THE UNIVERSITY
of EDINBURGH

Biophysical, biochemical and inhibition studies of hexokinases

ANDROMACHI XYPNITOU

DOCTOR OF PHILOSOPHY-UNIVERSITY OF EDINBURGH-2017

Abstract

Hexokinase is the first enzyme in glycolysis, a major pathway for the generation of energy in all eukaryotes. Mammalian cells have four isoforms (I, II, III, IV) that have different tissue distribution and kinetic properties. Among all isoforms, human hexokinase II (hHKII) has been found to be implicated in many cancers with an increased expression which serves a dual role. First, it maintains the high glycolytic rate of malignant cells (Warburg effect) and second it prevents apoptosis when is bound to mitochondria. *Trypanosoma brucei* is a parasite that causes Human African Trypanosomiasis (HAT) and has two isoforms with extensive sequence similarity (98%), TbHKI (active form) and TbHK2 (inactive form). The bloodstream-form parasites (BSF) depend exclusively on glycolysis for their survival. The enzyme from both organisms is a validated target for drug-discovery against both cancer and HAT. The aim of the present study is the discovery of novel and specific inhibitors of the enzymes based on their structure. Structure-based drug discovery is commonly used in pharmaceutical companies to aid in the discovery of potent lead compounds. *In silico* studies were performed in this project using the known crystal structure of human hexokinase I and a model of TbHKI generated by the protein modelling tool Phyre2. The docking programs, AutoDock (AD) and AutoDock Vina (Vina), were chosen to perform the docking of ~3 million compounds to the target molecules and scoring functions calculated the predicted binding affinities of each compound. In total, 28 compounds were purchased to test on the target molecules.

In the experimental part of the project, the two enzymes were cloned, expressed and purified. hHKII was successfully purified giving a high yield of active and pure protein. The protein was characterised using many biophysical methods to establish the oligomeric state, the homogeneity and the secondary structure. Crystallisation trials failed and for this reason, N and C domains of the hHKII were purified separately. Unfortunately, the domains also failed to crystallise thus SAXS data were collected and analysed to gain information of their shape at low resolution. A novel inhibition assay was developed and used to identify four weak inhibitors against full length hHKII.

TbHKI was difficult to express in a soluble form as most of the protein was expressed in inclusion bodies. The purification resulted in a small amount of active protein that was used entirely for biochemical assays. Four compounds were purchased from the docking of the TbHKI model and one was found to inhibit the enzyme over 65% at 100 μM. Because the active site of both enzymes (hHKII, TbHKI) is well conserved the compounds from hHKII docking were also screened against the TbHKI. Four compounds were found to inhibit this enzyme while one of them was also an inhibitor for human isoform. The remaining three were specific for inhibition of TbHKI.

Declaration

The work presented in this thesis is the original work of the author. This thesis has been composed by the author and has not been submitted in whole or in part of any other degree.

Andromachi Xypnitou

February, 2017.

Acknowledgements

I feel grateful for studying for the degree of Doctor of Philosophy in the University of Edinburgh and I would like to thank my principal supervisor, Prof. Malcolm Walkinshaw for giving me this opportunity. It is an honor to work together and I deeply appreciate his guidance all this time. Apart from being the most incredible supervisor one can ask he is also a great person with kind feelings and deep understanding. Thank you for everything!

I would like to thank Dr. Martin Wear for his precious advices and inspiration. It would have been impossible to tackle many problems during my studies without your help. We had amazing moments; happy, exciting and awesome (given a new definition on that word). Thank you for everything!

I would like to thank Dr. Alan Wise, CEO of the Pharmaceutical Company IOmet Pharma. He has always been helpful and has shown a great interest in every problem I faced. His advices helped me to solve many of these problems. Nevertheless, Dr. Phill Cowley, Head of Chemistry from the same company has helped me to deal with Medicinal Chemistry and he has been always available to help me. Thank you both for everything.

Dr. Matthew Nowicki, Dr. Liz Blackburn and Sandra Bruce have been by my side whenever I needed their help and I would like to thank them all for sharing their knowledge with me. I admire all of you for your skills and knowledge and I am lucky that I have met you! Thank you for everything!

I would like to thank Dr. Douglas Houston for performing the *in silico* studies and for helping me to understand the theory behind this. His contribution to the present study is major. Thank you!

Iain McNae, Jaqueline Dornan and Paul Taylor thank you for your guidance on SAR analysis, crystallization trials and computational tools respectively! You were all available to help me whenever I was in the need of it. Thank you for everything!

I would also like to thank Dr. Li-Hsuan Yen for his help all these years and especially for his help during the writing of my thesis. Thank you for everything!

Of course a special thanks to all the students I met here, James Kinkead, Charis Georgiou, Peter Fernandes, Yiyuan Chen, Dr. Meng Yuan, Dr. Simon Varzandeh. Last but not least, I would like to thank Giullia Romanelli for being an excellent friend and an ideal company in the lab and outside the lab! My PhD studies added a friendship for life to me and I am so happy for that!

While I am writing the list keeps growing … I could not skip from my list Dr. Natassa Kantsadi for helping me when I had a problem with my research. Thank you for being an excellent friend for more than 12 years now and I owe you a big part of what I am now. Thank you Natassa for everything!

Lastly I would like to thank my husband Makis Papaioannou for being by my side all these years! He tolerated all my mood change as a part of the PhD process and I am grateful for sharing this journey with him. I would not have done this without you….

Finally, a big thanks to my family and my husband's family for helping me with my daughter. It is 100% true to say that I would not have done it without their help. They took the best care of my baby girl while I had to do the experiments and the writing of the thesis! Thank you mom, George, Stella, Nikos and Evi for everything!

I dedicate this to my beautiful, most amazing daughter, my little star, Stellina; for being the best baby a mother could ask; for giving me the strength to carry on and finish my studies so that one day you will be proud of your mom..

Forgive me for all the moments I missed during the first year of your life… I hope one day that you can understand. I love you to the moon and back.

Abbreviations

AEX	Anion Exchange Chromatography
ATP	Adenosine Triphosphate
BSF	Bloodstream Form
CD	Circular Dichroism
CV	Column Volume
DLS	Dynamic Light Scattering
DMSO	Dimethyl sulfoxide
Glc	Glucose
GBM	Glioblastoma multiforme
GF	Gel Filtration
G6P	Glucose-6-phosphate
G6PDH	Glucose-6-phosphate dehydrogenase
HAT	Human African Trypanosomiasis
HK	Hexokinase
hHKI/II/III	Human hexokinase I/II/III
IBs	Inclusion Bodies
NAD	Nicotinamide Adenine Dinucleotide
OXPHOS	Oxidative phosphorylation
RI	Refractive Index
PDB	Protein Data Bank
PF	Procyclic Form
PPP	Pentose Phosphate Pathway
RT	Room temperature
SAR	Structure-Activity Relationship
SAXS	Small-angle X-ray Scattering
SBVS	Structure-Based Virtual Screening
SEC	Size Exclusion Chromatography
SEC-MALS	SEC coupled to Multi-Angle Light Scattering
SGC	Structural Genomic Consortium
SPR	Surface Plasmon Resonance
TbHKI/II	*Trypanosoma brucei* hexokinase isoform I/II
TCA	Tricarboxylic acid cycle
VDAC	Voltage Dependent Anion Channel
Ve	Elution Volume

Vo	Void Volume
VS	Virtual Screening
2DG	2-deoxyglucose
3BrPa	3-bromopyruvate

Contents

Table of Contents

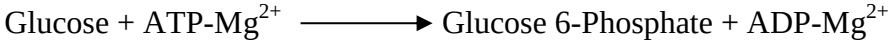
Table of Figures

Table of Tables

1. Chapter 1: Introduction

Glycolysis and reaction catalysed by hexokinase

Hexokinase (ATP:D-hexose 6-phosphotransferase, EC 2.7.1.1) is the first enzyme in glycolysis, a major pathway that generates ATP by catabolism of hexoses. Hexokinase, catalyses the reaction:



The ATP-dependent phosphorylation of glucose (Glc) is the first step in glycolysis. Hexokinase converts the nonionic Glc to ionic glucose 6-phosphate (G6P) which cannot exit the cells. The glycolytic metabolism of glucose occurs in all eukaryotic organisms, while many but not all prokaryotic organisms also use a similar glycolytic pathway (Cárdenas et al., 1998), (Pelicano et al., 2006). Glc is the preferred substrate of the hexokinases, but as the name implies, they can phosphorylate also other hexoses e.g. mannose, 2-deoxyglucose, fructose and galactose (Grossbard and Schimke, 1966). Negatively charged G6P fuels both glycolysis and the pentose phosphate pathway (PPP) (Fig. 1.1) and can also be used to synthesise polysaccharides like glycogen (not shown). Thus, hexokinase plays an important role in glucose metabolism.

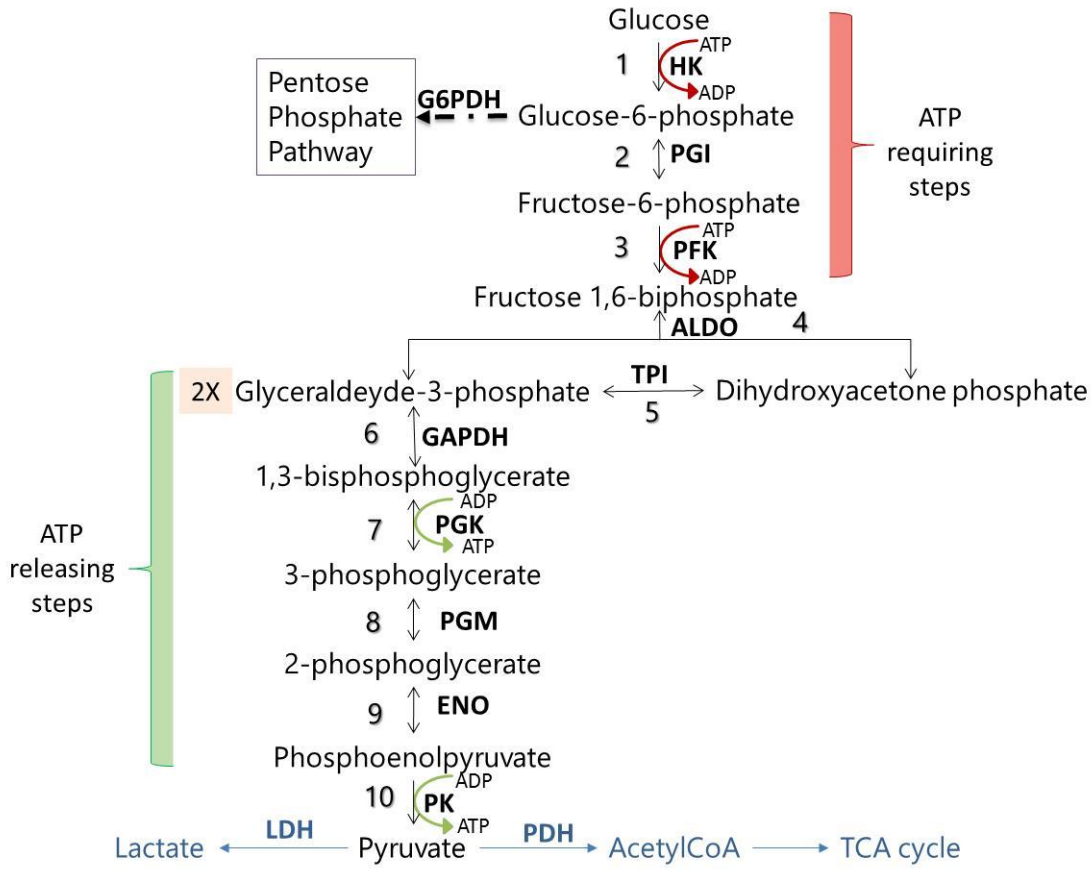


Figure 1.1: Glycolytic pathway consists of 10 reaction steps

The solid arrows indicate glycolytic reactions, whereas the dashed arrow indicates the interconnection with the pentose phosphate pathway (PPP). The blue arrows show the further metabolism of pyruvate. The first part of glycolysis (specifically in reactions 1 and 3) is the ATP requiring part, where two molecules of ATP are required to split the six-carbon sugar molecule into two three-carbon molecules. The red arrows show the ATP consumption steps. The second half of glycolysis (specifically in reactions 7 and 10) will produce a net gain of two ATP molecules by substrate phosphorylation. The green arrows show the energy release steps. HK, hexokinase; PGI, phosphoglucose isomerase; PFK, phosphofructokinase; ALDO, aldolase; TPI, triosephosphate isomerase; GAPDH, glyceraldeyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PK, pyruvate kinase; PDH, pyruvate dehydrogenase; LDH, lactate dehydrogenase; G6PDH, glucose-6-phosphate dehydrogenase.

Mammalian hexokinases and cancer

1.1 Isozymes of mammalian hexokinases

Four isozymes of hexokinase are found in a variety of mammalian tissues. One of the first studies on hexokinase categorised these isoforms in the rat as I, II, III and IV based on their electrophoretic pattern in a starch gel (Katzen and Schimke, 1965). The presence of four hexokinases appears to be characteristic of all animals, including the human, with each type being different from the other with regard to its kinetic properties and tissue distribution. HKI-III (100 kDa) are known as the "low-K_m hexokinases" with HKI mainly found in brain, hexokinase II in skeletal muscle and adipose tissue, whereas HKIII is found in small amounts in all tissues. HKIV (50 kDa) is also called Glucokinase, with the name being unfortunate as this is known as the "high-K_m hexokinase", so the glucose affinity is lower. Product inhibition (G6P) is a characteristic of 100 kDa hexokinases, while glucokinase does not seem to share this regulation. It is believed that hexokinases have evolved by a gene duplication and fusion of an ancestral hexokinase with a similar size of yeast hexokinase and HKIV (Katzen and Schimke, 1965, Cárdenas et al., 1998, Grossbard and Schimke, 1966). The 100 kDa HKI, HKII and HKIII contain two domains (N and C domains containing the residues 1-475 and 476-917, respectively). Studies on each domain of these isoforms have shown that only HKII has two active sites (Tsai and Wilson, 1996). The N regulatory domain has been found inactive in HKI and HKIII (Tsai, 1999). Two different evolutionary hypotheses exist about the origin of the mammalian glucokinase. Whereas the most popular one suggests that glucokinase diverged from a lineage leading to HKI-III before their ancestral gene underwent the duplication, Irwin and Tan 2008, propose that glucokinase evolved from a hexokinase already containing two active domains, but secondarily lost its N domain (Irwin and Tan, 2008). The evolutionary analysis of the hexokinase gene family in verterbrates, performed by the same group, led also to the discovery of a fifth hexokinase-like gene (HKDC1). HKDC1 genes were found in all genomes examined implying that it could be functional (Irwin and Tan, 2008). To date all published

papers and reviews refer to just four hexokinases (HKI-IV) which suggests that HKDC1 does not have sugar-phosphorylation activity.

The isoforms (HKI-IV) show an extensive sequence similarity and particularly HKI-HKII show 73% identity while the percentage drops to ~52-55% for HKIII with the other three isoforms. Figure 1.2 shows the sequence alignment performed with the Clustal Omega (ClustalW) online tool (Larkin et al., 2007). The sequences were obtained from UniprotKB (http://www.uniprot.org/) (2017).

```
Human_glucokinase        ------------------------------------------------------------
Human_hexokinase_III     MDSIGSSGLRQGEETLSCSEEGLPGPSDSSELVQECLQQFKVTRAQLQQIQASLLGSMEQ
Human_hexokinase_II      --MIASH-------LLAYFFTE--LNHDQVQKVDQYLYHMRLSDETLLEISKRFRKEMEK
Human_hexokinase_I       --MIAAQ-------LLAYYFTE--LKDDQVKKIDKYLYAMRLSDETLIDIMTRFRKEMKN

Human_glucokinase        ------------------------------------------------------------
Human_hexokinase_III     ALRGQASPAPAVRMLPTYVGSTPHGTEQGDFVVLELGATGASLRVLWVTLTGIEGHRVEP
Human_hexokinase_II      GLGATTHPTAAVKMLPTFVRSTPDGTEHGEFLALDLG--GTNFRVLWVKVTDNGLQKVEM
Human_hexokinase_I       GLSRDFNPTATVKMLPTFVRSIPDGSEKGDFIALDLG--GSSFRILRVQVNHEKNQNVHM

Human_glucokinase        ------------------------------------------------------------
Human_hexokinase_III     RSQEFVIPQEVMLGAGQQLFDFAAHCLSEFLDAQPVNKQGLQLGFSFSFPCHQTGLDRST
Human_hexokinase_II      ENQIYAIPEDIMRGSGTQLFDHIAECLANFMDKLQIKDKKLPLGFTFSFPCHQTKLDESF
Human_hexokinase_I       ESEVYDTPENIVHGSGSQLFDHVAECLGDFMEKRKIKDKKLPVGFTFSFPCQQSKIDEAI

Human_glucokinase        ------------------------------------------------------------
Human_hexokinase_III     LISWTKGFRCSGVEGQDVVQLLRDAIRRQGAYNIDVVAVVNDTVGTMMGCEPGVRPCEVG
Human_hexokinase_II      LVSWTKGFKSSGVEGRDVVALIRKAIQRRGDFDIDIVAVVNDTVGTMMTCGYDDHNCEIG
Human_hexokinase_I       LITWTKRFKASGVEGADVVKLLNKAIKKRGDYDANIVAVVNDTVGTMMTCGYDDQHCEVG

Human_glucokinase        ------------------------------------------------------------
Human_hexokinase_III     LVVDTGTNACYMEEARHVAVLDEDRGRVCVSVEWGSFSDDGALGPVLTTFDHTLDHESLN
Human_hexokinase_II      LIVGTGSNACYMEEMRHIDMVEGDEGRMCINMEWGAFGDDGSLNDIRTEFDQEIDMGSLN
Human_hexokinase_I       LIIGTGTNACYMEELRHIDLVEGDEGRMCINTEWGAFGDDGSLEDIRTEFDREIDRGSLN

Human_glucokinase        ------------------------------------------------------------
Human_hexokinase_III     PGAQRFEKMIGGLYLGELVRLVLAHLARCGVLFGGCTSPALLSQGSILLEHVAEMEDPST
Human_hexokinase_II      PGKQLFEKMISGMYMGELVRLILVKMAKEELLFGGKLSPELLNTGRFETKDISDIEGEKD
Human_hexokinase_I       PGKQLFEKMVSGMYLGELVRLILVKMAKEGLLFEGRITPELLTRGKFNTSDVSAIEKNKE

Human_glucokinase        ------------------------------------------------------------
Human_hexokinase_III     GAARVHAILQDLGLSPGASDVELVQHVCAAVCTRAAQLCAAALAAVLSCLQHSREQQTLQ
Human_hexokinase_II      GIRKAREVLMRLGLDPTQEDCVATHRICQIVSTRSASLCAATLAAVLQRIKENKGEERLR
Human_hexokinase_I       GLHNAKEILTRLGVEPSDDDCVSVQHVCTIVSFRSANLVAATLGAILNRLRDNKGTPRLR
```

```
Human_glucokinase         ----------------------------------------------MLDDRARMEAAKK
Human_hexokinase_III      VAVATGGRVCERHPRFCSVLQGTVMLLAPECDVSLIPSVDGGGRGVAMVTAVAARLAAHR
Human_hexokinase_II       STIGVDGSVYKKHPHFAKRLHKTVRRLVPGCDVRFLRSEDGSGKGAAMVTAVAYRLADQH
Human_hexokinase_I        TTVGVDGSLYKTHPQYSRRFHKTLRRLVPDSDVRFLLSESGSGKGAAMVTAVAYRLAEQH

Human_glucokinase         EKVEQILAEFQLQEEDLKKVMRRMQKEMDRGLRLETHEEASVKMLPTYVRSTPEGSEVGD
Human_hexokinase_III      RLLEETLAPFRLNHDQLAAVQAQMRKAMAKGLRGEA---SSLRMLPTFVRATPDGSERGD
Human_hexokinase_II       RARQKTLEHLQLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGD
Human_hexokinase_I        RQIEETLAHFHLTKDMLLEVKKRMRAEMELGLRKQTHNNAVVKMLPSFVRRTPDGTENGD

Human_glucokinase         FLSLDLGGTNFRVMLVKVGEGEEGQWSVKTKHQMYSIPEDAMTGTAEMLFDYISECISDF
Human_hexokinase_III      FLALDLGGTNFRVLLVRVTTG------VQITSEIYSIPETVAQGSGQQLFDHIVDCIVDF
Human_hexokinase_II       FLALDLGGTNFRVLLVRVRNGKWGG--VEMHNKIYAIPQEVMHGTGDELFDHIVQCIADF
Human_hexokinase_I        FLALDLGGTNFRVLLVKIRSGKKRT--VEMHNKIYAIPIEIMQGTGEELFDHIVSCISDF

Human_glucokinase         LDKHQMKHKKLPLGFTFSFPVRHEDIDKGILLNWTKGFKASGAEGNNVVGLLRDAIKRRG
Human_hexokinase_III      QQKQGLSGQSLPLGFTFSFPCRQLGLDQGILLNWTKGFKASDCEGQDVVSLLREAITRRQ
Human_hexokinase_II       LEYMGMKGVSLPLGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRRE
Human_hexokinase_I        LDYMGIKGPRMPLGFTFSFPCQQTSLDAGILITWTKGFKATDCVGHDVVTLLRDAIKRRE

Human_glucokinase         DFEMDVVAMVNDTVATMISCYYEDHQCEVGMIVGTGCNACYMEEMQNVELVEGDEGRMCV
Human_hexokinase_III      AVELNVVAIVNDTVGTMMSCGYEDPRCEIGLIVGTGTNACYMEELRNVAGVPGDSGRMCI
Human_hexokinase_II       EFDLDVVAVVNDTVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCV
Human_hexokinase_I        EFDLDVVAVVNDTVGTMMTCAYEEPTCEVGLIVGTGSNACYMEEMKNVEMVEGDQGQMCI

Human_glucokinase         NTEWGAFGDSGELDEFLLEYDRLVDESSANPGQQLYEKLIGGKYMGELVRLVLLRLVDEN
Human_hexokinase_III      NMEWGAFGDDGSLAMLSTRFDASVDQASINPGKQRFEKMISGMYLGEIVRHILLHLTSLG
Human_hexokinase_II       NMEWGAFGDNGCLDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRG
Human_hexokinase_I        NMEWGAFGDNGCLDDIRTHYDRLVDEYSLNAGKQRYEKMISGMYLGEIVRNILIDFTKKG

Human_glucokinase         LLFHGEASEQLRTRGAFETRFVSQVESDTGDRKQIYNILSTLGLRPSTTDCDIVRRACES
Human_hexokinase_III      VLFRGQQIQRLQTRDIFKTKFLSEIESDSLALRQVRAILEDLGLPLTSDDALMVLEVCQA
Human_hexokinase_II       LLFRGRISERLKTRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTV
Human_hexokinase_I        FLFRGQISETLKTRGIFETKFLSQIESDRLALLQVRAILQQLGLNSTCDDSILVKTVCGV

Human_glucokinase         VSTRAAHMCSAGLAGVINRMRESRSEDVMRITVGVDGSVYKLHPSFKERFHASVRRLTPS
Human_hexokinase_III      VSQRAAQLCGAGVAAVVEKIRENRGLEELAVSVGVDGTLYKLHPRFSSLVAATVRELAPR
Human_hexokinase_II       VARRAAQLCGAGMAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPK
Human_hexokinase_I        VSRRAAQLCGAGMAAVVDKIRENRGLDRLNVTVGVDGTLYKLHPHFSRIMHQTVKELSPK

Human_glucokinase         CEITFIESEEGSGRGAALVSAVACKKACMLGQ
Human_hexokinase_III      CVVTFLQSEDGSGKGAALVTAVACRLAQLTRV
Human_hexokinase_II       CDVSFLQSEDGSGKGAALITAVACRIREAGQR
Human_hexokinase_I        CNVSFLLSEDGSGKGAALITAVGVRLRTEASS
```

Figure 1.2: Sequence alignment of all four mammalian hexokinase isoforms (HKI-HKIV)

Sequence alignment of all mammalian hexokinases shows an extensive identity between isoforms I and II reaching 73%. HKIII does not hold the same identity with HKI or II as the percentage of identity ranges between 52-55%.Glucokinase (HKIV) shows approximately 52-55% identity with the other three isoforms.

1.2 Cancer metabolism and hexokinase II

In 1931 Otto Warburg was attributed the Nobel Prize in Physiology and Medicine mainly for his research on the metabolism of tumours and the respiration of cells. Warburg observed that the metabolism of cancer cells is different from that of normal adult cells. In order for normal cells to replicate, energy is required which is acquired as follows. Glucose upon entering inside the cells through glucose transporters (GLUTs) is metabolised to pyruvate after sequential reactions of the

glycolytic pathway. The glycolytic process results in the production of just two ATP molecules per glucose. In normal tissues the tricarboxylic acid (TCA) cycle and oxidative phosphorylation (OXPHOS) follow, resulting in an additional 36 molecules of ATP. Cancer cells have increased energy requirements to maintain their increased proliferative rate thus the glucose use is increased. Cancer cells show the characteristic increase of glycolysis even in the presence of oxygen, a phenomenon called aerobic glycolysis, the Warburg effect (Marie and Shinjo, 2011). The molecular basis of this shift in phenotype remains elusive.

Pedersen and his colleagues published a minireview about HKII which plays a pivotal role in fast growing cancer cells maintaining their growth and survival, thus HK isoform II may be considered as a target for cancer therapy (Pedersen et al., 2002). It is known that the genes for each of the four hexokinase isozymes are localised on different chromosomes; more specifically the HKI, HKII, HKIII and HKIV genes are localised on 10q22, 2p13, 5q35, and 7p15, respectively. None of the hexokinases derive from alternate exon splicing events from a single chromosomal locus or due to chromosomal rearrangements or deletions. This could indicate that epigenetic events and/or gene amplification play a significant role in the up-regulation of HKII gene expression during tumorigenesis (Mathupala et al., 2009).
Among the four hexokinase isoforms (HKI-IV) HKII has been reported since early years to be overexpressed in fast growing tumours. More specifically it has been demonstrated that AS-30D rat hepatoma cells contain 5-fold more HKII gene copies than normal hepatocytes with no observed structural differences in the HKII gene locus (Rempel et al., 1996).

Hexokinase isoforms from highly glycolytic tumours have been sequenced and found to contain overexpressed HKII, even 100-fold higher, than found in normal cells. This is possibly not the case for brain cancer where HKI expression may be higher (Pedersen et al., 2002). Nevertheless, HKII was recently found to be implicated in human glioblastoma multiforme (GBM), the most common brain cancer. More specifically an increased expression of HKII transcript and protein, instead of HKI which is mostly predominant in brain, was found in this type of cancer which correlated with the worse overall survival of GBM patients. Knockdown of HKII in

GBM cells led to decreased proliferation; all these findings suggest that HKII could be a target for GBM therapy (Wolf et al., 2011).

Apart from the elevated expression of this isoform, HKII in rapidly growing cancers is bound to the mitochondrial voltage dependent anion channel (VDAC) (Nakashima et al., 1986) resulting in relief from its product inhibition (G6P) while helping cancer cells to immortalise, likely by inhibiting the proapoptotic factor Bax to bind to the mitochondria (Mathupala et al., 2009), (Pastorino et al., 2002), (Robey, 2005). HKII when bound to mitochondria also benefits from preferential access to ATP produced during OXPHOS (Arora and Pedersen, 1988). The combination of the above results in G6P accumulation which has a dual role: it serves as a carbon source both for entrance in the pentose phosphate pathway (PPP), and as the initial substrate for glycolysis (Pedersen et al., 2002). The binding/ detachment of HKII in normal cells might be controlled by growth-related signalling pathways that are unable to "unlock" the enzyme in cancer cells. This can result in inhibition of apoptosis, thus the cells show increased survival (Pedersen et al., 2002).

All the above-mentioned characteristics and findings regarding hexokinase isoform II suggest that the latter is a major contributor of the immortalised profile that cancer cells exhibit (Figure 1.3). An excellent review by Lis, 2016, has demonstrated why HKII stands out from other targets of cancer treatment. Specifically the ability of HKII to bind to mitochondria is referred as the "Achilles heel" of cancer cells thus making it a promising cancer target (Lis et al., 2016).

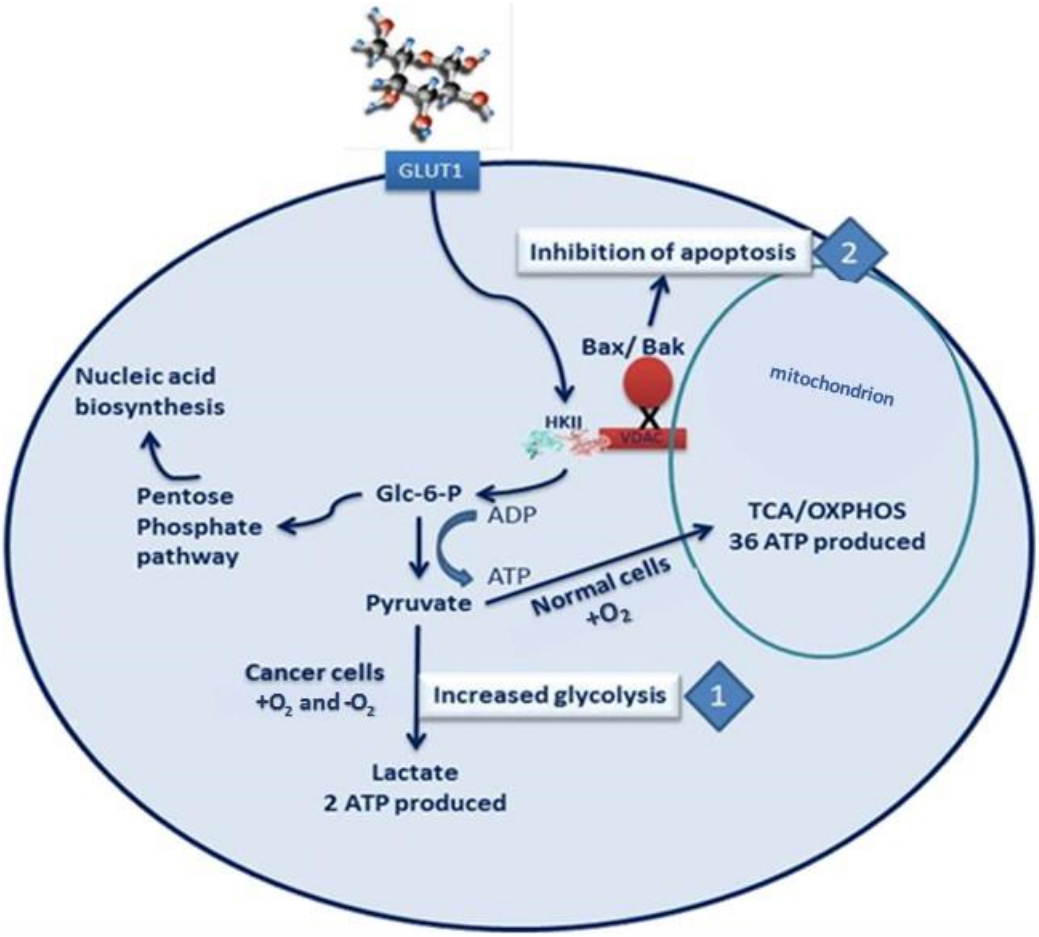


Figure 1.3: HKII, a key intermediate in cancer cell immortalisation

Glucose is brought inside the cell across the plasma-membrane by GLUTs. HKII is mainly bound to the outer mitochondrial membrane (OMM) via VDAC. G6P produced from HKII reaction can either enter the PPP for nucleic acid biosynthesis or can either be converted to pyruvate following the glycolytic pathway. Most of the pyruvate is converted to lactate instead of proceeding to the TCA cycle/ OXPHOS which takes place in normal cells in the presence of O_2. This results in the production of two ATPs in cancer cells instead of 38 ATP molecules per molecule of glucose consumed. The aerobic glycolysis is characteristic of most cancer cells and is known as the "Warburg effect". Moreover, HKII is strategically located on the OMM thus it gains a preferential access to ATP generated in the mitochondria and it becomes less sensitive to G6P inhibition. Also, HKII when bound to mitochondria can inhibit Bax induced cytochrome c release, thus it prevents apoptosis leading to cancer cell immortalisation. Adapted from (Mathupala et al., 2009).

