

**Structure-based design of pantothenate
kinase inhibitors as lead structures for new
antibiotics**

Ying Ho Leon Lee



This thesis is submitted in partial fulfillment of the requirements for the degree
of Master in Medical Biology

Department of Biomedicine

University of Bergen

Bergen, Norway

Contents

Acknowledgement	9
Abstract	9
1 Introduction	13
1.1 Need for new antibiotics	13
1.2 Pseudomonas aeruginosa	13
1.3 Biosynthesis of CoA	15
1.4 Pantothenate kinase	16
1.5 Structure-based virtual screening	20
1.6 FlexX	22
1.7 Protein crystallization	23
1.8 Aim of the study	24
2 Materials and Methods	25
2.1 Transformation of cells and overexpression of PanK	25
2.2 Purification of PanK	26
2.3 Gel electrophoresis	27
2.4 Dynamic light scattering	28
2.5 Protein thermal shift assay	28
2.6 Crystallization of PanK	29
2.7 Setup for docking	29
2.8 Database for virtual screening	30
2.9 Docking	31

3	Results	33
3.1	Overexpression of PanK	33
3.2	Purification of PanK	33
3.3	Dynamic light scattering	37
3.4	Thermal shift assay	40
3.5	Crystallization	51
3.6	Virtual screening	52
4	Discussion	57
4.1	Overexpression of PanK	57
4.2	Purification of PanK	57
4.3	Dynamic light scattering and thermal shift assay	58
4.4	PanK crystallization	59
4.5	Virtual screening	60
5	Conclusion and future perspective	63

List of Abbreviations

ATP Adenosine Triphosphate

BLI BioLayer Interferometry

CoA Coenzyme A

CTP Cytidine Triphosphate

CV Column Volume

DLS Dynamic light scattering

DPCK Dephospho-CoA Kinase

dPCoA dephospho-CoA

DTT Dithiothreitol

E. coli *Escherichia coli*

EDTA Ethylenediaminetetraacetic acid

ESBL Extended spectrum beta-lactamases

IPTG Isopropyl- β -D-1-thiogalactoside

LB Lysogeny Broth

LPS Lipopolysaccharide

MDR Multidrug resistant

MW Molecular Weight

P. aeruginosa *Pseudomonas aeruginosa*

PanK Pantothenate Kinase

PDB Protein Data Base

PEG Polyethylene Glycol

PPan 4'-Phosphopantothenate

PPantSH 4'-Phosphopantetheine

PPAT Phosphopantetheine Adenylyltransferase

PPC 4'-Phosphopantothenoylcysteine

PPCDC Phosphopantothenoylcysteine Decarboxylase

PPCS Phosphopantothenoylcysteine Synthetase

rpm rounds per minute

S. aureus *Staphylococcus aureus*

SDS-PAGE Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

TCEP tris(2-carboxyethyl)phosphine

TEV Tobacco Etch Virus

WHO World Health Organization

Acknowledgement

I would like to express my deepest appreciation to my supervisor, Ruth Brenk, for her continuous support, guidance and patience for my work. Your mentorship and advice have helped me to grow as a person and scientist and has prepared me for future challenges. I would like to thank all members of Brenks lab for their friendship, help and creating an enjoyable lab environment. My special thanks to Illimar Hugo Rekand for his help and guidance that helped me to understand the computational work and to Khanh Kim Dao for her help and guidance in the lab. My sincere thanks goes to Charis Georgiou who accepted the role as my co-supervisor. His endless guidance, advice, support and discussions have been a great help in my work and I am very grateful for that. Also, I want to thank NFR BedreHelse INDONOR program for funding this project. Finally, I want to thank my family for their unconditional love, support, patience and understanding.

Abstract

Antibiotic resistance is an increasing problem across the world. The number of infections caused by multidrug resistant bacteria is constantly increasing and the need for new antibiotics is very urgent. *Pseudomonas aeruginosa* is a bacteria that has been listed as one of the most critical bacteria that urgently need new antibiotics against due to the broad spectrum of infections it can cause and its resistance against available antibiotics. Type 3 pantothenate kinase (PanK) from *P. aeruginosa* was chosen as the target for the discovery of starting points for new antibiotics. Protocols for expression and purification of PanK were established. Crystals of PanK were successfully formed but no data could be collected from them. Virtual screening identified compounds that could potentially bind to the pantothenate binding site of PanK. A total amount of 20 compounds were selected as potential inhibitors for PanK.

List of Figures

1.1	The biosynthetic pathway of CoA	16
1.2	Phosphorylation of pantothenate	17
1.3	Type 3 PanK structure	19
1.4	The binding mode of pantothenate	19
1.5	Illustration of structural-based virtual screening	20
2.1	Sequence for expression	26
2.2	Binding site of PanK	30
3.1	SDS-PAGE from the overexpressed PanK	35
3.2	SDS-PAGE from His-tag affinity chromatography and TEV-site cleavage	36
3.3	SDS-PAGE from size exclusion chromatography and His-tag affinity chromatog- raphy using phosphate as lysis and elution buffer	36
3.4	Dynamic light scattering data using Tris-HCl as buffer	38
3.5	Dynamic light scattering data using citrate as buffer	39
3.6	Crystal of PanK	52
3.7	UV picture of the crystal of PanK	52
4.1	Example of a good binding mode	61

Chapter 1

Introduction

1.1 Need for new antibiotics

Antibiotic resistance has been identified as a serious threat to human health and is a rising problem all over the world. The emergence and spreading of new resistance mechanisms are threatening our ability to treat common infections. The frequent use of antibiotics in hospitals, community and agriculture have contributed to the development of resistance. In addition to the cost of human lives, antibiotic resistance has caused countries with high income to use more expensive antibiotics which has caused high economic costs for health care. Furthermore, medical interventions such as surgery, transplantation and chemotherapy can also be hindered due to antibiotic resistant bacteria [1].

1.2 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a rod-shaped, gram-negative bacterium commonly found in soil, water plants and animals. The bacteria species causes a variety of difficult to treat infections. Infections from *P. aeruginosa* can be acquired from a stay in a healthcare facility such as hospital or from the community. Patients with burns or cystic fibrosis are particularly susceptible to nosocomial infections. Infections caused by these bacteria are associated with the biofilms that this bacteria forms on the ventilators. Growing as a complex mass of cells attached to a surface greatly increases the survival of *P. aeruginosa* since the resistant to biocides increases

when compared to growing in a free-floating state [2] [3].

Antimicrobial therapy is becoming more problematic due to the resistance mechanism the bacteria have. There are a number of factors that contribute to the resistance. The cell wall of *P. aeruginosa* has low permeability which provides protection against antimicrobial agents. The genome of *P. aeruginosa* is able to express a wide range of resistance factors and mutations in chromosomal genes which regulate resistance genes can also contribute to resistance. Additionally, other organisms can provide additional resistance genes through plasmids, transposons and bacteriophages [4]. Frequent use of antibiotics in setting where patients are vulnerable to *P. aeruginosa* infection has contributed to the development of resistance [2]. The emergence of carbapenemases in multidrug resistant (MDR) strains of the bacteria has been very concerning in recent years. This is due to the fact that now the MDR strains of the bacteria have become resistant to carbapenems, which were used for the treatment of infections caused by MDR strains. Moreover, colistin resistance in carbapenem resistant *P. aeruginosa* has been observed, where colistin is an antibiotic used when there is no option left [2].

Furthermore, *P. aeruginosa* has several other mechanisms of resistance. One of the mechanisms involves the outer membrane that restricts the rate of penetration of small hydrophilic molecules and excluding larger molecules. β -lactams are small hydrophilic antibiotics that can only cross the outer membrane by passing through the aqueous channels provided by porin proteins. A specialized porin called oprD, is responsible for the uptake of positively charged amino acids and loss of this porin is associated with resistance to the antibiotic imipenem. This is because oprD is required for imipenem to cross the outer membrane [4]. The permeability of the outer membrane itself is already very restricted. The permeability is 12- to 100-fold lower than that of *E. coli* and consists of a bilayer of phospholipid and lipopolysaccharide (LPS), embedded with protein channels [5]. LPS are a target of a group of antibiotics called aminoglycosides. Aminoglycosides work by binding to LPS in order to increase the permeability of the membrane. This will promote the uptake of antibiotics and once they are inside of the cell, protein synthesis will be interfered. Although rarely observed, resistance against these aminoglycosides has been observed in *P. aeruginosa* [4]. Due to the fact that *P. aeruginosa* has become resistant to a large number of antibiotics and its ability to cause deadly infections, World Health Organization

(WHO) has included the species as amongst the most critical group of bacteria for which new antibiotics are urgently needed for research and developing new antibiotics (Table 1.1).

Table 1.1: The priority list of pathogens for research and development of new antibiotic from WHO

Priority category	Pathogens	Antibiotic resistance
1: Critical	<i>Acinetobacter baumannii</i>	Carbapenem-resistant
1: Critical	<i>Pseudomonas aeruginosa</i>	Carbapenem-resistant
1: Critical	<i>Enterobacteriaceae</i>	Carbapenem-resistant, Extended spectrum beta-lactamases (ESBL) producing
2: High	<i>Enterococcus faecium</i>	vancomycin-resistant
2: High	<i>Staphylococcus aureus</i>	methicillin-resistant, vancomycin-intermediate and resistant
2: High	<i>Helicobacter pylori</i>	clarithromycin-resistant
2: High	<i>Campylobacter</i>	fluoroquinolone-resistant
2: High	<i>Salmonellae</i>	fluoroquinolone-resistant
2: High	<i>Neisseria gonorrhoeae</i>	cephalosporin-resistant and fluoroquinolone-resistant
3: Medium	<i>Streptococcus pneumoniae</i>	penicillin-non-susceptible
3: Medium	<i>Haemophilus influenzae</i>	ampicillin-resistant
3: Medium	<i>Shigella</i>	fluoroquinolone-resistant

1.3 Biosynthesis of CoA

Coenzyme A (CoA) is an important cofactor in the Krebs cycle and fatty acid metabolism and is essential for all living organisms [6]. The biosynthesis of CoA (Figure 1.1) is a five step process where the enzyme pantothenate kinase (PanK) is responsible for the first step which will be highlighted in the next section. Phosphopantothenoylcysteine synthetase (PPCS) is the enzyme responsible for the second step in CoA biosynthesis. In this step, PPCS catalyzes the Mg^{2+} -dependent formation of 4'-phosphopantothenoylcysteine (PPC) by using the product from the previous step, 4'-phosphopantothenate (PPan), and L-cysteine. This step differs between eukaryotes and bacteria. In bacteria, cytidine triphosphate (CTP) is used for the activation of the reaction whereas ATP (Adenosine Triphosphate) is used in eukaryotes. This difference has also made PPCS an attractive target for the development of antimicrobial drugs. The enzyme responsible for the third step is called phosphopantothenoylcysteine decarboxylase (PPCDC). In bacteria, PPCS is fused to PPCDC and together they form a bifunctional CoaBC protein. The eukaryotic counterpart of the enzymes does not have this function. The reaction that PPCDC

catalyses is the decarboxylation of the cysteine moiety from PPC which gives the product 4'-phosphopantetheine (PPantSH) [7]. In the fourth step, PPantSH undergoes an adenylation to form dephospho-CoA (dPCoA) and pyrophosphate. Just like the second step, this reaction is dependent on Mg^{2+} . The enzyme responsible for this reaction is phosphopantetheine adenylyl-transferase (PPAT). The eukaryotic version of PPAT is normally fused with dephospho-CoA kinase (DPCK) which is the last enzyme of this pathway. This fusion protein is called CoA synthase and does not exist in bacteria. DPCK catalyzes the phosphorylation of the 3' hydroxy-group of the ribose moiety of dPCoA which results in ultimately CoA. This reaction is dependent on $MgATP$. All enzymes in this pathway except DPCK are considered to be an attractive target for drug development. Bacterial DPCK has high structural and sequence homology to the DPCK domain of eukaryotic CoASy protein, suggesting that developing a selective inhibitor for bacterial DPCK is unlikely [7].

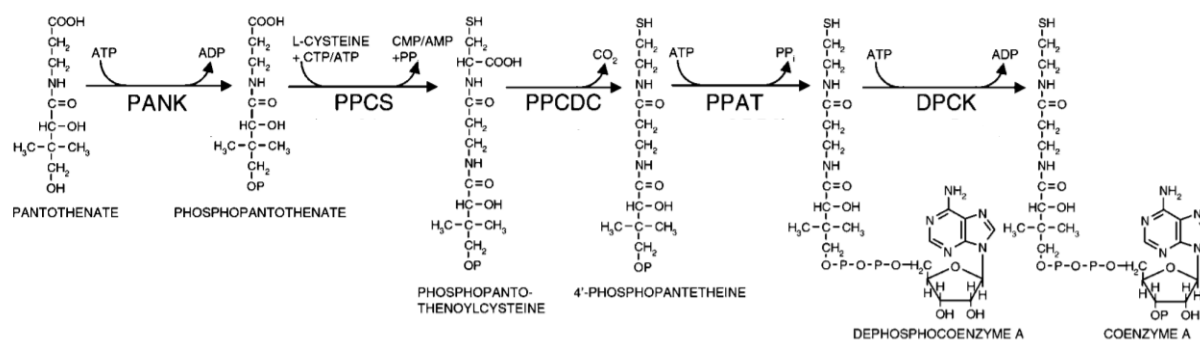


Figure 1.1: The biosynthesis of CoA. Figure is taken from [8] (edited).

1.4 Pantothenate kinase

PanK is an enzyme that catalyzes the ATP-dependent phosphorylation of pantothenate to PPan (Figure 1.2) and is considered as a potential target for the development of new antibiotic. This 1st step of the CoA biosynthesis pathway has been seen as a potential target for antimicrobial drug development for several reasons [7]. This step is believed to be the rate-limiting step of the pathway. Furthermore, there are three different types of PanKs all of which are diverse in terms of sequence, structural fold and properties in catalysis and inhibition. All these factors suggest that developing an inhibitor that selectively binds to the PanK of the pathogenic microorganism of interest is possible [7].

