAGCATGGGATGCACTGATTGTTCGTCCGGTTCGTCGTTGGCGTCGTGTTGCAGTTGGTGTTAATGCATGTGTTGA
TGTTGTTCTGAGCGGTGTTAAACTGCTGCAGGCACTGGGTCTGAGTCCGGGTAATGGTAAAGATCATTCAATTCTGCATAGCCGCA
ATGATCTGGAAGAAGCCTTTATTCATTTTATGGGTAAAGGTGCAGCAGCCGAACGTTTTTTTAGCGATAAAGAAACCTTTCATGAT
ATTGCACAGGTGGCAAGCGAATTTCCGGGTGCACAGCATTATGTTGGTGGTAATGCAGCACTGATTGGTCAGAAATTTGCAGCAAA
TAGCGATCTGAAAGTTCTGCTGTGTGGTCCGGTTGGTCCGAAACTGCATGAACTGCTGGATGATAATGTTTTTGTTCCTCCGGAAA
GCCTGCAAGAAGTTGACGAATTTCATCTGATTCTGGAATATCAGGCAGGCGAAGAATGGGGTCAGCTGAAAGCACCGCATGCAAAT
CGTTTTATTTTTAGCCATGATCTGAGCAATGGTGCCATGAATATGCTGGAAGTTTTTGTTAGCAGCCTGGAAGAATTTCAGCCGGA
TCTGGTTGTGCTGAGCGGTCTGCATATGATGGAAGGTCAGAGCAAAGAACTGCAGCGTAAACGTCTGCTGGAAGTTGTTACCAGCA
TTAGCGATATTCCGACCGGTATTCCGGTTCATCTGGAACTGGCAAGCATGACCAATCGTGAACTGATGAGCAGCATTGTTCATCAG
CAGGTTTTTCCGGCAGTTACCAGCCTGGGTCTGAATGAACAAGAACTGCTGTTTCTGACCCAGAGCGCAAGCGGTCCGCATAGCAG
CCTGAGCAGCTGGAATGGTGTTCCGGATGTTGGTATGGTGAGCGATATTCTGTTTTGGATTCTGAAAGAACATGGTCGTAGCAAAA
GCCGTGCAAGCGATCTGACCCGTATTCATTTTCATACCCTGGTGTATCATATTCTGGCAACCGTTGATGGCCATTGGGCAAATCAG
CTGGCAGCAGTTGCAGCCGGTGCACGTGTTGCAGGCACCCAGGCATGTGCAACCGAAACCATTGATACCAGCCGTGTTAGCCTGCG
TGCACCGCAAGAATTTATGACCAGCCATAGCGAAGCAGGTAGCCGTATTGTTCTGAATCCGAATAAACCGGTTGTTGAATGGCATC
GTGAAGGTATTAGCTTTCATTTTACACCGGTTCTGGTTTGCAAAGATCCGATTCGTACCGTTGGTCTGGGTGATGCAATTAGCGCA
GAAGGTCTGTTTTATAGCGAAGTTCATCCGCATTATTAATAATAACTCGAGGTACC
```

Table 6.3: Nucleotide sequence of *H. sapiens* ADPGK (wt)

```
ATGGCGCTGTGGCGCGGCTCCGCGTACGCGGGCTTCCTGGCGCTGGCCGTGGGCTGCGTCTTCCTGCTGGAGCCAGAGCTGCCAGG
CTCGGCGCTGCGCTCTCTCTGGAGCTCGCTGTGTCTGGGGCCCGCGCCTGCGCCCCCGGGACCCGTCTCCCCCGAGGGCCGGTTGG