1.3 Therapeutic studies targeting HKII

Considering the multiple roles that HKII plays in tumours, it can be considered as an attractive target for therapeutic intervention. Agents showing glycolytic inhibition (2-deoxyglucose (2-DG), and citrate) have already been tested as potential anticancer targets but have never shown to eradicate real cancers in animals. 2-DG is a glucose analogue, which enters the cancer cells to be phosphorylated by hexokinase. The product, 2-DG-6P cannot be further metabolised and it accumulates in a greater yield in cancer cells than in normal cells. The technology positron emission tomography (PET) uses the ^{18}fluorine labelled radioisotope 2-DG to image the solid tumours and is the most effective method for cancer detection (Lis et al., 2016).

Lonidamine ($C_{15}H_{10}C_{l2}N_2O_2$) is a known specific inhibitor of mitochondrial HKII since the early eighties when it was shown that 5 μM of lonidamine can decrease HK activity by 66% (Floridi et al., 1981). However, when lonidamine went to Phase II trials for GBM treatment combined with diazepam no therapeutic benefit was observed. The same drug reached Phase II/III trials for the treatment of benign prostatic hyperplasia but it was suspended after severe hepatic adverse effects (Porporato et al., 2011).

Methyl jasmonate ($C_{13}H_{20}O_3$), a plant stress hormone, can detach HKII when is bound to mitochondria although a high dose is required to have a significant effect (Goldin et al., 2008).

3-Bromopyruvate ($C_3H_3BrO_3$, 3BrPa) is a structural analogue of pyruvic acid, highly reactive with alkylating properties which covalently modifies cysteines with consequences for the conformation/ activity of the proteins (Lis et al., 2016). Studies have demonstrated that 3BrPa inhibits HKII in an *ex vivo* model of rabbit liver cancer (VX2 tumour model) (Ko et al., 2001). In a follow-up study on rats, 3BrPa was shown to selectively deplete ATP when animals were treated with this compound. All animals with advanced hepatoma were completely cancer free after 3BrPa treatment without reoccurrence of cancer (Ko et al., 2004). However, the mechanism of 3BrPa is not completely understood and it is possible, because it is a pyruvate analogue, that it can possibly affect other targets in the cells (Chen et al., 2009).

The need for a more effective anticancer drug that would selectively inhibit HKII is still required and under research, since all previous drug candidates have not been successful at the clinical trials.

1.4 Structural studies on HKII

The crystal structure of hHKII has been determined by the Structural Genomic Consortium (SGC) in 2006, with Glc (substrate) and G6P (product, allosteric inhibitor) present (PDB code: 2NZT) in both domains. In contrast to HKI and HKIII, where many studies have been performed on the catalytic C domain and the inactive N domain (Arora et al., 1993), (Tsai and Wilson, 1997), (White and Wilson, 1989), HKII contains two functional domains with comparable catalytic activities (Tsai and Wilson, 1996), (Ardehali et al., 1996).

The interaction of ATP with the active site of hexokinase II is still unknown since there is a lack of available crystal structures with the Hexokinase-ATP complex. However the putative ATP binding site has been studied by molecular modelling and site-directed mutagenesis in brain hexokinase (Zeng et al., 1996). Based on the latter studies ATP is predicted to interact with D532, R539, K621, D657, T680, E783, T863. Figure 1.4 shows the active site of hHKII with the exact positions of the Glc and G6P as identified by the crystal structure and the putative site for ATP binding.

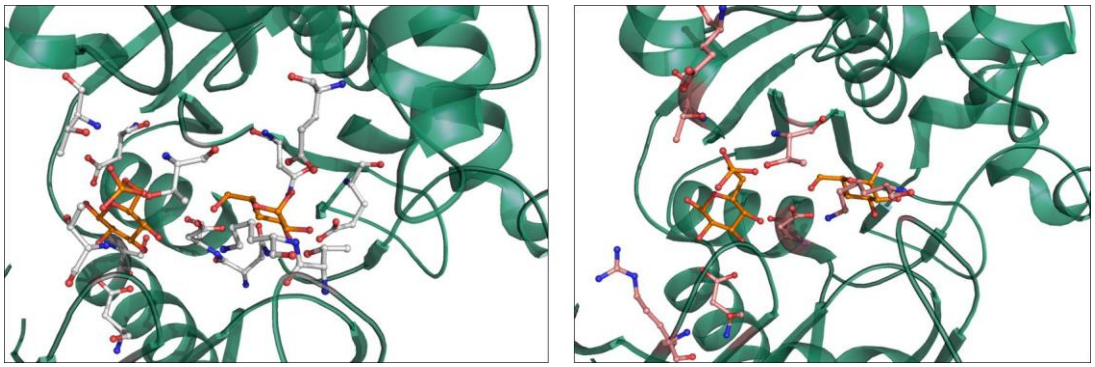


Figure 1.4: Active site of hHKII

The left panel shows the binding site of Glc and G6P shown as orange sticks. The white residues are the residues in hydrogen bond distance. The right panel shows the residues (pink sticks) which are predicted to interact with the ATP. As can be seen

the ATP putative binding site is in the same pocket as for Glc/G6P and in close proximity to both ligands

Figure 1.5 shows hexokinase II (Chain A) consisting of two domains along with a closer representation of residues of N and C domains which interact with Glc and G6P, respectively.

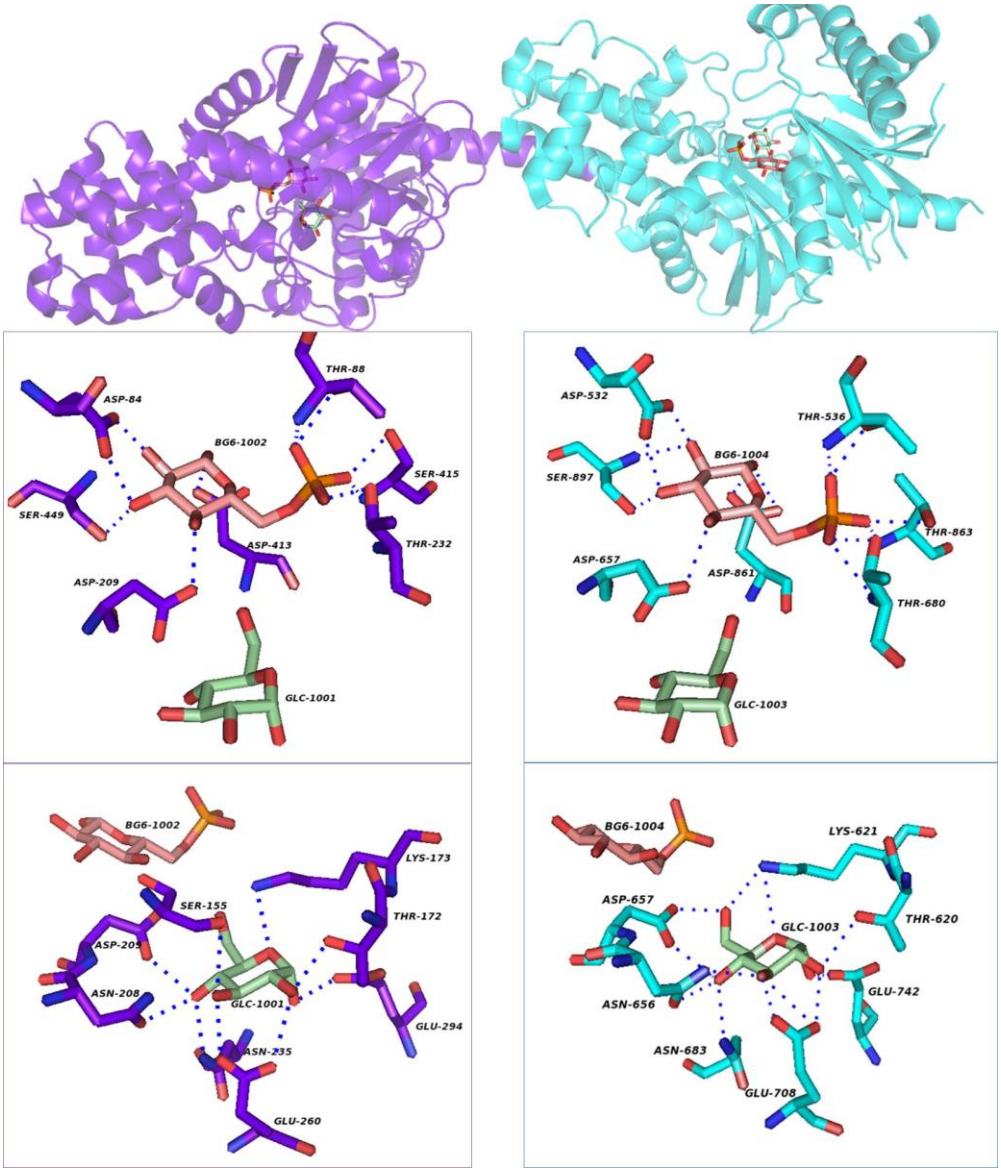


Figure 1.5: Monomer of hHKII and active site occupied by Glc and G6P

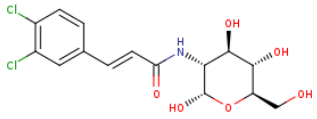
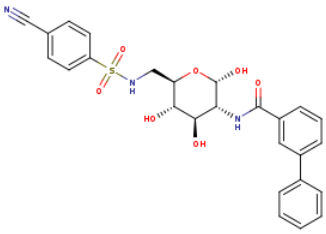
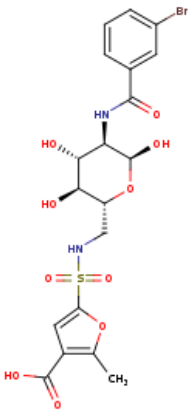
Cartoon representation of monomer hHKII (2NZT) consisting of two domains, the N (purple-blue) and C (cyan) domain. Both domains bind the ligands, Glc and G6P, which are shown as pale green and salmon sticks, respectively.

It is observed that Glc forms one contact less in the C-terminal domain and T863 substitutes for the N-terminal domain residue S415, maintaining, however, two hydrogen bonds to the G6P atom O7. The active site between the two domains are highly conserved but not identical.

HKII is considered as a potential target for cancer treatment; however, the extremely polar active site, the sequence similarity with HKI (73% identical and 84% similar) and the conserved glucose binding sites among all hexokinases have made it less attractive for drug discovery projects. Nevertheless, a group from the USA recently published a very interesting article regarding the discovery of a novel 2,6-disubstituted glucosamine series which selectively inhibited HKII in the nanomolar range (Lin et al., 2016). The following table contains the compounds discovered which were successfully co-crystallised with HKII, revealing for the first time an inhibitor-bound conformation of HKII. Unfortunately, there are no crystal structures for the compounds which showed an improved HKII selectivity. However, a Structure Activity Relationship (SAR) analysis revealed the following: a) 2,3-disubstituted sulphonamides at the 6-position, b) less-bulky amides at the 2-position and c) sulphonamide at the 2-position could improve the IC_{50} for HKII by 500-fold compared to the IC_{50} for HKI (Lin et al., 2016).

Table 1.1: Compounds which were co-crystallised with HKII. The first reported inhibitor-bound HKII

Compound_1 was one of the glucosamine derivatives discovered in a High-Throughput Screen (HTS). The co-crystal of compound_1 with HKII was rationalised to develop compound_30 which shows an improved potency for HKII with nanomolar affinity (10 nM).

Compound ID	Enzyme IC_{50} HKII (μM)	Enzyme IC_{50} HKI (μM)	Crystal structure of ligand bound HKII (PDB code)
Compound_1 	6.3	2.0	5HG1
Compound_27 	0.13	0.0079	5HFU
Compound_30 	0.010	0.020	5HEX

The very first crystallographic data for HKII bound with inhibitory ligands were obtained with compound_1 and G6P. Compound_1 is a weak inhibitor of HKII with no specificity over HKI. All donor-acceptor interactions of the hydroxyl groups observed when Glc binds to HKII are present upon compound_1 binding to HKII. The comparison with crystal structure 2NZT shows that the enzyme is flexible and an "induced-fit" mode is seen for HKII in order to accommodate the bulky compound (Fig. 1.6.A). In more detail compound_1 places its glucosamine ring in the glucose-binding pocket, which is consistent with the glucose competitive mode of these compounds. The cocrystal structure of compound_1 with HKII reveals a flexible loop in the active site (residues 616-633) which does not close over the pyranose ring, thus a wider pocket is formed which can accommodate bigger compounds.

Compounds 27 and 30 were subsequently co-crystallised with HKII without any other ligand present on the crystal (Fig. 1.6 B,C). The potency of compound_1 was improved by introduction of polar substitutions in the 6-position to mimic hydrogen bond interactions between the protein and G6P, since these new analogues were designed to extend into the G6P pocket. These modifications resulted in compound_27. Finally, a meta-carboxylic acid modification of the analogue resulted in compound_30 that was found to be more potent against HKII. However, the crystal structures of both ligands (27 and 30) did not show any G6P interactions as expected.

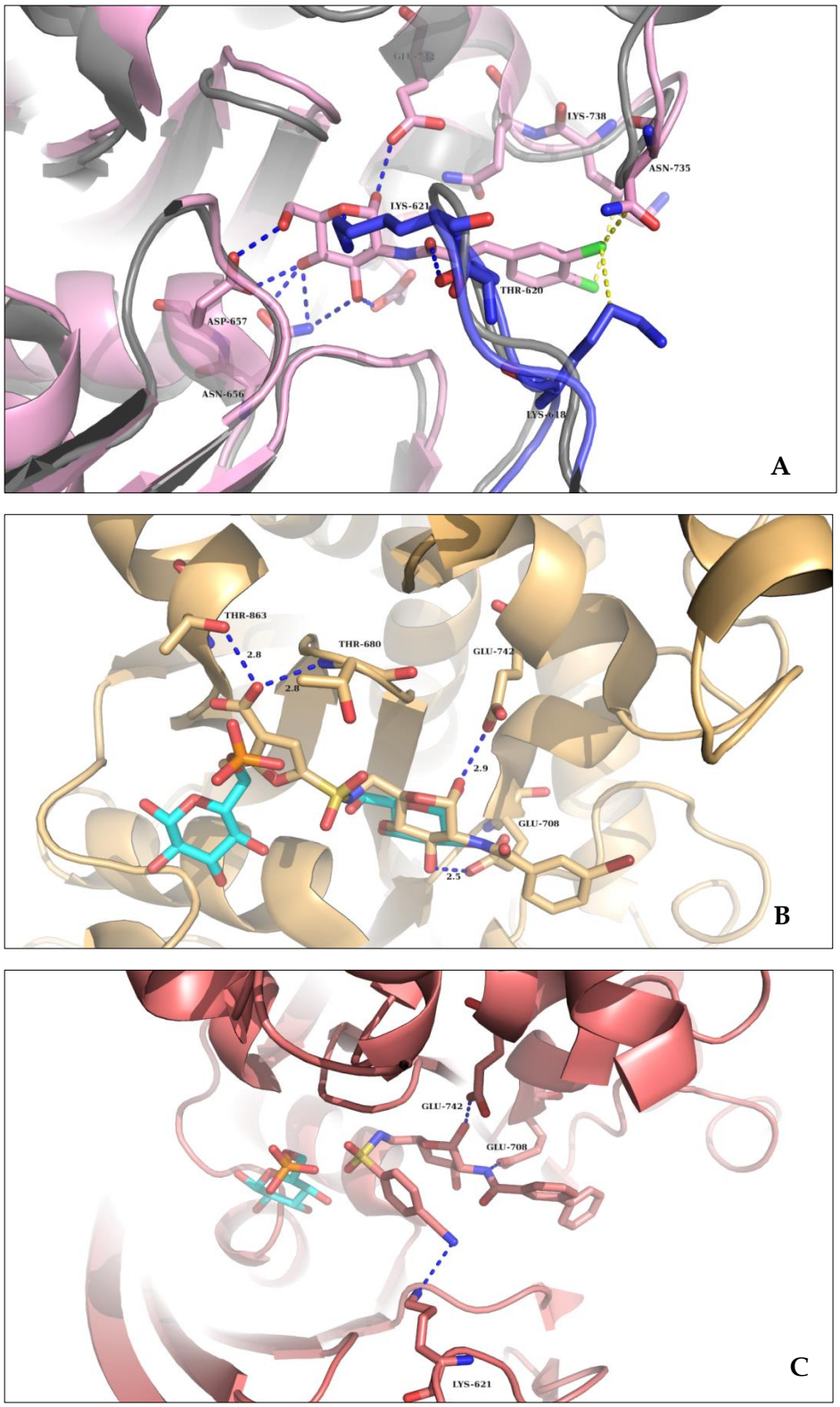


Figure 1.6: 2,6-disubstituted glucosamine analogues which were co-crystallised with HKII. The first reported inhibitor bound form of the enzyme reveals a flexible binding site.

A. 2NZT is shown as a grey cartoon and 5HG1 as a pink cartoon. The superposition of the two structures is shown. The flexible loop in the active site (residues 616-633, shown as blue loop for 5HG1 structure) does not close over the pyranose ring. The pink loop is shifted 3Å in the free HKII (without ligand).
B. 5HFU is shown as a light orange cartoon. Compound_27 is shown as light orange stick. The interactions are shown along with the distances from each residue.
C. 5HEX is shown as salmon cartoon. Compound_30 binding orientation is shown as salmon stick along with the interacting residues.

The present study from (Lin et al., 2016) shows that a specific inhibitor for isoform II can be found with activity in the nanomolar range. These results suggest that the HKII target might be a difficult target but should not be considered as unattractive and more effort should be done to discover more inhibitors. Finally, it is very interesting and important that some of the compounds from the same study were tested against a tumour cell line (UM-UC-3) which mostly expresses HKII (over HKI) and the results revealed a) inhibition of G6P production, b) reduction of glucose consumption and finally c) reduced production of lactate (Lin et al., 2016).

1.5 Trypanosoma brucei and HAT

The African trypanosome *Trypanosoma brucei* is known to be the causative agent of human sleeping sickness or Human African Trypanosomiasis (HAT), a fatal disease if left untreated. *T. brucei* has traditionally been grouped into three subspecies: *T. b. brucei, T. b. gambiense and T. b. rhodesiense*. Parasites of the latter two subspecies affect humans, while *T. b. brucei*, together with other species (*T. congolense* and *T. vivax*) cause a similar disease in cattle called 'nagana' causing major economic losses (up to 4 billion US dollar annually in sub-Saharan countries). Unfortunately, HAT is a neglected disease, since the affected people belong to deprived populations which do not provide an economic market for the pharmaceutical industry.

In the last century several major epidemics of sleeping sickness occurred each with thousands of victims. By better surveillance and vector control in the last two decades the number of annually infected people has now come down to less than 3000 (Drugs for Neglected Diseases initiative, DND*i*, 2017, http://www.dndi.org/diseases-projects/hat/). However, there is a serious risk of increased number of infected people yet again when health control breaks down in the politically unstable African countries.

HAT consists of two stages: during the first stage the parasite spreads in the blood and the lymphatic system of the human host before the parasite crosses the blood-brain barrier (stage II). Treatment of HAT is difficult especially in the CNS stage (stage II) and it is fatal if left without medical treatment (Russell et al., 2016).

Figure 1.7 shows the life cycle of *Trypanosoma brucei* parasites. When a tsetse fly takes a blood meal it injects metacyclic trypomastigotes which are transformed to bloodstream-form (BSF) trypomastigotes in the human host while they are spreading into its blood and lymphatic system. Asexual multiplication of trypomastigotes takes place and a tsetse fly during a blood meal is infected from the host. In the infected tsetse fly the BSF trypomastigotes transform into procyclic-form (PF) trypomastigotes followed by transformation to epimastigotes ending in the metacyclic transformation.

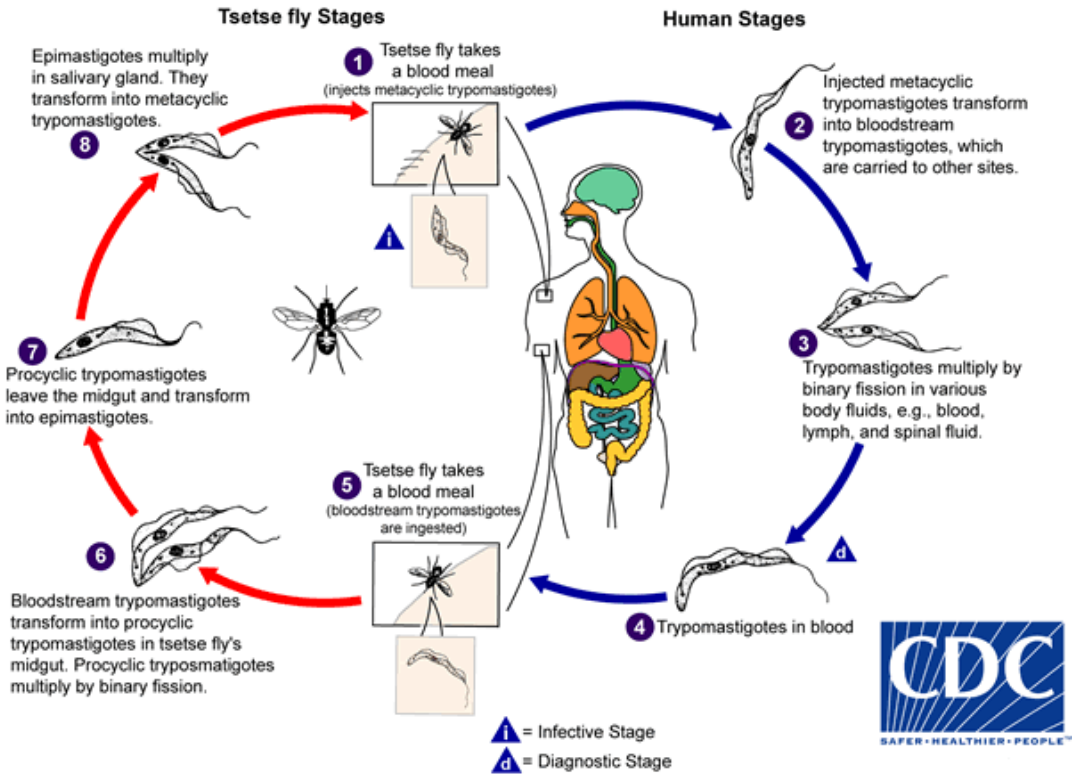


Figure 1.7: Life cycle of *Trypanosoma brucei* parasites

Picture taken from https://www.cdc.gov/parasites/sleepingsickness/biology.html

The parasites, when in bloodstream form (BSF), rely exclusively on glycolysis for ATP production and it is that form that is pathogenic to the human host. Otherwise, when they are found in the midgut of the insect vector they belong to the procyclic form (PF), most of the time catabolising amino acids to produce ATP (Chambers et al., 2008b). Only after a blood meal, they preferentially consume glucose from the blood that is however exhausted within 15-30 minutes. Trypanosome glycolysis has a unique difference compared to human as the majority of the enzymes of the glycolytic pathway are compartmentalised in peroxisomes called glycosomes (Fig.1.8) (Chambers et al., 2008b).

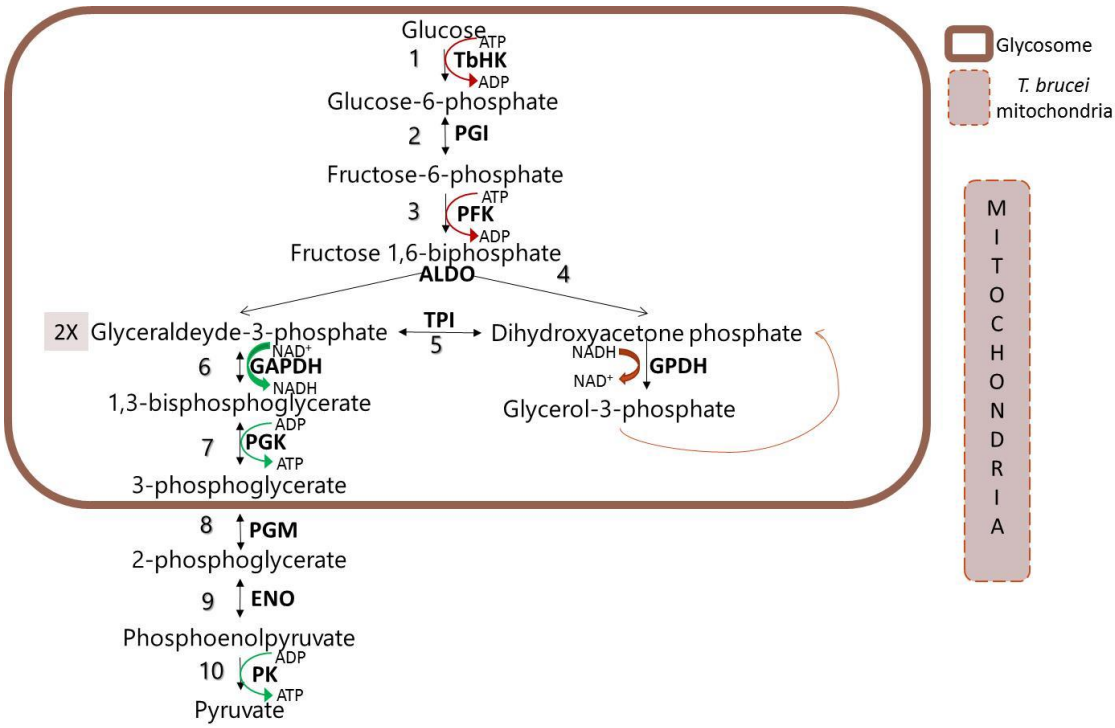


Figure 1.8: Glycolysis in the BSF of the African Trypanosome

The first seven enzymes of the glycolytic pathway are organised inside glycosomes. Under aerobic conditions glucose is converted to 3-phosphoglycerate which then is further converted to pyruvate in the cytosol.

Since BSF parasites depend entirely on the glucose metabolism for energy production, the glycolytic pathway could be exploited as a therapeutic target. Moreover, the unique organisation of glycolysis within glycosomes has led the trypanosomatid enzymes to adopt distinct kinetic, regulatory and structural properties. Thus, there is a good potential that a drug designed against the trypanosomatid enzymes will be selective, not affecting the human host enzymes (Albert et al., 2005).

T. brucei expresses two hexokinases (TbHKI, TbHKII) which show 98% sequence identity. Most of the difference is located at the end of the amino acid sequence as shown in Figure 1.9. Recombinant TbHKI is active with activity levels similar to that of other hexokinases, while recombinant TbHKII lacks any detectable activity (Morris et al., 2006). Later, it was shown that parasites express active TbHKI which forms homohexameric complexes, not covalently linked (Chambers et al., 2008b). It

is not clear why *T. brucei* expresses two almost identical proteins with one of them being in an inactive form when forming a homomeric complex, in contrast to the active hexameric TbHKI. However, TbHKII can be activated when mixed with TbHKI and for this reason it might play a regulatory role (Chambers et al., 2008b).

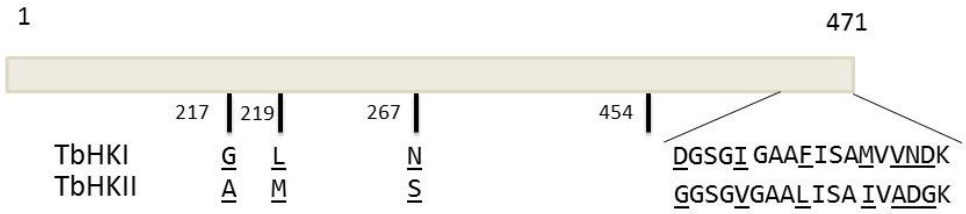


Figure 1.9: Sequence alignment between the two isoforms of T. brucei hexokinase
Comparison of amino acid sequences between TbHKI and TbHKII. The different residues are underlined. The enzyme contains 471 amino acids with apparent mass at 50 kDa. Adapted from (Morris et al., 2006).

TbHKI shares a low sequence identity with mammalian hexokinases (36-37%) and due to the unique organisation of glycosomes, TbHKI has distinct kinetic and regulatory properties like most of the other trypanosomatid glycosomal enzymes when compared with their counterparts in non-trypanosomatid organisms. For example, TbHKI is not regulated by its product G6P and it has a low specificity for ATP while it can also use ITP, UTP and GTP. Thus, the discovery of a selective inhibitor of trypanosomatid hexokinase could be considered. Moreover, TbHKI has been chemically and genetically validated as a target for therapeutic development. Morris and coworkers have shown the efficacy to kill BSF *T. brucei* with various HK inhibitors (see next section, **1.6**) and Albert et al (2005) have proved that depletion of TbHKI by RNAi is lethal to BSF trypanosomes (Morris et al., 2002, Albert et al., 2005).