As mentioned in the previous paragraph, there are three different types of PanK. Type 1 PanK is typically found in bacterial species such as *Escherichia coli* and *Mycobacterium tuberculosis*, and is encoded by the *coaA* gene. Type 1 PanK is regulated through feedback inhibition by CoA and its thioesters [9].

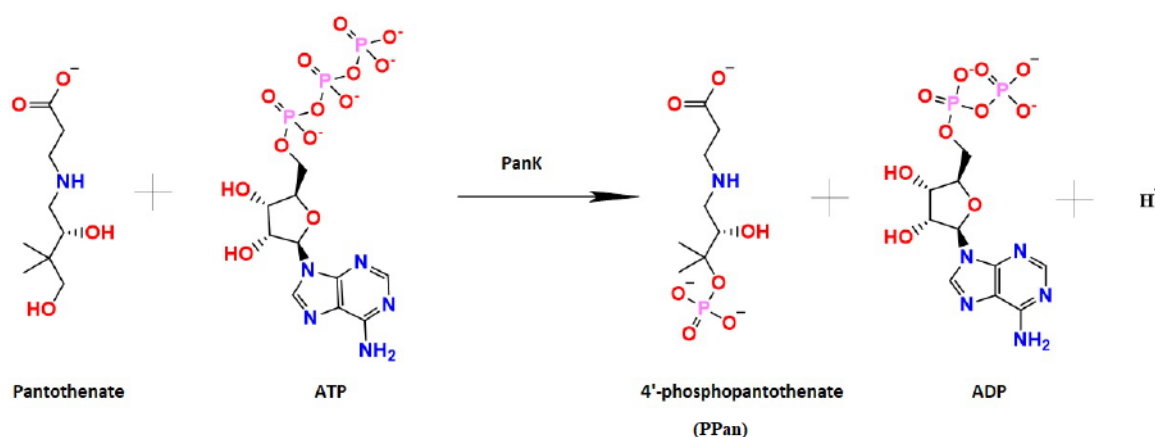


Figure 1.2: The reaction that PanK catalyzes. The phosphoryl group from ATP is transferred to pantothenate yielding PPan.

Type 2 PanK is typically found in *Staphylococcus aureus* and eukaryotic cells. While the amino acid homology between type 1 and type 2 is not very high, they do share the same feedback mechanism. The exception from this rule is *S. aureus* PanK where the feedback regulation of CoA and its thioesters is lacking [9][10]. *N*-alkylpantotheamides belongs to a class of pantothenate analogues that both type 1 and type 2 PanK are able to phosphorylate. The phosphorylation will transform the *N*-alkylpantotheamides into inactive Coa analogues via the organism's native CoA biosynthetic pathway. These CoA analogues will act as antimetabolites and inhibitors of bacterial growth [11].

Type 3 PanK, which occurs in a number of bacteria such as *Pseudomonas aeruginosa* and *Helicobacter pylori*, lacks any amino acid similarity to the other types. This enzyme is also not feedback inhibited by CoA and its thioesters. This PanK forms a dimer (Figure 1.3). Unlike

type 1 and type 2 PanK, type 3 PanK is not able to use *N*-alkylpantotheamides as substrates and it has a very high K_m for ATP (~ 10 mM) [11]. Sequence alignment of the type 3 PanK crystal structures from *P. aeruginosa* (*PaPanK*), *Bacillus anthracis* (*BaPanK*) and *Burkholderia cenocepacia* (*BcPanK*) has shown a near perfect alignment of all major secondary structural features (α -helices, β -sheets and loop regions) suggesting that the structure of type 3 PanK is very similar across bacteria. However, the alignment of the nucleotide binding site of type 3 PanK from these bacteria has shown an alteration in the substrate-interacting loop region in *BaPanK* when compared to the other two bacteria species. Inhibitors of *BaPanK* are also lacking activity against *PaPanK* and *BcPanK* [11] [6].

Structural comparison between type 1 PanK from *E.coli* (*EcPanK*) and type 3 PanK from *Thermotoga maritima* (*TmPanK*) may provide an explanation on why type 3 PanK does not have the ability to use *N*-alkylpantotheamides as substrates and why it is lacking feedback inhibition by CoA. A hypothesis claimed that a hydrophobic pocket that is mostly lined with aromatic residues, accommodates the hydrophobic tail of the *N*-alkylpantothenamides because this pocket was shown to be able to accommodate the additional thiol group of CoA. Therefore, the *N*-alkylpantothenamides can be used as a substrate by *EcPanK* and converted to antimetabolites of CoA. Inspection of *TmPanK* did not show any hydrophobic pocket that could do the same, which may explain why *N*-alkylpantothenamides are not inhibitors or substrates of type 3 PanK and why type 3 PanK lacks feedback inhibition of CoA and its thioesters [12].

Pantothenate has shown to bind to PanK through multiple hydrogen bonds (Figure 1.4). Several inhibitors for type 1 and type 2 PanK have been reported but very few inhibitors of type 3 PanK have been reported. One reported inhibitor for type 3 PanK is a ATP mimic with a K_i of $164 \pm 3 \mu\text{M}$, which is 3-fold lower than the K_m value of ATP ($510 \mu\text{M}$). This inhibitor works as a competitive inhibitor on *BaPanK* but no bacterial growth inhibition was reported [13] [14]. The structure of *PaPanK* is known, therefore, structure-based methods for hit discovery can be applied. One such methods is structure-based virtual screening.

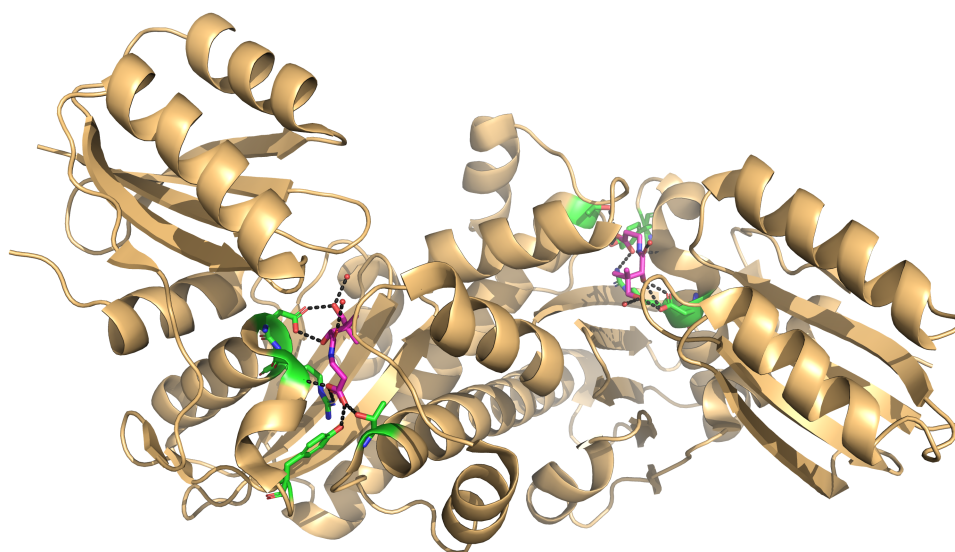


Figure 1.3: Structure of type 3 PanK from *P. aeruginosa*. The substrate, pantothenic acid (pantothenate), is shown in purple sticks. Hydrogen bonds are indicated as black dotted lines. PDB code: 2F9W

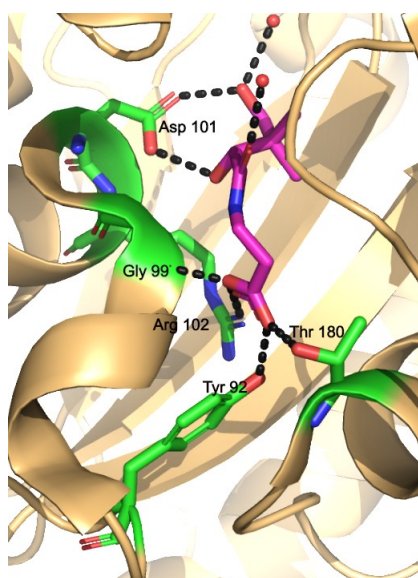


Figure 1.4: A zoomed in image of the structure of *PaPanK* (Figure 1.3) that shows the binding mode of pantothenate. Pantothenate binds to PanK by forming hydrogen bonds with the residues Gly 99, Asp 101, Arg 102, Tyr 92 and Thr 180.

1.5 Structure-based virtual screening

Structure-based virtual screening is a computational method to screen large libraries of chemicals against a biological target. The predicted compounds can then be experimentally tested against the target in order to evaluate the predictions from the virtual screening (Figure 1.5). This form of screening is not without challenges, such as sampling the different conformations of flexible molecules and calculation of the absolute energy for binding in aqueous environment. Regardless, this field has shown success by predicting new ligands with their receptor-bound structures, and in some cases, giving us hit rates that are greater than with high-throughput screening [15].

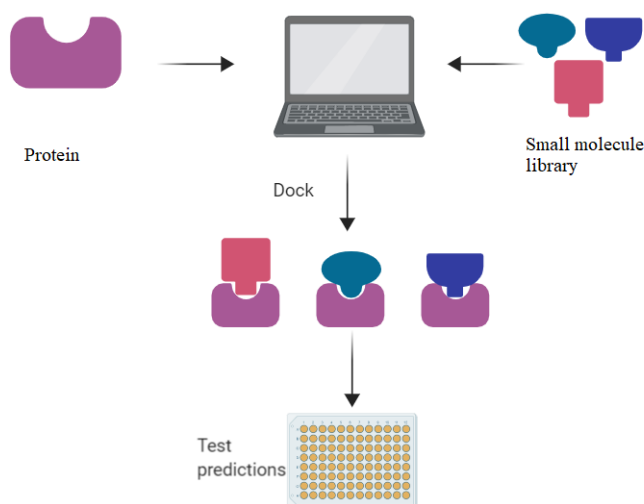


Figure 1.5: In structural-based virtual screening, libraries of compounds are docked into the binding site of the target molecule by a docking computer program. Each compound is sampled in thousands to millions of possible configurations. Each configurations is then scored based on its complementarity to the binding site. Ligands can then be ranked based on various factors such as torsion angles, inter- and intra clashes and the interactions that the ligands form with the target. The best ranked ligands are then chosen for testing for activity in an experimental assay.

In the initial phases of a virtual screening a database or a list of compounds needs to be selected and then prepared. The database usually comes from suppliers with a collection of compounds that are physically available, or are readily synthesizable and can be produced easily once they have been selected from a virtual screen [16]. There are several rules applied to guide

the selection of molecules that can be used in the early stages of drug discoveries. The most popular rule is the Lipinski Rule of Five. The rule states that drug-like compounds are more likely to be orally available if they have a molecular weight lower than 500 g/mol, lipophilicity (logP) lower than 5, less than five hydrogen bond donors and less than 10 hydrogen acceptors [17]. Astex's rule of 3 is an alternative guideline that reduces the various thresholds from 5 to 3. Astex's rule of 3 states that the molecular weight of the compounds should be lower than 300 g/mol, the number of hydrogen bond donors and acceptors is ≤ 3 , XlogP is ≤ 3 , the number of freely rotatable bonds is ≤ 3 and the polar surface area is $\leq 60 \text{ \AA}^2$ [18]. Other guidelines that can be worth following in the early stages of drug discoveries is to remove compounds that have unwanted functionalities before screening. Compounds that possess potentially mutagenic groups or are likely to have unfavourable pharmacokinetic properties are usually not desired. These groups include nitro groups, thiols, sulfates and phosphates [19].

In structure-based virtual screening, the essence is to simulate molecular recognition events. Therefore, it is necessary to prepare the database from its initial 2D form to a database with 3D coordinates. Compounds in a database are often registered in only one tautomer that is not necessarily correct at the experimental pH, where a balanced tautomeric equilibria might exist. Also, that tautomer might not be able to bind on the target molecule. Therefore, it is necessary to consider all relevant tautomers and assign them to the compounds in the database since there is no way of knowing beforehand which tautomer will most likely bind to the target molecule [16].

Just like the database, the active site of the target molecule also needs preparation. Ionizable residues in the active site have protonation states that need to be determined. In addition, the right histidines tautomers must also be assigned. The next step is to dock each molecule in the database into the binding site of the target. The binding mode is predicted by sampling the coordinate space of the binding site and the quality of the generated poses are ranked based on scoring functions [16].

A challenge with docking is the flexibility of the protein target. The binding pocket of the target usually adapts to fit the ligands upon binding through various conformations and majority

of docking tools assumes that the protein target is held fixed in its conformation. This assumption is not accurate but because of the increased complexity of the binding, this assumption is necessary. Water molecules also play a role in the binding between the ligand and the target protein. Water molecules can form hydrogen bonds at the interface between the target and the ligands and mediate ligand-target interaction. The location of water molecules may vary largely among ligands. The role of water molecules in docking is very important but the accurate prediction of the contribution from the water molecule is a long-standing challenge [20] [16].

1.6 FlexX

FlexX is a docking program that uses a docking algorithm based on an incremental construction strategy. The first step of the docking algorithm is base selection. This step selects a fragment used as a starting point and this fragment is called the base fragment. The next step is called base placement. The main goal of this step is to dock the base fragment into the active site of the protein and to find favorable interactions between the fragment and the protein. The third and final step is reconstruction of the ligand in an incremental way [21].