CGGCAGCCTGGGACGCGCTTATCGTGCGGCCAGTCCGGCGCTGGCGCCGCGTGGCAGTGGGAGTCAATGCATGTGTTGATGTGGTG
CTCTCAGGGGTGAAGCTCTTGCAGGCACTTGGCCTTAGTCCTGGGAATGGGAAAGATCACAGCATTCTGCATTCAAGGAATGATCT
GGAAGAAGCCTTCATTCACTTCATGGGGAAGGGAGCAGCTGCTGAGCGCTTCTTCAGTGATAAGGAAACTTTTCACGACATTGCCC
AGGTTGCGTCAGAGTTCCCAGGAGCCCAGCACTATGTAGGAGGAAATGCAGCTTTAATTGGACAGAAATTTGCAGCCAACTCAGAT
TTAAAGGTTCTTCTTTGCGGTCCAGTTGGTCCAAAGCTACATGAGCTTCTTGATGACAATGTCTTTGTTCCACCAGAGTCATTGCA
GGAAGTGGATGAGTTCCACCTCATTTTAGAGTATCAAGCAGGGGAGGAGTGGGGCCAGTTAAAAGCTCCCCATGCCAACCGATTCA
TCTTCTCTCACGACCTCTCCAACGGGGCCATGAATATGCTGGAGGTGTTTGTGTCTAGCCTGGAGGAGTTTCAGCCAGACCTGGTG
GTCCTCTCTGGATTGCACATGATGGAGGGACAAAGCAAGGAGCTCCAGAGGAAGAGACTCTTGGAGGTTGTAACCTCCATTTCTGA
CATCCCCACTGGTATTCCAGTTCACCTAGAGCTGGCCAGTATGACTAACAGGGAGCTCATGAGCAGCATTGTCCATCAGCAGGTCT
TTCCCGCGGTGACTTCCCTTGGGCTGAATGAACAGGAGCTGTTATTTCTCACCCAGTCAGCCTCTGGACCTCACTCTTCTCTCTCT
TCCTGGAACGGTGTTCCTGATGTGGGCATGGTCAGTGACATCCTCTTCTGGATCTTGAAAGAACATGGGAGGAGTAAAAGCAGAGC
CTCGGATCTCACCAGGATCCATTTCCACACGCTGGTCTACCACATCCTGGCAACTGTGGATGGACACTGGGCCAACCAGCTGGCAG
CCGTGGCTGCAGGAGCTCGTGTGGCTGGGACACAGGCCTGCGCCACAGAAACCATAGACACCAGCCGAGTGTCTCTGAGGGCACCC
CAAGAGTTCATGACTTCCCATTCGGAGGCAGGCTCCAGGATTGTATTAAACCCAAACAAGCCAGTAGTAGAATGGCACAGAGAGGG
AATATCCTTCCACTTCACACCAGTATTGGTGTGTAAAGACCCCATTCGAACTGTAGGCCTTGGAGATGCCATTTCAGCCGAAGGAC
TCTTCTATTCGGAAGTACACCCTCACTATTAG
```

Table 6.4: Nucleotide sequence of *M. musculus* ADPGK

```
ACATGGCGCTGTGGCGCGGCTCTGCGTGCGCCGGCTTCCTGGCGCTGGCCGTGGGCTGCGTGTTCCTGCTGGAGCCGGAGCTGCCG
GGCACGGCGCTGCGCTCGCTCTGGAGCTCCCTGAGGCTGGGACCCGCGCCAGTGCCCGTGGGACCTCTGTCCCCGGAGAGCCGGCT
GGCAGCCGCTTGGGACGCCCTCATCGCGCAGCCCGCCCGGCGCTGGCGCCGCGTGGCTGTGGGAGTCAACGCCTGTGTGGATGTGG