1.6 Inhibition studies on TbHKI, a therapeutic approach for sleeping sickness

The currently available treatments need optimisation since they have toxic side effects, are not fully available, or are difficult to administer. For most drugs the exact mode of action is not clear. Figure 1.10 shows the structure of the established therapeutics and the drug candidates in clinical trials. Suramin and pentamidine fail to treat the disease when it is in the neurological stage and melarsoprol can cause death to 5-10% of patients receiving this drug. Finally, eflornithine has three drawbacks as it is expensive, is only effective against *T. brucei gambiense* subspecies and requires a long period (14 days) of many doses (4 times per day) of intravenous administration. However, it is successful in both stages of the disease (Sharlow et al., 2011). Recently, a combination of nifurtimox with eflornithine (NECT = nifurtimox, eflornithine combination therapy) has been introduced to lower and shorten the dose of eflornithine; nifurtimox is a drug frequently used in treatment American trypanosomiasis or Chagas disease. Like eflornithine, NECT is effective for stage II of the disease but only the form caused by *T. brucei gambiense* (Russell et al., 2016*).* Pafuramidine was suspended from clinical trials as it causes renal toxicity. Clinical phase I trials of the oxaborole compound SCYX-7158 have been successfully completed in 2016 and fexinidazole is undergoing phase II/III clinical trials (Russell et al., 2016). Because none of the above mentioned drugs have been effective it becomes clear that there is a need for development of novel compounds for future therapeutic success.

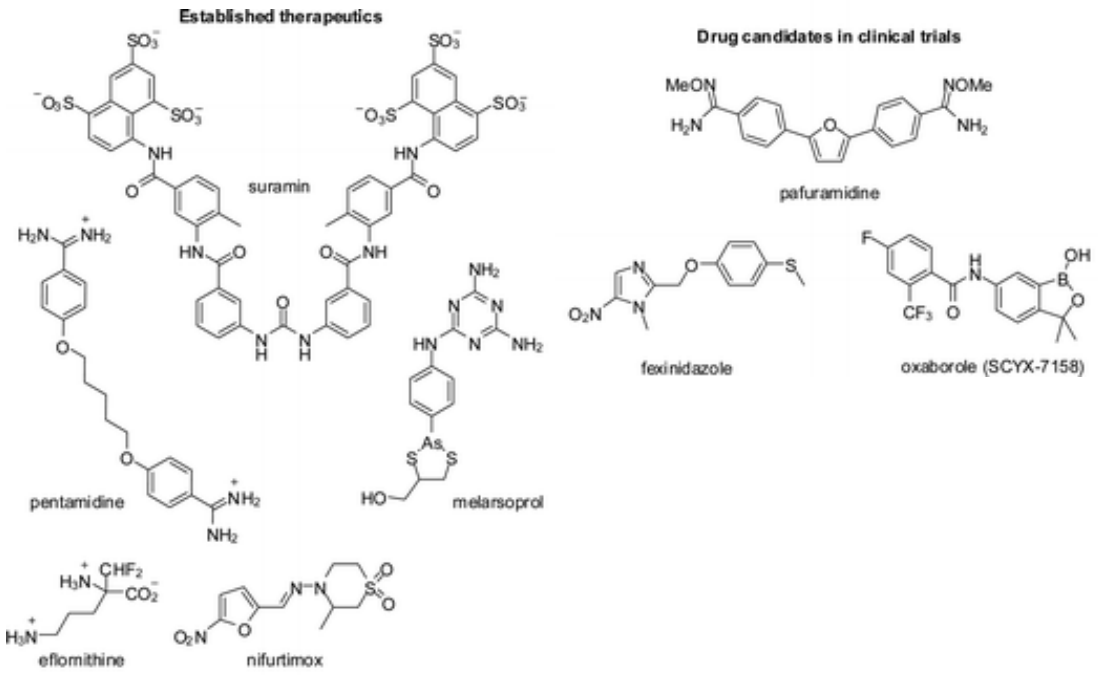


Figure 1.10: Currently available drugs and drug candidates for treatment of HAT

Suramin and pentamidine are effective for the stage I of the disease. Melarsoprol, eflornithine, and its combination therapy with nifurtimox (NECT) are designated for stage II. Pafuramidine was recently taken to clinical trial but failed because of renal toxicity. Oxaborole is scheduled to enter clinical trials phase II/III in the near future, after completion of phase I in 2016 and a phase II/III clinical trial on fexinidazole is underway. Adapted from (Russell et al., 2016).

A number of known inhibitors against TbHKI have been developed as potential drugs for HAT treatment (Table 1.2). However, they have not yet been tested in preclinical trials. Querquetin (QCN) is a natural flavanol found in plants such as apples, onions and capers. QCN was found to be a mixed inhibitor with respect to ATP against recombinant TbHKI. The compound potency did not result from dissociating the hexamer formation (Dodson et al., 2011).

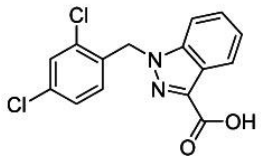
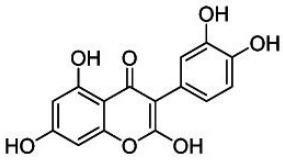
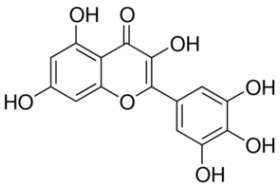
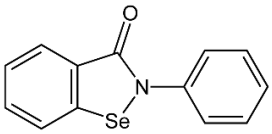
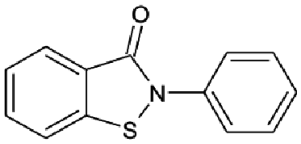
Lonidamine, a known inhibitor of mitochondria-bound mammalian HK, has been also investigated as a potential anti-parasitic compound. The compound was found to inhibit recombinant TbHKI and TbHKI purified from parasites, with the inhibition being non-competitive with respect to ATP. Moreover, lonidamine was shown to be toxic against both BSF and PF parasites cultured in vitro. However, when PF

parasites were grown in low-glucose medium, and rely heavily on amino acid oxidation for their ATP supply, the toxicity was overcome suggesting that the compound indeed inhibits glycolysis of the parasites. TbHKII $^{-/-}$ homozygous trypanosomes did not show sensitivity to lonidamine and finally TbHKI overexpression limited the toxic effects of the compound; all findings suggesting that TbHKI is the target (Chambers et al., 2008a).

A HTS campaign to identify potential inhibitors of TbHKI led to the discovery of ten inhibitors (Sharlow et al., 2010). The most potent among these inhibitors were Ebselen (EbSe) and EbS (the latter differs from EbSe by replacement of the selenium atom with sulphur). These structurally related inhibitors were found to be mixed inhibitors with respect to ATP (Sharlow et al., 2010). Ebse was further studied to reveal that it inhibits TbHKI by oxidising a single critical Cys residue (Cys327) (Joice et al., 2012). It is suggested that the BSF parasite toxicity is because of specific TbHKI inhibition (Sharlow et al., 2010).

Table 1.2: Selective known inhibitors of TbHKI

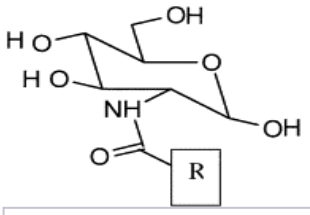
Listed are some known inhibitors of TbHKI with IC_{50} values determined in the low-high micromolar range. (R) indicates the IC_{50} of recombinant TbHKI while (L) is the IC_{50} for lysates of BSF *T. brucei* parasites.

Name	Structure	IC_{50}	Reference
Lonidamine		850µM (R) 965µM (L)	(Chambers et al., 2008a)
Querquetin		4.1±0.8µM	(Dodson et al., 2011)
Myricetin		48.9±0.7µM	
Ebselen (EbSe)		0.05±0.03µM (R) 0.43±0.02µM (L)	(Sharlow et al., 2010)
Ebsulfur (Ebs)		2±0.5µM (R) 1.2±0.12µM (L)	

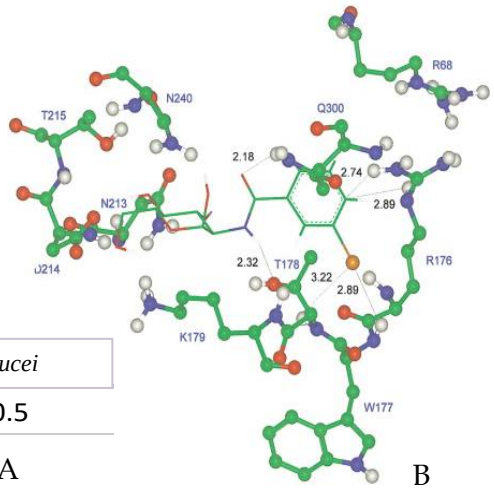
1.7 Structural studies on TbHKs

The crystal structure of *T. brucei* hexokinase is not yet known. The only published structural study involves a model of *T. brucei* hexokinase by Willson et al., 2002. During this study a series of analogues of glucosamine were tested on TbHKI purified from BSF parasites and potent specific inhibitors (with selectivity over yeast hexokinase) were identified.

The most potent compound (Compound 8, Fig. 1.11 A) is an m-bromophenyl glucosamide whose binding mode was predicted by molecular dynamics simulation performed with the structure TbHKI model (Willson et al., 2002). The compound was predicted to bind close to the glucose binding site (Fig.1.11 B) and in close proximity to ortho-toluyl-glucosamide (OTG) that has been modelled in yeast hexokinase (Steitz et al., 1977). The trans conformer of the bromine atom yielded lower interaction energy and both the aromatic and amidic bonds were coplanar compared to perpendicular for yeast hexokinase bound to ortho-toluyl-glucosamide (OTG). The bromine atom interacts with R176, T178 and Q300, whereas the close distance of both the terminal and internal NH of R176 could induce a positive π interaction which is stronger than common hydrogen bonds which could account for the improved potency of the compound 8 (Willson et al., 2002).



R	Compound ID	Yeast	*T. brucei*
m-Br-C6H4	8	–	0.5



A B

Figure 1.11: Inhibition value IC_{50} (mM) for most potent glucosamine derivative on HK of yeast and *T. brucei*. The ligand is modelled to propose the binding mechanism

A series of glucosamine analogues were studied on both the yeast and *T. brucei* hexokinase **A:** The R substitution of the glucosamine analogue is shown along with the IC_{50} value for yeast and *T. brucei* respectively. The dash indicates no effect at 20mM.

B: The modelling studies with compound 8, the most potent and selective inhibitor, show that the selectivity is provided by the coplanar orientation of the aromatic ring attached at the glucose moiety which makes a π^{+}-NH_3R interaction with the Arg176 which is unique to parasites. Adapted from (Willson et al., 2002).

No crystallographic studies have been reported for TbHKI and the binding mode of all of the above mentioned compounds is still unknown. Moreover, the glucosamine compounds need further optimisation since the IC_{50} is in millimolar range. However, the fact that the glucosamine analogues are specific for TbHKI gives hope for the discovery of potent and selective inhibitors for future therapeutic development of HAT.

1.8 Aims of the project

From a review of the literature, it is clear that mammalian and parasite hexokinases are involved in two major diseases, cancer and HAT, respectively. The enzymes are validated targets against both diseases although the current compounds/ drugs tested

have not succeeded to reach the market. Both diseases are in need of a therapeutic method which will be effective, without toxic effects on the human host and, particularly for the neglected parasitic disease, without significant high cost of administration. Therefore, the aim of this project is the discovery of novel compounds that will effectively inhibit the recombinant hexokinases from both organisms. There is also a lack of biophysical characterisation of the aforementioned proteins thus the implementation of this is another aim of this project.

2. Chapter 2: Protein expression and purification of hHKII and TbHKI

The expression and purification strategies of hHKII and TbHKI will be discussed in this chapter. The plasmid containing the ORF of human hexokinase II was purchased from Addgene (Cambridge, USA, plasmid no 25529). This construct, Thr hHKII, is lacking the first 16 amino acids of the human hexokinase isoform II and has a 6His tag with a Thrombin cleavage site at the N terminal site. During the project 4 different constructs of hHKII were made. Cloning by restriction enzyme digests was performed to create these using the NdeI as the 5' cloning site and and EcoRI as the 3' site respectively. The Addgene plasmid DNA was used as the template DNA for all cloning methods. The different constructs created are numbered here:

1) hHKII (17-917) with a TEV recognition site (ENLYFQ/S) on the N-terminal site **(TEV hHKII)**
2) hHKII (17-917) without an affinity tag **(untagged hHKII)**
3) N domain of hHKII with a TEV recognition site on the N-terminal (17-475) **(N hHKII)**
4) C domain of hHKII with a TEV recognition site at the beginning of C domain (476-917) **(C hHKII)**

A sequence alignment of the constructs is shown in Figure 2.1.

```
FLhHKII_Uniprot      ---MIASHLLAYFFTELNHDQVQKVDQYLYHMRLSDETLLEISKRFRKEMEKGLGATTHP
TEV_hHKII            ----HHHHHHGAENLYFQSDQVQKVDQYLYHMRLSDETLLEISKRFRKEMEKGLGATTHP
Thr_hHKII            MGSSHHHHHHSSGLVPRGSDQVQKVDQYLYHMRLSDETLLEISKRFRKEMEKGLGATTHP
Untagged_hHKII       -------------------DQVQKVDQYLYHMRLSDETLLEISKRFRKEMEKGLGATTHP
N_hHKII              ----HHHHHHGAENLYFQGDQVQKVDQYLYHMRLSDETLLEISKRFRKEMEKGLGATTHP
C_hHKII              ------------------------------------------------------------

FLhHKII_Uniprot      TAAVKMLPTFVRSTPDGTEHGEFLALDLGGTNFRVLWVKVTDNGLQKVEMENQIYAIPED
TEV_hHKII            TAAVKMLPTFVRSTPDGTEHGEFLALDLGGTNFRVLWVKVTDNGLQKVEMENQIYAIPED
Thr_hHKII            TAAVKMLPTFVRSTPDGTEHGEFLALDLGGTNFRVLWVKVTDNGLQKVEMENQIYAIPED
Untagged_hHKII       TAAVKMLPTFVRSTPDGTEHGEFLALDLGGTNFRVLWVKVTDNGLQKVEMENQIYAIPED
N_hHKII              TAAVKMLPTFVRSTPDGTEHGEFLALDLGGTNFRVLWVKVTDNGLQKVEMENQIYAIPED
C_hHKII              ------------------------------------------------------------

FLhHKII_Uniprot      IMRGSGTQLFDHIAECLANFMDKLQIKDKKLPLGFTFSFPCHQTKLDESFLVSWTKGFKS
TEV_hHKII            IMRGSGTQLFDHIAECLANFMDKLQIKDKKLPLGFTFSFPCHQTKLDESFLVSWTKGFKS
Thr_hHKII            IMRGSGTQLFDHIAECLANFMDKLQIKDKKLPLGFTFSFPCHQTKLDESFLVSWTKGFKS
Untagged_hHKII       IMRGSGTQLFDHIAECLANFMDKLQIKDKKLPLGFTFSFPCHQTKLDESFLVSWTKGFKS
N_hHKII              IMRGSGTQLFDHIAECLANFMDKLQIKDKKLPLGFTFSFPCHQTKLDESFLVSWTKGFKS
C_hHKII              ------------------------------------------------------------

FLhHKII_Uniprot      SGVEGRDVVALIRKAIQRRGDFDIDIVAVVNDTVGTMMTCGYDDHNCEIGLIVGTGSNAC
TEV_hHKII            SGVEGRDVVALIRKAIQRRGDFDIDIVAVVNDTVGTMMTCGYDDHNCEIGLIVGTGSNAC
Thr_hHKII            SGVEGRDVVALIRKAIQRRGDFDIDIVAVVNDTVGTMMTCGYDDHNCEIGLIVGTGSNAC
Untagged_hHKII       SGVEGRDVVALIRKAIQRRGDFDIDIVAVVNDTVGTMMTCGYDDHNCEIGLIVGTGSNAC
N_hHKII              SGVEGRDVVALIRKAIQRRGDFDIDIVAVVNDTVGTMMTCGYDDHNCEIGLIVGTGSNAC
C_hHKII              ------------------------------------------------------------

FLhHKII_Uniprot      YMEEMRHIDMVEGDEGRMCINMEWGAFGDDGSLNDIRTEFDQEIDMGSLNPGKQLFEKMI
TEV_hHKII            YMEEMRHIDMVEGDEGRMCINMEWGAFGDDGSLNDIRTEFDQEIDMGSLNPGKQLFEKMI
Thr_hHKII            YMEEMRHIDMVEGDEGRMCINMEWGAFGDDGSLNDIRTEFDQEIDMGSLNPGKQLFEKMI
Untagged_hHKII       YMEEMRHIDMVEGDEGRMCINMEWGAFGDDGSLNDIRTEFDQEIDMGSLNPGKQLFEKMI
N_hHKII              YMEEMRHIDMVEGDEGRMCINMEWGAFGDDGSLNDIRTEFDQEIDMGSLNPGKQLFEKMI
C_hHKII              ------------------------------------------------------------

FLhHKII_Uniprot      SGMYMGELVRLILVKMAKEELLFGGKLSPELLNTGRFETKDISDIEGEKDGIRKAREVLM
TEV_hHKII            SGMYMGELVRLILVKMAKEELLFGGKLSPELLNTGRFETKDISDIEGEKDGIRKAREVLM
Thr_hHKII            SGMYMGELVRLILVKMAKEELLFGGKLSPELLNTGRFETKDISDIEGEKDGIRKAREVLM
Untagged_hHKII       SGMYMGELVRLILVKMAKEELLFGGKLSPELLNTGRFETKDISDIEGEKDGIRKAREVLM
N_hHKII              SGMYMGELVRLILVKMAKEELLFGGKLSPELLNTGRFETKDISDIEGEKDGIRKAREVLM
C_hHKII              ------------------------------------------------------------

FLhHKII_Uniprot      RLGLDPTQEDCVATHRICQIVSTRSASLCAATLAAVLQRIKENKGEERLRSTIGVDGSVY
TEV_hHKII            RLGLDPTQEDCVATHRICQIVSTRSASLCAATLAAVLQRIKENKGEERLRSTIGVDGSVY
Thr_hHKII            RLGLDPTQEDCVATHRICQIVSTRSASLCAATLAAVLQRIKENKGEERLRSTIGVDGSVY
Untagged_hHKII       RLGLDPTQEDCVATHRICQIVSTRSASLCAATLAAVLQRIKENKGEERLRSTIGVDGSVY
N_hHKII              RLGLDPTQEDCVATHRICQIVSTRSASLCAATLAAVLQRIKENKGEERLRSTIGVDGSVY
C_hHKII              ------------------------------------------------------------
```

```
FLhHKII_Uniprot      KKHPHFAKRLHKTVRRLVPGCDVRFLRSEDGSGKGAAMVTAVAYRLADQHRARQKTLEHL
TEV_hHKII            KKHPHFAKRLHKTVRRLVPGCDVRFLRSEDGSGKGAAMVTAVAYRLADQHRARQKTLEHL
Thr_hHKII            KKHPHFAKRLHKTVRRLVPGCDVRFLRSEDGSGKGAAMVTAVAYRLADQHRARQKTLEHL
Untagged_hHKII       KKHPHFAKRLHKTVRRLVPGCDVRFLRSEDGSGKGAAMVTAVAYRLADQHRARQKTLEHL
N_hHKII              KKHPHFAKRLHKTVRRLVPGCDVRFLRSEDGSGKGAAMVTAVAYRLADQHRARQKTLE--
C_hHKII              ------------------------------------------HHHHHHGAENLYFQGHL

FLhHKII_Uniprot      QLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGDFLALDLGGTN
TEV_hHKII            QLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGDFLALDLGGTN
Thr_hHKII            QLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGDFLALDLGGTN
Untagged_hHKII       QLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGDFLALDLGGTN
N_hHKII              ------------------------------------------------------------
C_hHKII              QLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGDFLALDLGGTN

FLhHKII_Uniprot      FRVLLVRVRNGKWGGVEMHNKIYAIPQEVMHGTGDELFDHIVQCIADFLEYMGMKGVSLP
TEV_hHKII            FRVLLVRVRNGKWGGVEMHNKIYAIPQEVMHGTGDELFDHIVQCIADFLEYMGMKGVSLP
Thr_hHKII            FRVLLVRVRNGKWGGVEMHNKIYAIPQEVMHGTGDELFDHIVQCIADFLEYMGMKGVSLP
Untagged_hHKII       FRVLLVRVRNGKWGGVEMHNKIYAIPQEVMHGTGDELFDHIVQCIADFLEYMGMKGVSLP
N_hHKII              ------------------------------------------------------------
C_hHKII              FRVLLVRVRNGKWGGVEMHNKIYAIPQEVMHGTGDELFDHIVQCIADFLEYMGMKGVSLP

FLhHKII_Uniprot      LGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRREEFDLDVVAVVND
TEV_hHKII            LGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRREEFDLDVVAVVND
Thr_hHKII            LGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRREEFDLDVVAVVND
Untagged_hHKII       LGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRREEFDLDVVAVVND
N_hHKII              ------------------------------------------------------------
C_hHKII              LGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRREEFDLDVVAVVND

FLhHKII_Uniprot      TVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCVNMEWGAFGDNGC
TEV_hHKII            TVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCVNMEWGAFGDNGC
Thr_hHKII            TVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCVNMEWGAFGDNGC
Untagged_hHKII       TVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCVNMEWGAFGDNGC
N_hHKII              ------------------------------------------------------------
C_hHKII              TVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCVNMEWGAFGDNGC

FLhHKII_Uniprot      LDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRGLLFRGRISERLK
TEV_hHKII            LDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRGLLFRGRISERLK
Thr_hHKII            LDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRGLLFRGRISERLK
Untagged_hHKII       LDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRGLLFRGRISERLK
N_hHKII              ------------------------------------------------------------
C_hHKII              LDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRGLLFRGRISERLK

FLhHKII_Uniprot      TRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTVVARRAAQLCGAG
TEV_hHKII            TRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTVVARRAAQLCGAG
Thr_hHKII            TRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTVVARRAAQLCGAG
Untagged_hHKII       TRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTVVARRAAQLCGAG
N_hHKII              ------------------------------------------------------------
C_hHKII              TRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTVVARRAAQLCGAG

FLhHKII_Uniprot      MAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPKCDVSFLQSEDGS
TEV_hHKII            MAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPKCDVSFLQSEDGS
Thr_hHKII            MAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPKCDVSFLQSEDGS
Untagged_hHKII       MAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPKCDVSFLQSEDGS
N_hHKII              ------------------------------------------------------------
C_hHKII              MAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPKCDVSFLQSEDGS

FLhHKII_Uniprot      GKGAALITAVACRIREAGQR
TEV_hHKII            GKGAALITAVACRIREAGQR
Thr_hHKII            GKGAALITAVACRIREAGQR
Untagged_hHKII       GKGAALITAVACRIREAGQR
N_hHKII              --------------------
```

Figure 2.1: Sequence alignment of all hHKII constructs created for the project

FL_hHKII_Uniprot is the protein sequence taken from Uniprot database (http://www.uniprot.org/uniprot/P52789). The coloured sequence in FL-hHKII shows the

residues that are absent from all other constructs. The coloured sequence for the rest shows the affinity tags at the beginning of each construct.

T. brucei hexokinase I ORF was purchased from Geneart. According to TriTrypDB database (Aslett et al., 2009) the amino acid sequence for the active form of hexokinase (TbHKI) is the following:

```
                10        20        30        40        50        60        70        80        90
TbHK  MSRRLNNILEHISIQGNDGETVRAVKRDVAMAALTNQFTMSVESMRQIMTYLLYEMVEGLEGRESTVRMLPSYVYKADPKRATGVFYALD

               100       110       120       130       140       150       160       170       180
TbHK  LGGTNFRVLRVACKEGAVVDSSTSAFKIPKYALEGNATDLFDFIASNVKKTMETRAPEDLNRTVPLGFTFSFPVEQTKVNRGVLIRWTKG

               190       200       210       220       230       240       250       260       270
TbHK  FSTKGVQGNDVIALLQAAFGRVSLKVNVVALCNDTVGTLISHYFKDPEVQVGVIIGTGSNACYFETASAVTKDPAVAARGSALTPINMES

               280       290       300       310       320       330       340       350       360
TbHK  GNFDSKYRFVLPTTKFDLDIDDASLNKGQQALEKMISGMYLGEIARRVIVHLSSINCLPAALQTALGNRGSFESRFAGMISADRMPGLQF

               370       380       390       400       410       420       430       440       450
TbHK  TRSTIQKVCGVDVQSIEDLRIIRDVCRLVRGRAAQLSASFCCAPLVKTQTQGRATIAIDGSVFEKIPSFRRVLQDNINRILGPECDVRAV

               460       470
TbHK  LAKDGSGIGAAFISAMVVNDK
```

The DNA sequence was optimised for *E.coli* expression system to overcome the codon usage bias. The NdeI and EcorI restriction sites were used as the 5’ and 3’ cloning sites. The 5’ cloning site was designed in order to attach the 6His tag with a TEV cleavable site (HMHHHHHHGAENLYFQG).

2.1 Methods

2.1.1 Cloning of human HKII constructs

For cloning we used the DNA coding sequence for Thr hHKII (17-917) that was obtained from Addgene. The following set of primers were designed and purchased from Integrated DNA Technologies (IDT) to amplify the DNA coding sequence.

Table 2.1: All the primers used for the cloning procedures

The below set of primers were designed and purchased from Integrated DNA Technologies (IDT) to amplify the DNA coding sequence (ORF of hHKII from Addgene).

Primer name	Description	5'-3' Sequence
TEV For	Replaces Thr site for TEV site (ENLYFQ/S**)** Attaches NdeI restriction site at the beginning	GGTTTACATATGCATCATCATCATC ATCACGGTGCTGAGAACCTATATTT CCAAAGTGACCAAGTGCAGAAGGTT GAC
hHKII Rev	Attaches EcoRI restriction site at the end of the full length protein	CGGAATTCTCATCACTGTCCAGCCT CACGGATG
N For	Attaches NdeI restriction site at the beginning and TEV site and 6xhis tag	GGTTTACATATGCATCATCATCATC ATCACGGTGCTGAAAACCTGTATTT TCAGGGCGACCAAGTGCAGAAGGTT GAC

Primer name	Description	5'-3' Sequence
N Rev	Attaches EcorI restriction site at the end of N domain	CGGAATTCTCATCACTCTAATGTCTTCTGGCGGGCACGG
C For	Attaches NdeI restriction site at the beginning and TEV site and 6xhis tag	GGTTTACATATGCATCATCATCATCATCACGGTGCTGAAAACCTGTATTTTCAGGGCCATCTGCAGCTGAGCCATGAC
C Rev	Attaches EcorI restriction site at the end of C domain	CGGAATTCTCATCACTGTCCAGCCTCACGGATG
Untag For	Attaches NdeI restriction site at the beginning	GGTTTACATATGGACCAAGTGCAGAAGGTTGAC

For all PCR reactions the Elongase enzyme mix kit (Invitrogen) was used. The components of the PCR reaction are shown in the table below.

Table 2.2: Reaction set up for PCR Elongase enzyme mix

Elongase enzyme mix was used for the DNA amplification. The recommended protocol was followed. Mix 1 is added to Mix 2 in an amplification tube on ice.

Mix 1	**Volume**	**Mix2**	**Volume**
dNTP mix (10mM)	1 µl	5x Buffer A	5 µl
Forward primer (10µM)	1 µl	5x Buffer A	5 µl
Reverse primer (10µM)	1 µl	Elongase	1 µl
DNA template (>100nM)	1-2 µl (depending the concentration of the template)	dH_2O	Up to 30 µl
dH_2O	Up to 20 µl		

The annealing temperature was not the same for all cloning as the primers had different annealing temperatures (Ta) based on their length and composition.

The PCR reaction conditions for the *untagged and TEV hHKII* are shown below:

1) Pre-amplification denaturation — 94^{o}C for 30sec, 1cycle
2) Thermal cycling — Denaturation: 94^{o}C for 30sec; Annealing: 55-65^{o}C for 30sec (30 cycles); Extension: 68^{o}C for 2min:15sec

For the amplification of the *N and C domains of hHKII* the following cycling conditions were performed:

1) Pre-amplification denaturation 94°C for 30sec, 1cycle

2) Thermal cycling

Denaturation: 94°C for 30sec	10 cycles
Annealing: 50°C for 30sec	
Extension: 68°C for 2min:15sec	
Denaturation: 94°C for 30sec	20 cycles
Annealing: 60°C for 30sec	
Extension: 68°C for 2min:15sec	

The PCR cycling conditions are done in two steps for the N and C domains. The reason for this is because the Tm of the part of the primer complementary only to the gene is lower compared to the Tm of the entire primer (Restriction site + His tag + gene). In this case the first 10 cycles performed are specific for the part of the primer complementary to the gene (55 degrees Celsius). Then the Tm is increased to 65 degrees Celsius to factor in the whole length of the primer, thereby making the PCR product more likely to be highly specific.

The PCR products were verified on a 0.7% agarose gel. The band at the expected size was cut out from the agarose gel and purified using a PCR purification kit (QIAquick).

After the purification of the DNA, the overnight restriction digests for the donor and recipient plasmid were performed separately. All restriction enzymes were purchased from New England Biolabs and the protocol performed is shown in Table 2.3.

Table 2.3: Set up reaction for double restriction digest

The overnight digests of vector and DNA insert were performed at 37°C using both restriction enzymes (1 µl each is added to the mix). The mix volume is made up to 50 µl following the instructions below.

Component	**Volume**
Restriction enzyme 1	1 µl
Restriction enzyme 2	1 µl
Vector or Insert	~1 µg
10X NEBuffer	5 µl
dH_2O	Up to 50 µl

The overnight incubation took place at 37°C. The restriction digest was verified in a 0.7% agarose gel and the desired bands were excised from the gel and purified. Finally, we conducted a DNA ligation to fuse the insert with the recipient plasmid, pET3a. T4 DNA ligase from New England Biolabs was used for this step and the following reaction was set up in a microcentrifuge tube on ice:

Table 2.4: Set up reaction for DNA ligation

DNA ligation was performed following the titration below, to ligate pET3a with the insert DNA.