The ranking of the generated solutions is performed by estimating the free binding energy ΔG of the protein-ligand complex. This scoring function is based on an empirical scoring function developed by Böhm [22]. The ranking considers hydrogen bonds (ΔG_{hb}), ionic interactions (ΔG_{io}), lipophilic protein-ligand contact surfaces (ΔG_{lipo}), number of rotatable bonds, the loss of entropy due to translational and rotational hindrance and aromatic interactions (ΔG_{aro}). $f(\Delta R, \Delta \alpha)$ is a scaling function that penalizes deviation from the ideal geometry and N_{rot} is the number of free rotatable bonds that are immobilized in the complex [21].

$$\begin{aligned} \Delta G = & \Delta G_0 + \Delta G_{\text{rot}} \cdot N_{\text{rot}} \\ & + \Delta G_{\text{hb}} \sum_{\text{neutral H-bonds}} f(\Delta R, \Delta \alpha) \\ & + \Delta G_{\text{io}} \sum_{\text{ionic int.}} f(\Delta R, \Delta \alpha) \\ & + \Delta G_{\text{aro}} \sum_{\text{aro.int}} f(\Delta R, \Delta \alpha) \end{aligned} \quad (1.1)$$

$$+ \Delta G_{\text{lipo}} \sum_{\text{lipo.count}} f^*(\Delta R)$$

HYDE is a scoring function developed by Reulecke et al. [23] and is based on the estimation of the free energies emerging from HYdrogen bonding and DEhydration during the binding between protein and ligand. It describes the binding energies contributions of hydrogen bonding and the hydrophobic effect that are energetically favorable, as well as polar desolvations that energetically unfavorable. Overall, the purpose of the HYDE scoring function is to estimate the change in saturation and desolvation between the bound and unbound molecules [24]. The HYDE scoring function is formulated like this:

$$\Delta G_{\text{HYDE}} = \sum_{\text{atoms}i} \Delta G_{\text{desolvation}}^i + \Delta G_{\text{saturation}}^i \quad (1.2)$$

HYDE scoring can be visualized in the software Seesar. Atoms with the most favorable contributions are coloured green, yellow for neutral and red for unfavorable contributions. For example, atoms involved in a interaction with a good geometry are coloured green and atoms in unfavorable regions are coloured red [25].

1.7 Protein crystallization

To determine the atomic and molecular structure of a protein, X-ray crystallography is often utilized. A protein crystal is a necessity for this method. There is a variety of ways to form protein crystals but finding the best conditions for crystallization for new protein is a difficult task. The first step is to set up screening trials by exposing the protein to a number of different precipitant agents. The screening can help to find hits that can be indicative for protein crystallization. In such conditions protein may form crystals, crystalline precipitate or phase separation, and those conditions are worth pursuing. Once a possible crystallization condition has been found, optimization can be performed by varying the protein concentration, pH, temperature or by adding additives [26]. Commercially available crystallization screening kits can be used as the first choice when crystallization conditions are not known [27].

Vapour diffusion is the most common method for crystallization. The sample is either a hanging or sitting drop containing a fixed ratio between both protein and reservoir solution. This drop is then equilibrated against the reservoir solution that contains crystallization agents

at higher concentration than the drop. To achieve equilibration, water vapour will diffuse between the drop and the reservoir promoting the supersaturation of the protein drop with protein and crystal formation [26].

The information extracted from X-ray crystallography is a requirement for a structure-based drug discovery project. This is because this information provides knowledge about the protein structure which can be used to develop and optimize the inhibitors of that protein. A crystal structure of a protein-inhibitor complex can pinpoint key interactions and this information can be used for designing compounds that can improve the affinity and selectivity as well as compound pharmacokinetics [28].

1.8 Aim of the study

Antibiotic resistance is a serious problem around the world and the problem is constantly increasing. Therefore, the need of new antibiotics is very urgent, in particular for *P.aeruginosa*. The overall aim of the project is to discover starting point for new antibiotics. PanK is an attractive target for developing new antibiotics as PanK is essential for bacteria. Further, differences between the isoforms makes it likely that selective inhibitor can be developed. Virtual screening will be used to identify novel ligands that bind into the active site of PanK. Hit compounds from the virtual screening will then be selected and tested experimentally, using X-ray crystallography to verify their binding to the target and to guide the design of new PanK inhibitors. To enable the experiments, a robust protein expression and purification protocol will be established.

Chapter 2

Materials and Methods

Chemicals

All chemicals were purchased from Sigma-Aldrich unless otherwise stated.

2.1 Transformation of cells and overexpression of PanK

E. coli C41 (DE3), C43 (DE3), BL21 (DE3), Rosetta (DE3), C41 pLYsS (DE3) and C43 pLYsS (DE3) competent cells (Lucigen) were used to express the recombinant type 3 PanK from *P. aeruginosa*. The vector pET-28a(+) (Genscript) was used for the expression of the protein. This vector contains a kanamycin resistance gene as well as a gene sequence for the expression of a N-terminal 6x histidine tag followed by a Tobacco Etch Virus (TEV) cleavage site (Figure 2.1). PanK gene was inserted into the plasmid using the NheI/BamHI cloning sites.

1 μL of stock plasmid (10-50 ng) was added to the competent cells and incubated on ice for 30 minutes. Cells were then heat shocked at 42°C for 45 seconds followed by incubation on ice again for 2 minutes. 950 μL of Lysogeny Broth (LB) medium was added to the cells and the cells were incubated at 37°C for 1 hour at 250 rounds per minute (rpm) . 200 μL of the cells were plated on LB agar plates containing 50 $\mu\text{g ml}^{-1}$ kanamycin and incubated overnight at 37°C.

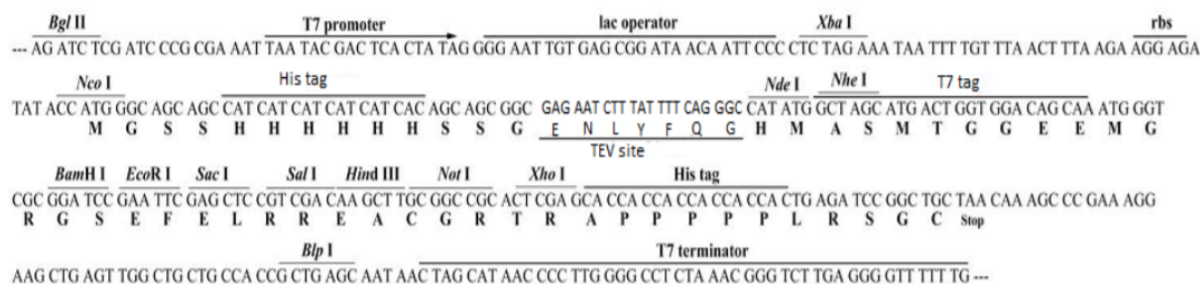


Figure 2.1: Vector used to express PanK with 6x histidine tag and a TEV site. The cloning sites of PanK is on NheI/BamHI (sequence of PanK not shown).

A single colony from freshly transformed cells was incubated in 10 mL of LB-medium containing the same concentration of kanamycin while shaking at 37°C overnight and then aliquoted and stored in -80°C for future use. 100 µL of the stock of transformed cells were incubated in 100 mL of LB-medium containing the same concentration of kanamycin while shaking at 37°C overnight. The overnight culture was then transferred to 900 mL of LB-medium containing the same concentration of kanamycin and incubated until the OD₆₀₀ reached 0.8-1.0. Isopropyl-β-D-1-thiogalactoside (IPTG) induction agent with a final concentration of 1 mM was added to the cultures which was further incubated for 3 hours at 37°C while shaking. The culture was centrifuged at 5000 x g for 10 minutes at 4°C. Supernatant was removed and the cell pellet was flash frozen with liquid nitrogen and stored at -80°C.

2.2 Purification of PanK

Purification of PanK was achieved in a similar way to [9] using affinity chromatography, 6-His tag cleavage with TEV protease and finally size exclusion chromatography. Cell pellets from the cell line *E. coli* C41 (DE3) were first thawed on ice and resuspended in lysis buffer containing 20 mM Tris pH 7.9, 500 mM NaCl, 20 mM imidazole and 10% (v/v) glycerol supplemented with 2 tablets of Complete EDTA-free protease inhibitor cocktail (Roche) and 40 µL of DNase 1. The cells were sonicated for a total of 2 minutes with 10 seconds pulses followed by centrifugation at 15000 rpm for 30 minutes. The supernatant was collected and filtered through a 0.2 µm filter.

All purifications were performed on an ÄKTA pure instrument (GE Healthcare) at room temperature. The protein was purified by binding to a HisTrap HP 5 mL column (GE Healthcare). After loading, the column was washed with 3 column volumes (CVs) of lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole 10% (v/v) glycerol, pH 7.9) and then the bound protein was eluted by using 0-100% gradient on 15 CVs of elution buffer (20 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, 10% (v/v) glycerol, pH 7.9).

The imidazole was removed by buffer exchange against gel filtration buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) by using Amicon[®] Ultra centrifuge filters (Merck) with a molecular weight (MW) cut-off of 10 kD. The protein fractions were concentrated using the same filters and concentration was measured by using a Nanodrop. Purified TEV protease with a ratio of 1:100 (mg of TEV: protein) was added to the protein and incubated at 4°C with shaking in order to cleave the TEV site in the protein. The same column as in the previous step was washed with 3 CVs of gel filtration buffer and followed by 5 CVs of the same elution buffer. His-tag cleaved protein was collected in the flow through fractions. The fractions were concentrated to a final volume \approx 2 mL by centrifugation at 4000 rpm for 20 min. Size exclusion chromatography was performed by loading the fractions onto a HiLoad 26/600 Superdex 75 column equilibrated with gel filtration buffer. The column was eluted with 1.2 CV with gel filtration buffer. The concentration of the protein was determined by a NanoDrop[®] 1000 Spectrophotometer (Thermo Fisher Scientific) using an extinction coefficient of $0.677 \text{ (mg/mL)}^{-1} \text{ cm}^{-1}$. In addition, A His-tag affinity chromatography purification attempt replacing Tris-HCl with 100 mM sodium phosphate buffer (pH 6.5) with the same additives was all also conducted.

2.3 Gel electrophoresis

The purity of the PanK after each purification step was checked by using SDS-PAGE (Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis) using Any kD[™] Mini-PROTEAN[®] TGX[™] Precast Protein Gels (Biorad) with 2 M of dithiothreitol (DTT) . The MW of protein bands was assigned by using Precision Plus Protein[™] Dual Color Standards (Biorad). Electrophoresis was run for 30 minutes at 200 V. Gels were stained using InstantBlue[™] Ultrafast Protein Stain

(Merck). A tiny piece of gel containing the protein was then cut off and send to the University of Oulu for a MALDI-TOF mass spectrometry analysis in order to confirm the identity of the protein.

2.4 Dynamic light scattering

Dynamic light scattering (DLS) is a technique used to determine distribution profile of polymers in a solution. This technique can determine if the purified protein aggregates in the solution. Before the DLS measurement, the refractive index of the sample was measured in triplicate using Abbemat 500 refractometer (Anton Paar). 70 μL of purified PanK solution was pipetted into a disposable cuvette. The sample was measured in the DLS instrument (Malvern). The measurement was conducted at 4°C, in triplicate and with a delay of 10 seconds between each measurement.

2.5 Protein thermal shift assay

A protein thermal shift assay was conducted in order to find out which buffers and additives stabilize or destabilize PanK. This method monitors the temperature dependent protein unfolding by letting the fluorescent dye SYPRO orange bind to the hydrophobic region of the protein. When the protein is unfolded due to the increased temperature, SYPRO orange will bind to the exposed hydrophobic region which will result in a large increase in fluorescence.

The Rubic buffer screen kit (Molecular Dimension) was used for the first experiment. 21 μL of buffer and 2 μL of purified PanK with a initial concentration of 2.88 mg/ml was pipetted to a 384-well plate. Stock solution of SYPRO orange (Merck) was prepared by diluting 15.5 μL of 5000X to a final concentration of 155X with water, making a 50 μL stock solution. 2 μL of the prepared stock solution of SYPRO orange was added to each well that contained buffer and PanK. The plate was sealed using the semi automated plate sealer from Axygen® PlateMax® and the plate was heated from 20°C to 95°C with a heating rate of 0.04°C/s by using the LightCycler® 480 II (Roche). The Rubic additive screen kit (Molecular Dimension) was used for the second

experiment. 16 μL of Rubic additive screen and 5 μL of buffer chosen from the first experiment were pipetted to a 384-well plate. PanK and SYPRO Orange were kept at the same concentration as in the first experiment. 2 μL of both solutions were pipetted to the plate making a 25 μL solution in the wells. The plate was sealed and heated in the same way as in the first experiment.

2.6 Crystallization of PanK

Purified PanK was concentrated to around 5 mg/ml in 100 mM sodium citrate pH 6.0, 250 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 5 % glycerol (v/v) for the crystallization trials. Trials were set up with the sitting-drop vapor diffusion method using Mosquito[®] HV (Sptlabtech) in Swissci 96-Well 3-Dropa well plates. Drops containing 2 μl of PanK were mixed with reservoir solution in a ratio of 1:1, 1:2 and 1:3 (volume PanK: volume reservoir solution). As reservoir solutions the following crystallization kits were used: JCSG-*plus*[™](Molecular Dimension), Ligand-Friendly screen (Molecular Dimension), Pact *premier*[™](Molecular Dimension) and AmSo4 (Qiagen). All crystals were grown in 20°C with the exception of one plate that contained His-tag PanK that were grown in 8°C. Crystals typically appeared within one week in form of needles and prior to data collection, crystals were transferred to cryoprotectant solutions containing 50% glycerol before flash-freezing in liquid nitrogen. Crystals were sent to Diamond synchrotron facility (United Kingdom) and Max IV laboratory (Sweden) for diffraction.