TAATCTCTGGCGTGAAACTCCTGCAGGCCCTCGGCCTCAGTCCTGGGAGTGGAAAAGATCATGCTATTCTACACTCGAGAAGTGAT
CTGGAAGAAGCCTTCCTCTACTTCATGGGGAAGGGAGCCGCTGCTGAGCGCTTCTTCAGTGATAAAGAGACCTTTCATGACATTGC
CCAGGCTGCATCTGAGTTTCCAGGAGCTCAGCATTATGTAGGAGGAAATGCTGCTTTAATTGGACAGAGGTTTGCTGCCAACACAG
ACTTAAAGGTTCTTCTTTGTGGTCCAATTGGCCCAAAACTGCATGAACTTCTTGATGACAATGTGTTTGTTCCACCAGAGTCACTG
CAGGAAGAGGATGAGTTCCACCTTATTTTAGAGTACCTGGCAGGGGAGGAGTGGGGCCCGTTTAAAGCTCCCCATGCCAACCGCTT
CATCTTCTCTCACGACCTCTCCAACGGGGCCATGAATATGCTGGAGGTGTTTGTGTCTAGCCTGGAGGAGTTCCAGCCTGACCTGG
TGGTCCTTTCTGGATTGCACATGATGGAGGGACAAAGCAAGGAGCTCCAGAGGAAGAGGCTCTTGGAGGTTGTGACCGCCATTTCT
GACATCCCCACTGGTATTCCAGTTCACCTAGAGCTGGCCAGTATGACCAACAGGGAGCTCATGAGCAGTATTGTGCATCAGGTCTT
TCCCACGGTGGCTTCTCTTGGGCTGAATGAACAGGAGCTGCTGTTTCTCAGCCAGTCGGCATCTGGACCCCATGCTTCTCTCTCCT
CCTGGGATGGTGTTCCTGATGTGGGCATGGTCAGTGACATCCTCTTCTGGATCTTGAAAGAACATGGAAGGAGTGAAAACAGAGCC
TCGGATCTCACCAGGATCCATTTCCACACGCTGGTCTACCACATTCTGGCAACTGTGGACGGACACTGGGCCAACCAGCTCGCAGC
CGTAGCTGCGGGAGCTCGTGTAGCTGGGACACAAGCCTGTGCCACAGAAACCATAGACACCAACCGAGTGTCTCTAAGAGCACCTC
AGGAGTTTACGACATCCCATTTGGAATCCGGTTCCAGGATTGTATTGAACCCAGATAAGCCAGTGGTGGAGTGGCACAGGGAGGGA
ATAACTTTCCACTTCACACCTGTGATCAGCAGGTCTTTCCCACGGTGGTTGGTGTGTAAAGACCCCGTTCGAACTGTAGGCCTTGG
TGATGCCATTTCCGCTGAAGGCCTCTTCTATTCAGAAGCACGCCCTGATTAGACACGTTGTGTGAGGAATGAAAGTCACCCATCTT
AAGAATTCCAGCAAGAACATAGCTTGGGAACTA
```

Table 6.5: Nucleotide sequence of *D. rerio* ADPGK

ATGTGGAGGAAGGCTGTGCTGGTGGCGGTGTTGGCGCTGGGGATGGGTTATCTCTACCACACTGACCCAGAGCTGCCGGCCCGCAT
GCTCAGCTACGTCAGCCAATCACTGCCACCGATGGAGGCTGGTGGAGGATCAGAACCCCGCCCACCCACACTGGAGGAGGCCATCG