Component	**Volume**
10X T4 DNA ligase Buffer	2 µl
Vector DNA (4.64 kb)	70 ng or 30 ng
Insert DNA (2.7 and 1.35kb)	30 ng or 35 ng
T4 DNA ligage	1 µl
Nuclease free H2O	Up to 20 µl

The ligation was made with a 4:1 molar ratio of insert to vector taking into account the DNA sizes. For whole hHKII (2.7 kb) the ratio was 30 ng:70 ng of insert:vector and for the N and C domains (1.35 kb) respectively, the ratio was 35 ng:30 ng. The ligation mix was incubated for 20 min at room temperature (RT), then cooled on ice for a couple of minutes. DH5α cells were then transformed following the common protocol. Plates were left overnight at 37^{o}C and next day were examined for single colonies. Mini-preps were set up overnight in 5 ml LB containing 100 µg/ml carbenicillin and plasmid DNA was purified using QIAprep Spin Miniprep kit from QIAGEN. All DNA sequencing was carried out by either the GenePool service of University of Edinburgh or the Dundee sequencing facility within the Medical Research Council Protein Phosphorylation and Ubiquitylation Unit (MRC PPU) in Dundee, Scotland. DNA chromatograms were viewed with FinchTV (http://www.cambridgesoft.com/services/SupportNews/details/?SupportNews=124) and sequences were analysed with Clustal Omega software (Larkin et al., 2007).

2.1.2 TbHKI subcloning to expression vector

The plasmid containing the gene of interest is shown in Figure 2.2. The cloning plasmid has a size of 3754 bp from which 1476 bp belong to the gene of interest. The plasmid was transformed to DH5a cells following the common protocol.

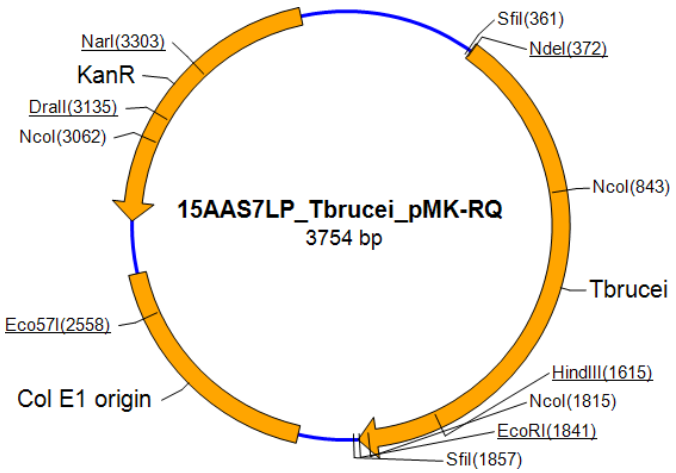


Figure 2.2: Cloning plasmid containing the ORF of TbHKI

EcorI and NdeI are the restriction recognition sites added to the 5' and 3' coding sequence.

Following the transformation of DH5a cells the DNA of interest was cloned to the expression vector pET3a. Restriction reactions were set up for both cloning and expression vectors. EcorI and NdeI were purchased from New England Biolabs and the recommended protocol was used. To maximise the capacity of the restriction enzyme the reaction was left overnight at 37^{o}C.The reactions were run in a 0.7% agarose gel, stained with the Safeview dye, for 50 min at 100 mAU. The expression vector pET3a with TbHKI ORF was kept at -20^{o}C for future transformation of different cell lines for protein expression trials.

2.1.3 Transformation and expression of hHKII and TbHKI constructs

2.1.3.1 hHKII

Plasmid DNA was transformed to *E.coli* BL21 Codon Plus (DE3) RIL competent cells (Agilent). Protocols from suppliers were followed. All agar plates and mediums contained the appropriate antibiotics.

Construct	***Antibiotics used in LB plates and mediums***
Thr hHKII (pET28-LIC)	50 μg/ml Kanamycin (plasmid), 34 μg/ml Chloramphenicol (cells)
Untagged hHKII, TEV hHKII, N hHKII, C hHKII (pET-3a)	100 μg/ml Carbenicillin (plasmid), 34 μg/ml Chloramphenicol (cells)

Single colonies of transformants were picked from LB plates and used to inoculate 50 ml of LB and left shaking overnight at 250 rpm, 37^{o}C. 50ml of the overnight culture was used to inoculate each of 500 ml of Terrific broth (TB) medium, LB medium, 2xTY medium and SOC medium. When OD_{600}= 3.0, protein expression was induced with 1 mM IPTG and the cultures transferred to 15^{o}C and left shaking overnight. For N and C domains only TB and LB medium were used. For cold shock treatment, when OD_{600}= 0.8, cultures were transferred to 4^{o}C for 1 hour. Finally

protein expression was induced with 1mM IPTG at 20°C overnight. Induction performed at 250 rpm unless otherwise stated.

2.1.3.2 TbHKI

The following table contains in summary the cell lines, media and conditions tested in order to get soluble expression of TbHKI. For every cell line the suggested protocol was followed. BL21 star (DE3) and C43 (DE3) competent cells were available within the group. BL21-CodonPlus RIL (DE3) and ArcticExpress competent cells were purchased from Agilent. Overnight cultures (obtained as in section 2.1.3.1) were inoculated to the appropriate medium for induction.

Table 2.5: Conditions screened for expression of TbHKI

Cell line	**Antibiotics used**	**Media**	**Description**
BL21-CodonPlus (DE3) RIL strain	Carbenicillin and Chloramphenicol	EnPresso growth system	1 mM IPTG induction, Overnight at 30°C, 250 rpm
BL21-CodonPlus (DE3) RIL strain	Carbenicillin and Chloramphenicol	LB, SB, 2xTY	1 mM IPTG induction, Overnight at 18°C, with and without cold shock, 250 rpm
BL21-CodonPlus (DE3) RIL strain	Carbenicillin and Chloramphenicol	LB, TB	1 mM IPTG induction, 3h at 37°C, 250 rpm
BL21-CodonPlus (DE3) RIL strain	Carbenicillin and Chloramphenicol	LB, TB	Cold shock, 1 mM IPTG induction, 20°C overnight, 250 and 100 rpm
C43 (DE3)	Carbenicillin	LB, TB	1 mM IPTG induction, Overnight at 18°C, 3h at 37°C, 250 rpm

Cell line	Antibiotics used	Media	Description
BL21 star (DE3)	Carbenicillin	LB, SB, 2xTY	1 mM IPTG induction, Overnight at 18^{o}C, 250 rpm, with and without cold shock
BL21 star (DE3)	Carbenicillin	LB, TB	1 mM IPTG induction, Overnight at 18^{o}C, 3h at 37^{o}C, 250 rpm
BL21 star (DE3)	Carbenicillin	TB	Coexpression with GroEL/ GroES, 1 mM IPTG induction, Overnight at 18^{o}C, 250 rpm
ArcticExpress (DE3)	Carbenicillin and 20 μg/ml Gentamycin	LB, SB	1 mM IPTG, 24h at 13^{o}C, 250 rpm

2.1.4 Common steps before any purification strategy

All purifications were performed using AKTA purifier systems (10 ml/min or 100 ml/min) at 6^{o}C. All buffers were prepared at 4-6^{o}C unless otherwise stated. The cellular pellet was resuspended in 5 ml of Lysis Buffer (i.e the buffer used to equilibrate the first chromatographic column containing a protease inhibitor cocktail tablet, unless otherwise stated) per 1 gram of cell pellet. The resuspension was passed through a Constant Systems cell disruptor to break open the cells. The lysate from this was then centrifuged at 20,000xg for 45 min at 4-8^{o}C. The supernatant was filtered using a 0.22 μm filter. If the cell pellets were treated otherwise it is stated where appropriate.

2.1.5 Three step purification of Thr hHKII

2.1.5.1 Anion Exchange Chromatography (AEX)

The supernatant was loaded into the column HiPrep 16/10 DEAE FF, pre-equilibrated with Buffer A (10 mM Tris pH 7.5, 0.5 M NaCl, 5% glycerol, 5 mM imidazole) with a stable flowrate set at 5ml/ min. Molecules that are bound to the resin elute with 0.5 M NaCl. The collected fractions were pooled together for the next step.

2.1.5.2 Immobilised Metal Ion Affinity Chromatography (IMAC)

The fractions from previous step were loaded onto a 5 ml affinity column (HiTrap FF 5 ml, charged with 0.1 M $NiCl_2$, pre-equilibrated with Buffer A) at 5 ml/min flowrate. The column was washed with 30 column volumes (CV) of Buffer B (10 mM Tris pH 7.5, 0.5 M NaCl, 5% glycerol, 30 mM imidazole). The target protein was eluted with 10 CV of Buffer C (10 mM Tris, pH 7.5, 0.5 M NaCl, 5% glycerol, 250 mM imidazole). The purity of the fractions was confirmed in SDS-PAGE gel and the fractions containing the desired protein were pooled together. The pooled fractions were concentrated to ~10 ml for the next step using a Vivaspin column (molecular weight cut-off= 30 kDa).

2.1.5.3 Size exclusion chromatography (SEC)

HiLoad 26/60 Superdex 200pg was equilibrated with 2 CV of Gel Filtration Buffer (10 mM Tris, pH 8.0, 0.5 M NaCl, 5 mM $MgCl_2$ and 10 mM DTT). The sample was injected through a 10 ml loop, the flowrate was set to 2.5 ml/ min and fractions were collected and resolved in an SDS-PAGE. Pure hHKII was pooled and concentrated to 1 mg/ ml.

2.1.6 Cleavage of 6His tag using Thrombin protease

1) hHKII in Gel Filtration Buffer was incubated with Thrombin (GE healthcare) at 4^oC overnight (10 U/mg of protein). SDS-PAGE analysis was performed.

2) As Thrombin could be more active in a lower concentration of NaCl and in the presence of $CaCl_2$, hHKII in Gel Filtration Buffer was diluted 1:10 in Thr Buffer: 50 mM Tris, pH 8.0 (adjusted at RT), 150 mM NaCl, 2.5 mM $CaCl_2$. 50 μg of protein was incubated overnight at RT with 2 U and 4 U of Thrombin. Samples were analysed by SDS-PAGE (200 V, 55 min) and Western Blot.

2.1.7 Solubility assay

10 μl of the protein (initial concentration 10 μM) was diluted to 90 μl of buffer under study and left for incubation on the bench for 1 hour. Using a Vivaspin concentrator with a molecular weight cut-off= 300 kDa (GE Healthcare), soluble protein was separated from the aggregates (after centrifugation at 16,000x g for 15 minutes) and collected in the flow through. The aggregated protein was retained in the membrane and to retrieve it 30 μl of dH_2O was pipetted across the membrane. 30 μl of soluble and 30 μl of aggregated protein were mixed separately with 10 μl of *4x Sample Buffer* (250 mM Tris-HCl, 40% glycerol, 140 mM SDS, 0.6M β-mercaptoethanol, pH 6.8) and heated to 90^{o}C for 10 minutes prior to SDS-PAGE.

2.1.8 Three step purification of N and C domain of hHKII

The purification protocol for both domains of hHKII is exactly the same as described at section 2.1.5 for Thr hHKII. All columns and buffers used were the same as previously mentioned apart from Gel Filtration Buffer: 10 mM Tris, pH 8.0, 0.5 M NaCl, 5 mM $MgCl_2$, 2 mM DTT, 2% glycerol. All fractions after each step were analysed with SDS-PAGE gel to monitor the presence and the purity of the target protein. Both proteins were concentrated to 1 mg/ml using a Vivaspin concentrator with molecular weight-cut off= 30 kDa.

2.1.9 Three step purification of TEV hHKII

The purification process for the 6His tag TEV cleavable protein was the same as mentioned at section 2.1.5 for the hHKII.

2.1.10 Cleavage of 6His tag using TEV protease

TEV cleavage screening conditions were set up as follows:

Protein was desalted in TEV Buffer (50 mM Tris, pH 8.0 adjusted at RT, 1 mM DTT, 150 mM NaCl) using a HiTrap Desalting 5ml column.

a) In 50 μg of desalted hHKII add 2.5 U of TEV and leave for overnight incubation at 4^{o}C and RT (21^{o}C) 6His tag Cyclophilin A used as a positive control).
b) In 50 μg of desalted hHKII add 5 U of TEV and leave for overnight incubation at 4^{o}C and RT (21^{o}C). (Cyclophilin A as a positive control). SDS-PAGE was used to analyse results.

2.1.11 Three step purification of Untagged hHKII

The AEX was chosen as the first step for the purification of the hHKII lacking any affinity tag. Resource 30Q (2 ml/ min) was used to perform the pH screening from 7.0-8.0 (0.5 intervals). The Akta automated buffer preparation system was used to make the buffers that are differing in their pH value and ionic strength during the elution steps. Buffer A: 0.1 M Hepes and Buffer B: 0.1 M Hepes. 2 M NaCl is made up by the user and the Akta system mixes them along with 2 M NaOH to create the desired pH and anionic strength. Lysis Buffer: 10 mM Tris, 50 mM NaCl, 5% glycerol, pH 7.5 (+ protease inhibitor tablets) was used to resuspend the cell pellets. After cell disruption and centrifugation the soluble supernatant was loaded to the Resource 30Q (2 ml/min) pre-equilibrated with Buffer A. The elution protocol was the same for all runs:

Elution protocol

(20 ml of lysate loaded to column)

1) Linear gradient: 0-60% Buffer B in 60 CV
2) Linear gradient: 60-100% Buffer B in 5 CV
3) Step: 100% Buffer B for 5 CV

All fractions were analysed with SDS-PAGE to monitor the purity of the target protein under different pH conditions.

Resource 6Q (20 ml/min) was used when purification performed in bigger scale (~100 ml of lysate loaded into the column). The elution profile was the same as previously mentioned.

The pooled fractions from this step were loaded into the affinity column HiTrap blue HP (1 ml/ min). The unbound material was washed with 15 CV of Buffer A and target protein eluted with a gradient of 0-100% B in 10 CV. SDS-PAGE were run to determine the presence of the target protein. The fractions containing a band at the expected size were pooled together and finally loaded to the SEC Superdex 26/600 200pg pre-equilibrated with the Gel Filtration Buffer: 10 mM Tris, 0.5 M NaCl, 5 mM $MgCl_2$, pH 8.0. Fractions were analysed with SDS-PAGE gel and protein was stored for further analysis.

2.1.12 TbHKI purification from inclusion bodies

The protein was captured and eluted as unfolded in an affinity column and the refolding process took place during the SEC step. The Buffers used for this protocol are as follows:

Buffer A: 50 mM HEPES, pH 7.8, 6 M GuHCl, 1 mM DTT, 20 mM Imidazole
Buffer B: 50 mM HEPES, pH 7.8, 6 M GuHCl, 1 mM DTT, 500 mM Imidazole
Buffer C: 50 mM HEPES, pH 7.8, 500 mM NaCl, 2 mM DTT, 250 mM L-Arginine, 45 mM NDSB-211, 2% glycerol, 0.005% Tween 20.

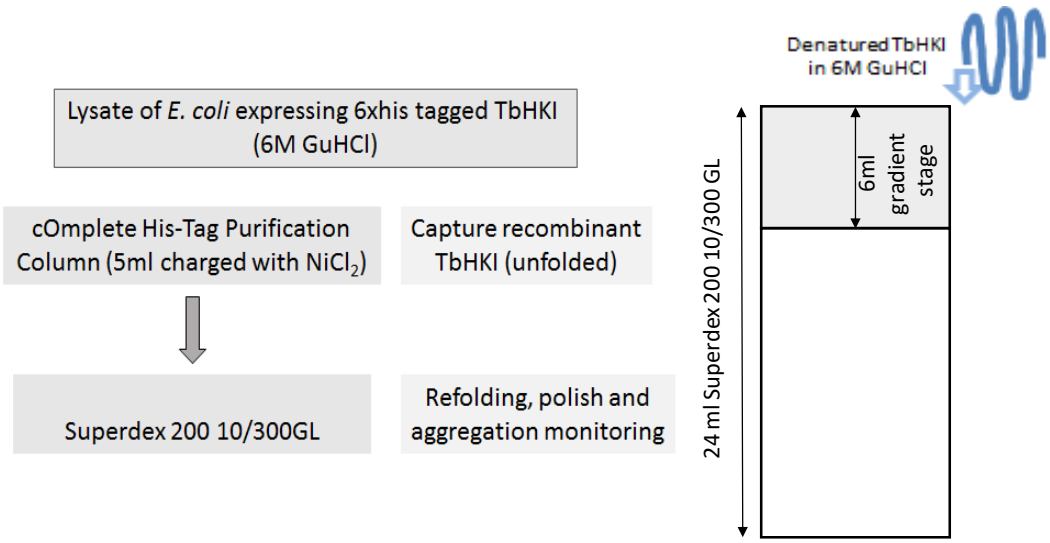


Figure 2.3: Schematic representation of refolding protocol of TbHKI in SEC column
The diagram is a representation of the steps performed for the isolation of inclusion bodies and refolding of the denaturing protein. A two-step purification protocol was performed with an affinity step to capture the unfolded TbHKI followed by in SEC column refolding. The unfolded TbHKI enters the Superdex 200 10/300GL which has a 6ml inverse gradient from 6 M Guanidine Hydrochloride (GuHCl) to 0 M GuHCl.

Cell pellet from 0.5 L cell culture was resuspended in Buffer A (+ protease inhibitor cocktail tablet), containing 6 M of the denaturant GuHCl. The cells were disrupted by sonication at 25 kPsi and the lysate was left stirring at RT for 1 hour. The lysate was then centrifuged at 15,000g for 30 min and the supernatant was filtered using a 0.22 μm filter. The supernatant was loaded into a 5ml cOmplete His-Tag purification column (Roche, 5 ml/min), pre-equilibrated in Buffer A. The unbound proteins were eluted with 25 CV of Buffer A and the target protein eluted with 10 CV of Buffer B (2 ml fractions collected). The fractions were not analysed with SDS-PAGE as the protein and the SDS were precipitating even after 1:1 dilution in dH_2O. The fractions from the elution step that had a high absorbance at Abs280 were pooled together for the next stage.

Superdex 200 10/300GL was used for the in-column refolding of TbHKI. The column was equilibrated with 2 CV of Buffer C and then an inverse linear gradient of 6 M to 0 M of GuHCl was implemented for 6 ml (1/4 of the total column volume) so

that protein gradually inserts into the refolding Buffer C. One ml concentrated, pooled fractions from previous step (Vivaspin column molecular weight cut-off= 30 kDa) was loaded manually from a 1 ml loop to the system. The flow rate was 0.5 ml/min so the passage of the protein from 6 M to 0 M of guanidine (6 ml inverse gradient) lasted for 12 min. The fractions were analysed with SDS-PAGE.

2.1.13 TbHKI purification from soluble fraction

Buffer A: 20 mM $NaHPO_4$, pH 7.5, 250 mM NaCl, 5 mM imidazole, 2% glycerol, 150 mM NDSB 201, 0.1% Tween 20 (+ protease inhibitor tablet)
Buffer B: 20 mM $NaHPO_4$, pH 7.5, 250 mM NaCl, 5 mM imidazole, 2% glycerol, 150 mM NDSB 201, 0.1% Tween 20, 0.1% Reduced Triton
Buffer C: 20 mM $NaHPO_4$, pH 7.5, 250 mM NaCl, 20 mM imidazole, 2% glycerol, 150 mM NDSB 201, 0.1% Tween 20
Buffer D: 20 mM $NaHPO_4$, pH 7.5, 250 mM NaCl, 250 mM imidazole, 2% glycerol, 150 mM NDSB 201, 0.1% Tween 20

Cell pellets from 4 L of cell culture in BL21 star (DE3) were resuspended in Buffer A, and centrifuged at 30,000xg for 45 min. The supernatant was sonicated and filtered before loading into 1 ml cOmplete His-Tag resin (1 ml/min). The unbound material was washed with 10 CV of Buffer B, the lightly bound protein on the column was removed by passing 20 CV of Buffer C and finally the target protein was eluted with 10 CV of Buffer D (2 ml fractions).

The fractions containing the target protein, as seen from SDS-PAGE, were concentrated and 0.5 ml was loaded into the SEC column, Superdex 200 10/300GL (pre-equilibrated with Gel Filtration Buffer: 20 mM $NaHPO_4$, 250 mM NaCl, 2 mM DTT, 2% glycerol). 0.5 ml fractions were collected and analysed.

2.2 Results and Discussion

2.2.1 Expression and purification of Thr hHKII

With the vector pET28a-LIC the biosynthesis of proteins is driven by the T7 lac promoter. The T7 expression system is also supported from the BL21-CodonPlus (DE3) RIL *E. coli*. Overexpression of Thr hHKII protein induced by IPTG (final concentration 1 mM) was proven by the detection of a thick band at the expected mass (~100 kDa) in SDS-PAGE gel. It was concluded that the protein is expressed in a better yield when grown in TB medium compared to 2xTY and SOC medium, whereas cultures grown in LB medium showed no expression at all.

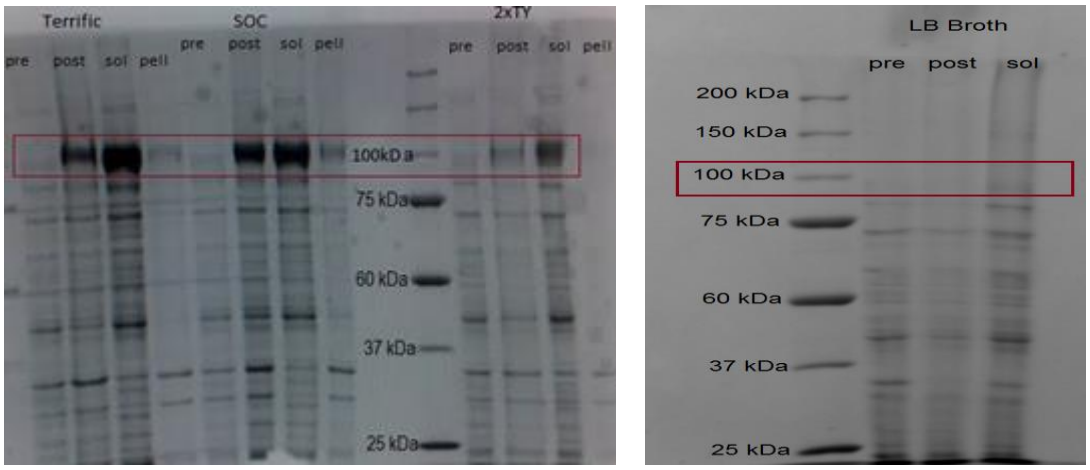


Figure 2.4: SDS-PAGE after expression of Thr hHKII

SDS-PAGE gel after expression trials of Thr hHKII shows that the over-expression of soluble protein was successful in TB, SOC and 2xTY medium. The red box highlights the thick bands that correspond to the ~100 kDa Thr hHKII. The fractions run in the gel are separated as pre-induced sample (pre), after induction sample (post), soluble fraction (sol) and insoluble fraction (pell). The band in the soluble fraction of TB medium is more intense than for SOC and 2xTY. The LB broth did not show any overexpression.

1 L of pellets were purified and passed through the cell disruptor at 25 kPSI. The filtered supernatant was loaded into the AEX HiPrep 16/10 DEAE FF at 0.5 ml/min. Ion exchange chromatography (IEX) separates proteins that differ in their net surface charge. Each protein has its own unique net charge change when pH changes but the general rule that applies to the IEX is that a protein at a pH above its pI will bind to a

positively charged medium (anion exchanger) and at a pH below its pI a protein will interact with a negatively charged medium (cation exchanger). hHKII has a theoretical pI of 5.85 (estimated by Expasy Protparam tool, http://web.expasy.org/protparam) so it is negatively charged in the Buffer A *(pH 7.5)*. It ies expected therefore to interact with the positively charged resin HiPrep 16/10 DEAE FF. SDS-PAGE gel comparing the cell lysate and the flow through after the DEAE shows little difference so we could speculate that all proteins elute from resin at the same time apart from other negatively charged molecules such as DNA and RNA nucleotides. All proteins elute in one broad peak (Fig. 2.5).

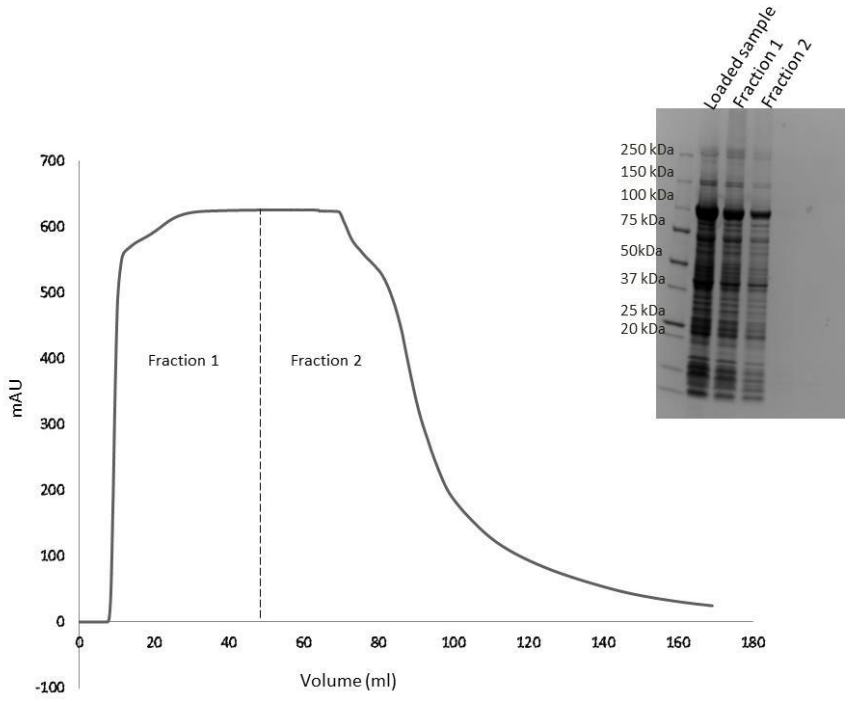


Figure 2.5: First step of purification of Thr hHKII (AEX)

The soluble fraction of the cells was loaded to the HiPrep 16/10 DEAE FF. The elution profile of AEX step contains a broad peak. All proteins elute in this peak according to SDS-PAGE. Two fractions (50 ml each) were pooled together for the next step.

The fractions collected from the AEX are loaded onto a 5 ml affinity column HiTrap FF (5 ml/min). Histidine tagged proteins have a high selective affinity for Ni^{2+} and other metals (i.e Co, Zn, Fe and Cd) that can be immobilised on chromatographic media using chelating ligands and for this reason a protein containing a histidine tag

will be the strongest binder among all the rest proteins in a crude sample extract. The target protein was eluted with a step of 100% Buffer C (250 mM imidazole, 10 CV) resulting in a sharp peak (Fig. 2.6). The 2 ml fractions of sharp peak were analysed in an SDS-PAGE gel. The fractions containing the desired protein were pooled together and concentrated for the final step.

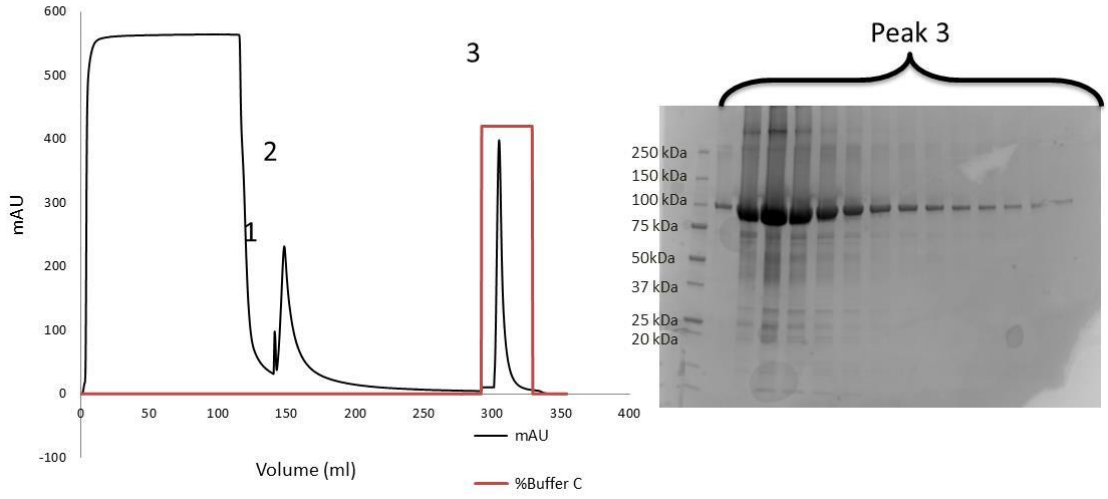


Figure 2.6: Second step of purification of Thr hHKII (IMAC)

14 fractions (28 mls) were analysed by SDS-PAGE. As seen from SDS-PAGE gel all fractions contain a band at ~100 kDa so Thr hHKII elutes as expected when high imidazole (step of 100% of Buffer C) is used. Peaks 1 and 2 contain also Thr hHKII (data not shown) but in very low concentration and fractions were discarded.

The final polishing step was performed using the HiLoad 26/60 Superdex 200 pg. Size exclusion chromatography can separate molecules based on their molecular weight in solution. The concentrated pooled sample from IMAC was injected using a 10 ml loop. Thr hHKII elutes in a big symmetrical peak (peak 3) with elution volume ~180 ml (Fig. 2.7). SDS-PAGE also shows the presence of Thr hHKII in peak 2. Peaks 1 and 4 when resolved in an SDS-PAGE did not show anything, possibly because the concentration is very low (SDS-PAGE data not shown here). Peak 2 probably corresponds to aggregated Thr hHKII.