2.7 Setup for docking

The crystal structure of the PanK in complex with pantothenate (PDB code: 2F9W) was prepared with the software SeeSAR ver. 9.2 (BiosolveIT, Germany) in order to use it as a receptor for docking. The binding site was defined as the residues around pantothenate plus the adjacent pocket for the binding of ATP in chain A as defined by the SeeSAR pocket finder algorithm (Figure 2.2). Two different setups were used. In one setup, a carboxylate group was required and in the other setups, a hydrophobic group was required. Both groups were placed inside a sphere with a radius of 2.5 Å around C8 of pantothenate in chain A. This was done to ensure

that the ligands are placed in the pan binding site in order to make sure that the ligands are docked into the pantothenate binding site. Both setups were validated by redocking pantothenate a hydrophobic group requirement with a radius of 2.5 Å was generated around C8 of chain A.

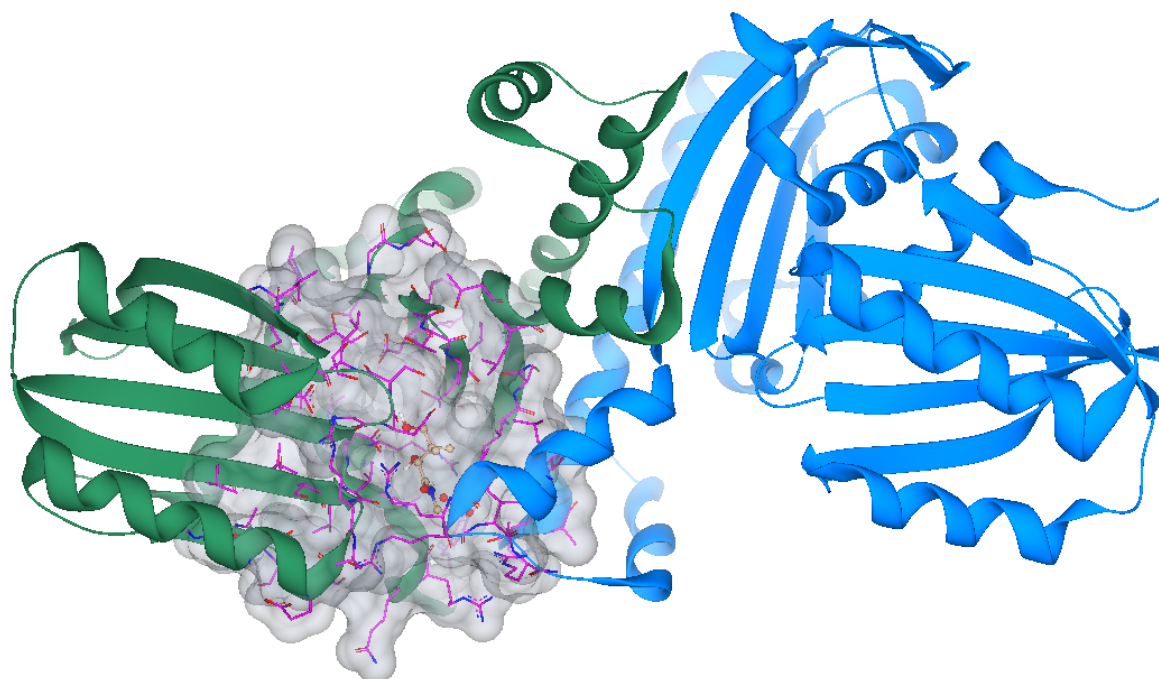


Figure 2.2: Binding site of PanK as defined by SeeSAR pocket algorithm. The binding site is located at the cloudy space.

2.8 Database for virtual screening

An in-house MySQL database of around 6 million commercially available compounds was prepared. The compounds were obtained from Otava chemicals, Asinex, Chembridge, Chemdiv, Enamine, Interbioscreen, Key Organics, LifeChemicals, Specs, Timtec and Vitas. Using in-house scripts based on the RDkit, unique smiles for all compounds, stereoisomers and physico-chemical properties were calculated. From these compounds, the libraries were prepared, one containing carboxylates and one containing diverse molecules. Both libraries were obtained using the following filters: 28 heavy atoms or less, 1 or more hydrogen bond donor and acceptor, 7 or more rotatable bonds, 3 ring systems or less, maximum 2 fused rings, 2 or more atoms that are not part of a ring system, maximum 1 nitrile group, 2 ether group or less, a total charge be-

tween -2 and 2 and xlogp value between 2 and -3. This resulted in 975408 unique compounds in both libraries. For the carboxylate library, compounds were filtered for those that contain a carboxylate group using in-house scripts based on RDKit. This resulted in 89305 compounds with carboxylate group. Using in-house scripts based on RDKit, a total amount of 162888 stereoisomers for docking were generated for these compounds. For the diverse library, a total amount of 1977327 stereoisomers for docking were generated.

2.9 Docking

For the docking, FlexX 4.0 was used for filtering the HYDE scoring. For each compound the best binding pose was saved. The pose which got a score of ≤ -20.0 were subsequently scored with HYDE. Docked ligands that had a predicted ligand-lipophilicity efficiency (LLE) of ++ were kept. If the ligand had more than one pose that scored with LLE of ++, the one with the better predicted binding affinity was kept. The compounds were then loaded into the software Maestro ver. 11.9 for visual inspection. Pharmacophore hypotheses were derived based on the retrieved structures. The pharmacophore hypothesis derived from the carboxylate group featured one negative ionic group, one positive ionic group and one hydrogen bond donor that points Asp101. The pharmacophore hypothesis derived from the diverse group used the same features as the carboxylate group, except the negative ionic group was replaced with a hydrogen bond acceptor. A second pharmacophore hypothesis was developed in the diverse group by removing the hydrogen bond acceptor and include an excluded group with a radius of 1.0 Å derived from the carboxylate group from the pantothenate in chain A. All pharmacophore features have a radius of 2.0 Å. A total amount of 20 compounds were selected for future hit validation (Table 3.18 and Table 3.19).

Chapter 3

Results

3.1 Overexpression of PanK

First, suitable conditions to overexpress PanK were established. Six different *E.coli* cell lines were tested, namely C41 (DE3), C43 (DE3), C41 pLYsS (DE3), C43 pLYsS (DE3), BL21 (DE3) and Rosetta (DE3). All cell lines were transformed with the plasmid of interest and 100 mL test cultures were grown as described in the previous chapter. After sonication and centrifugation, the supernatant and cell lysates were tested for the expression of desired protein via SDS electrophoresis.. As control, the corresponding samples of cell cultures that were not induced for protein expression were also added to the gels. The main goal of this experiment is to find out which cell lines are most suitable to express PanK. His-tag PanK has a MW of 29.45 kD. From the SDS-page (Figure 3.1), it can be observed that C41 pLYsS and C43 pLYsS do not seem to express much PanK as the other cell lines, therefore, not suitable for expressing PanK. The other cell lines showed a better expression of PanK, therefore, they were used for subsequently used to overexpress PanK for protein purification.

3.2 Purification of PanK

Next, a suitable purification protocol was established. The purification strategy is adapted from [9]. The strategy of the first purification (His-tag affinity chromatography) is to isolate His-tag PanK from other undesired proteins by letting it bind to the nickel resin via the His-tag in

the column. Once the other undesired proteins have passed the column and ended up in the flowthrough, the bound His-tag PanK can then be eluted from the column by letting the elution buffer to pass through the column. The elution buffer contains a high concentration of imidazole which competes with His-tag PanK for binding to the nickel resin in the column. The next purification step is to cleave the TEV -site and the His-tag from PanK. After PanK was incubated with TEV-protease, the His-tag and the TEV -site were cut from PanK and PanK can be isolated by loading the sample onto the same column. The TEV-cleavage site and TEV will attach to the nickel resin in the column via the His-tag and PanK will pass through the column. As a final step of the purification protocol, a size exclusion chromatography was conducted. This step is for polishing the purity of PanK and to separate PanK from potential protein aggregate that may have formed during the purification.

Between each purification step, the fractions were analyzed using gel electrophoresis in order to confirm the purity of the fractions. The samples on the first gel (Figure 3.2 A) are from the first purification (his-tag affinity chromatography). Lane 2 is the elution fraction that contained the purified his-tag PanK. As mentioned before His-tag PanK has a MW of 29.45 kD and proteins can be observed between the 25 kD and 37 kD mark which indicates that PanK was located in the fraction where it is expected to be in. Lane 3-6 are the flow-through fractions. Undesired proteins can be observed in these fractions and most of them cannot be observed in the elution fraction which is expected. The second gel (Figure 3.2 B) contains samples from the second purification (Cleavage of TEV -site). Lane 2 is the TEV-protease used to cut the TEV-cleavage site, lane 3 is the loading sample that was loaded into the HisTrap HP 5 mL column, lane 4 is His-tag PanK before incubation with TEV-protease, lane 5-6 are the flowthrough and lane 7 is the elution. Protein can be observed around the 25 kD mark in lane 6 as expected for PanK. No other proteins can be observed in lane 6 which indicates that the purification managed to separate PanK from other undesired proteins. The third and final gel (Figure 3.3 A) contains the sample from the size exclusion chromatography. Lane 4 is the fraction collected from the purification. A simple band was visible around 25 kD corresponding to PanK. A typical yield of PanK after size exclusion chromatography was around 0.82 mg/l cell culture. Results from the MALDI-TOF mass spectrometry confirmed the molecular mass of the to be 27.241002 kD which matches with theoretical value (27.24151 kD) of PanK. Purification attempt using 100

mM sodium phosphate (pH 6.5) was conducted because this buffer appeared to be able to stabilize PanK better than Tris-HCl which will be discussed later (Table 3.1). Purified His-tag PanK is most likely located in lane 9 due to a simple band was visible around 25 kD in that lane (Figure 3.3 B). Compared to the His-tag affinity chromatography using Tris-HCl as buffer (Table 3.2 A), the purity of the protein appeared to be better as no other proteins can be observed in lane 9.

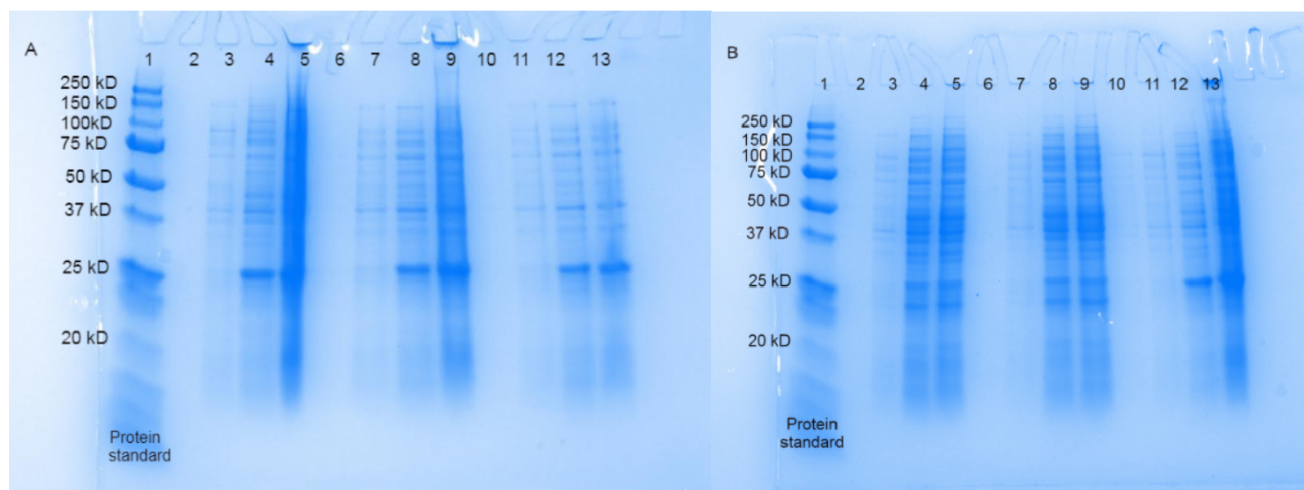


Figure 3.1: SDS-PAGE from the test cultures. His-tag PanK has a MW of 29.45 kD, which means PanK is located between the 25 kD and 37 kD mark. A: SDS-PAGE that contained samples from the cell line C43 (DE3), BL21 (DE3) and Rosetta (DE3). Lane 2-5 contained the cell line C43 (DE3). Lane 2 was the pre-induced supernatant, lane 3 was the pre-induced cell lysate, lane 4 was the induced supernatant and lane 5 was the induced cell lysate. Lane 6-9 contained the cell line BL21 (DE3). Lane 6 the was pre-induced supernatant, lane 7 was the pre-induced cell lysate, lane 8 was the induced supernatant and lane 9 was the induced cell lysate. Lane 10-13 contained the cell line Rosetta (DE3). Lane 10 was the pre-induced supernatant, lane 11 was the pre-induced cell lysate, lane 12 was the induced supernatant and lane 13 was the induced cell lysate. B: SDS-PAGE that contained samples from C41 pLYsS (DE3), C43 pLYsS (DE3) and C41 (DE3). Lane 2-5 contained the cell line contained the cell line C41 pLYsS (DE3). Lane 2 was the pre-induced supernatant, lane 3 was the pre-induced cell lysate lane 4 was the induced supernatant and lane 5 was the induced cell lysate. Lane 6-9 contained the cell line C43 pLYsS (DE3). Lane 6 the was pre-induced supernatant, lane 7 was the pre-induced cell lysate, lane 8 was the induced supernatant and lane 9 was the induced cell lysate. Lane 10-13 contained the cell line C41 (DE3). Lane 10 was the pre-induced supernatant, lane 11 was the pre-induced cell lysate, lane 12 was the induced supernatant and lane 13 was the induced cell lysate.