CTCGCGCCTGGCAGACCCTCATCACGGCTCCGGCCCGCCGCTGGGGGAGAGTTGCAGTGGGGGTGAACGCCTGTGTGGATGTGGTG
GTGTCCGGCGTGGGGCTGCTGCAGGCGCTCGCTCTAGAACCAGGCTCCGGCAGAGACCACGAGGTGCTACACAGCAAAGAGGACCT
GAAGGAGGCCTTCATCCACTACATGGGGAAGGGCGCAGCAGCAGAGCGTTTCTTCTCAGATAAGGAGGTTTTCCAGAGGATCGCAC
GAGCCGCTGCAGAGTATCCCGGCGCTCAGCTTTATGTGGGCGGAAACGCTGCTCTGATTGGCCAGAAGCTGGCATCTAACCCACAA
CTAGTGGTCCTATTGTGTGGTCCGGTGGGGCCGAAGCTTCACGAGATGCTGGATGAACAGATCGTGGTGCCGCCAGAATCTCTGCA
GGAAACGGATGAGTACCACCTGATCCTGGAGTACAAGTCAGGTGAACAGTGGGGTCCTTCACGAGCTCCTCAGGCGAATCGCTTCA
TTCTCTCCCACGATGTCTCCAATGGGGAAATGAGCGCGCTGGAGATGTTCGTGGCCAGTCTGGAGGAGTTCCAGCCGGATCTGGTG
GTTCTATCCGGACTGCATATGATGGAAGGAATGGGCCGGGAACTCTGGGAGGAACGACTCAAAGAGGCTGTGGTGGCCATCTCAGA
CGTGAGGAACGAGGTGCCCATCCACCTGGAGCTGGCCAGTATGACGGACAGAGACTACATGAACCGAATCCTGCAGGAGCAGGTCA
TGCCCATCGTGAACTCCATCGGTCTGAACGAGCAGGAGCTTTTATTCCTGACCCAATCAGGAGGAGGCCCTCACTCCGACCTCACT
TCATGGGACGGCGTACCAGACGTGGGCCGAGTCACAGACATCCTCCTCTGGGTCCTAGAGCAACACGGCCGGGCGGACGCACAATC
AGAAGCCGACCTCACCCGCATCCACTTCCACACACTAGCATACCACATCCTGGCTACGGTGGACGGGCACTGGGGGAACCAGGCGG
CGGCGGTGGCAGCGGGCGCGCGTGTGGCTAGCAGTCAAGCGTGCGGTTTGCAAACCATCGATATCTCAAAAGTGCTTCTTAAGGCA
CCGTTAAACTTCCACAGCTCCTTCTCGGAGCCGAGAGAGAGTCTAAAAGTGGAGCCGTCCCGACCGGTGACCGTGTGGCGGAGAGG
AAACGTTTCTTTCCATCTCACACCCGTGCTGGTCTGCAAACAGCCGCTGAGGACTGTGGGATTGGGCGACGCTATTTCAGCCGAGG
GACTCGTGTTTTCAGAGCTCACGTCTGAGGTTTAG

6.3 Standard and calibration curves

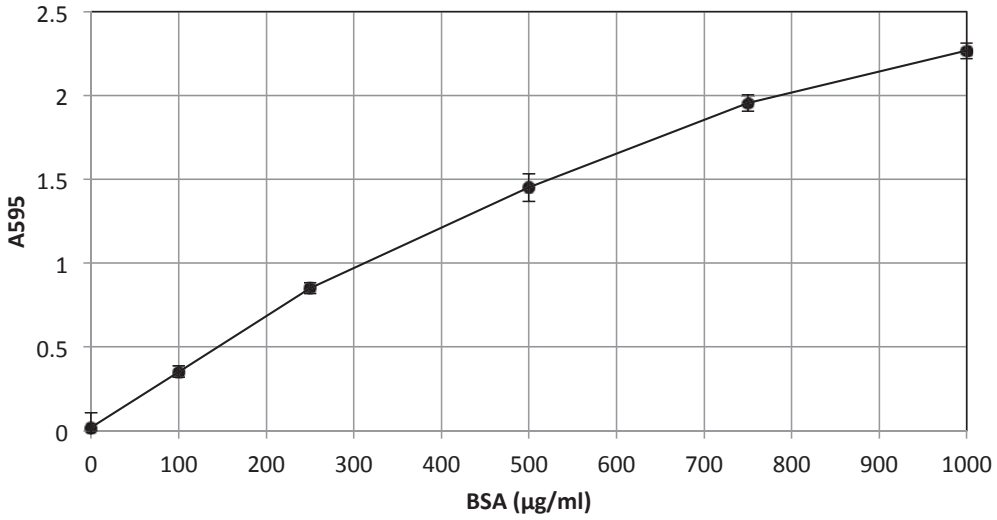

Figure 83: Standard curve for Bradford protein assay.

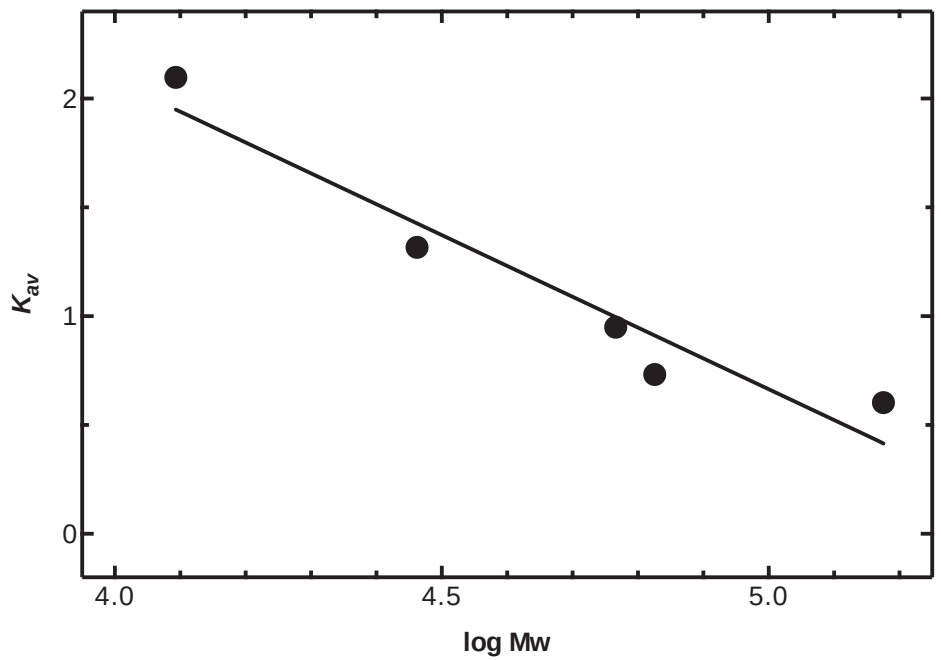

Figure 84: Calibration curve for Superdex S200 size-exclusion chromatography column.