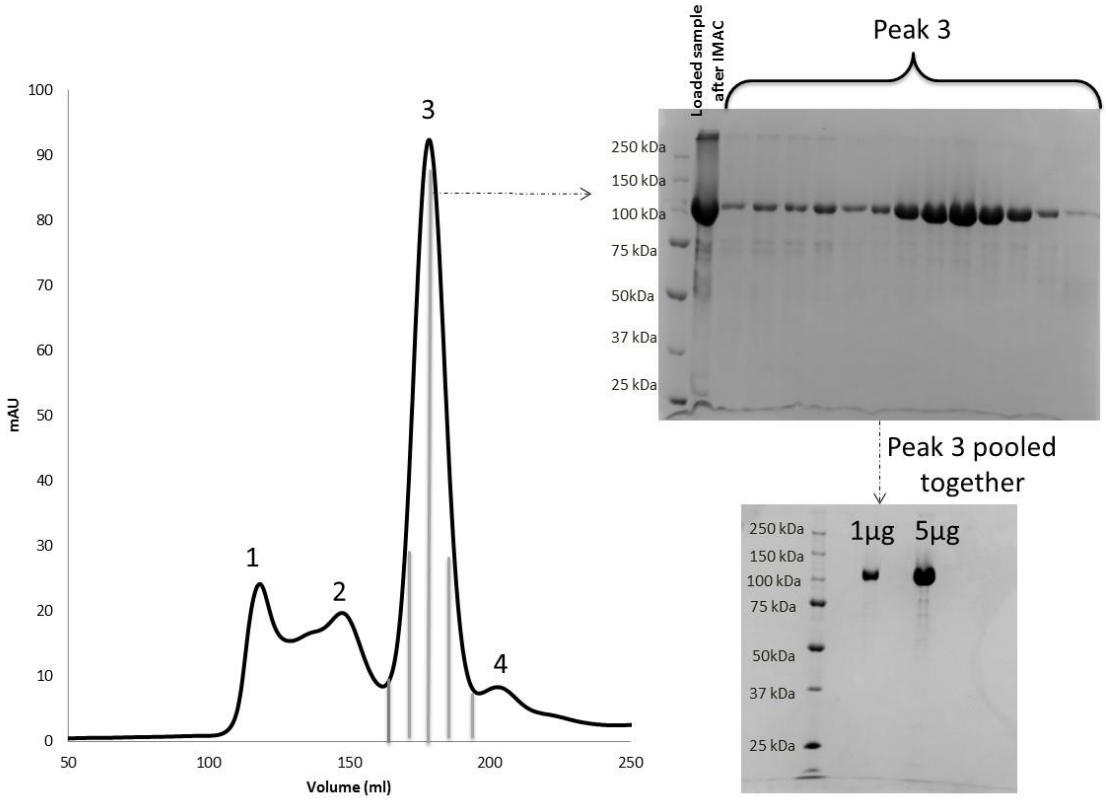


Figure 2.7: Third step of purification of Thr hHKII

Protein after IMAC was concentrated to ~10ml and loaded to SEC column. Protein elution volume is ~180 ml. All fractions of the symmetrical peak 3 were analysed and Thr hHKII was present in all with a small number of impurities at very low concentration. Thr hHKII at 1 mg/ml was found to be >95% pure based on SDS-PAGE where 1 µg and 5 µg of protein were analysed.

All fractions from peak 3 were pooled together and the concentration measured using a NanoVue spectrophotometer reading the absorbance at 280 nm using the molar extinction coefficient of the protein, 55360 M^{-1} cm^{-1}. Protein was concentrated to 1 mg/ml. The final yield is estimated to be 34 mg from 1 L of cell culture. Aliquots of the protein were stored at 4^{o}C, -20^{o}C and -80^{o}C for further analysis.

Thrombin cleavage trials were set up:

1) hHKII in Gel Filtration Buffer was incubated with Thrombin at 4^{o}C (24 hours) (10 U/mg of protein).

2) Purified protein was diluted 1:10 in Thr Buffer: 50 mM Tris, pH 8.0 (adjusted at RT), 150 mM NaCl, 2.5 mM $CaCl_2$. 50 µg of protein was incubated overnight at RT with 2 U and 4 U of Thrombin. Samples were analysed by SDS-PAGE (200 V, 55min).

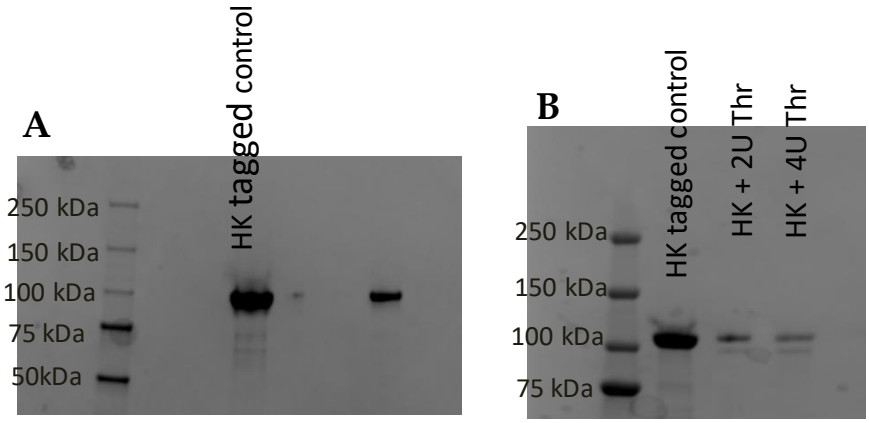


Figure 2.8: Tag removal using Thrombin protease

A. The first cleavage trial did not succeed.

B. The optimised Buffer conditions showed a sign of cleavage. HK with the addition of 2 and 4 U of Thrombin seem to have a lower faint band that could correspond to the cleaved protein. The lower band is more obvious when 4 U of Thrombin are added. However the cleaved: uncleaved ratio is very small.

(Bondos and Bicknell, 2003) , developed a solubility assay which was performed to help determine in which buffers the protein aggregates less. Six different buffers were tested which had different NaCl concentrations supplemented with or without DTT.

- Buffer 1: 10 mM Tris, pH 8.0, 5 mM $MgCl_2$, 0.15 M NaCl
- Buffer 2: 10 mM Tris, pH 8.0, 5 mM $MgCl_2$, 0.25 M NaCl
- Buffer 3: 10 mM Tris, pH 8.0, 5 mM $MgCl_2$, 0.5 M NaCl
- Buffer 1 + 5 mM DTT
- Buffer 2 + 5 mM DTT
- Buffer 3 + 5 mM DTT

As explained in 2.1.7 the aggregated protein (agg) was retained in the membrane and to retrieve this 30 μl of dH_2O was pipetted across the membrane. 30 μl of soluble (sol) and 30 μl of aggregated protein were mixed separately with *4x Sample Buffer* and heated to 90^{o}C for 10 minutes prior to SDS-PAGE. The gel results (Fig. 2.9) show that protein in high concentration of NaCl (0.5 M) is more stable since agg band is less intense than band for Buffers 1 and 2 (0.15 M and 0.25 M NaCl respectively). On the other hand the addition of 5 mM DTT did not seem to reduce the aggregation rate. For this reason DTT was omitted from the future purification.

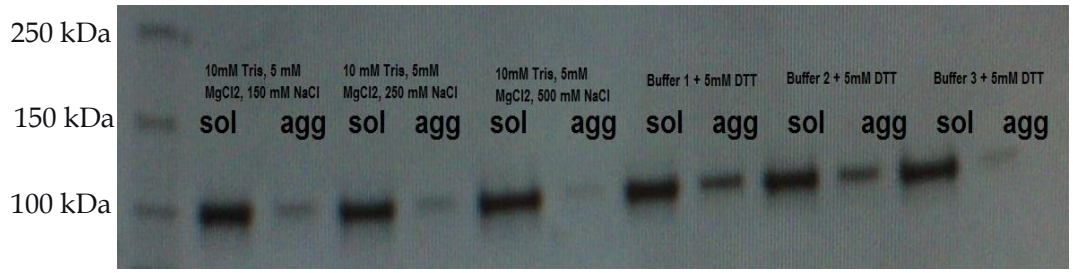


Figure 2.9: Solubility assay for Thr hHKII

2.2.2 Expression and purification of TEV hHKII

The percentage of cleavage of the 6His tag from Thr hHKII was not good enough and for this reason we decided to create a new construct with a different cleavable site. The new construct is identical to the Thr hHKII with the only difference in the N-terminal end of the protein. The Thrombin cleavage site is replaced by a TEV cleavable site (Fig. 2.1). TEV protease is considered to be more specific than Thrombin. The cloning of TEV hHKII was successful and plasmid DNA was sent for sequencing to Genepool and the correct sequence verified. The same cell line and medium were screened for the expression of the TEV hHKII construct (2.1.3.1). As seen from SDS PAGE (Fig. 2.10) the protein is overexpressed in all medium apart from LB. The same results were seen previously for Thr hHKII. However, the overall yield for all conditions is less than before.

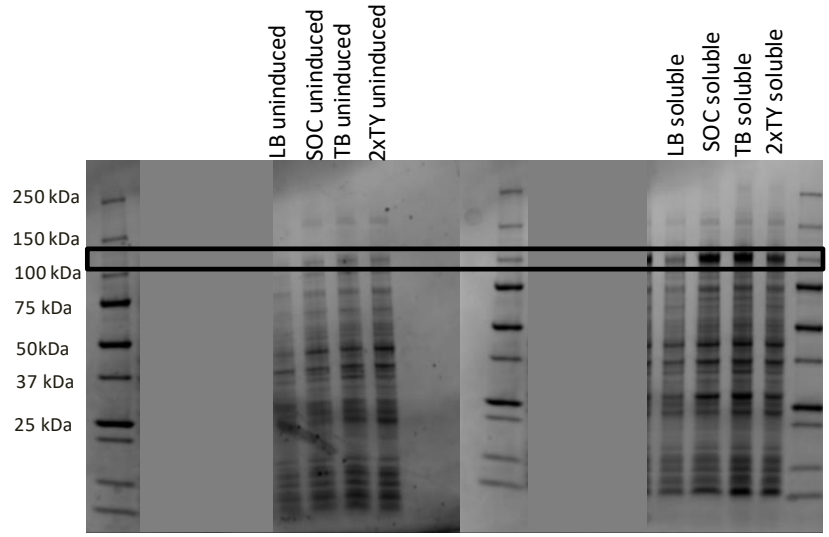


Figure 2.10: SDS-PAGE after expression of TEV hHKII

TEV hHKII was overexpressed in SOC, TB and 2xTY medium. LB medium did not lead to a significant soluble expression of the protein.

The three step purification of TEV hHKII gave similar peaks as previously for Thr hHKII. The protein comes off the Superdex 26/60 200pg at a consistent elution volume and protein is expected to be a monomer. The protein was concentrated to 1 mg/ml and kept in -80^{o}C freezer, with addition of 10% of glycerol, for future studies.

TEV cleavage trials were again unsuccessful (Fig. 2.11). TEV protease did not manage to cleave the tag so previous cleavage failures cannot be attributed to lack of Thrombin specificity. All of the studies on hHKII were performed using the 6his tagged constructs.

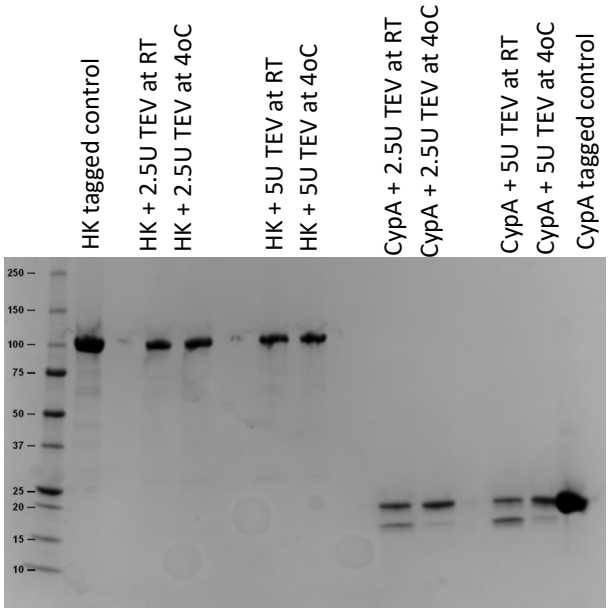


Figure 2.11: Tag cleavage trials with TEV protease

SDS-PAGE gel shows that TEV protease is active since control Cyclophilin A (CypA) is cleaved when 2.5 and 5 U of TEV are incubated with the protein at RT. There is not a lower band for hHKII even with 5 U of TEV suggesting that the cleavage was not successful.

2.2.3 Expression and purification of N and C domains of hHKII

The cloning of N and C domains was successful and both DNA sequences were verified. Successful overexpression was observed for C domain when plasmid DNA was transformed to BL21 plus (RIL) competent cells and single colonies were left shaking in TB medium until OD_{600}=0.8 before cold shock for 1 h at 4^{o}C. The cell expression was induced with 1 mM IPTG and cells were transferred at 20^{o}C and left shaking overnight. N domain of hHKII was overexpressed following the same protocol as for C domain but LB medium was also screened. SDS-PAGE gel shows that both TB and LB medium produced similar yields of soluble protein.

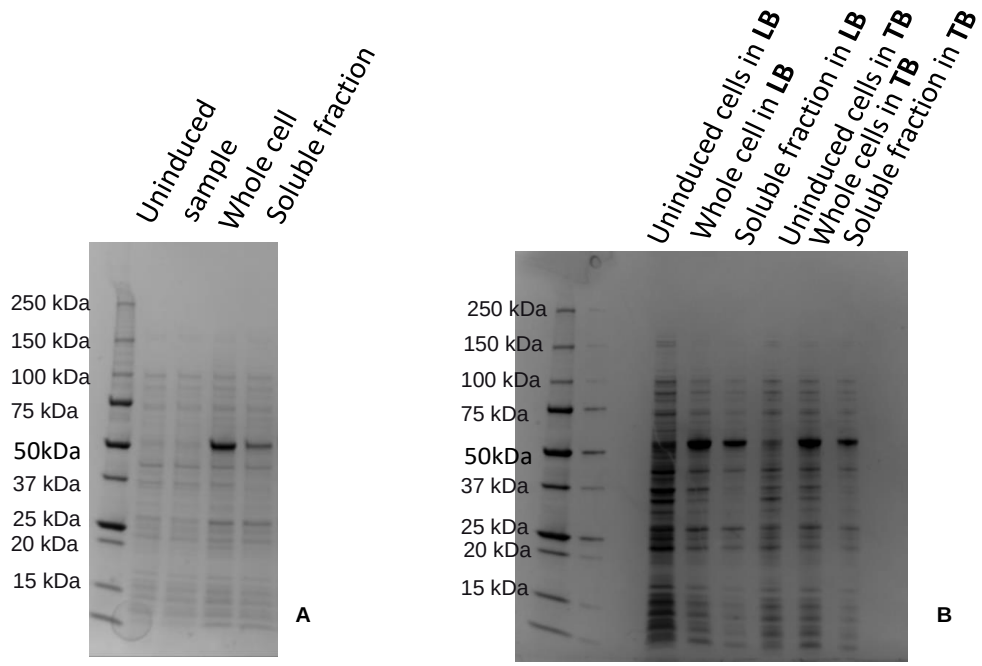


Figure 2.12: Overexpression of N and C domain hHKII

The yield of overexpression from both domains was analysed in an SDS-PAGE where the uninduced, and induced whole and soluble cell fractions were compared (Panel A: C domain samples, Panel B: N domain fractions). For both constructs we could observe that the soluble amount is almost 50% less than the amount of protein in the whole cell fraction. The other 50% of the protein is probably expressed in inclusion bodies.

The purification protocol for both domains was kept the same as for Thr hHKII (2.1.5). For both constructs the fractions after IMAC that contained the target protein were loaded into a Superdex 26/600 200 pg pre equilibrated with Gel Filtration Buffer (2.1.8). The three-step purification protocol gave almost identical peaks during the AEX and IMAC steps but the SEC step resulted in two different elution profiles as seen at Figure 2.13. The N domain elutes in a main symmetrical peak that corresponds to a monomer of ~50 kDa. SDS-PAGE analysis proves the presence of a protein close to 50 kDa. The C domain elutes in a double peak with low resolution between the two species that could be a ratio of monomer: dimer.

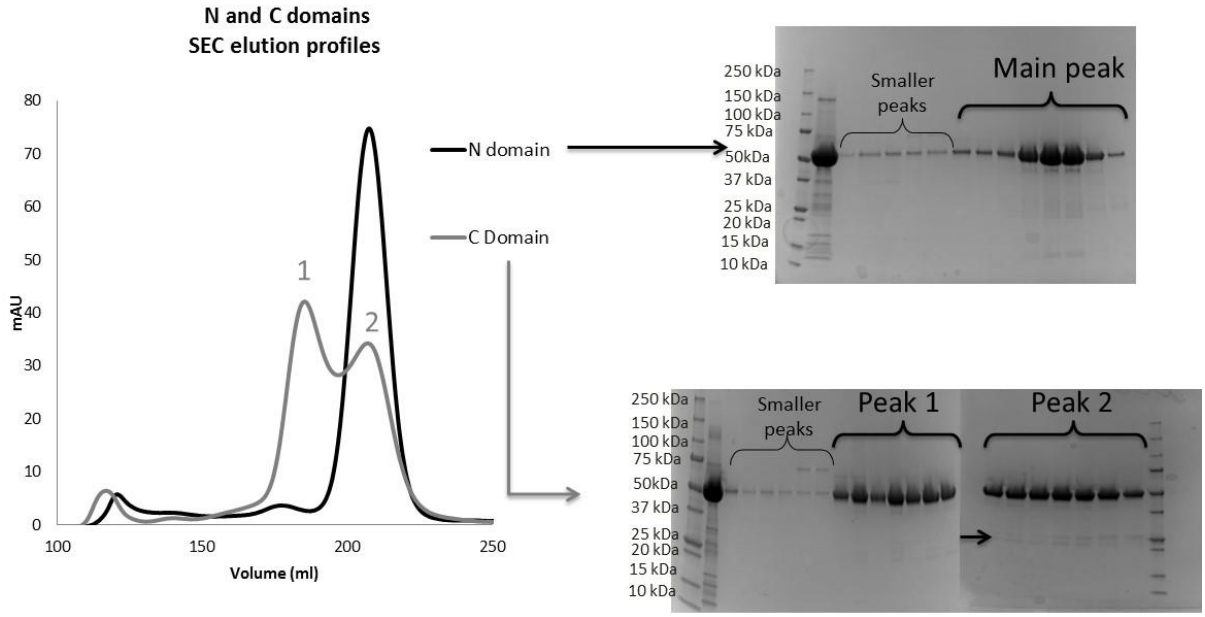


Figure 2.13: N and C domain elution profiles after SEC step

The elution profile after the SEC step differs between the N and C domain. SDS-PAGE across the two peaks showed the presence of a single band at 50 kDa for peak 1. Peak 2 also contains a thick band at the right size with a small number of impurities present at ~25 kDa (arrow). There are 2 more small peaks for both domains before the elution of the main big peaks which correspond to aggregates of the N and C domain respectively as seen from the SDS-PAGE.

The main peak of N domain was pooled together and concentrated to 1 mg/ml (Fig. 2.14).

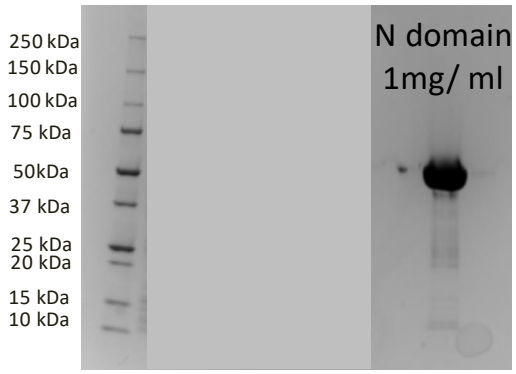


Figure 2.14: N domain at 1 mg/ml

The N domain was found to be >95% pure according to the SDS-PAGE.

The two peaks corresponding to the C domain were pooled separately (Refer to as pool 1 and pool 2 for the combined fractions of peak 1 and peak 2 respectively). All purified proteins, with the addition of 8% glycerol, were kept at -80^{o}C for future use.

2.2.4 Expression and purification of Untagged hHKII

The cloning process was successful and sequencing confirmed the correct sequence for the Untagged HKII. The plasmid DNA was transformed in BL21 plus RIL (DE3) cells and expression trials were set up as mentioned in section 2.1.3.2. The protein shows soluble expression under all conditions with significant overexpression in TB and 2xTY medium. LB medium resulted in a lower yield of soluble protein and even lower for SOC medium.

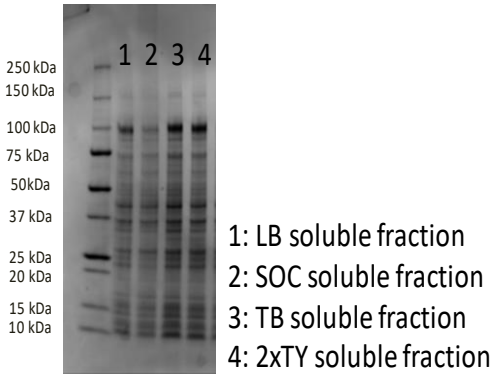


Figure 2.15: Overexpression of Untagged hHKII

The overexpression trials screening 4 different medium were successful for all, nevertheless TB and 2xTY medium were the optimal medium for this construct as the expression yield is significantly better on these two compared to LB and SOC.

Resource 30Q (2 ml/min) was used to perform the screening of the optimal pH for the purification of the untagged hHKII as the selectivity and capacity of a weak anion exchanger column changes depending on the pH during the elution of the protein. This could lead to a cleaner target protein, less contaminated with bulk impurities, when the right pH is chosen. Based on the elution chromatograms and SDS-PAGE gels it was shown that the best separation of target protein from impurities occurs when purification process was performed at pH 7.0.

The process was scaled up after optimal pH had been decided. Resource Q 6 ml was equilibrated with 10 CV of Buffer A, the flow rate was kept at 20 ml/min and 1.5 ml fractions were collected. Fractions were analysed with SDS-PAGE. Pooled fractions containing the target protein were then loaded to an affinity column, the Hi Trap Blue HP (1 ml/min). The column is prepacked with Blue Sepharose High Performance and via this matrix, the dye ligand, Cibacron Blue F3G-A is covalently attached via the triazine part. The dye has been exploited as a chromatographic medium to separate and purify a variety of proteins, such as dehydrogenases, kinases, serum albumins, interferons, several plasma proteins etc. There are studies regarding the chromophore itself and the immobilised ligand to fully understand the interactions that it makes with the proteins (Subramanian and Ross, 1984). Some proteins could interact with this dye because of the structural similarity with nucleotide cofactors so since hexokinase has a binding site for ATP it has a good potential to bind to the column.

The unbound material was washed with 15 CV of Buffer A and the target protein was eluted with a linear gradient to 100% in 10 CV. SDS-PAGE showed that protein eluted at 61.6% B and above. The fractions that contained the target protein were pooled together. Finally SEC was performed and the protein elution volume is consistent with Thr hHKII elution volume. This result was expected since both constructs have the same molecular weight. It is expected that Untagged hHKII is a monomer. Protein was concentrated to 1 mg/ml (using molar extinction coefficient: 55360 M^{-1} cm^{-1}); 1 μg and 5 μg were run in SDS-PAGE to check the purity of purified enzyme. Figure 2.16 shows all chromatograms and gels described in this section.

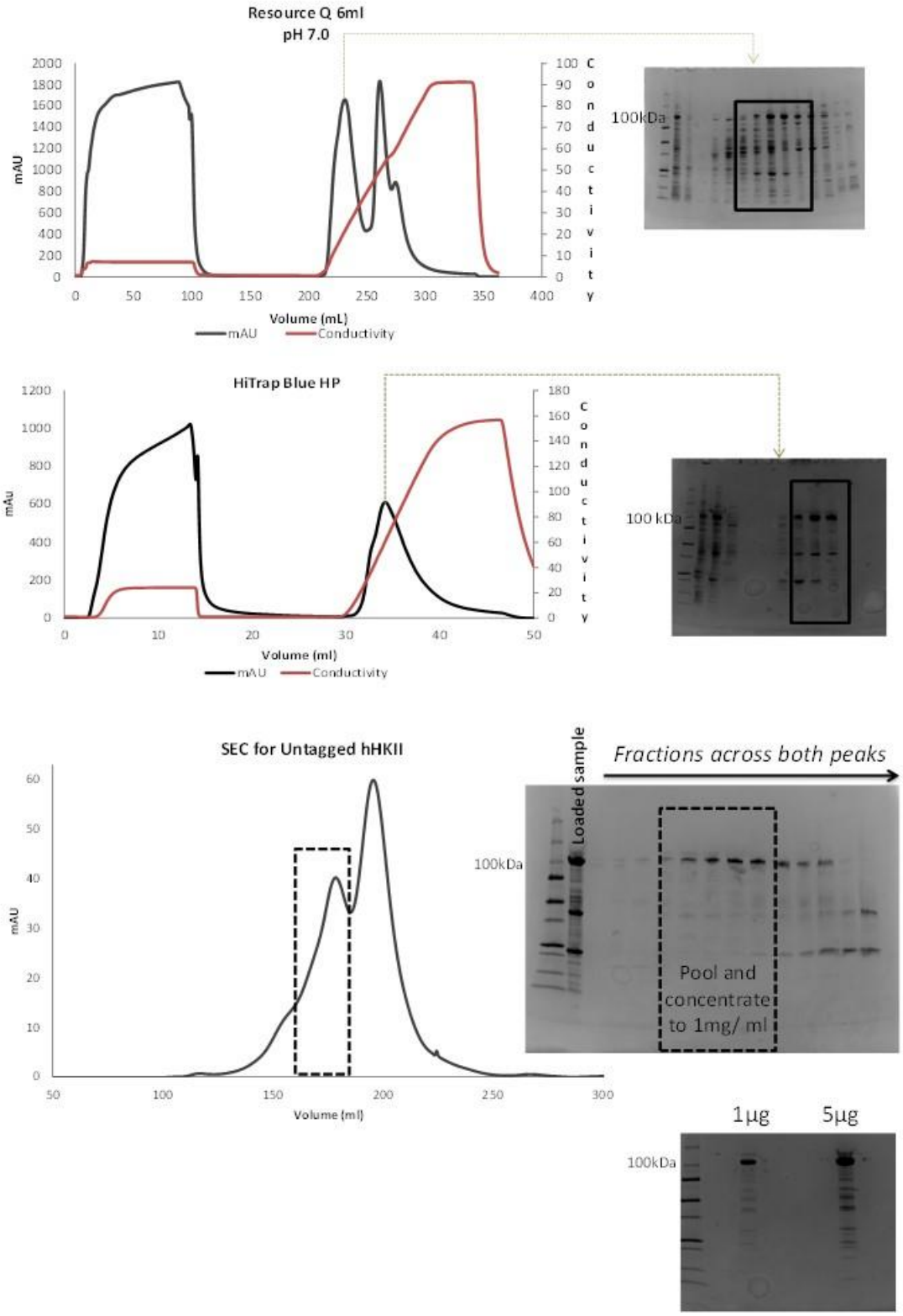


Figure 2.16: Three step purification of Untagged hHKII

The AEX step resulted in the elution of 3 peaks. All peaks were analysed with SDS-PAGE and a band around 100 kDa was present in all. The highest amount was found in peak 1 and fractions in the black box were pooled together for next step. The target protein was then

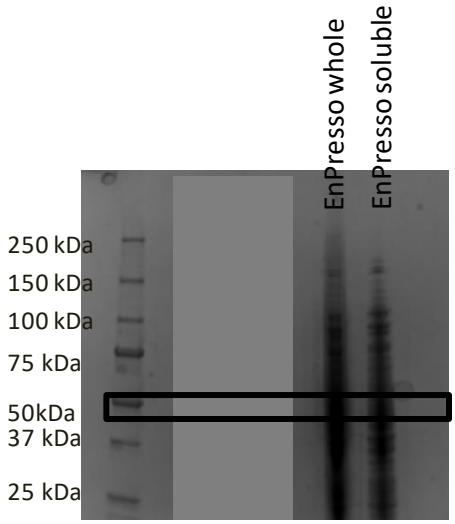
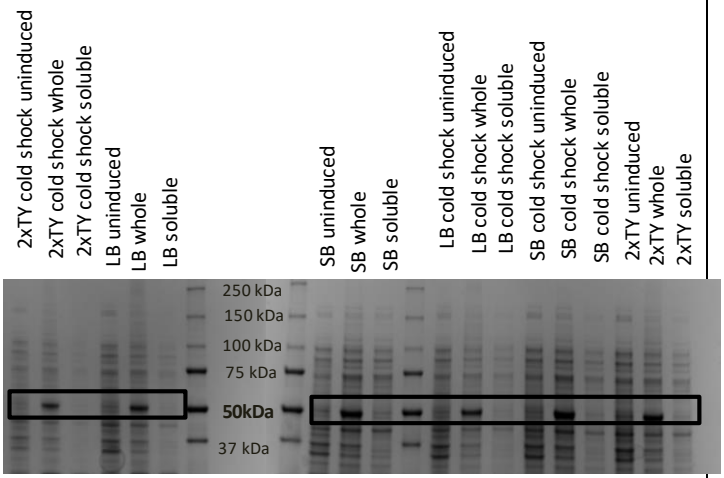
bound to HiTrap Blue HP column and eluted when linear gradient performed (elution started when %B was ~62%). Only the fractions in the black box proceeded to SEC (Superdex 26/60 200pg). When final protein was concentrated and run at 1μg and 5μg did not run as a pure band. Instead many bands appear within the range from 25 kDa to 75 kDa). The protein is estimated to be ~90% pure.