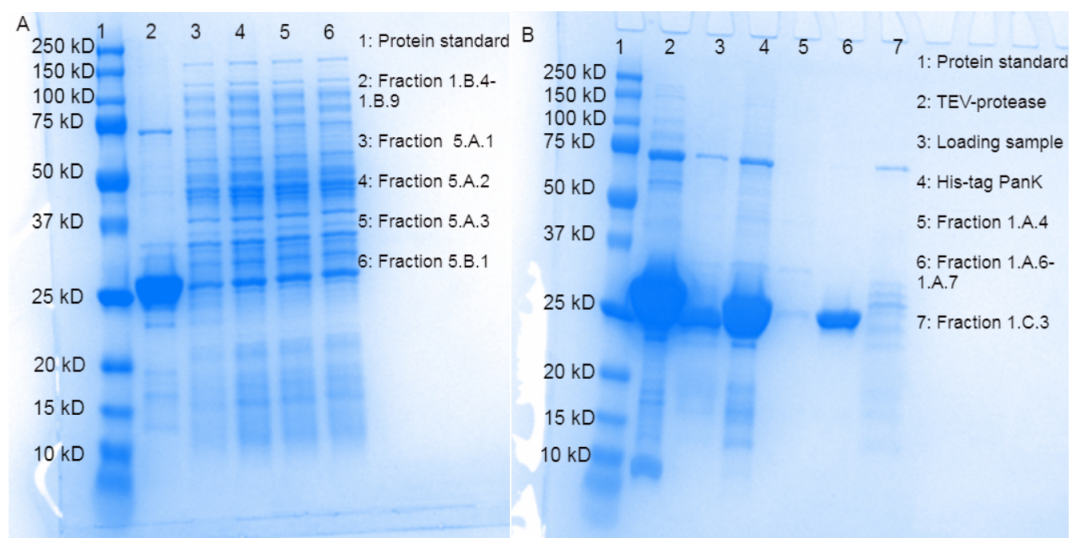


Figure 3.2: A: SDS-PAGE from the His-tag affinity chromatography. Lane 1 is the protein standard while lane 2 and lane 3-6 are the elution and flow-through fractions respectively. B: SDS-PAGE from the cleavage of the TEV site. Lane 1 is the protein standard, lane 2 is the TEV protease used to cut the TEV-site in PanK, lane 3 is the loading sample that was purified, lane 4 is His-tag PanK before incubation with TEV protease and column and lane 5-6 and 7 are the elution and flowthrough respectively.

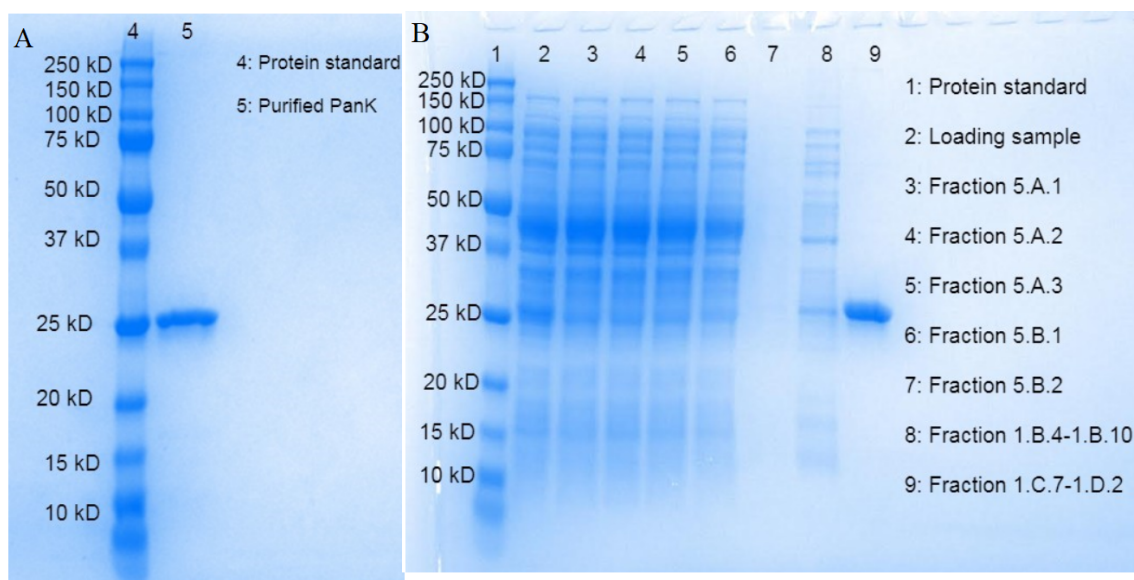


Figure 3.3: SDS-PAGE from the size exclusion chromatography and His-tag affinity chromatography using phosphate as lysis and elution buffer. Gel on the right was to the size exclusion chromatography while the gel to the left is from the His-tag affinity chromatography. A: Lane 4 was the protein standard while lane 5 was the elution. B: Lane 1 was the protein standard, lane 2 was the loading sample, lane 3-7 were the flowthrough while lane 8 and 9 were the elutions.

3.3 Dynamic light scattering

For the protein crystallization, the protein should not form aggregates. Therefore, DLS used for measuring the size distribution of PanK when it is dissolved in the gel-filtration buffer in order to find any potential protein aggregations. The sample was measured in triplicates with a delay of 10 seconds between each measurements. The results are presented in the form of an intensity curve and a correlogram. Multiple peaks intensity curve were observed (Figure 3.4 A). The peak with the highest intensity is located around the 10 nm mark and corresponded likely to non-aggregated PanK. In addition, peaks at 100 nm and 1000 nm can be observed, indicating aggregates. The correlogram (Figure 3.4 B) is also not very ideal as a "bump" can be observed which could indicate aggregation in the sample. Because of the potential aggregates discovered in Tris-HCl buffer, a new buffer was needed. A thermal shift assay was therefore conducted. Results from the thermal shift assay suggested that citrate buffer at pH 6.0 was a better buffer for PanK (Table 3.1). A new DLS experiment was conducted using citrate as a buffer for PanK (Figure 3.5). The highest peak can again be observed around 10 nm which further suggested that that this peak was non-aggregated PanK. Additional peaks could also be observed around 1 nm and between 100 nm and 1000 nm (Figure 3.5 A). These peaks could indicate aggregates but compared to intensity curve from Tris-HCl (Figure 3.4 A), citrate looked much better. The correlogram also appeared to have improved as no "bump" could be observed (Figure 3.5 B).

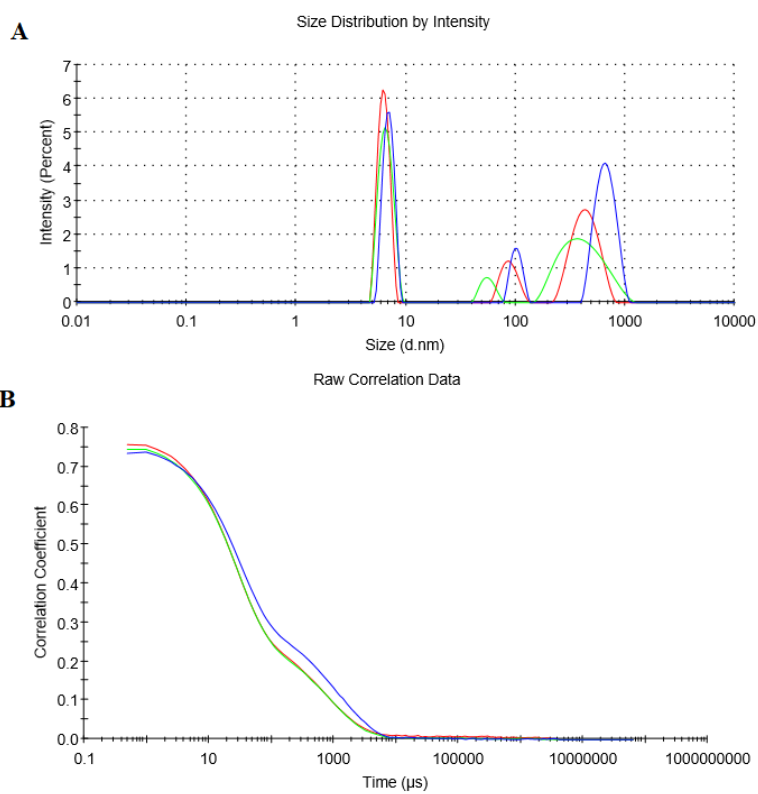


Figure 3.4: Obtained data from the DLS experiment using Tris-HCl as the buffer for PanK. The experiments were conducted in triplicates. A: Intensity curve. B: Correlogram.

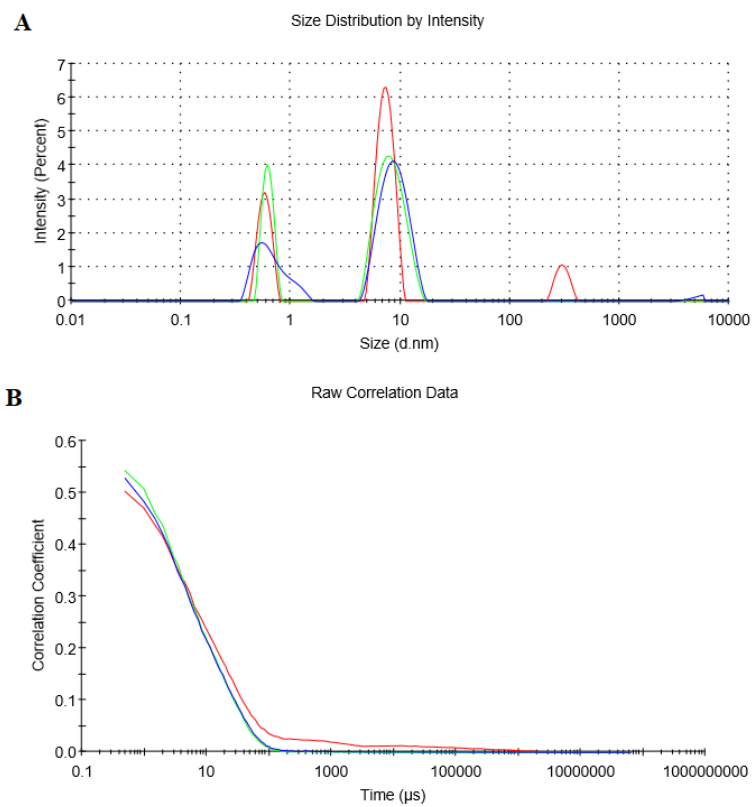


Figure 3.5: Obtained data from the DLS experiment using citrate as the buffer for PanK. The experiment was conducted in triplicates. A: Intensity curve. B: Correlogram

3.4 Thermal shift assay

In the next experiment, buffer screen with 96 different conditions was conducted in order to find which buffer stabilizes PanK. This section shows the shift of the melting temperatures (ΔT_m) of PanK in the different buffers. The reference temperature was 56°C and was determined with PanK dissolved in water. The buffers are sorted according to the different buffer groups provided by the supplier.

The main purpose of the buffer and pH screens (Table 3.1) is to find what combination of buffer and pH is most stabilizing for the protein. Phosphate buffers and citric acid buffer stabilized PanK most among all buffers that were listed in this table. The pH of the buffer had also an impact on the stability of the protein. This could be observed from the by using citric acid buffer with a different pH. While a citric acid buffer with a pH of 4.0 destabilizes PanK by ΔT_m -12.19, the same buffer at pH 6.0 stabilized the protein by ΔT_m 2.81. Tris-HCl with a pH-value of 7.5 is the buffer used as gel filtration buffer for PanK. While this buffer stabilized the PanK by ΔT_m by 1.77, aggregation has been observed when using this buffer (Figure 3.4). Phosphate buffers and citric acid at a pH -value of 6.0 appeared to be a better choice for PanK.

The purpose of the next set of buffers (Table 3.2) was to observe the change in ΔT_m when NaCl was added. The addition of salt in the citric acid buffers was less stabilizing for PanK with the exception when the pH was 6.0. For the phosphate buffers, the addition of NaCl improved the stability of PanK by ΔT_m 4.20 for di-sodium phosphate and 4.12 for potassium phosphate. The result from determining the thermal shift effect caused by different salt concentration (Table 3.4) supported the finding that NaCl contributed to the stabilization of PanK as ΔT_m increases as a function of NaCl concentration.

The result from the SPG buffer (Table 3.3) showed the effect of pH on PanK stability, uncoupled to differences on ionic strength, salt or buffer systems used. From pH 4 to 4.5 it could be observed that PanK was less stable while from pH 5.0 to pH 6.5, the stability of PanK increased with the pH -value. A pH of 6.0 to 6.5 appeared to be the best for the stability of PanK. Increasing the pH further decreased the ΔT_m of PanK.

The final set (Table 3.6) was carried out to determine the thermal shift effect by experimenting with different buffers and imidazole. Most buffer systems did not seem to be able to stabilize PanK with the exception of the buffer system 50 mM sodium phosphate/citric acid pH 7.5, 50 mM imidazole + 100 mM NaCl pH 7.5 and 125 mM imidazole + 100 mM NaCl pH 7.5 which slightly improved the stability. However, ΔT_m decreased as a function of the concentration of imidazole, which indicates that a high amount of imidazole was not suitable for the stabilization of PanK.

Next, an additive screen was conducted in order to find out which additives can contribute to the stabilization of PanK. The experiment was conducted with two buffers based on the buffer screen. The two buffers chosen for this experiment were Tris-HCl pH 7.5 and citric acid pH 6.0. A reference melting temperature was made for each buffer. The reference melting temperature for PanK in citric acid was 56.87°C and in Tris-HCl 55.44°C.

The result from testing the thermal shift effect from salts (Table 3.7) showed that most tested salts were able to stabilize PanK though there were some salts that appeared to have a greater buffer depending thermal shift effect on PanK, such as magnesium sulfate as the results from this salt were different between the two buffers that were tested. The results from the monovalent ions (Table 3.8) showed that most salts in this category contributed to the stabilization of PanK with the exception of the iodide salts.