$Y = -1.417 \times x + 7.749$. $R^2 = 0.93$.

6.4 Signal peptide and transmembrane domain predictions

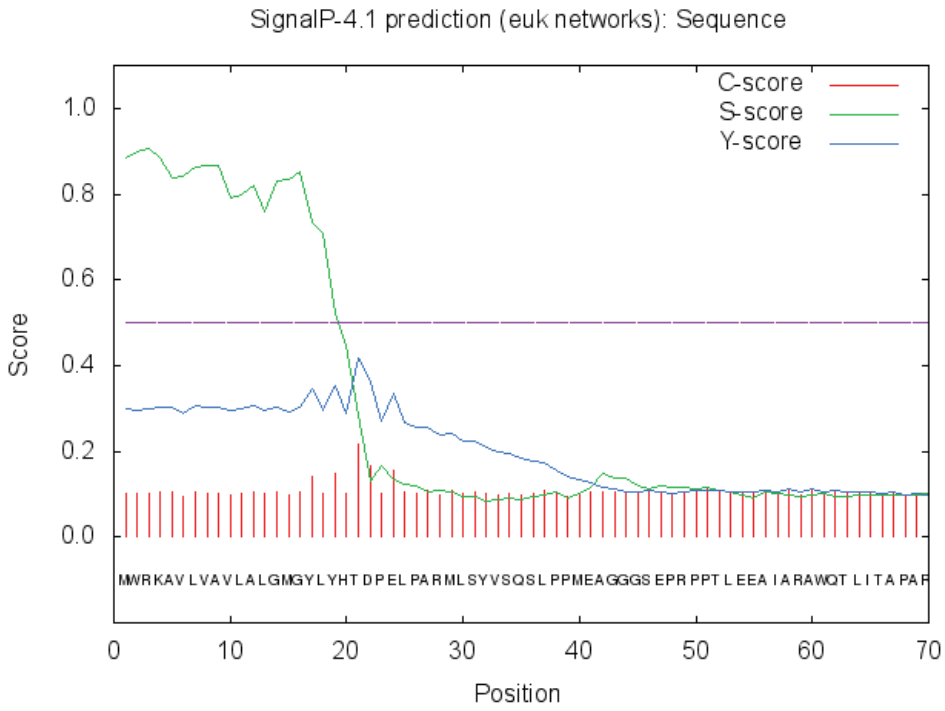

Figure 85: *D. rerio* ADPGK signal peptide prediction from SignalP 4.1.

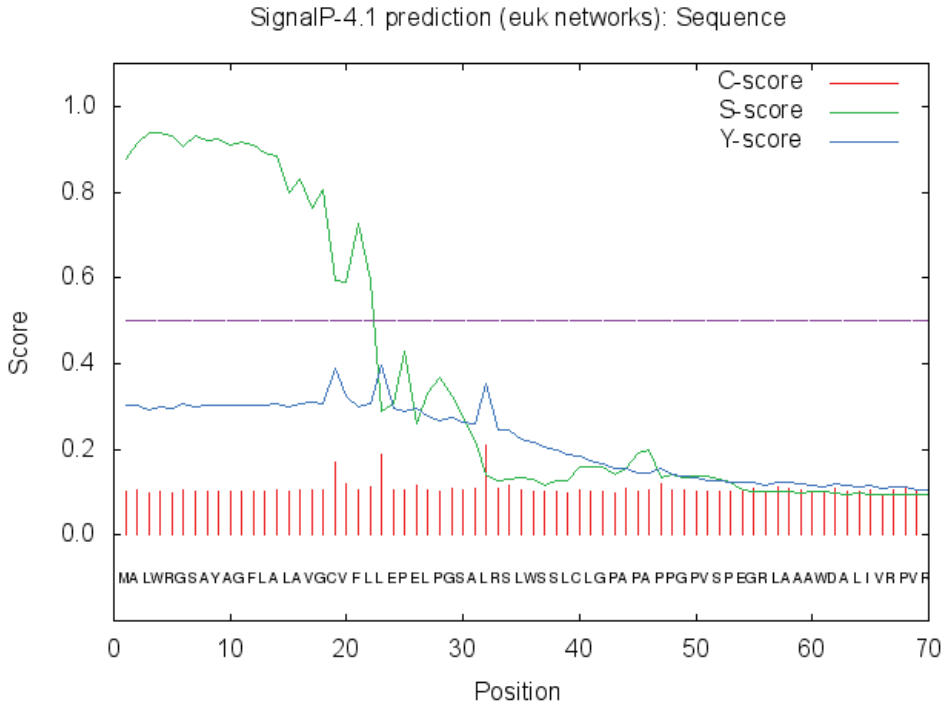

Figure 86: *H. sapiens* ADPGK signal peptide prediction from SignalP 4.1.