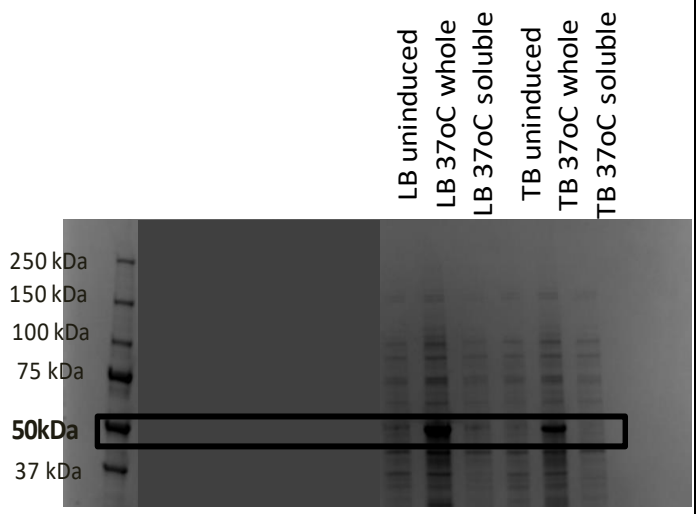
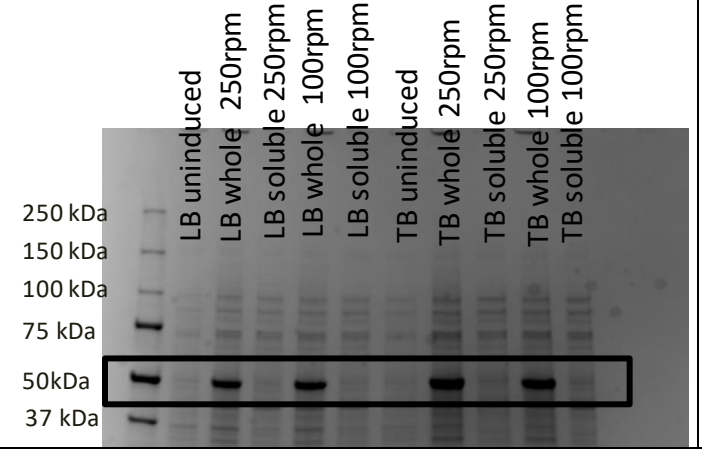
2.2.5 Expression and Purification of TbHKI

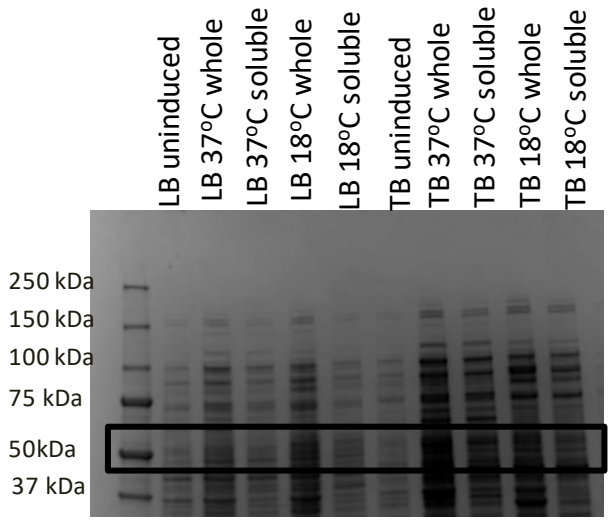
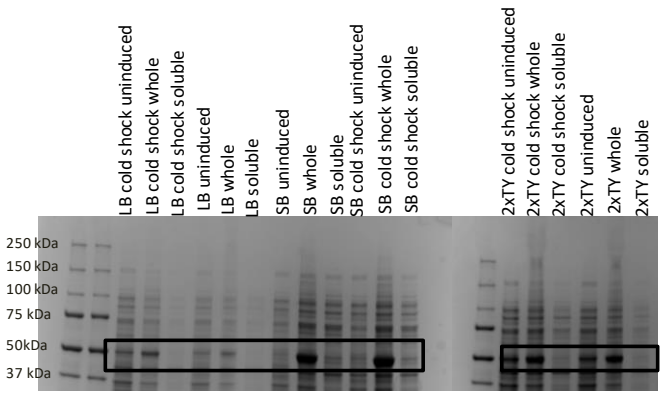
TbHKI has proven a difficult target since many different conditions were tested without any significant increase of the protein's soluble expression. The DNA sequence is codon optimised for *E.coli* system as the tRNA levels between species can be very different and potentially affect the expression levels. However, the codon optimisation performed (GeneOptimizer, GeneArt) did not seem to improve the expression process even with the use of special host strains such as BL21-CodonPlus (RIL) competent cells. These cells contain a ColE1-compatible, pASYC-based plasmid containing extra copies of the argU, ileY, and leuW tRNA genes. By comparison of the SDS-PAGE in Table 2.6 we could observe that the latter cell line gave the best expression yield of protein along with the host strain BL21 star (DE3). These are also designed for applications that require high-level expression of recombinant proteins from low copy number, T7 promoter-based plasmids. There was no overexpression (soluble or in inclusion bodies (IBs)) of TbHKI in C43 (DE3) and ArcticExpress (DE3) cell lines. The latter cells have been engineered for improved protein processing at low temperatures, since low temperature cultivation could lead to an increased recovery of soluble protein (Schein, 1989). Using this cell line, the induction with IPTG was performed at 13^{o}C for 24 hours. Under these conditions no protein expression was observed. However, the cold-shock expression at BL21 star (DE3) cells resulted in an increased expression of insoluble TbHKI. We could say that the cold-shock for a shorter time could result in a high-yield of protein expression while a cold-shock period of 24 hours could minimize it. On the other hand, protein is overexpressed at high temperature of 37^{o}C after 3 hours, although insoluble. The recombinant GroEl/ GroES co-expression chaperone system was used which is known to prevent protein aggregation and promote the right folding (Hartl et al., 2011). Unfortunately, it did not show an increase in the soluble protein

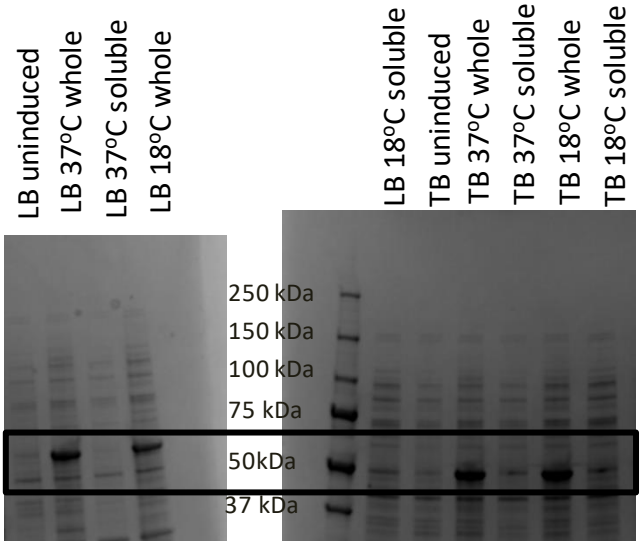
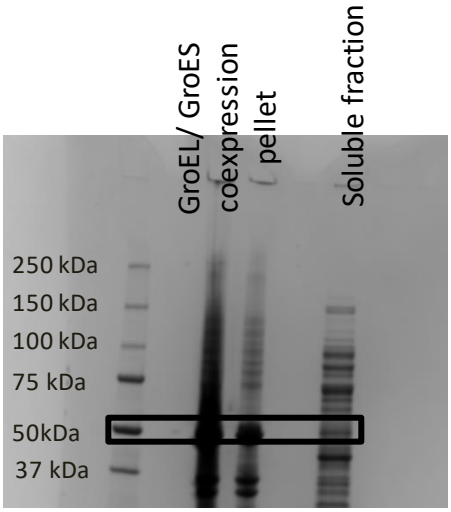
expression of the transformed cells. However, the analysis of the SDS-PAGE does not show any big band in the expected size of GroEL/ GroES (60/ 10 kDa) so the failure of the soluble expression of TbHKI could be a result of the low inadequate expression level of the chaperone system.

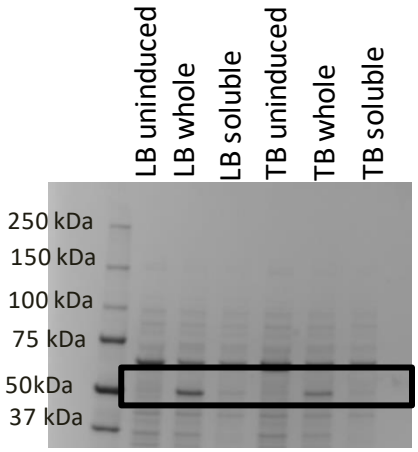
Table 2.6: TbHKI expression resulted in no or very small amount of soluble protein

Cell line	SDS-PAGE	Comments
BL21-CodonPlus (DE3) RIL strain 1 mM IPTG induction, Overnight at 30°C, 250 rpm		Background expression is very high, but no significant overexpression can be seen at 50 kDa
BL21-CodonPlus (DE3) RIL strain 1 mM IPTG induction, Overnight at 18°C, with and without cold shock, 250 rpm		Very little/ No soluble expression

Cell line	SDS-PAGE	Comments
BL21-CodonPlus (DE3) RIL strain 1 mM IPTG induction, 3h at 37^{o}C, 250 rpm		Very little soluble expression
BL21-CodonPlus (DE3) RIL strain Cold shock, 1 mM IPTG induction, 20^{o}C overnight, 250 and 100 rpm		Very little soluble expression

Cell line	SDS-PAGE	Comments
C43 (DE3) 1 mM IPTG induction, Overnight at 18°C, 3h at 37°C, 250 rpm		No overexpression
BL21 star (DE3) 1 mM IPTG induction, Overnight at 18°C, 250 rpm, with and without cold shock		No or very little soluble expression

Cell line	SDS-PAGE	Comments
BL21 star (DE3) 1 mM IPTG induction, Overnight at 18°C, 3h at 37°C, 250 rpm		No or very little soluble expression
BL21 star (DE3) Coexpression with GroEL/ GroES, 1mM IPTG induction, Overnight at 18°C, 250 rpm		Very little soluble expression

Cell line	SDS-PAGE	Comments
ArcticExpress (DE3) 1 mM IPTG, 24h at 13^{o}C, 250 rpm		No soluble expression

Despite all efforts the target protein is still expressed in IBs so the final effort was to denature and refold the protein *in vitro*. Refolding of IBs is not a straightforward process and IBs need to be solubilised and then refolded into an active conformation. The choice of solubilising agents is crucial as well as the rate of refolding (Tsumoto et al., 2003). In this case guanidine hydrochloride (GuHCl) was chosen as the solubilising agent and The TbHKI purification from the IBs was performed in two steps where the affinity step (Fig. 2.17) was initially performed to separate the 6His tagged unfolded TbHKI from the other *E.coli* impurities. The elution profile when high imidazole buffer was used leads to a small peak that was pooled together and loaded to the SEC column.

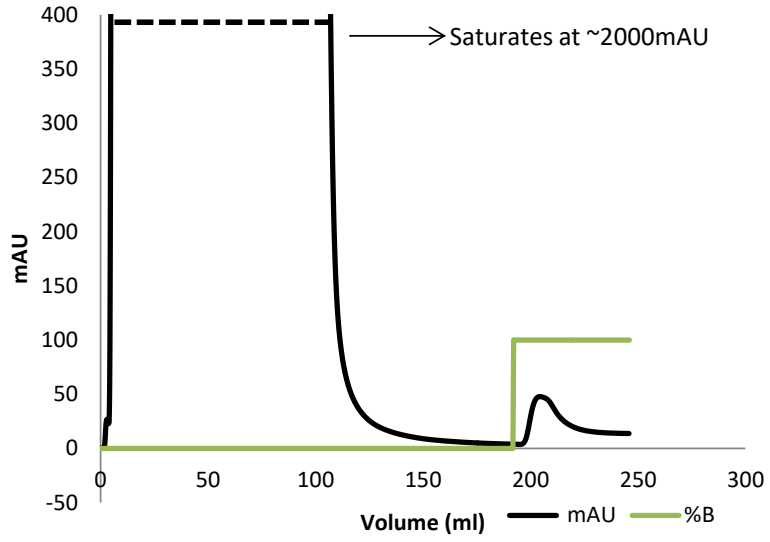


Figure 2.17: Elution profile of unfolded TbHKI with high imidazole

The unfolded TbHKI is captured in a 1 ml cOmplete His-Tag resin (1 ml/min). The small peak is not resolved in an SDS-PAGE gel as high concentration of GuHCl precipitated the SDS (added in the SDS-PAGE Running Buffer). The fractions under the small peak were pooled together and concentrated for the next step.

The denaturant gradient procedure was performed during the size exclusion step in Superdex 200 10/300GL. An inverse linear gradient of 6 M to 0 M of GuHCl was implemented for 6 ml (1/4 of the total column volume) so that protein gradually inserted into the refolding Buffer C in 12 min. Another key parameter in the refolding process is the effect of small additives. Two different Refolding Buffers were screened here:

Optimised Buffer: 50 mM HEPES, pH 7.8, 500 mM NaCl, 2 mM DTT, 250 mM L-Arginine, 45 mM NDSB-211, 2% glycerol, 0.005% Tween 20

Refolding Buffer 1: 50 mM HEPES, pH 7.8, 500 mM NaCl, 2 mM DTT; 250 mM L-Arginine, 150 mM NDSB-201, 2% glycerol, 0.01% Tween 20

The two buffers have a different type of non-detergent sulfobetaine (NDSB) and a different concentration of Tween 20. The elution profiles of the two different

refolding buffers are almost identical with same peaks (Fig. 2.18), with an earlier elution for Optimised buffer (~0.5 ml faster elution).

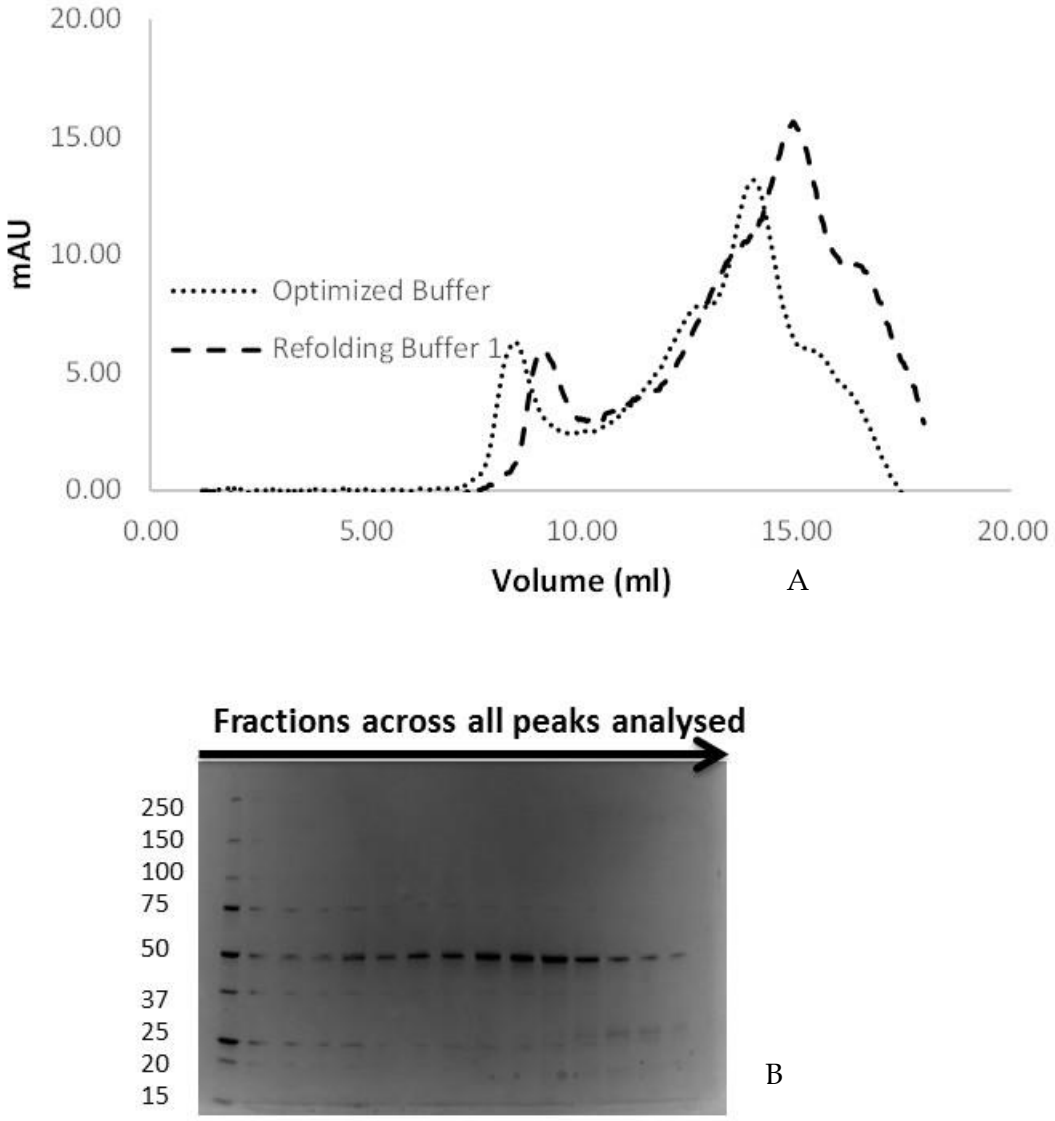


Figure 2.18: Overlay of SEC profiles between two different Refolding Buffers

An overlay of the two chromatograms shows that the elution profile stays unaffected between the two runs. SDS-PAGE across all peaks shows the elution of a macromolecule at the expected size.

Activity assays across the peaks show that Optimised Buffer increased the yield of activity even up to 10 times for specific fractions. However the protein was unstable during concentration and/or buffer exchange to buffer without L-arginine and NDSB

so no further studies were performed. For this reason, since we could not verify if protein was properly folded after the denaturation, we did not proceed in screening the compounds against this purified protein.

Instead, we expressed the TbHKI in 4 L of cell culture and purified the soluble fraction in 2 step purification process. The protein after the affinity column precipitated during concentration so serial injections of 0.5 ml were performed with diluted protein. All chromatograms were very consistent to each other and looked like the chromatogram on Figure 2.19.

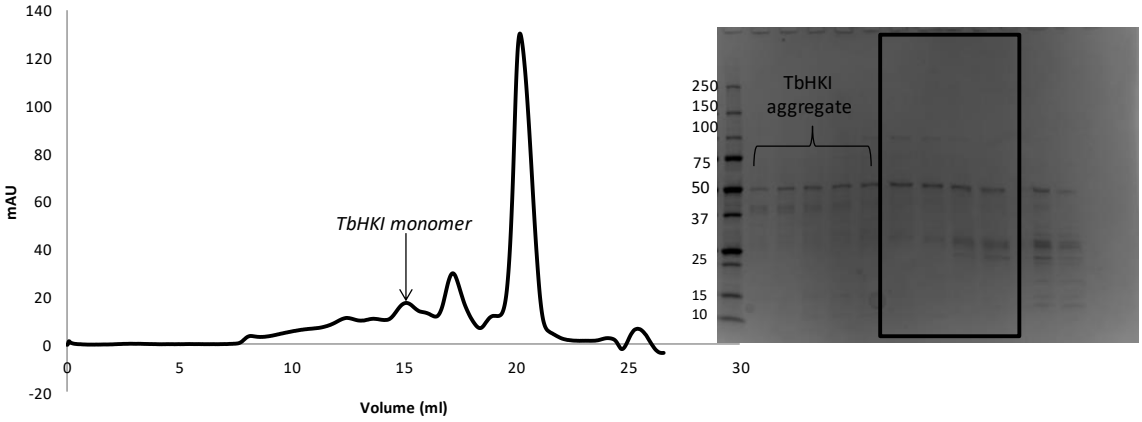


Figure 2.19: Chromatogram of TbHKI after SEC step

The peak with elution volume ~15 ml corresponds to the monomer TbHKI as seen from SDS-PAGE. Protein is also present in previous peaks that could represent TbHKI in dimer-tetramer formation. The fractions that pooled together are shown in the black box.

The final yield was ~2.5 mg from 4 L of culture and protein concentration was estimated at 0.3 mg/ml after SEC (using NanoVue spectrophotometer measuring the absorbance at 280 nm using the molar extinction coefficient of the protein, 22390 M^{-1} cm^{-1}). The purified protein was assayed and K_m values for the substrates (Glc and G6P) were determined. This protein was used for screening the compounds as will be discussed in Chapter 7.

2.3 Summary and Conclusions

The hHKII was successfully expressed and purified in 5 different constructs. Thr hHKII resulted in high yield of soluble protein. However the tag removal using the Thrombin protease was not successful and another construct was created with a difference in the N terminal site. This site had a TEV cleavable site but again, after successful purification, the tag was unable to be removed. As a last attempt the untagged hHKII was created and purified but the purity is estimated ~90% and the activity of the protein was dramatically reduced (Chapter 3, section 3.3.3). The biophysical and biochemical analysis performed on both tagged constructs did not show any signs of tag interference with the active site of the protein. However the crystallisation trials were not successful and for this reason N and C domains were also purified separately. The purification process resulted in interesting results as N domain elutes as a monomer but C domain elutes as a mixture of monomer and dimer. Further studies show that both domains are active and their functional parameters were determined (K_m for ATP and Glc). A low resolution structure was determined through SAXS (Chapter 4) for both domains.

TbHKI expression trials did not result in a high yield of soluble fraction and two different refolding protocols were tested. Both resulted in an active protein but the activity was improved with the Optimised Refolding Buffer. This purified protein was unstable in solution and it precipitated during concentration and/or buffer exchange so we could not continue the studies in order to investigate the oligomerisation state of the protein. As it was essential to produce an active and properly folded protein to screen against the hits from virtual screening we purified the soluble fraction of TbHKI from 4 L of culture. The final yield was 2.5 mg, the protein was active and it was used for the biochemical inhibition assays.

3. Chapter 3: Biophysical and Biochemical characterisation of hHKII and TbHKI

3.1 Introduction

The biophysical and biochemical analysis of proteins has the ultimate goal of obtaining meaningful parameters to enable a characterisation of the biochemical function of the molecule. An understanding of the oligomeric state, homogeneity and secondary structure of the proteins is fundamental in aspects of drug discovery and development. Structural elucidation should follow after proper characterisation of the macromolecule. A proper strategy for this should involve a combination of techniques that show reliable and consistent information about the protein under study (Malik and Shrivastava, 2013). To this end, I performed dynamic light scattering (DLS), circular dichroism (CD), SEC coupled to multi angle light scattering (SEC-MALS), surface plasmon resonance (SPR) and enzymatic assays to determine the homogeneity, oligomeric state, activity and functional parameters of the purified enzymes.

Dynamic Light Scattering (DLS) is a light scattering technique which can be used to determine any polydispersity and aggregation state/ propensity of the protein. The basic principle is simple: the sample in solution is illuminated by a laser beam and the fluctuations of the scattered light are detected by a fast photon detector at a known scattering angle θ. The Brownian motion of particles or molecules in suspension causes laser to be scattered at different intensities. Analysis of these intensities yields the diffusion coefficient of the particles which may be related to the hydrodynamic radius (Rh) using the Stokes-Einstein equation:

$$\mathrm{Rh} = \frac{k\,T}{6\,\pi\,\eta\,D} \quad \text{(Equation 3.1)}$$

Rh= hydrodynamic diameter,

k= Boltzmann's constant,

T= absolute temperature,

η= viscosity,

D= diffusion coefficient

Circular Dichroism (CD) is a biophysical method used to analyse the secondary structure component of a protein in solution. The CD effect relies on the differential absorption of linear polarised light, between the left handed (L) and right handed (R) components. A CD signal will be observed when chiral molecules are studied (i.e proteins) because L and R are absorbed to different extents (Kelly et al., 2005). CD signals from proteins arise from the peptide bond (absorption below 240nm), aromatic amino acid chains (absorption in the range 260-320nm) and disulphide bonds (weak absorption bands centered around 260nm). The different types of secondary structure (helix, sheet, turns) give rise to characteristic UV spectra in the far UV. These properties are used from various algorithms which analyse the data to provide an estimation of the secondary structure composition of the protein (Kelly et al., 2005).

SEC-MALS is a method combining Size Exclusion Chromatography to separate a protein based on its hydrodynamic size and shape coupled to Multi Angle Light Scattering for absolute molar mass determination. The Light Scattering (LS), Refractive Index (RI), and Ultraviolet Detection (UV) combined give important information on the molar mass and oligomeric state in solution (Nedelkov et al., 2006). The light-scattering detectors can provide absolute measurement of molecular weight and are very sensitive at detecting aggregates even at low concentrations. The relationship between light scattering and molecular weight is defined by the Rayleigh equation:

$$\frac{KC}{R_\theta} = \left(\frac{1}{Mw} + 2A_2\, C\right)\frac{1}{P_\theta}, \text{ (Equation 3.2)}$$

Where

C: sample concentration

θ: the measurement angle

Rθ: the Rayleigh ratio (the ratio of scattered light intensity to incident light intensity)

Mw: the molecular mass

$A_{2:}$ the second virial coefficient

Pθ: a term that defines angular dependence

K: a constant which is system, solvent and sample dependent

K is defined in the equation below:

$$K = \frac{4\pi^2}{\lambda_0^4 N_A} (n_0 \frac{dn}{dc})^2, \text{ (Equation 3.3)}$$

Where:

λ: laser wavelength in a vacuum

N_A: Avogadro's number

n_0: refractive index of the solvent

dn/dc: the change in refractive index of the solution with change in concentration

The angular dependence of the scattered light can give information regarding the molecular dimensions of the polymer, i.e the radius of gyration (Rg) (Tarazona and Saiz, 2003).

In SEC we are interested in the molecular mass of the sample. The distribution obtained from SEC is typically a molecular weight distribution describing how much material there is present of the various molecular weight "slices." The distribution is traditionally described by two numbers derived from it: M_w and M_n, where M_w is the weight-average molar mass and M_n is the number-average molar mass. The molecular weight is measured in each data slice from the SEC and the dn/dc and dA/dc and the MALS intensity.

In SEC-MALS an important derived parameter is the dispersity (M_w/M_n). The Mw/Mn is a statistical analysis that relates the distribution of mass weight and mass number. For a perfectly uniform sample consisting of only one molecular species the M_w and M_n are the same and M_w/M_n equals 1. However, this is not the case for real samples and this ratio is used to describe the oligomeric state of the molecule under study. Basically, it shows how far away the encountered distribution is from a perfectly monodisperse species.

Surface Plasmon Resonance (SPR) is a biophysical technique used to characterise macromolecular interactions in a label-free way. In general, the interaction is studied between a ligand in solution, called the analyte in SPR terminology, with an immobilised partner on an SPR active sensor surface, the ligand in SPR terminology (usually the protein). Basically, SPR is an optical method which measures the increase of the refractive index (RI), expressed as Response Units (RU), as the analyte binds to the ligand leading to an accumulation of protein on the sensor surface. One can monitor the interaction in real time and measure the rates of

association (k_{on}) and dissociation (k_{off}) precisely and use these values to calculate the corresponding affinity constants (McDonnell, 2001).

Enzymatic assays: Two different enzymatic assays were used for hexokinase biochemical characterisation; the commercial available from BioVision (Hexokinase Colorimetric Assay kit) and a coupled enzyme assay that I developed during the present studies.

In the BioVision assay kit, glucose is converted to G6P by HK, G6P is oxidised from glucose-6-phosphate dehydrogenase (G6PDH) to form NADH which reduces a colourless probe that will absorb at 450 nm. The kit is generally used with sample from serum, tissues or cell culture but it was also successful with purified protein (5 nM final concentration). However, the kit contains undisclosed reagents in the assay buffer. The concentration of the substrates (Glc, ATP) and cofactors (Mg^{2+}) is not known and this makes it impossible to determine the K_m values of HK substrates and to measure the IC_{50} of potential inhibitors. For an initial screen of inhibitors the substrate concentration is usually at or below the K_m of the protein. If the ATP concentration in the assay buffer is higher than the K_m, then ATP competitive inhibitors would not show any inhibition using these specific assay conditions.

To this end I developed a coupled enzymatic assay, the resazurin assay (also known as Alamar Blue® Assay) which is mostly used to assess the cell proliferation of various human and animal cell lines, bacteria and fungi and cell cytotoxity (Bonnier et al., 2014). The assay relies on the reduction of the non-fluorescent resazurin dye (blue colour) to the highly fluorescent molecule resorufin (pink-red colour), catalysed either by reductases or dehydrogenases (Guilbault, 1975) in the presence of the Nicotinamide Adenine Dinucleotide (NAD) and diaphorase. During this assay excitation at 530 nm and emission at 590 nm is monitored.

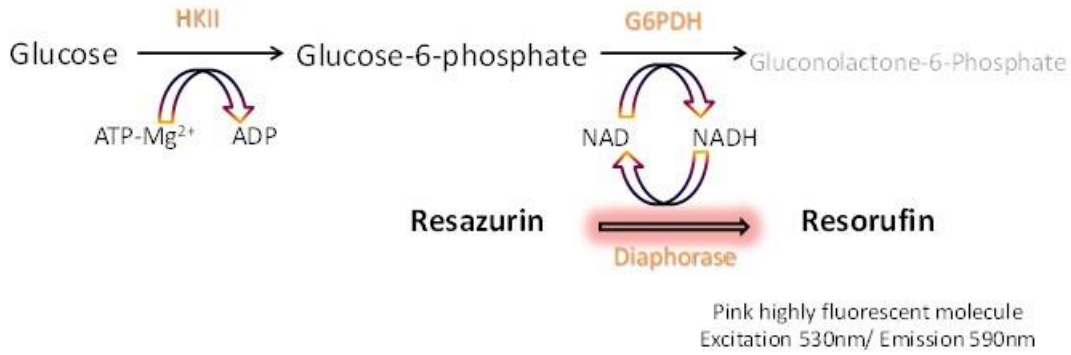


Figure 3.1: Resazurin assay, a coupled enzymatic assay for Hexokinase activity measurement

Glc is converted to G6P from hexokinase which is further metabolised from G6PDH in the presence of NAD. The reduction of NAD to NADH is brought about by Diaphorase which converts resazurin to the fluorescent molecule resorufin (excitation 530nm/ emission 590nm).

3.2 Materials and Methods

3.2.1 DLS

DLS was performed using a Zetasizer Auto Plate Sampler (Malvern) using 384 polypropylene plates (Corning). Protein concentration was typically 1 mg/ ml (unless otherwise stated) and 60 μl added to each well. Each sample was measured three times with an equilibration time at 120 sec (delay time between each measurement 10sec). The experimental temperature was set at 10^{o}C. All analysis was performed using Malvern Zetasizer APS software, version 7.11.

3.2.2 CD

The protein was buffer exchanged, prior to analysis with CD, to remove the high concentration of chloride ions (0.5 M NaCl) which show high absorbance in the high UV region. The HiTrap 5 ml desalt column, GE Healthcare, was used for the protein desalting. 500 μl of protein was added in the column at 10 ml/ min with CD buffer. Fractions of 250 μl were collected. The desalt process was successful and the UV peak was fully separated from the conductivity peak. The CD buffer contains: 10 mM Tris, pH 8.0 (sulphuric acid) and 250 mM NaF or 50mM NaF. The far UV

circular dichroism spectrum for 1 μM of hHKII was recorded at 20 nm/ min, data pitch 0.1 nm, response time 1 sec between 185 and 265 nm in a 0.1 cm path-length cuvette at 25^{o}C (JASCO-810 spectrometer). Spectra were corrected by subtracting a buffer baseline recorded at the same temperature. Spectra were recorded in triplicate. Secondary structure was estimated using the Dichroweb CD secondary structure analysis server using the methods CONTIN, SELCON3 and CDSSTR.