The results from the chaotropic agents (Table 3.9) showed that urea was not a good candidate for stabilization of PanK. Urea has been widely used for protein denaturation [29]. The fact that urea destabilized PanK should therefore be expected. Guanidine-HCl is also known to denature proteins [30] though the results suggested that a concentration of 150 mM could actually stabilize PanK to some degree.

Both TCEP and DTT (Table 3.16) are known for being reducing agents for protein. DTT is a reducing agent normally used in protein research but it is also known for being very unstable. TCEP is a reducing agent used for the same purpose as DTT but TCEP is more stable. Both reducing agents stabilized PanK slightly but TCEP is better. Imidazole (Table 3.17) was already

known for being unable to stabilize PanK at higher concentration based on previous results (Table 3.6). As for the other additives, with a few exceptions, most of them are able to stabilize PanK to some degree.

Table 3.1: The thermal shift effect of buffer and pH (low ionic strength)

Buffer	ΔT_m (in °C)
100 mM Citric acid pH 4.0	-12.19
100 mM Sodium acetate pH 4.5	-6.20
100 mM Citric acid pH 5.0	0.75
100 mM MES pH 6.0	-1.64
100 mM Potassium phosphate pH 6.0	2.42
100 mM Citric acid pH 6.0	2.81
100 mM Bis-Tris pH 6.5	0.87
100 mM Mes pH 6.5	-0.01
100 mM Di-sodium phosphate (monobasic) pH 7.0	2.70
100 mM Potassium phosphate (dibasic) pH 7.0	2.16
100 mM HEPES pH 7.0	0.18
100 mM MOPS pH 7	-0.13
100 mM Ammonium acetate pH 7.3	2.42
100 mM Tris-HCl pH 7.5	1.77
100 mM Imidazole pH 7.5	-0.69
100 mM HEPES pH 8.0	-1.24
100 mM Tris-HCl pH 8.0	0.67
100 mM Tricine pH 8.0	-0.52
100 mM Bicine pH 8.0	-0.55
100 mM Bicine pH 8.5	-1.96
100 mM Tris-HCl pH 8.5	0.53
100 mM CHES pH 9.0	-5.02

Table 3.2: The thermal shift effect of buffer and pH (high ionic strength)

Buffer	ΔT_m (in °C)
Water + 250 mM NaCl	2.98
100 mM Citric acid + 250 mM NaCl pH 4.0	-15.69
100 mM Sodium acetate + 250 mM NaCl pH 4.5	-7.41
100 mM Citric acid + 250 mM NaCl pH 5.0	-0.37
100 mM MES + 250 mM NaCl pH 6.0	1.99
100 mM Potassium phosphate + 250 mM NaCl pH 6.0	4.57
100 mM Citric acid + 250 mM NaCl pH 6.0	4.77
100 mM Bis-Tris + 250 mM NaCl pH 6.5	3.30
100 mM Mes + 250 mM NaCl pH 6.5	3.42
100 mM Di-sodium phosphate + 250 mM NaCl pH 7.0	4.20
100 mM Potassium phosphate + 250 mM NaCl pH 7.0	4.12
100 mM HEPES + 250 mM NaCl pH 7.0	3.20
100 mM MOPS + 250 NaCl pH 7.0	2.93
100 mM Ammonium acetate + 250 mM NaCl pH 7.3	3.73
100 mM Tris-HCl + 250 mM NaCl pH 7.5	2.83
100 mM Di-sodium phosphate + 250 mM NaCl pH 7.5	3.61
100mM Imidazole + 250 mM NaCl pH 7.5	1.73
100 mM HEPES + 250 mM NaCl pH 8.0	1.40
100 mM Tris-HCl + 250 mM NaCl pH 8.0	2.47
100 mM Tricine + 250 mM NaCl pH 8.0	1.69
100 mM Bicine + 250 mM NaCl pH 8.0	2.19
100 mM Bicine + 250 mM NaCl pH 8.5	1.54
100 mM Tris-HCl + 250 mM NaCl pH 8.5	-1.50
100 mM CHES + 250 mM NaCl pH 9.0	2.79

Table 3.3: The thermal shift effect of pH (same buffer)

Buffer	ΔT_m (in °C)
100 mM SPG pH 4.0	-13.86
100 mM SPG pH 4.5	-5.14
100 mM SPG pH 5.0	2.12
100 mM SPG pH 5.5	2.12
100 mM SPG pH 6.0	2.10
100 mM SPG pH 6.5	2.28
100 mM SPG pH 7.0	1.14
100 mM SPG pH 7.5	0.61
100 mM SPG pH 8.0	0.31
100 mM SPG pH 8.5	-0.18
100 mM SPG pH 9.0	-0.33
100 mM SPG pH 10	-4.56

Table 3.4: The thermal shift effect of ionic strength (salt)

Buffer	ΔT_m (in °C)
50 mM HEPES + 50 mM NaCl pH 7.5	0.96
50 mM HEPES + 125 mM NaCl pH 7.5	-0.08
50 mM HEPES + 250 mM NaCl pH 7.5	1.84
50 mM HEPES + 500 mM NaCl pH 7.5	4.31
50 mM HEPES + 750 mM NaCl pH 7.5	2.80
50 mM HEPES + 1000 mM mM NaCl pH 7.5	6.60
50 mM Tris-HCl + 50 mM NaCl pH 8.0	1.00
50 mM Tris-HCl + 125 mM NaCl pH 8.0	0.39
50 mM Tris-HCl + 250 mM NaCl pH 8.0	2.33
50 mM Tris-HCl + 500 mM NaCl pH 8.0	4.26
50 mM Tris-HCl + 750 mM NaCl pH 8.0	5.51
50 mM Tris-HCl + 1000 mM NaCl pH 8.0	6.56

Table 3.5: The thermal shift effect of ionic strength (buffer)

Buffer	ΔT_m (in °C)
20 mM HEPES pH 7.5	0.18
50 mM HEPES pH 7.5	-0.04
125 mM HEPES pH 7.5	-0.53
250 mM HEPES pH 7.5	-0.48
20 mM Di-sodium phosphate (dibasic) pH 7.5	0.10
50 mM Di-sodium phosphate (dibasic) pH 7.5	0.81
125 mM Di-sodium phosphate (dibasic) pH 7.5	2.03
250 mM Di-sodium phosphate (dibasic) pH 7.5	5.21
20 mM Tris-HCl pH 8.0	0.09
50 mM Tris-HCl pH 8.0	0.34
125 mM Tris-HCl pH 8.0	0.97
250 mM Tris-HCl pH 8.0	1.90

Table 3.6: The thermal shift effect of buffer system and imidazole

Buffer	ΔT_m (in °C)
50 mM MES/Bis-Tris pH 6.0	-0.66
50 mM MES/ Imidazole pH 6.5	-0.96
50 mM Bis-Tris/PIPES pH 6.5	-0.58
50 mM MOPS/Bis-Tris propane pH 7.0	-0.31
50 mM Sodium phosphate/citric acid pH 7.5	0.59
50 mM MOPS/ HEPES-Na pH 7.5	0.02
0.1 M Bicine/ Trizma Base pH 8.5	-1.10
50 mM Imidazole + 100 mM NaCl pH 7.5	1.12
125 mM Imidazole + 100 mM NaCl pH 7.5	0.62
250 mM Imidazole + 100 mM NaCl pH 7.5	-0.47
350 mM Imidazole + 100 mM NaCl pH 7.5	-1.84
500 mM Imidazole + 100 mM NaCl pH 7.5	-3.80

Table 3.7: Thermal shift effect from salts

Salts	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
100 mM Sodium acetate	1.76	1.95
100 mM Calcium acetate	2.72	0.91
100 mM Potassium acetate	2.12	1.90
100 mM Ammonium acetate	1.75	1.79
100 mM Sodium sulfate	2.39	2.84
100 mM Magnesium sulfate	1.78	4.50
100 mM Potassium sulfate	0.68	0.93
100 mM Ammonium sulfate	2.51	2.76
100 mM Sodium phosphate (monobasic)	2.37	1.15
100 mM Sodium phosphate (dibasic)	1.07	0.80
100 mM Potassium phosphate (monobasic)	1.90	1.67
100 mM Potassium phosphate (dibasic)	0.93	1.31
100 mM Sodium tartrate	1.39	1.80
100 mM Sodium citrate (tribasic)	1.83	1.23
100 mM Sodium malonate	1.57	2.01
100 mM Sodium nitrate	0.53	0.60
100 mM Sodium formate	1.32	1.68
100 mM Potassium formate	1.30	0.69

Table 3.8: The thermal shift effect from monovalent ions

Monovalent ions	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
100 mM Sodium fluoride	1.63	1.89
100 mM Potassium fluoride	1.65	1.62
100 mM Ammonium fluoride	1.64	1.74
100 mM Lithium chloride	1.17	0.85
100 mM Sodium chloride	1.32	1.54
100 mM Potassium chloride	1.41	0.51
100 mM Ammonium chloride	1.13	1.27
100 mM Sodium iodide	-4.17	-1.00
100 mM Potassium iodide	-4.14	1.37
100 mM Sodium bromide	0.87	0.92

Table 3.9: The thermal shift effect from chaotropic agents

Chaotropic agents	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
0.1 M Urea	-0.10	0.13
0.5 M Urea	1.07	-0.84
1 M Urea	-2.35	-3.42
2 M Urea	-5.75	-9.89
4 M Urea	-21.09	-30.94
150 mM Guanidine-HCl	0.33	0.81
500 mM Guanidine-HCl	-0.13	-0.88

Table 3.10: The thermal shift effect from Non detergents and detergents

Non detergents, detergents	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
1 mM NDSB-195	0.27	0.37
1 mM NDSB-201	0.27	-0.07
1 mM Fos Choline 12	-10.16	-2.34
1 mM CHAPS	-0.11	0.02
1 mM CHAPSO	-0.07	0.02
1 mM OG	-0.18	-0.11
1 mM DM	2.50	-1.07
1 mM DDM	9.19	-5.82

Table 3.11: The thermal shift effect from Carbohydrates

Carbohydrates	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
25 mM Monosaccharides mix MD2-100-75	0.47	0.58
25 mM Glucose	0.33	0.35
25 mM Sucrose	0.62	0.13
25 mM Maltose	0.65	0.25

Table 3.12: The thermal shift effect from polyols

Polyols	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
5%(v/v) DMSO	-0.72	-1.19
5%(v/v) Ethylene glycol	-0.25	0.33
5%(v/v) Glycerol	1.50	0.92
20%(v/v) Glycerol	3.45	2.88
5%(v/v) PEG 400	-0.02	-1.34
5%(w/v) PEG 1000	0.26	0.58
5%(w/v) PEG 3350	0.25	-0.69

Table 3.13: The thermal shift effects from Carboxylic acids and amino acids

Carboxylic acids, amino acids (racemic)	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
50 mM Carboxylic acids mix MD2-100-76	1.28	1.01
50 mM Proline	0.56	0.42
50 mM Glycine	0.74	0.72
50 mM Glutamic acid	1.10	1.08
500 mM Glutamic acid	3.39	3.84
50 mM Arginine	-3.94	-0.36
500 mM Arginine	-14.15	-9.93
50 mM Arginine + 50 mM Glutamic acid	-2.19	0.01
500 mM Arginine + 500 mM Glutamic acid	-3.33	-0.76
50 mM Gly-Gly-Gly	0.51	0.15
5 mM Oxaloacetic acid	0.06	0.40

Table 3.14: The thermal shift effects from Co-factor and polyamines

Co-factor, polyamines	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
5 mM Biotin	0.28	0.49
5 mM Betaine	0.53	0.81
5 mM Coenzyme A	0.70	0.19
5 mM Nicotinic acid	0.32	0.55
1 mM Spermidine	0.30	0.28
1 mM Spermine	-0.07	0.31
1 mM Sarcosine	0.33	0.28

Table 3.15: The thermal shift effects Nucleotides

Nucleotides	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
~20 μ M Deoxyribonucleic acid library < 50 bp	0.65	0.49
1 mM ATP + 1 mM magnesium chloride	0.38	0.65
1 mM ATP γ S + 1 mM magnesium chloride	0.82	0.89
1 mM cAMP + 1 mM magnesium chloride	0.37	0.32
1 mM GTP γ S + 1 mM magnesium chloride	0.69	0.37
1 mM cGMP + 1 mM magnesium chloride	0.80	0.48
1 mM NADH + 1 mM magnesium chloride	0.16	0.48
1 mM NADPH + 1 mM magnesium chloride	0.41	0.47
5 mM Polyethylenimine	-2.40	-1.17

Table 3.16: The thermal shift effects from reducing agents

Reducing agents	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
5 mM DTT	-0.09	0.20
5 mM TCEP	0.81	0.51

Table 3.17: The thermal shift effects from imidazoles

Imidazoles	ΔT_m (in °C) in citric acid	ΔT_m (in °C) in Tris-HCl
200 mM Imidazole	-4.76	-3.78
400 mM Imidazole	-10.45	-8.36

3.5 Crystallization

In order to collect data from the structure of PanK through X-ray crystallography, a crystal of PanK is required. Based on the results from the thermal shift assay and DLS experiment, the gel filtration buffer chosen to be used for PanK was 100 mM sodium citrate with a pH 6.0 containing 250 mM NaCl, 1 mM TCEP and 5% glycerol. All crystallization screens that were used had 96 conditions each, meaning a total number of 384 conditions were tested for the crystallization of PanK. Crystals of PanK can be observed in all crystallization screens. However, Ligand friendly crystallization kit and Pact *premier*TM crystallization kit appeared to be the best among the kits that were used as PanK crystallized in many conditions provided by these kits. The crystal morphology was always in the form of thin needles as exemplified by PanK crystals in 0.1 M MES buffer containing 0.2 M magnesium chloride hexahydrate and 25 % (w/v) PEG 6000 at 20°C (Figure 3.6).