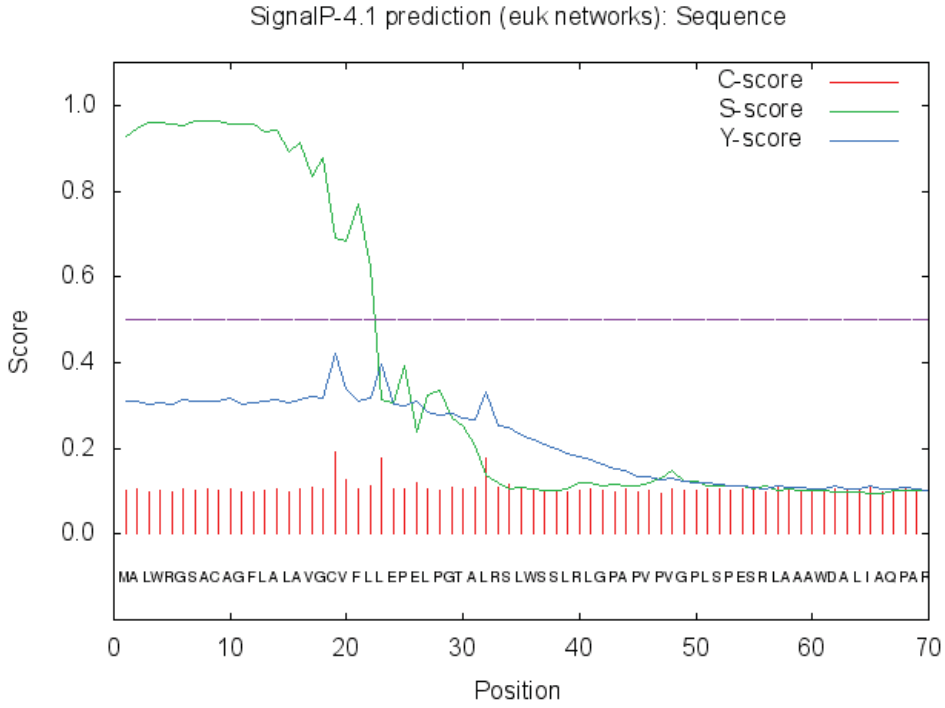

Figure 87: ***M. musculus*** **ADPGK signal peptide prediction from SignalP 4.1**

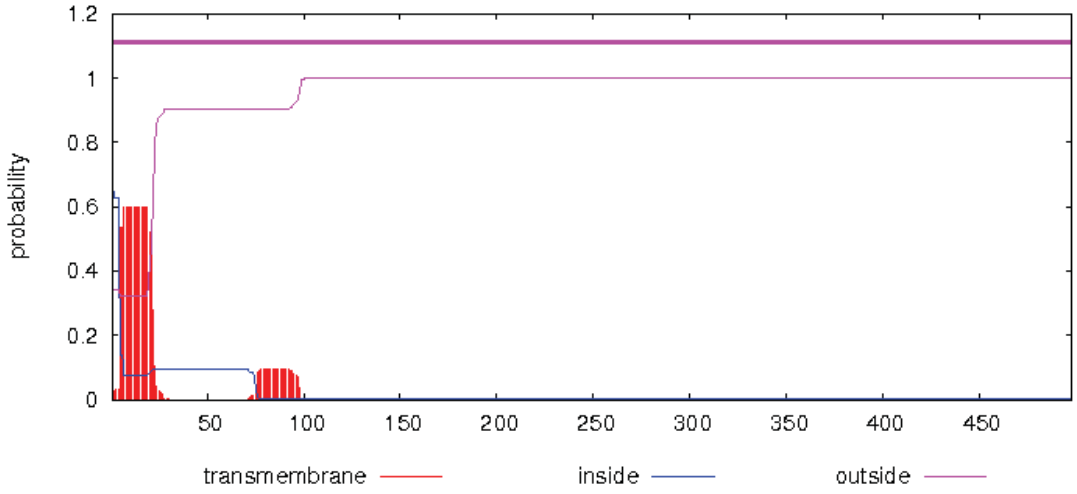

Figure 88: ***D. rerio*** **ADPGK transmembrane domain prediction from TMHMM2.0.**