3.2.3 SEC-MALS

Size-exclusion chromatography (ÅKTA-Micro, GE Healthcare) coupled to UV, static light scattering and refractive index detection (Viscotec SEC-MALS 20 and Viscotek RI Detector: VE3580, Malvern Instruments) were used to determine the molecular mass of protein in solution. 100 μL of 1 mg/mL (unless otherwise stated) was run on a Superdex-200 10/300 GL (GE Healthcare) size exclusion column pre-equilibrated in Gel Filtration Buffer (for each protein construct, unless otherwise stated), at 22˚C with a flow rate of 0.5 mL/ min. 100 μl BSA at 2 mg/ml was injected and the monomer peak (66.8 kDa) was used as the detector off-set and calibration standard, run under identical conditions. Light scattering, RI and Abs_{280} were analysed by a homo-polymer model (OmniSEC software, v 5.1, Malvern Instruments).

3.2.4 SPR

SPR measurement was performed on a BIAcore T200. His-tag hHKII was immobilised and covalently stabilised on an NTA sensor chip essentially as described before (Wear et al., 2005). Briefly, the sensor surface was primed with a 60 sec injection of 500 μM $NiCl_2$ at 5 μl/ min. The surface was then minimally activated with a 240 sec injection (at 5 μl/ min) of a mixture of N-hydroxysuccinimide (NHS) (115 mg/ ml) and 1-ethyl-3-(-3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) (750 mg/ ml). hHKII in 10 mM Hepes (pH 7.5), 500 mM NaCl, 1% DMSO, and 5 mM $MgCl_2$, at concentration 500 nM, was passed over the sensor surface at a flow rate of 30 μl/ min. Following saturation of the response units (RU) signal, this was followed by a 240 sec injection (at 5 μl/ min) of

1 M ethanolamine (pH 8.5) to quench the remaining active succinamide esters. The final amount of hHKII covalently immobilised on the surface was typically around 8,000 RU. A two-fold dilution series of glucose ranging from 2 mM to 0.0625 mM was run in this experiment. The binding curves were analysed for a one-to-one binding model using the analysis software provided by the instrument (v2.02, GE Healthcare).

3.2.5 Enzymatic assay

For hexokinase colorimetric assay the steps performed are as follows:

For the NADH Standard Curve: Add 0, 2, 4, 6, 8 and 10 µl of 1.25 mM NADH Standard into a series of wells in duplicate in 96 well plate to generate 0, 2.5, 5.0, 7.5, 10 and 12.5 nmol/well of NADH Standard. Adjust volume to 50 µl/ well with HK Assay Buffer.

For Sample Preparation: Add 10 µl of 50 nM purified enzyme and adjust final volume to 50 µl with HK Assay Buffer. Prepare a parallel sample well as the background control to avoid interference from the NADH in the sample. For the reaction mix, for each well, prepare 50 µl containing:

HK Assay Buffer 34 µl
HK Enzyme Mix 2 µl
HK Developer 2 µl
HK Coenzyme 2 µl
HK Substrate 10 µl (For background mix no substrate- 44µl of Assay Buffer). Add 50 µl of the reaction mix to each well containing the Standard and test samples and 50 µl of background control mix to each well containing the background control sample (final assay volume 100 µl). Measure OD 450 nm using the multi-detection microplate reader system, SpectraMax M5 (Molecular Devices) every 5 min for 30 min. To calculate the activity, subtract the 0 standard reading from all standard readings. Plot the NADH standard curve. Correct sample background by subtracting the value derived from the background control from all sample readings. Calculate the hexokinase activity of the test sample: $\Delta OD = A2 - A1$. Apply the ΔOD to the NADH standard curve to get B nmol of NADH generated by hexokinase during the reaction time ($\Delta T = T2 - T1$).

Sample Hexokinase activity = B/ ($\Delta T \times V$) × Dilution Factor = nmol/min/ml/ = mU/ml

Where: B is the NADH amount from standard curve (nmol). ΔT is the reaction time (min). V is the sample volume added into the reaction well (ml).

For resazurin assay:

For the G6P standard curve add 20 µl of 250 µM G6P in a Corning 96-solid black plate and take 10 µl to perform twofold dilutions in 10 µl Assay Buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$) to generate 125, 62.5, 31.25, 15.6, 7.8, 0 µM of G6P. Adjust volume to 100 µl/ well with Reaction mix (17 mM NAD, 100 µM resazurin, 0.01 U G6PDH, 0.02 U diaphorase). Incubate at RT for 10 min and take endpoint measurement with excitation 530 nm / emission 590 nm. Subtract the 0 standard reading from all standard readings. Plot the G6P standard curve.

For K_m measurements 10 µl of substrate (Glc or ATP in serial dilutions) was added to each well and 90 µl of reaction mix was added to initiate the reaction (17 mM NAD, 100 µM resazurin, 0.01 U G6PDH, 0.02 U diaphorase, 10 nM HKII, 2 mM Glc or 7 mM ATP). The mixture was rapidly mixed and fluorescence at 530/590nm was monitored using a, SpectraMax M5 (Molecular Dimensions), multi-detection microplate reader at 21^{o}C with a reading taken every 20-25 sec. The initial linear portion of the slopes was converted to an enzymatic rate. K_m and k_{cat} values were determined using a range of substrate concentrations. The initial reaction rates (*Vo*) were plotted against the concentrations of substrate and the data least squares fit to Equation 3.4 using Kaleidagraph v4.0 software:

$$Vo = (k\text{cat} \times [\text{HK}]) \times [\text{Sub}]0 \,/([\text{Sub}]0/\text{Km}) + (\text{Constant}) \quad \text{(Equation 3.4)}$$

Where [HK] is the concentration of the protein added to the enzymatic assay, [Sub]0 is the initial concentration of the substrate and Constant is an off-set correlation function to account for non-origin starting values for the assay out-put signal. k_{cat} is the turnover number and K_m the Michaelis-Menten constant (Wear et al., 2007).

3.3 Results and Discussion

3.3.1 Biophysical and Biochemical Characterisation of hHKII

3.3.1.1 Glycerol prevents aggregation of hHKII during cryo-freezing

DLS was performed on newly purified enzyme and enzyme after being frozen with liquid N_2 to determine the effect of freezing process. The scattering intensity is

proportional to the square of molecular weight and it can be misleading in that a small amount of aggregation can dominate the distribution. For this reason, intensity distribution can be converted to volume distribution, describing the relative proportion of multiple components based on their mass or volume. It is a good practice to report the size of the peak based on the intensity analysis and report the relative percentages from a volume distribution analysis.

It is known that the formation of ice crystals can interfere with correct folding of the protein (Hamada et al., 2009). Fresh hHKII is homogenous with a symmetrical peak with a diameter of 9.9 nm (Fig. 3.2 A). Particles with a diameter of 113nm are also present which could represent aggregates but they are in a low concentration making it insignificant.

It was shown that the enzyme clearly aggregates after the freezing process (Fig. 3.2 B) since the particles are estimated to have a diameter of 800 nm (~100 fold increase of protein size estimated at panel A). The conclusion is that hHKII cannot be kept at -20^{o}C or -80^{o}C without using an additive to prevent aggregation. Glycerol is shown to form an amphiphilic layer between the hydrophobic patches on the protein surface and the polar solvent. This helps to stabilise more compact conformations of the protein (Vagenende et al., 2009). DLS supports this notion; the addition of 10% glycerol in GF buffer (10 mM Tris, pH 8.0, 0.5 M NaCl, 5 mM $MgCl_2$) does appear to prevent the formation of large aggregates (Fig. 3.2 C). The most abundant peak is now symmetrical with the correct diameter while the other particles (65 and 289 nm) form a small proportion. In the future 10% glycerol was added to every purified protein before cryo-freezing.

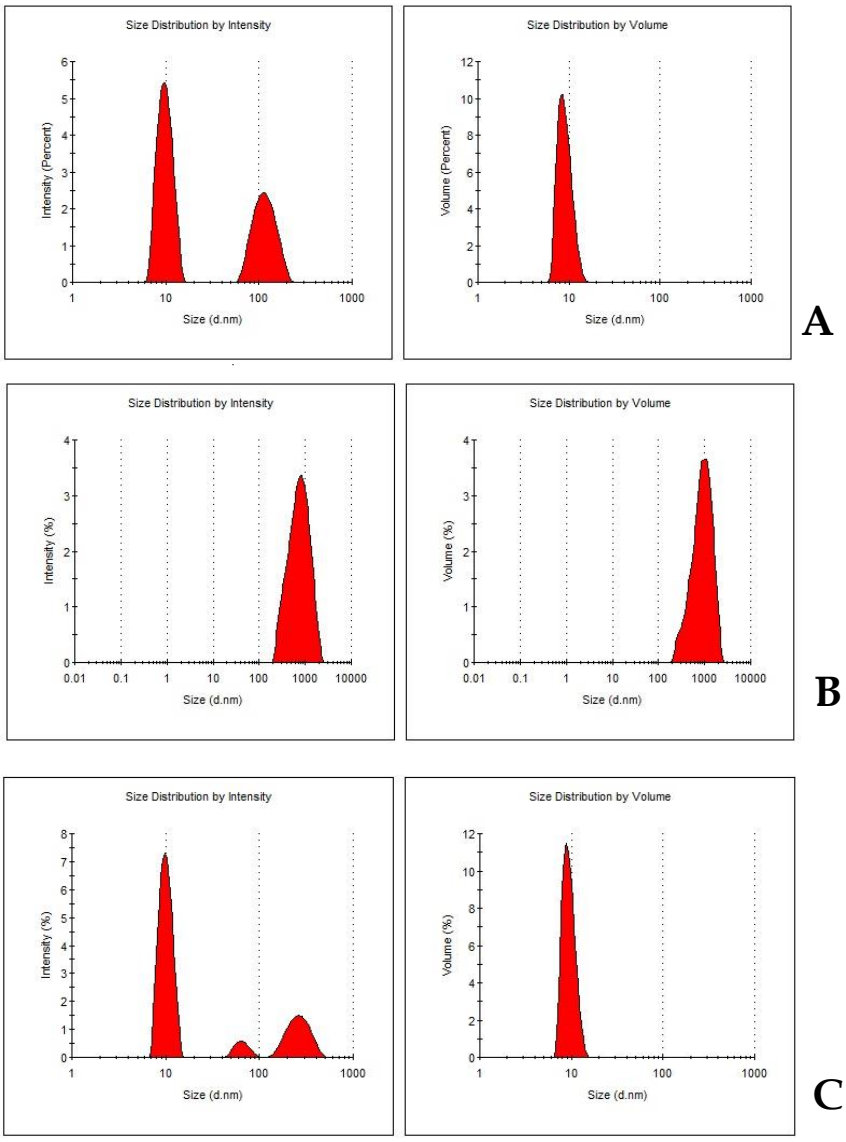


Figure 3.2: Size Distribution by Intensity and Volume of hHKII measured by DLS

Size Distribution by intensity corresponds to left panels and Size Distribution by Volume to right panels.

A. DLS was performed for fresh hHKII in GF Buffer and is estimated to be monodisperse with molecular weight of 144 kDa and Rh 9.9 nm. The theoretical MW is ~102 kDa but DLS estimates the size based on the hydrodynamic calculations for a sphere and this could explain the difference between estimated and real size. The freshly purified protein does not have a significant amount of aggregates based on the volume distribution.

B. After cryo-freezing the protein, DLS shows the formation of large aggregates. The diameter of the particles is estimated to be ~820 nm which corresponds to a 100 fold increase compared to symmetrical peak of fresh protein.

C. 10% of glycerol inhibits the formation of protein aggregates as the percentage of large particles is insignificant on the volume distribution and the estimated diameter and size agree with the values from panel A (10.1 nm and 149 kDa).

3.3.1.2 Enzymatic activity of hHKII is not dramatically altered after cryo-freezing

The hexokinase colorimetric assay was performed to compare the activity yield between fresh enzyme and enzyme after being frozen with liquid N_2. 5 µl of 100 nM stock protein in GF Buffer was added to 100 µl of final assay volume. The reaction was left for incubation for 25 min (measurements were taken every 5 min).

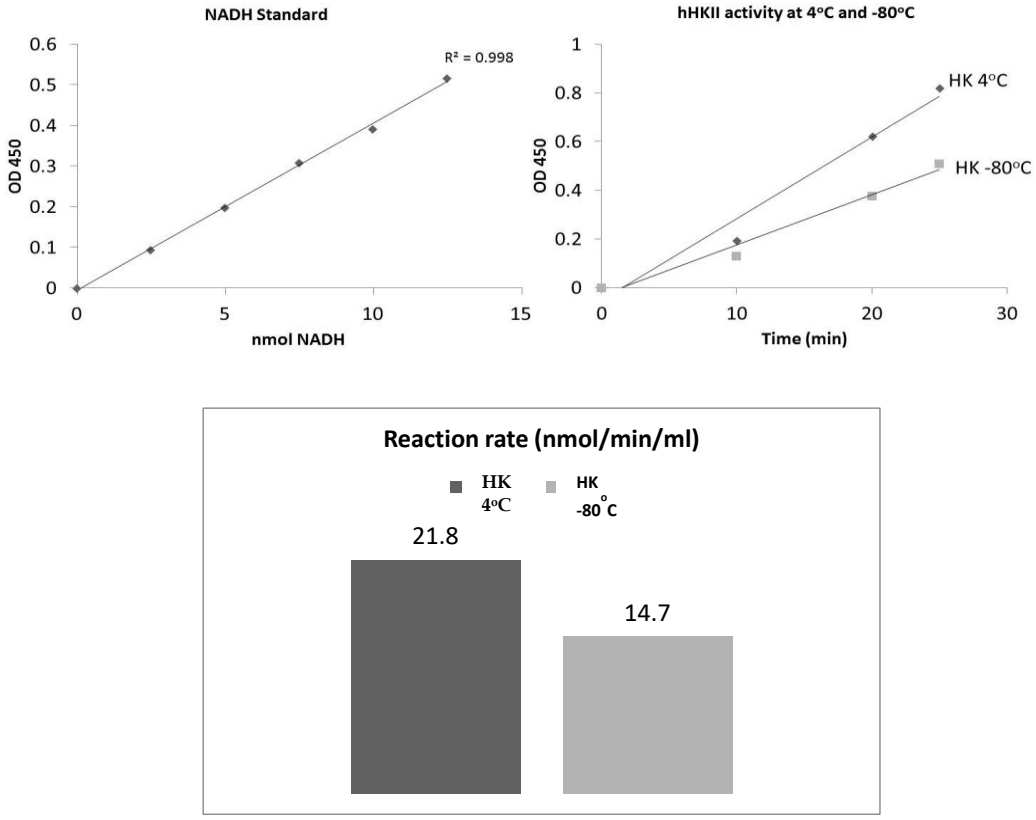


Figure 3.3: hHKII is active after cryo-freezing

The colorimetric assay was used to compare the activity rate between fresh hHKII and hHKII after cryo-freezing. NADH Standard was used to convert the hexokinase absorbance at 450 nm into hexokinase activity (nmol NADH/ min/ ml). Hexokinase stored at 4°C is more active (~33% more active) than hexokinase stored at -80°C so we could conclude that

the formation of minor aggregates after cryo-freezing interferes with the enzyme activity but the rate drop is not dramatic to prevent use of the frozen protein in enzymatic assays.

3.3.1.3 Biochemical characterisation of hHKII

The resazurin assay was performed to measure the binding affinities for Glc and ATP. Two different ionic strength buffers were compared. For the K_m determination, Glc concentration was titrated from 2 mM to 31.25 μM. For the K_m ATP, ATP titrated from 6 mM or 3 mM. 8 mM ATP was found to inhibit the reaction probably because of product (G6P) inhibition (Kosow and Rose, 1970). The initial linear portion of the slopes (100-225 sec) was converted to rate (RFU/sec). The table shows the K_m, k_{cat} values between the different buffers. The physiological concentration of NaCl in the assay buffer (150 mM) gave similar binding affinities as the ones reported previously (Ardehali et al., 1996, Tsai and Wilson, 1997). The binding affinities were improved without NaCl in the reaction buffer. The turnover number is also increased in the absence of NaCl. The present studies cannot reveal the mechanism behind this. However, a hypothesis is that the charged groups within the catalytic active site are influenced by the ionic composition of the medium or that the movement of charged molecules into the active site of the enzyme (i.e ATP-$Mg^{2+)}$ is favoured by a low ionic environment (less electrostatic forces).

Table 3.1: K_m and k_{cat} determination for hHKII under two different conditions of ionic strength

The K_m and k_{cat} values were determined in buffers with 150 mM NaCl and 0 mM NaCl. Purified protein was diluted 1:1000 in each buffer and added to the assay mix. It can be seen that K_m for both substrates decreased in 0 mM NaCl (~2 fold) while the turnover number increases significantly (2-4 times). The k_{cat}/K_m values for both substrates increase more than 3 fold for Glc and around 7 fold for ATP. The ionic strength of the solution seems to be an important parameter affecting hHKII activity.

Buffer composition	**K_m Glc (µM)**	**k_{cat} Glc (s-1)**	**K_m ATP (µM)**	**k_{cat} ATP (s-1)**	**k_{cat}/K_m Glc×10^6 ($M^{-1}s^{-1}$)**	**k_{cat}/K_m ATP×10^6 ($M^{-1}s^{-1}$)**
50mM Tris, pH 7.5, 150mM NaCl, 5mM $MgCl_2$	145±8	695±11	695±125	343±19	4.8	0.49
50mM Tris, pH 7.5, 5mM $MgCl_2$	87±11	1,302±42	391±56	1,343±61	14.9	3.4

3.3.1.4 Glucose K_D is estimated with SPR

SPR analysis of binding specificities for glucose to immobilised hHKII provided a useful insight into the binding mechanism. D-glucose was found to bind to hexokinase in a specific and dose dependent manner as shown from Figure 3.4 A. Steady-state binding response in SPR fits a 1:1 affinity model (3.4 B). No difference in K_D was observed between α and β D-glucose as K_D determined for both is

identical between each other and in the same range of K_D for D-glucose (Fig.3.4 C and D). The mean K_D calculated from 3 different experiments (different days and different protein immobilisation) is 217 ± 77μM. Glucose binds hHKII without the need for the presence of ATP or Mg^{2+}. Both are essential for enzyme catalysis but not essential for binding to the enzyme.

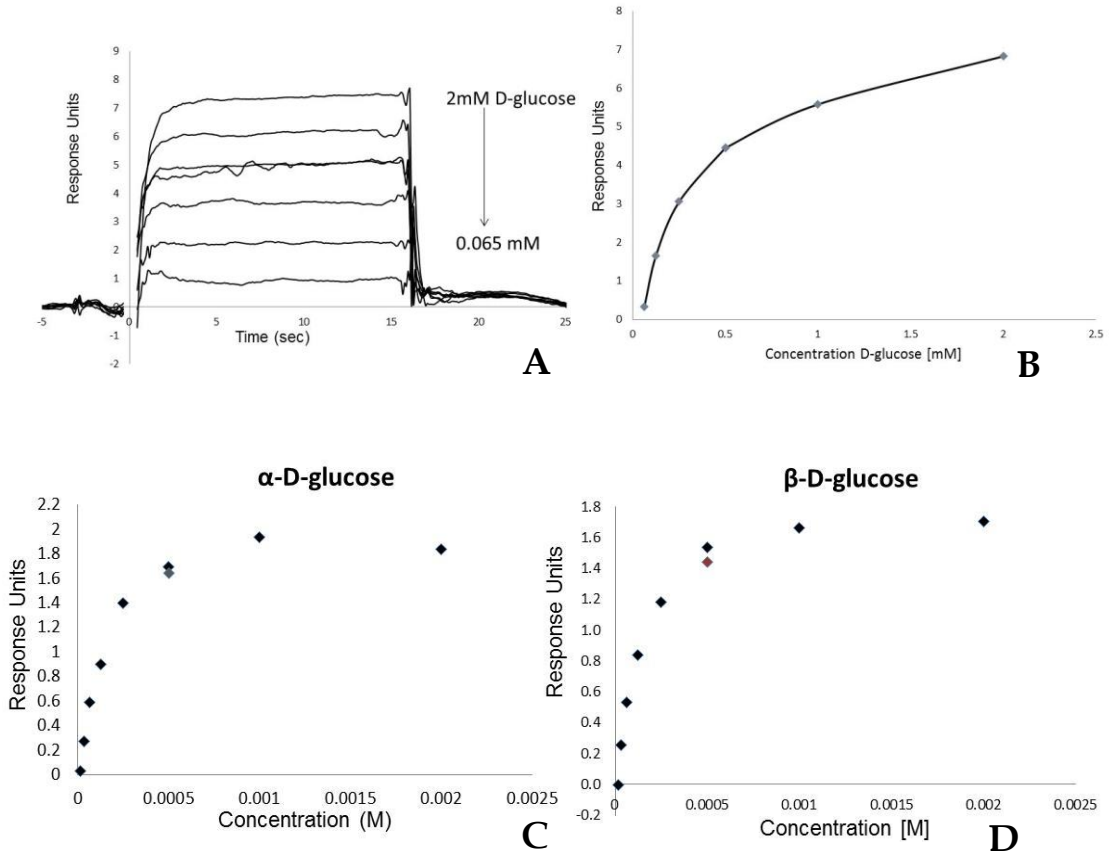


Figure 3.4: Glucose interacts with hHKII in a specific and dose dependent manner

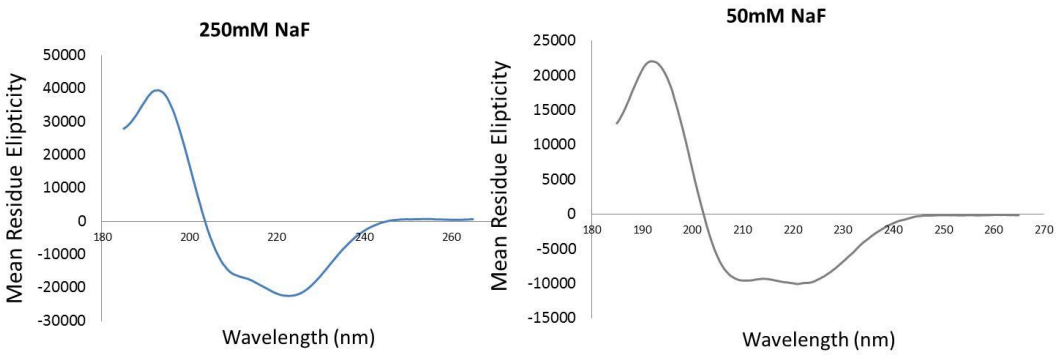
A. D-glucose is found to bind hHKII in a specific and dose dependent manner. A concentration series of 2 mM-0.065 mM (twofold serial dilutions) was tested.

B. Steady-state binding was fit to a 1:1 Lagmuir binding model and K_D is calculated to be 298 μM.

C, D. Steady state binding response for α-D-glucose and β-D-glucose was fit to a 1:1 binding model as previously and calculated to be 109 and 111 μM for α and β glucose respectively.

3.3.1.5 hHKII secondary structure is different in high and low anionic strength buffer

TEV hHKII secondary structure was analysed with CD in two buffers with different anionic strengths (low anionic strength: 50 mM NaF, and high anionic strength: 250 mM NaF). The protein was exchanged into each buffer and analysed immediately. NaF is used instead of NaCl as chloride ion has a strong UV absorbance at low wavelengths. The output data was converted from ellipticity to mean residue ellipticity that is independent of protein size and protein concentration and therefore more suitable for comparing different concentrations of a protein. The DichroWeb on-line resource (Whitmore and Wallace, 2004) was used to analyse the data using 3 different algorithms, SELCON3, CONTIN and CDSSTR. All experimental input data show a good fit to the calculated spectrum derived from the calculated output secondary structure. Protein in high ionic buffer seems to be properly folded since the predicted secondary structure is in agreement with the crystal structure (2NZT.pdb). The crystal structure of hHKII contains 43% alpha- helix and 17% beta-strand which are very close with the predicted average of all methods (50.5% helical and 14.4% strand respectively). However, the secondary structure seems to change when protein is exchanged to buffer with 50 mM NaF. The protein is predicted to be less helical (32.2%) with more beta-strand sheets (25%). Previously it was shown that hHKII is more active in 0 mM NaCl buffer. CD shows that there is also a structural change at 50 mM NaF. The change in secondary structure was further analysed with SEC-MALS to make a more robust conclusion from a combination of different techniques.



Method	% Helix	% Strand	% Turns
SELCON3	45.7	18.6	9.1
CONTIN	48.8	14.7	10.8
CDSSTR	57	10	9
Average	*50.5*	*14.4*	*9.6*

Method	% Helix	% Strand	% Turns
SELCON3	37.4	24.3	12.1
CONTIN	28.4	26.7	11.8
CDSSTR	31	24	12
Average	*32.2*	*25*	*12*

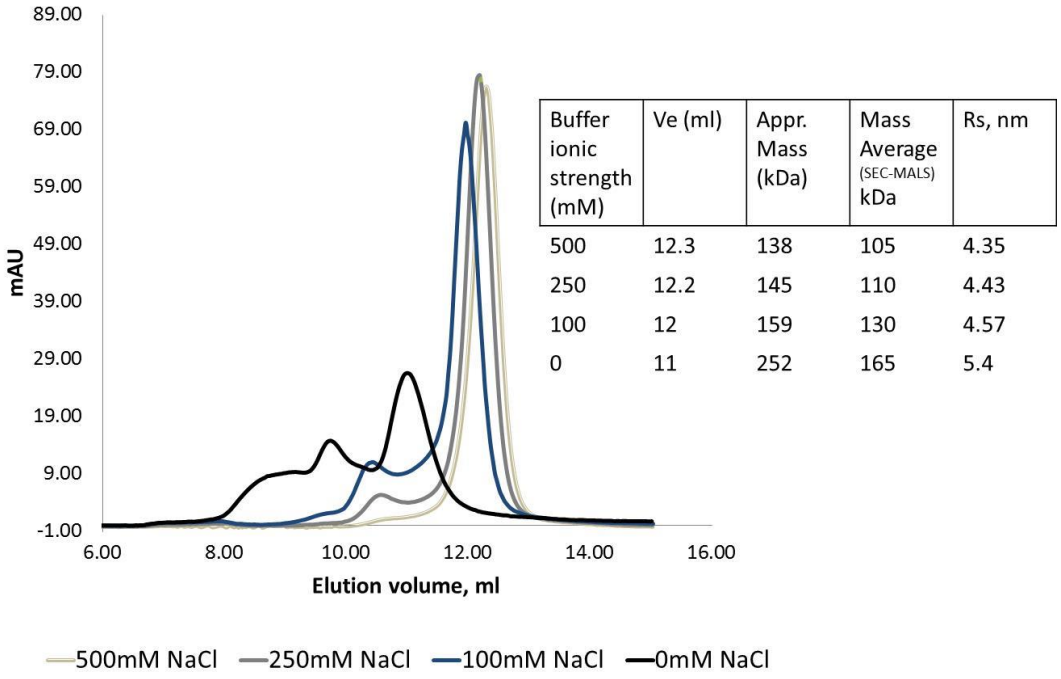
Figure 3.5: hHKII adopts a different secondary structure in low ionic strength buffer
CD analysis of hHKII in two buffers with low and high anionic strength (50 mM NaF and 250 mM NaF) indicates the adoption of a different secondary structure when protein is in 50 mM NaF. The protein in 250 mM NaF is as helical and stranded as crystal structure of the protein in the crystal structure (2NZT). However, when protein changes buffer the helical content drops ~17% and strands are increased by over 10%.

3.3.1.6 hHKII changes conformation when exchanged from high ionic strength to low ionic strength

SEC-MALS was used to determine the molar mass of protein in solution in four different buffers. Relation of the scattering intensity with accurate determination of concentration analysis allows an accurate measurement of the absolute molar mass of the molecules in solution (average mass accuracy error is ±1.97%) than SEC alone. Initially, protein in Gel Filtration Buffer (10 mM Tris, pH 8.0, 500 mM NaCl, 5 mM $MgCl_2$) was analysed and then chloride ion concentration was decreased to 250 mM NaCl, 100 mM NaCl and 0 mM NaCl. Protein mass change was monitored. 100 µL hHKII at 1 mg/ mL was run on a Superdex-200 10/300 GL (GE Healthcare) size exclusion column pre-equilibrated in each buffer under study. A plot of K_{av} versus the $\log_{10}$ of the molecular weight of 11 standards was used to estimate the molecular weight of the unknown sample based on the elution position. Basically K_{av} is

calculated based on elution volumes and elution times as defined from the following relationship: Kav= $(V_e-V_o)/(V_t-V_o)$. V_e is the elution volume, V_t the total volume and V_o the void volume of the column. Light scattering, RI and Abs_{280} were analysed by a homo-polymer model.

Figure 3.6 shows the SEC chromatograms and a table with the SEC-MALS analysis. hHKII at 500 mM Nacl elutes at 12.3 ml as a single peak with a molar mass of 105 kDa (theoretical mass is ~102 kDa). The hydrodynamic diameter, Rs, is estimated at 4.35 nm according to the V_e of the protein. When hHKII is run in 250 mM NaCl, the differences are small. The V_e is the same (12.2 ml) and the molar mass is slightly increased to 110 kDa with estimated Rs at 4.43 nm. These numbers could indicate that protein has started to aggregate. In even lower NaCl (100mM) the V_e of the main peak dropped to 12 ml and the molar mass is estimated at 130 kDa with the Rs at 4.57 nm. At 0 mM Nacl the V_e is 11 ml, a significant drop, with the Rs at 5.4 nm. The apparent mass and the average mass show a big difference (252- 165 kDa). These numbers do not just indicate the presence of aggregates but could represent an increased multimerisation and a conformational change of hHKII (as shown from increase of Rs).