Crystals were illuminated with UV radiation after 5 and 21 days for detection of protein crystals. UV fluorescence from the crystals was detected (Figure 3.7) which confirmed that the crystals in the drop was indeed protein crystals and not salt or buffer crystals. Crystals of His-tag PanK could be observed in the condition containing 0.1 M MMT pH 6 and 30 % (w/v) PEG 1000 (Ligand friendly) at 20°C. Crystals of His-tag PanK grown at 8°C could be observed in three conditions that shared the same chemicals namely 0.1 M Bis Tris Propane, 0.2 M sodium iodide, 20 % (w/v) PEG 3350 and 10 % (v/v) ethylene glycol (Ligand friendly). The only difference between these conditions was the pH. Unlike the crystals of PanK, the morphology of the His-tag PanK crystals appeared as small broken plates. Crystals of PanK grown in the conditions 2.2 M ammonium sulfate containing 0.2 M NaCl (AmSO₄), 0.1 M Bis-Tris propane with a pH 6.5 containing 0.2 M sodium nitrate and 20 % (w/v) polyethylene glycol (PEG) 3350 (Ligand friendly), 0.1 M Bis-Tris propane with a pH 6.5 containing 0.2 M sodium formate and 20 % (w/v) PEG 3350 (Ligand friendly) were tested for diffraction. In addition, His-tag PanK crystallized 0.1 M MMT pH 6 and 30 % (w/v) PEG 1000 was also tested for diffraction. None of the crystals were able to diffract.

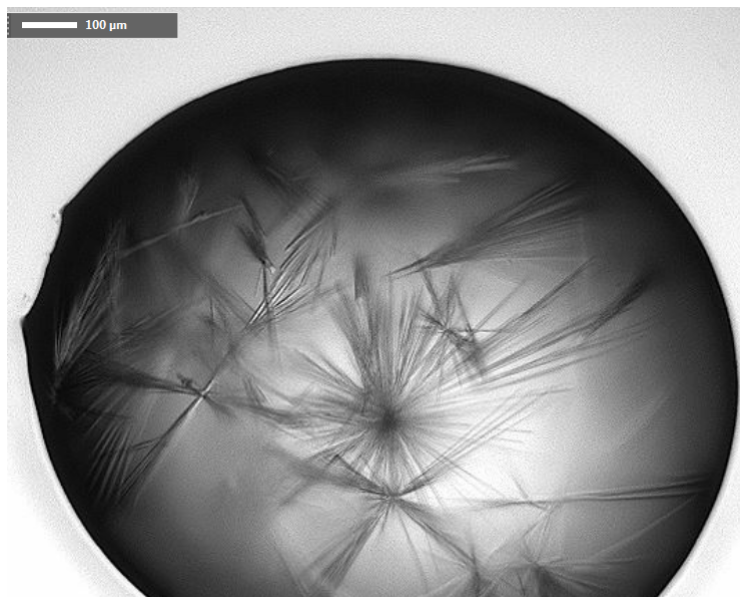


Figure 3.6: Crystals of PanK grown in the condition that contained 0.1 M MES pH 6.0, 0.2 M magnesium chloride hexahydrate and 20 % PEG 6000. The grown protein crystals could be observed after a week growing in 20°C .

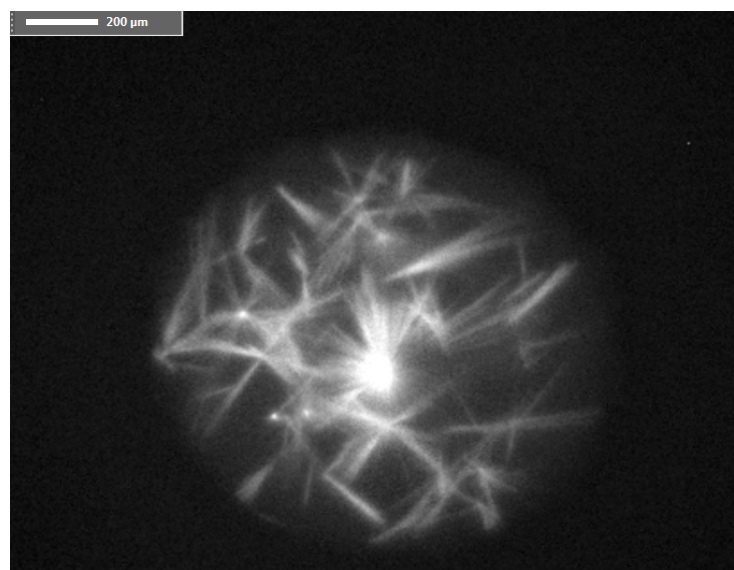


Figure 3.7: UV-picture of the same protein crystals (Figure 3.6). By illuminating the protein crystals with UV radiation, fluorescence can be detected from the protein crystals.

3.6 Virtual screening

In order to retrieve compounds relevant for the docking, compounds were first prepared by calculating unique smiles for all compounds, stereoisomers and physico-chemical properties. Libraries were then generated using the filters described in the section "Database for virtual

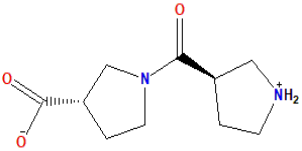
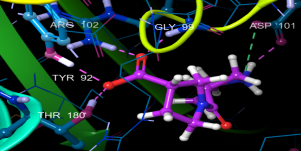
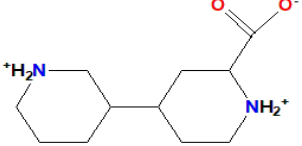
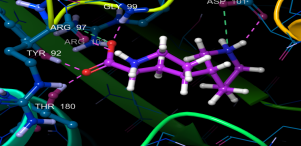
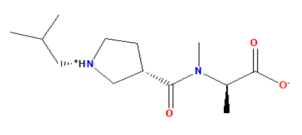
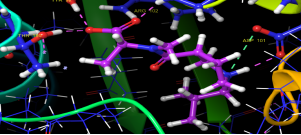
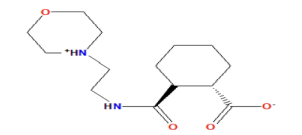
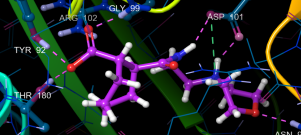
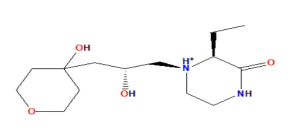
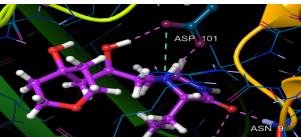
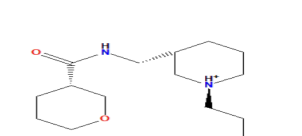
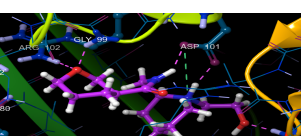
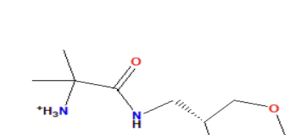
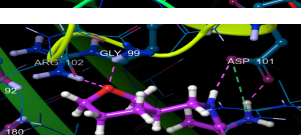
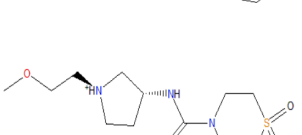
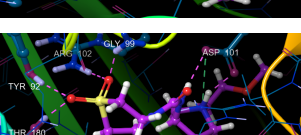
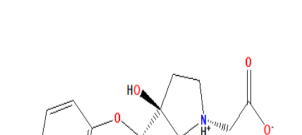
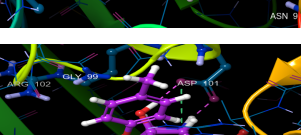
screening". Among the 162888 compounds in the carboxylate group and 1977327 compounds in the diverse group, 536 and 1281 compounds were able to dock into the binding site respectively. The list of docked compounds were filtered by generating pharmacophore hypotheses based on the bindings between the compounds and PanK. For the compounds containing carboxylate group, a pharmacophore hypothesis was generated by featuring one negative ionic group based on the carboxylate group on the compounds, one positive ionic group based on the charge group on the compounds and one hydrogen bond donor that points towards Asp 101. Visual inspection of the compounds in the carboxylate group has shown that many of them tend to bind to Asp 101 through the positive charged group in the form of hydrogen bonds and salt bridge. In addition, the oxygens on the carboxylate group also tend to form interact with other residues in the binding site. Therefore, the compounds in the carboxylate group were screened using the pharmacophore hypothesis described earlier. Compounds were screened by keeping their positions rigid. After screening, 344 compounds were accepted by the pharmacophore hypothesis and inspected by eye. Among the 344 compounds, 14 of them were considered for hit validation.

For the diverse group, two pharmacophore hypotheses were generated. The first pharmacophore hypothesis contained the same features as the one used for the carboxylate group except the negative ionic group is replaced by a hydrogen bond acceptor. Visual inspection of the compounds in the diverse group showed that many of them also tend to bind to Asp 101 through the positive charged group which was the reason why the hydrogen bond donor and positive ionic group were kept intact. As for the replacement of the negative ionic group with a hydrogen bond acceptor was to cover all functional groups that could act as a hydrogen bond acceptor. Among the 1281 compounds that were screened, 38 of them were accepted by the pharmacophore hypothesis and inspected by eye. The second pharmacophore was generated in order to see the importance the carboxylate group had for the binding. This pharmacophore replaced the hydrogen acceptor feature with an excluded carboxylate group. Among the 1281 compounds that were screened, 3 of them were accepted by the pharmacophore and all of them existed in the 38 compounds that were screened with the first pharmacophore. A total amount of 6 compounds from the diverse group were considered for hit validation.

Table 3.18: Structures of compounds and predicted binding mode

Structure id	Structure	Predicted binding mode
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		

Table 3.19: Structures of compounds and predicted binding mode

Structure id	Structure	Predicted binding mode
12		
13		
14		
15		
16		
17		
18		
19		
20		

Chapter 4

Discussion

4.1 Overexpression of PanK

The goal in this experiment is to find suitable cell lines that can express PanK. *E. coli* C41 (DE3), C43 (DE3), BL21 (DE3) and Rosetta (DE3) appeared to be suitable cell lines for the expression of PanK. *E. coli* C41 (DE3) and C43 (DE3) were chosen for the expression of PanK. Both cell lines were to be able to express His-tag PanK with a yield 7.9 mg/l cell culture. *E. coli* BL21 (DE3) and Rosetta (DE3) were not used for further expression. An interesting experiment that could be conducted is to purify His-tag PanK from these two cell lines and compare the yield between all four cell lines in order to see which cell line expresses most protein.

4.2 Purification of PanK

As mentioned in the result section, the goal of the of the purification experiment is to establish a purification protocol for PanK. In each purification step, SDS-gel electrophoresis was used to confirm the purity of the fractions. Each column from the flowthrough (lane. 3, 4, 5, and 6 Figure 3.2 A) indicated that some PanK did not bind to the nickel resin in the column. The reason could be because the nickel resin was saturated so that no more PanK could bind to it anymore. Another possible explanation is that these proteins were not PanK but protein with the same MW.

TEV-protease has a MW of 27.45 kD and from sample nr. 2 on the SDS-PAGE (Figure 3.2

B) it can be observed that the TEV-protease used for cleaving was not completely pure since there were other undesired protein in the TEV-protease sample. The reason why the loading sample was loaded into the gel (lane 3 Figure 3.2 B) is to compare it to the His-tag PanK (lane 4 Figure 3.2 B) since the loading sample is supposed to have a lower MW than the his-tagged PanK but due to column 4 being overloaded, this comparison was not possible. This comparison would have been possible if the samples were diluted before the SDS-PAGE gel electrophoresis. SDS-PAGE (Figure 3.3 A) from the size exclusion chromatography suggested that the PanK samples does not contain any other undesired proteins indicating a successful purification. As mentioned in the previous section, phosphate buffer appeared to be a better at stabilizing PanK than Tris-HCl. A His-tag affinity chromatography using phosphate buffer as lysis and elution buffer was conducted hoping that aggregation can be reduced during purifications. The purified His-tag PanK resulted in a very pure protein sample and no other undesired protein could be observed on lane 9 (Figure 3.3 B). The only main differences between these two His-tag affinity purification were that change of pH from 7.9 to 6.5 and 100 mM sodium phosphate was used instead of 20 mM Tris-HCl. The reason for the improvement of the purity could be because of the phosphate buffer and/or the change in pH but to be certain, the purification should be repeated.

4.3 Dynamic light scattering and thermal shift assay

A previously published protocol was used as guideline [9], therefore, Tris-HCl was initially chosen as the buffer for purification and gel-filtration of the PanK. However, PanK has shown that it may aggregates in this buffer (Figure 3.4). To find a more suitable buffer for crystallization a buffer screen was conducted. In order to cover a wide range of conditions, it was decided to use the Rubic buffer screen kit followed by Rubic additive screen kit. From the results of the buffer screen, the best buffer for PanK appeared to be phosphate buffer and citric acid buffer at a pH around 6 and 6.5 (Table 3.1 and 3.3). However, phosphate buffer is not suitable when crystallizing protein since phosphate tend to form salt crystals. For that reason, it was decided to use citric acid buffer for the additive screen. In addition, Tris-HCl buffer was also used as this buffer was used previously for the protein purification [9]. Another buffer that could be interesting to use as gel-filtration buffer for PanK is 100 mM ammonium acetate pH 7.3. The result

of this buffer from the buffer screen (Table 3.1) indicates that this buffer is also suitable for PanK.