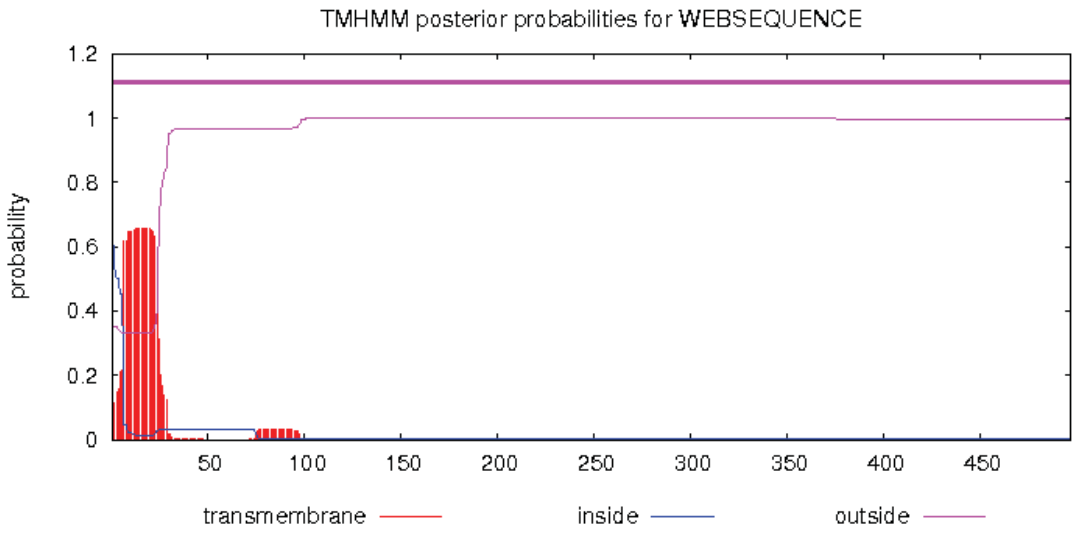

Figure 89: *H. sapiens* ADPGK transmembrane domain prediction from TMHMM2.0.

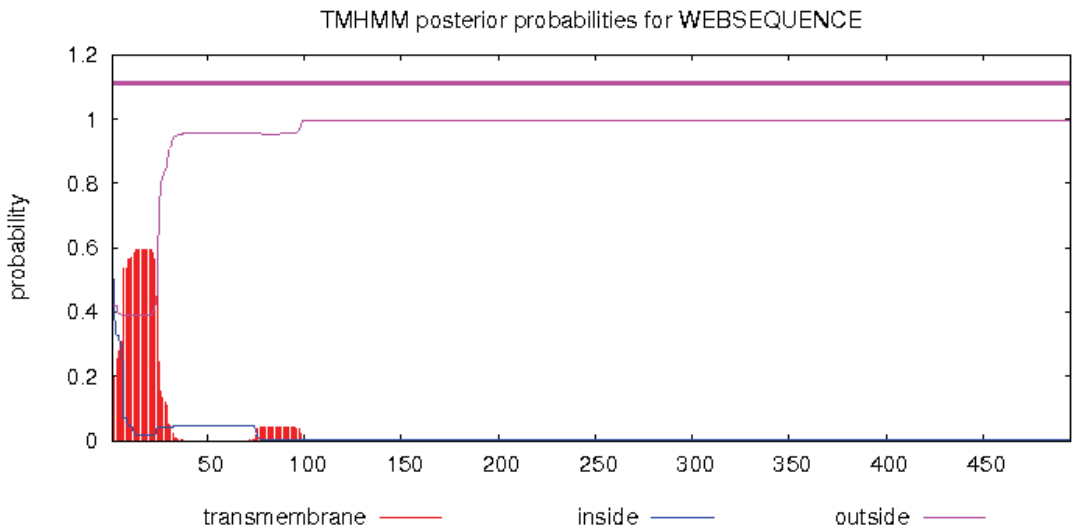

Figure 90: *M. musculus* ADPGK transmembrane domain prediction from TMHMM2.0.