As mentioned in the results section, protein should not form aggregates when it is used for crystallization. Therefore, DLS experiment was conducted in order to find potential PanK aggregates. Both the intensity curve and correlogram obtained from Tris-HCl (Figure 3.4) suggested that aggregates were present in the PanK solution as the results suggested that the solution was very polydisperse. Results from the thermal shift assay suggested that citrate is a better buffer for PanK crystallization (Table 3.1), thus, a DLS experiment was conducted using citrate as buffer. The intensity curve obtained with PanK purified with the citrate buffer for gel filtration (Figure 3.5 A) still showed multiple peaks at different sizes. However, the correlogram (Figure 3.5 B) from citrate appeared to have improved compared to the correlogram from Tris-HCl (Figure 3.4 B). The different peaks shown on the intensity curve (Figure 3.5 A) does not need to be protein aggregates, these peaks could possibly be the different additives that were added to the buffer (250 mM NaCl, 1 mM TCEP, 5% glycerol) since there was no option to instruct the instrument to ignore the additives. Overall, the results obtained from the buffer screen suggested that citrate and phosphate are more suitable buffers for PanK than Tris-HCl while results from the DLS experiment also appeared to support the statement that citrate was a more suitable buffer for PanK.

4.4 PanK crystallization

Due to the results achieved from the thermal shift assay experiment, it was decided to redesign the gel filtration buffer from Tris-HCl to citrate. The new buffer contained 100 mM sodium citrate pH 6.0, 250 mM NaCl, 1 mM TCEP and 5 % glycerol. For His-tag purification a phosphate buffer was used in hopes of getting a better yield of PanK and having it more stabilize. The purification resulted in very pure His-tag PanK, therefore, it was decided to attempt to crystallize His-tag PanK as well. Protein crystallization is a trial and error experiment, thus, it was attempted to crystallize PanK in a wide range of crystallization screen and in different temperatures to observe if PanK crystallizes in the different conditions that were used. PanK managed to crystallize in all crystallization screens in form of thin needles. The challenge of having pro-

tein crystals in this morphology is that taking them out from the plate without breaking them is very difficult. As mentioned in the result section, Ligand friendly crystallization kit and Pact *premier*TMcrystallization kit appeared to be the best among the kits for crystallization of PanK. For the crystallization using Ligand friendly kit, PanK usually crystallized in conditions with a pH from 6 to 8 containing sodium salt, and PEG and ethylene glycol as precipitants. The conditions where PanK crystallized using Pact *premier*TMkit shared a strong resemblances with Ligand friendly kit except ethylene glycol was not used in Pact *premier*TMcrystallization kit. This could indicate that PanK prefer to crystallize in these conditions and could serve as a starting point for future crystallization attempt. As for the other crystallization kits that were used, PanK crystallized in a very few conditions. Again, PanK crystals was observed in conditions that contained sodium salts and the precipitant PEG which further indicates that these chemicals are important for the crystallization. Crystals were not able to diffract, thus, data could not be collected from them.

4.5 Virtual screening

A library of 89305 carboxylates and a library of 975408 diverse small molecules were screened for PanK inhibitors. In total, 1817 compounds were suggested for testing. All selected compounds were predicted to have a LLE of ++ by the Hyde scoring function. In addition, it was required that the compounds form an interaction with Asp 101 as seen for pantothenate. Among the 20 compounds that were selected, 14 of them are from the carboxylate group (Table ??, compound id 1-11 and Table 3.19, compound id 12-14) while the rest is from the diverse group (Table 3.19, compound id 15-20).

As mentioned in the results section, pharmacophore hypotheses were generated for both groups in order to filter the docked compounds. The pharmacophore for the carboxylate group contained a negative ionic group, one positive ionic group and one hydrogen bond donor that points Asp101. These features were selected because inspection of compounds revealed that the compounds mainly formed interactions with PanK through the charged group and carboxylate group. Compounds were selected based on the bindings they form with PanK. As exemplified

by compound nr. 2 (Figure 4.1), compounds that formed hydrogen bonds with both oxygens on Asp 101 and both oxygens on the carboxylate group interacted with other residues were preferred.

Two pharmacophore hypotheses were generated for the diverse group. The first pharmacophore contained the feature types positive ionic group, a hydrogen bond donor and a hydrogen bond acceptor. The second one replaced the acceptor with an excluded carboxylate group. Among the 1281 compounds that were screened with the first pharmacophore hypothesis, 38 compounds were accepted by the pharmacophore. Out the 1281 compounds that were screened with the second pharmacophore, only 3 compounds were accepted. Compounds that contained a carboxylate group could still be found after screening with the first pharmacophore. This might indicate that carboxylate group is important for the binding to PanK and this idea was supported by the screening with the second pharmacophore. All selected compounds in this group (Table 3.19, compound id 15-20) interacted with Asp 101 using hydrogen bonds and salt bridges and two of them had a carboxylate group attached (Table 3.19. compound id 15 and 20). Overall, the docking of the two libraries has provided some compounds that could serve as starting point for discovering new inhibitor for PanK. In order to validate the hits, the compounds must be tested experimentally to confirm if they are actually active against PanK.

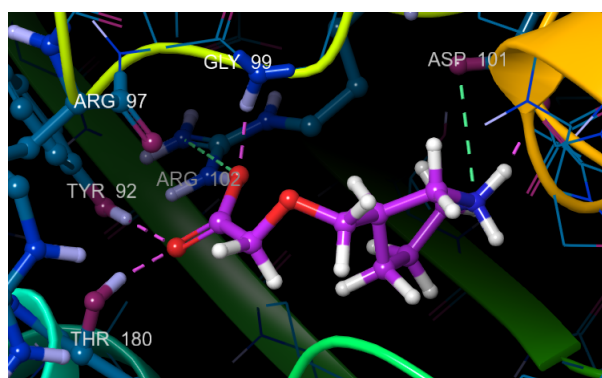


Figure 4.1: Binding mode exemplified by compound nr. 2. In this example, the charged group is forming salt bridge (green dotted lines) and hydrogen bond (pink dotted lines) with Asp 101. Additional interactions are also formed via the carboxylate group by interacting with Arg 102, Thr 180, Gly 99 and Tyr 92.

Chapter 5

Conclusion and future perspective

In this thesis type 3 PanK from *P. aeruginosa* was chosen as target for the discovery of new starting for developing new antibiotics. A purification protocol for the protein was established. A series of thermal shift assay showed that phosphate buffer at pH 6.0 appeared to be the best buffer for the protein. Purification attempt of the protein with this buffer for His-tag affinity chromatography resulted in very pure his-tag PanK. Further purification attempts using this buffer should be pursued in order to see if the results can be replicated and if the yield of protein is improved. Virtual screening identified compounds that theoretically bind to the pantothenate binding site of PanK. Based on the interactions with Asp 101, 20 compounds were chosen. In future experiments, binding assays using the identified compounds should be developed using for example BioLayer Interferometry (BLI) which allows screening the compounds with PanK and determining the binding constants of the screening hits. PanK was able to form crystals in form of needles. Unfortunately, did not diffract and no data could be collected from the them. A new crystallization approach should be conducted.

Bibliography

- [1] Laxminarayan, R, Duse, A, Wattal, C, Zaidi, A. K, Wertheim, H. F, Sumpradit, N, Vlieghe, E, Hara, G. L, Gould, I. M, Goossens, H, et al. Antibiotic resistance—the need for global solutions (2013) *The Lancet infectious diseases* **13**, 1057–1098.
- [2] Kerr, K. G & Snelling, A. M. *Pseudomonas aeruginosa*: a formidable and ever-present adversary (2009) *Journal of Hospital Infection* **73**, 338–344.
- [3] Moore, N. M & Flaws, M. L. Introduction: *Pseudomonas aeruginosa* (2011) *Clinical Laboratory Science* **24**, 41.
- [4] Lambert, P. Mechanisms of antibiotic resistance in *Pseudomonas aeruginosa*. (2002) *Journal of the royal society of medicine* **95**, 22.
- [5] Pang, Z, Raudonis, R, Glick, B. R, Lin, T.-J, & Cheng, Z. Antibiotic resistance in *Pseudomonas aeruginosa*: mechanisms and alternative therapeutic strategies (2019) *Biotechnology advances* **37**, 177–192.
- [6] Shapiro, J. A, Varga, J. J, Parsonage, D, Walton, W, Redinbo, M. R, Ross, L. J, White, E. L, Bostwick, R, Wuest, W. M, Claiborne, A, et al. Identification of specific and nonspecific inhibitors of *Bacillus anthracis* type III pantothenate kinase (pank) (2019) *ChemMedChem* **14**, 78–82.
- [7] Moolman, W. J, de Villiers, M, & Strauss, E. (2014) Recent advances in targeting coenzyme a biosynthesis and utilization for antimicrobial drug development.
- [8] Daugherty, M, Polanuyer, B, Farrell, M, Scholle, M, Lykidis, A, de Crécy-Lagard, V, & Osterman, A. Complete reconstitution of the human coenzyme a biosynthetic pathway via comparative genomics (2002) *Journal of Biological Chemistry* **277**, 21431–21439.

- [9] Hong, B. S, Yun, M. K, Zhang, Y.-M, Chohnan, S, Rock, C. O, White, S. W, Jackowski, S, Park, H.-W, & Leonardi, R. Prokaryotic type II and type III pantothenate kinases: the same monomer fold creates dimers with distinct catalytic properties (2006) *Structure* **14**, 1251–1261.
- [10] Awasthy, D, Ambady, A, Bhat, J, Sheikh, G, Ravishankar, S, Subbulakshmi, V, Mukherjee, K, Sambandamurthy, V, & Sharma, U. Essentiality and functional analysis of type I and type III pantothenate kinases of *Mycobacterium tuberculosis* (2010) *Microbiology* **156**, 2691–2701.
- [11] Yang, K, Strauss, E, Huerta, C, & Zhang, H. Structural basis for substrate binding and the catalytic mechanism of type III pantothenate kinase (2008) *Biochemistry* **47**, 1369–1380.
- [12] Yang, K, Eyobo, Y, Brand, L. A, Martynowski, D, Tomchick, D, Strauss, E, & Zhang, H. Crystal structure of a type III pantothenate kinase: insight into the mechanism of an essential coenzyme a biosynthetic enzyme universally distributed in bacteria (2006) *Journal of bacteriology* **188**, 5532–5540.
- [13] Guan, J, Barnard, L, Cresson, J, Hoegl, A, Chang, J. H, Strauss, E, & Auclair, K. Probing the ligand preferences of the three types of bacterial pantothenate kinase (2018) *Bioorganic & medicinal chemistry* **26**, 5896–5902.
- [14] Rowan, A. S, Nicely, N. I, Cochrane, N, Wlassoff, W. A, Claiborne, A, & Hamilton, C. J. Nucleoside triphosphate mimicry: a sugar triazolyl nucleoside as an atp-competitive inhibitor of b. anthracis pantothenate kinase (2009) *Organic & biomolecular chemistry* **7**, 4029–4036.
- [15] Shoichet, B. K. Virtual screening of chemical libraries (2004) *Nature* **432**, 862.
- [16] Lyne, P. D. Structure-based virtual screening: an overview (2002) *Drug discovery today* **7**, 1047–1055.
- [17] Lionta, E, Spyrou, G, K Vassilatis, D, & Cournia, Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances (2014) *Current topics in medicinal chemistry* **14**, 1923–1938.

- [18] Köster, H, Craan, T, Brass, S, Herhaus, C, Zentgraf, M, Neumann, L, Heine, A, & Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes (2011) *Journal of medicinal chemistry* **54**, 7784–7796.
- [19] Brenk, R, Schipani, A, James, D, Krasowski, A, Gilbert, I. H, Frearson, J, & Wyatt, P. G. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases (2008) *ChemMedChem: Chemistry Enabling Drug Discovery* **3**, 435–444.
- [20] Cheng, T, Li, Q, Zhou, Z, Wang, Y, & Bryant, S. H. Structure-based virtual screening for drug discovery: a problem-centric review (2012) *The AAPS journal* **14**, 133–141.
- [21] Rarey, M, Kramer, B, Lengauer, T, & Klebe, G. A fast flexible docking method using an incremental construction algorithm (1996) *Journal of molecular biology* **261**, 470–489.
- [22] Böhm, H.-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure (1994) *Journal of computer-aided molecular design* **8**, 243–256.
- [23] Reulecke, I, Lange, G, Albrecht, J, Klein, R, & Rarey, M. Towards an integrated description of hydrogen bonding and dehydration: decreasing false positives in virtual screening with the hyde scoring function (2008) *ChemMedChem: Chemistry Enabling Drug Discovery* **3**, 885–897.
- [24] Schneider, N, Hindle, S, Lange, G, Klein, R, Albrecht, J, Briem, H, Beyer, K, Claußen, H, Gastreich, M, Lemmen, C, et al. Substantial improvements in large-scale redocking and screening using the novel hyde scoring function (2012) *Journal of computer-aided molecular design* **26**, 701–723.
- [25] Schneider, N, Lange, G, Hindle, S, Klein, R, & Rarey, M. A consistent description of hydrogen bond and dehydration energies in protein–ligand complexes: methods behind the hyde scoring function (2013) *Journal of computer-aided molecular design* **27**, 15–29.
- [26] Chayen, N. E & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal (2008) *Nature methods* **5**, 147.
- [27] McPherson, A. Introduction to protein crystallization (2004) *Methods* **34**, 254–265.

- [28] Hosfield, D, Palan, J, Hilgers, M, Scheibe, D, McRee, D. E, & Stevens, R. C. A fully integrated protein crystallization platform for small-molecule drug discovery (2003) *Journal of structural biology* **142**, 207–217.
- [29] Stumpe, M. C & Grubmüller, H. Interaction of urea with amino acids: implications for urea-induced protein denaturation (2007) *Journal of the American Chemical Society* **129**, 16126–16131.
- [30] Monera, O. D, Kay, C. M, & Hodges, R. S. Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions (1994) *Protein Science* **3**, 1984–1991.