6.5 Sequence alignments

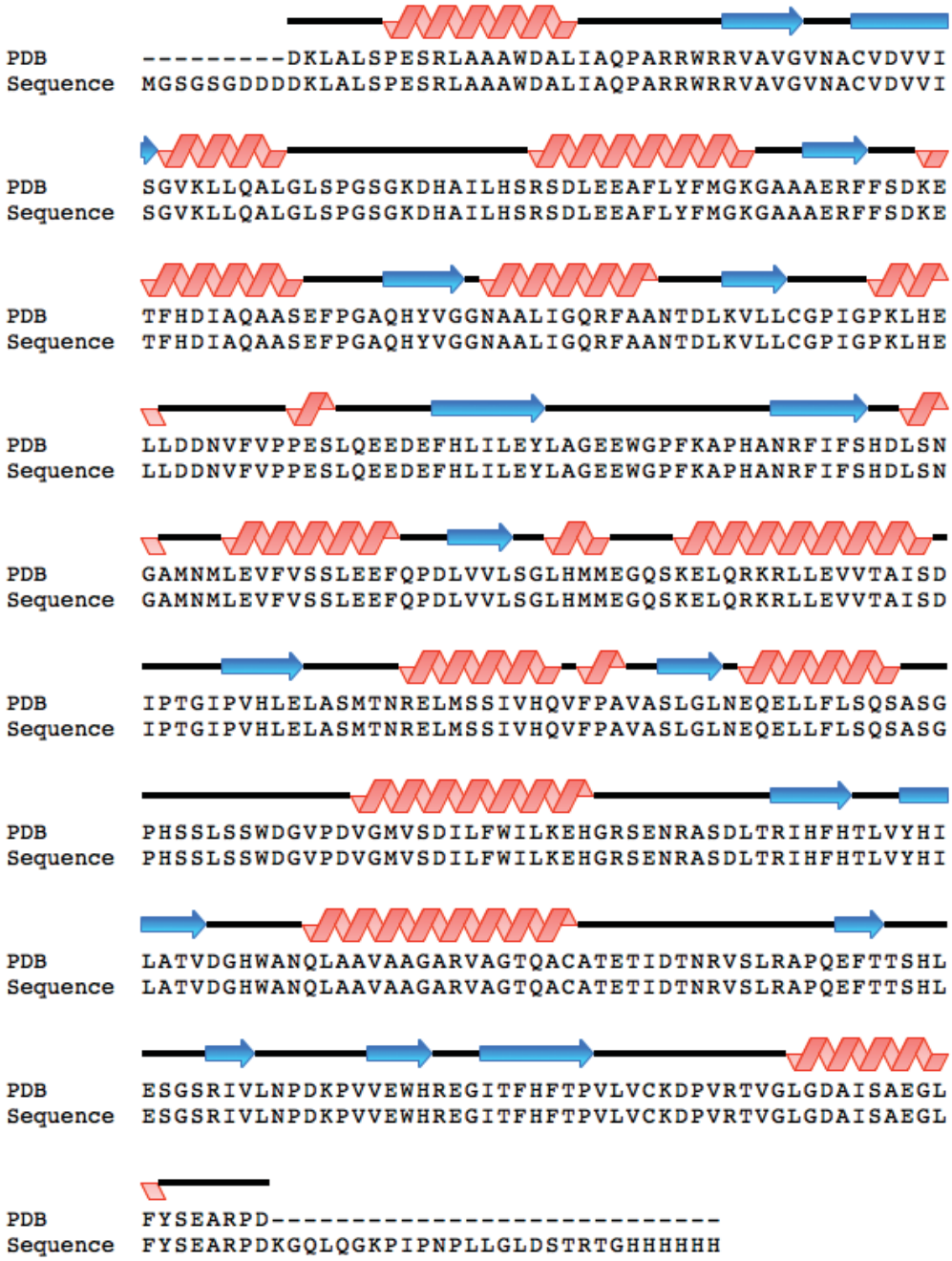

Figure 91: Alignment of sequence of rmADPGKΔ51 and sequence from the PDB file.

Sequence alignment generated with the Phenix validation program (Adams et al. 2010) of rhADPGKΔ51, from vector pBAD-rmADPGKΔ51 (Sequence) and the sequence visible in the electron density of the crystallised protein (PDB). Secondary structure is indicated as red helices, blue strands and other regions are black.