Buffer ionic strength (mM)	Ve (ml)	Appr. Mass (kDa)	Mass Average (SEC-MALS) kDa	Rs, nm
500	12.3	138	105	4.35
250	12.2	145	110	4.43
100	12	159	130	4.57
0	11	252	165	5.4

Figure 3.6: SEC-MALS for hHKII in different ionic strength in solution

SEC-MALS analysis of hHKII shows that protein starts to aggregate when concentration of NaCl is lower than 500 mM as smaller peaks start to elute sooner. However, the main peak that corresponds to the monomer is fully analysed in the insert table. The V_e decreases as ionic strength decreases and mass average increases too. When concentration of NaCl is 0 mM the apparent mass increases almost twofold compared to apparent mass in 500 mM NaCl but the calculated mass from SEC-MALS does not increase within the same range. The average mass increase in combination with the Rs estimation could represent a conformational change.

3.3.2 Biophysical and Biochemical characterisation of N and C domains of hHKII

Chapter 2, section 2.2.3 explains the purification results for N and C domains of hHKII. For C domain the elution profile after GF is not a single peak that could correspond to a monomer, but protein elutes in two peaks which could represent a mixture of monomer and dimer since the resolution between the two species is poor. Further studies were performed on C domain to reveal that both species have the same activity and same functional parameters. DLS did not show any significant

difference in the size between the two peaks. SEC-MALS as a more powerful technique in molar mass estimation shows that both peaks indeed are a mixture of dimer: monomer.

3.3.2.1 DLS performed for pool 1 and pool 2 of C domain did not show any significant difference in their size

Pool 1 (0.3mg/ ml) and Pool 2 (0.25mg/ ml) were analysed with DLS. Both are estimated to be monodisperse with similar size. For pool 1 the R_h=7.28±1.37 nm, 69.2 kDa and for pool 2 R_h= 7.08±1.07 nm, 64.9 kDa. According to this, protein in pool 1 is estimated to be slightly larger than pool 2 but no further information can be obtained from this technique. SEC-MALS is therefore performed to investigate the oligomeric state of both pools.

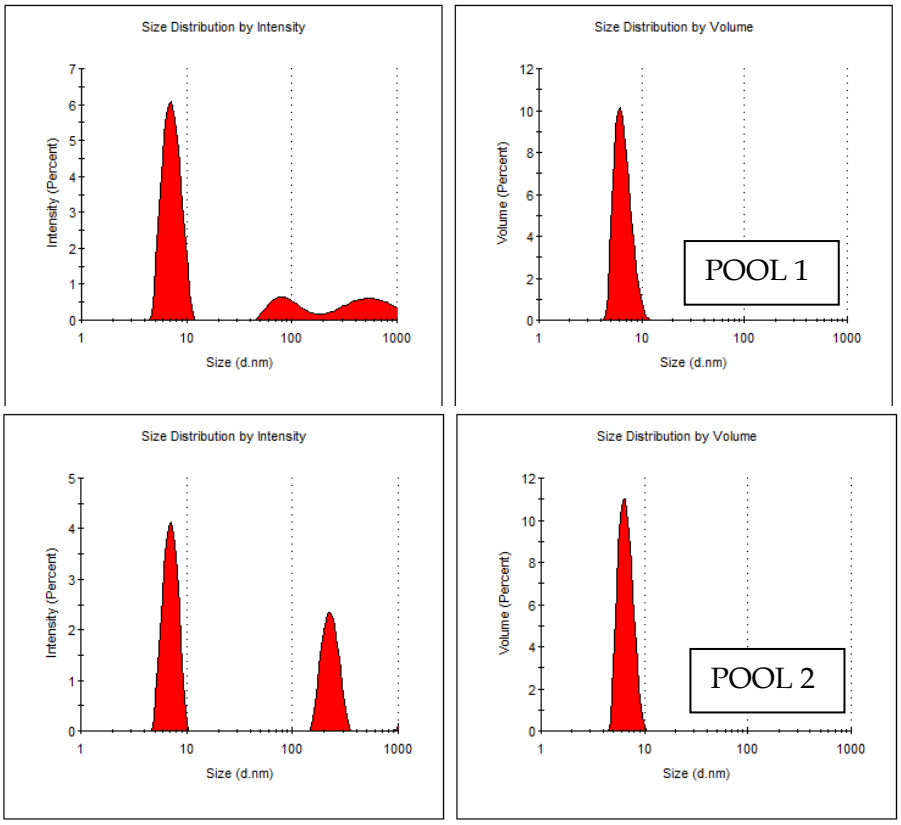


Figure 3.7: DLS for two pools of C domain cannot give enough information regarding the oligomeric state of the protein

3.3.2.2 SEC-MALS shows that both pools are a mixture of monomer and dimer

C domain pool 1 and 2 were further analysed with SEC-MALS for a more accurate estimation of the molar mass. A concentration series was tested for each pool. SEC-MALS shows that both pools are a mixture of dimer: monomer in a ratio ranging from ~20:80 to 30:70. The pools eluted in 2 peaks which are analysed at Table 3.2. The dimerisation does not seem to be concentration dependent or disulphide bond dependent since when 10 mM DTT was incubated with protein no significant change was observed in the ratio of dimerisation. Table 3.2 contains a summary of all conditions tested with SEC-MALS; for all samples tested the two peaks eluted from Superdex 200 13/300 GL which were both analysed.

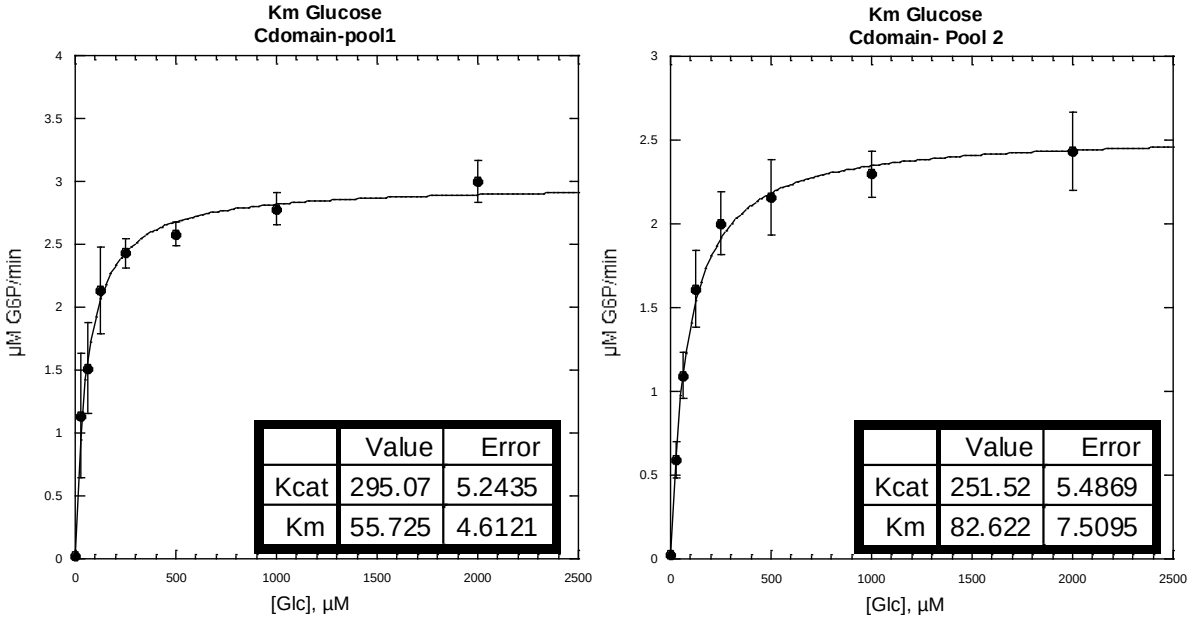
Table 3.2: SEC-MALS analysis on C domain (Pool 1 and 2)

C domain pools were analysed separately with SEC-MALS to determine their oligomeric state and molecular mass. Two peaks elute from SEC for all samples tested (with a good monodispersity Mw/Mn, around 1.001-1.005) which are estimated around 100kDa and 50kDa (dimer and monomer of C domain).

C domain Pool 1	**SEC-MALS Analysis**	**Peak 1**	**Peak 2**	**Comments**
0.5mg/ ml	Mw	98.2	51.5	Dimer:monomer 29:71
	Mw/Mn	1.003	1.005	
1mg/ ml	Mw	91.7	52.3	Dimer:monomer 26:74
	Mw/Mn	1.005	1.003	
5mg/ ml	Mw	97.8	53	Dimer:monomer 30:70
	Mw/Mn	1.003	1.002	
C domain Pool 2	**SEC-MALS Analysis**	**Peak 1**	**Peak 2**	**Comments**
1 mg/ml	Mw (kDa)	95.6	50.2	Dimer:monomer 16:84
	Mw/Mn	1.001	1.001	
5 mg/ml	Mw (kDa)	98.7	52	Dimer:monomer 22:78
	Mw/Mn	1.003	1.003	
5 mg/ml + 10mM DTT (10min incubation)	Mw (kDa)	97.0	52.4	Dimer:monomer 20:80
	Mw/Mn	1.004	1.001	

3.3.2.3 C domain pools are both active and K_m values for Glc and ATP are determined

The resazurin assay was performed to determine the K_m values for Glc and ATP for each pool separately. The initial rate (μM/ min) was plotted against the concentration of the substrate and K_m, k_{cat} values were calculated. Glucose concentration was kept stable at 2 mM when titration of ATP performed (7 mM-0.109 mM, two-fold dilutions) and ATP concentration was kept at 7 mM when K_m of Glc (2 mM-0.031 mM, twofold dilutions) was estimated. However, 7 mM of ATP seems to inhibit significantly the reaction so the rate of reaction was omitted from the plot.



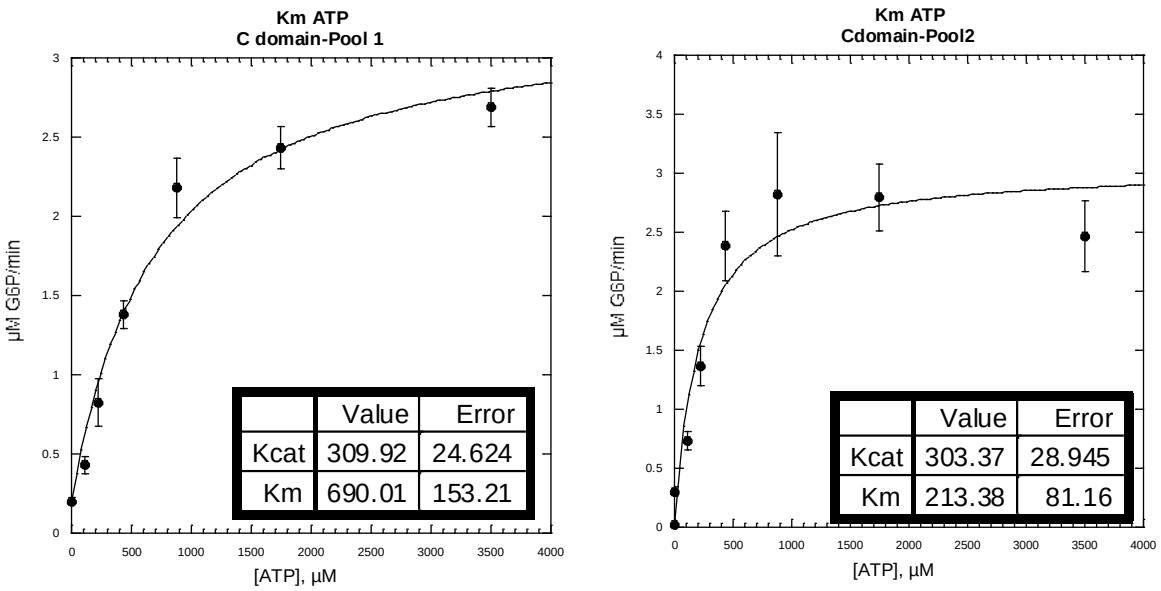


Figure 3.8: K_m values for Glc and ATP for two pools of C domain

Two pools of C domain were characterised in an assay to determine K_m and k_{cat}. The K_m values for Glc were 55 µM (pool 1) and 82 µM (pool 2). These values agree with the reported ones (Ardehali et al., 1996). However, the K_m for ATP for both pools is lower than those reported. Ardehali et al., predict K_m ATP to be ~4 times higher than for the intact enzyme. In our studies, the K_m ATP for Pool 1 is in the same range as for K_m ATP for full length hHKII and the K_m ATP for Pool 2 is ~3 times lower. In our case the affinity of ATP for C domain seems to be as reported elsewhere (Arora et al., 1993).

3.3.2.4 N domain is an active monomer of the right mass

Two different methods (DLS, SEC-MALS) were used to analyse N domain. This domain behaves as a single monodisperse species with apparent molecular mass of 51.5 kDa (theoretical mass is 53.4 kDa). The k_m values for Glucose and ATP were determined with the resazurin assay.

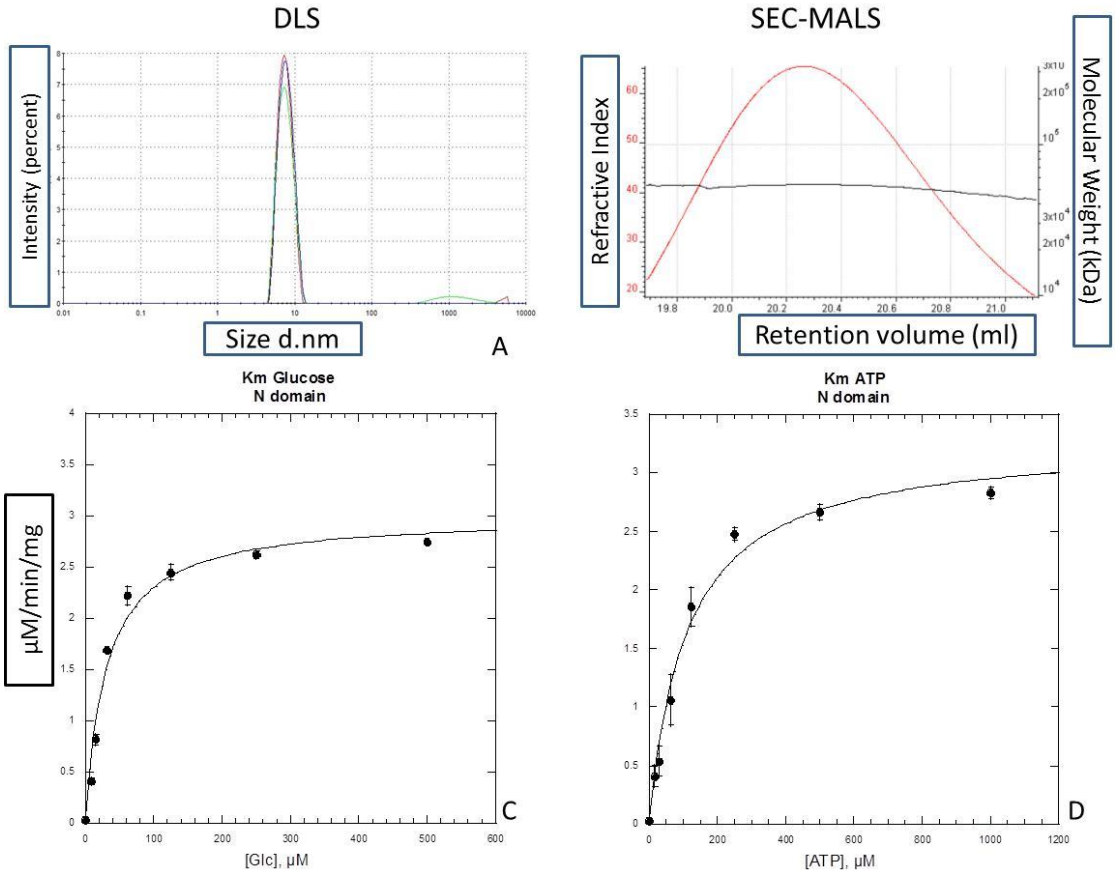


Figure 3.9: Biochemical and biophysical characterisation of N domain

According to DLS, (Intensity Distribution) N domain is a monodisperse species with R_h around 7.6 nm without any significant aggregation. SEC-MALS confirmed the presence of a monomer with the expected mass as protein elutes in a single broad peak with the molecular mass average across the elution profile at 51.5 kDa with good monodispersity (Mw/ Mn=1.003). The K_m for both substrates are lower than expected (Ardehali et al., 1996). The regulatory site alone seems to have a higher affinity for Glucose and ATP in our studies under our enzymatic conditions (K_m Glc 30µM, K_m ATP 113µM).

3.3.3 Untagged hHKII is not active

Untagged hHKII activity was investigated with hexokinase colorimetric assay kit and was found to be ~80 times less active than hHKII (based on the initial rate between 60 and 120 sec).

It is not clear why untagged protein has such a low level of activity but one hypothesis is that the catalytic site of protein, after interaction with HiTrap Blue affinity column, is affected. As mentioned previously some proteins could interact

with this dye because of the structural similarity with nucleotide cofactors. Hexokinase, having a binding site for ATP, has a good potential to bind to the column through the active site of the protein that may lead in a conformational change of it or a saturation of the active site with this ligand leading to a protein that is not able to bind ATP at the concentration tested.

We were unable to purify the untagged protein in satisfactory yield or in an active form. Appendix B shows a summary of all protocols tested for untagged hHKII purification with some comments for each one. None was successful, so untagged hHKII was considered as a difficult target for purification without a tag label.

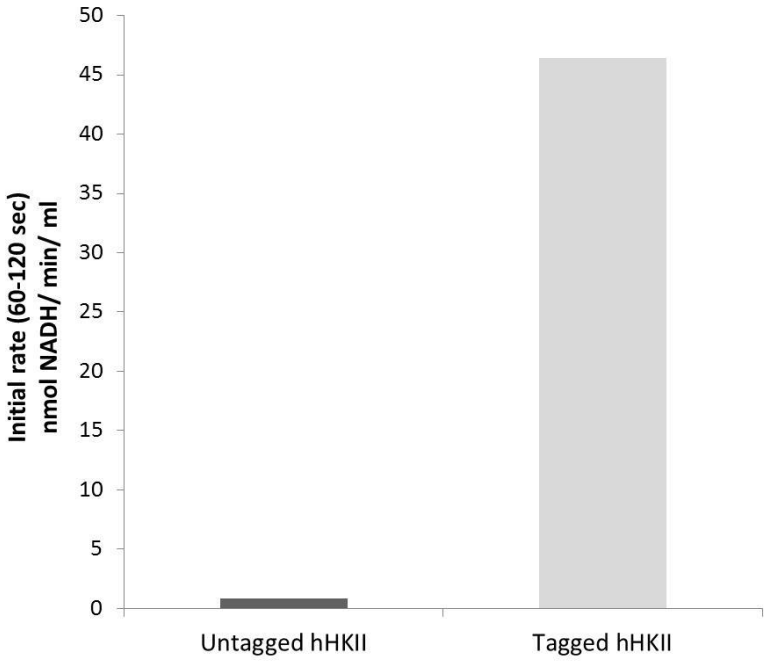


Figure 3.10: Untagged hHKII was not active after three step purification

The activity of the purified protein was measured with the assay kit from BioVision and compared to the purified hHKII (tagged). The untagged hHKII V_o was estimated at 0.853 nmol/min/ml where the tagged protein had V_o= 46.4 nmol/min/ml. The activity level of untagged hHKII is too low in order to use for further studies (structural elucidation, functional parameters, etc.). The purification strategies for untagged hHKII were all unsuccessful.

3.3.4 Biophysical and biochemical characterisation of TbHKI

Chapter 2 (section 2.2.5) refers to three different purification protocols for TbHKI: 2 refolding protocols (screening two different refolding buffers) and one purification protocol from soluble fraction of TbHKI.

The enzymatic activity of similar fractions between the refolding protocols was tested. It was found that specific fractions had up to 10 times higher V_o when the Optimised Buffer (chapter 2, section 2.2.5) was used. The Optimised Buffer has 45 mM NDBS211 (instead of 150 mM NDSB201) and 0.005% Tween 20 (instead of 0.01%). The non-ionic detergent Tween 20 has a CMC of 0.0074%, thus one concentration tested was above CMC and the other below CMC. Even though Tween 20 is considered a mild surfactant that will not affect protein activity when the concentration added is below CMC the protein seems more active. The refolding buffer 1 contained NDSB 201 which has a strong absorbance at 280 nm due to the ring structure in the molecule. For this reason, NDSB 201 was replaced with NDSB 211 in the Optimised refolding buffer to make the protein concentration determination at 280 nm more precise. The resazurin enzymatic assay was used to compare the activity of refolded protein that elutes after SEC. The concentration of each fraction from refolding buffer 1 (with 150 mM NDSB which has a strong absorbance at 280 nm) was measured with Pierce™ BCA Protein Assay Kit (ThermoFisher Scientific). The kit is a high-precision, detergent compatible assay reagent set to measure total protein concentration compared to a standard at 562nm.

The reaction was initiated by addition of 90 µl reaction mix (50 mM Hepes, pH 7.8, 50 mM KCl, 17 mM NAD, 0.02 U diaphorase, 0.01 U G6PDH, 50 mM Tris, pH 7.5, 5mM $MgCl_2$, 10 mM Glc, 1 mM ATP) to 10 µl of TbHKI (elution fractions in triplicate). The reaction rate (V_o) was determined from the initial linear portion of the slopes which was converted to nmol/min/mg.

However, the refolded protein was not studied extensively since it precipitated during concentration and/ or buffer exchange. It seems that protein is very unstable after denaturation and therefore instead of continuing to improve the refolding protocol, TbHKI was purified from the soluble fraction.

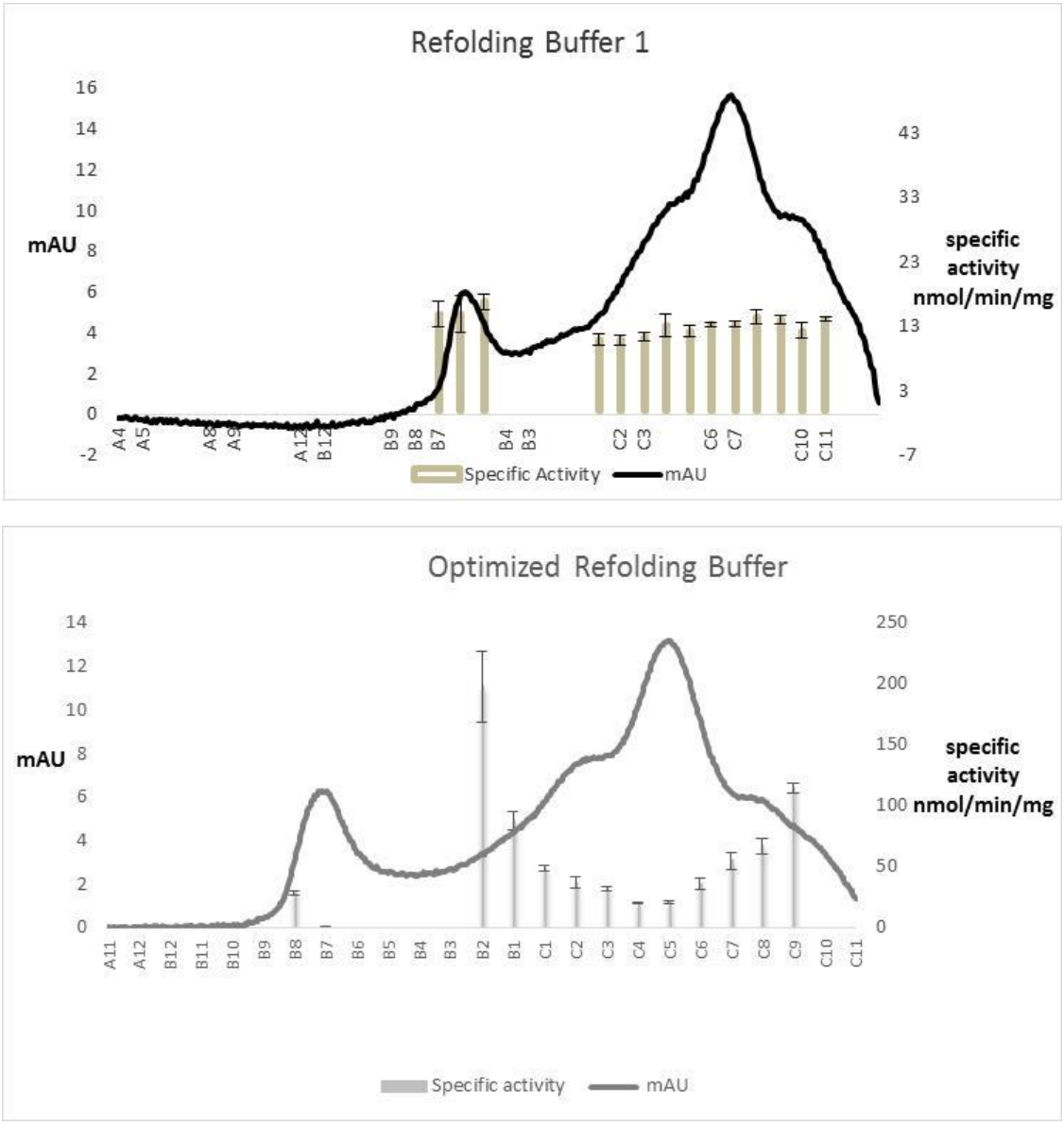


Figure 3.11: Comparison of activity in eluted fractions from two refolding protocols

The initial rate of the refolded fractions was determined with the resazurin assay. The fractions in Optimized Refolding Buffer are more active with activities ranging from 30nmol/min/mg to 198nmol/min/mg compared to an average V_o of 15nmol/min/mg for the refolding buffer 1.

4L of cell culture where TbHKI is slightly overexpressed in the soluble fraction (BL21 star (DE3), TB medium, 18^{o}C, 1hour cold shock) was purified successfully as explained in Chapter 2, 2.2.5. The K_m values for both substrates were determined with the resazurin assay. The concentrations of Glc and ATP were variable and the

reaction buffer consists (50 mM TEA, pH 8.0, 5 mM $MgCl_2$, 17 mM NAD, 100 μM resazurin). The initial rate in RFU/sec was plotted against substrate concentration to estimate the K_m values. TbHKI did not behave like hHKII as the apparent K_m for ATP is much lower, 6± 1.6 μM. The K_m for glucose is in the same range as for hHKII (65 ± 5 μM) and agrees with reported values (Morris et al., 2006). No further biophysical characterisation was carried out on the protein as the final purified amount was not enough for both inhibition studies and biophysical studies. The protein eluted as a monomer after SEC and is active therefore was used for compound screening.

3.4 Summary and Conclusions

hHKII was extensively analysed and functional parameters were determined. Two different constructs (TEV and Thr hHKII) have the same sequence (apart from the 6His tag at the N-terminus). These are expressed in the same organism (BL21 plus RIL (DE3) *E.coli* competent cells) and the elution profiles after SEC, DLS analysis and SEC-MALS analysis performed looked the same. We assume that both constructs will behave identically though the analysis performed above was performed in part for Thr hHKII and TEV hHKII. In general, hHKII could be characterised as a well behaved enzyme which could be stored for a long period at -80^oC without loss of activity. The protein elutes as a monomer with a calculated R_h from DLS at ~10 nm which is in agreement with the crystal structure of hHKII (2NZT). Using PyMOL(TM) 1.8.2.3 the distance from the N-domain to the end of C-domain of 2NZT is around ~120 Å, close to the distance estimated from DLS. The addition of 10% glycerol prevented the aggregation of protein and for this reason all future purified proteins were supplemented with 10% glycerol before cryo-freezing. The K_m, k_{cat} values were determined for both substrates (glucose/ ATP). K_m Glc is 145 μM and K_m ATP is 695 μM. Both numbers are in good agreement with reported values; K_m Glc= 340 μM, K_m ATP= 1.02 mM, (Ardehali et al., 1996), Km Glc= 150 μM, K_m ATP= 420μM,(Tsai and Wilson, 1997), K_m Glc=150 μM, K_m ATP= 700 μM (Wilson, 2003). It was found that hHKII activity was improved in low ionic environment, i.e at zero salt concentration as shown from k_{cat}/ K_m values. More

specifically, the Km values for both substrates decreased up to 1.7 times. On the other hand the k_{cat}, the catalytic turnover number increases (k_{cat} ATP increases from 343 sec^{-1} to 1,343 sec^{-1} and k_{cat} Glc increased from 695 sec^{-1} to 1302 sec^{-1}). The mechanism behind that was not further investigated but there are a number of reviews that have studied the effect of ionic strength on enzyme catalysis (Goldstein, 1972, Nørby and Esmann, 1997).

The K_D for glucose binding to hHKII was determined for the first time using SPR and it was shown that the substrate binds to the protein without the presence of ATP or Mg^{2+}. CD and SEC-MALS were implemented in different ionic strength medium and we could conclude that hHKII changes secondary structure (alpha-helix decreases- beta-strands increase) and that these conformational changes result in an increased hydrodynamic radius. Two hypotheses to explain these findings are, that the long helix which holds the N and C domain might be flexible resulting in a more elongated protein in low salt or alternatively, the two domains are flexible and flap-around in solution and an environment with low salt favours a more open-domain conformation.

The above studies have been implemented using the tagged constructs (TEV or Thr hHKII) and no comparison with the untagged hHKII could be performed. The untagged protein when purified was found to be almost inactive as activity rate was 80 times lower than for the tagged protein. The C and N domains were purified separately and C domain was found to be a mixture of monomer:dimer in equilibrium. Both oligomers were active and the K_m values for both substrates were determined. N domain is a monodisperse species of the expected mass with improved binding affinities for both substrates compared to the full length hHKII and C domain. Purified TbHKI from soluble fraction is active and K_m values were also determined.

4. Chapter 4: A low resolution structure of hHKII

4.1 Introduction

The human hexokinase II constructs purified during the present studies were investigated with Small Angle X-ray Scattering (SAXS). Crystallisation trials have been unsuccessful for tagged hHKII full length and N domain (Appendix A shows the crystallisation trials screened for hHKII), thus an alternative method was required to obtain information regarding the molecular dimensions, the radius of gyration (Rg), the oligomerisation state, the flexibility and if possible the 3D envelope.

To interpret scattering data directly in terms of structural parameters, the sample needs to be a monodisperse species or a mixture of oligomers with defined shapes (Jacques and Trewhella, 2010). Full length hHKII and N domain hHKII have been extensively characterised in solution and seem ideal candidates for SAXS. Full length hHKII and N domain hHKII are monomeric in solution, so it was possible to construct 3D envelopes using *ab initio* methods. However, C domain hHKII studies have shown that it is polydisperse in solution containing both monomer and dimer in equilibrium. We do not know the molecular architecture of the dimer. However, SAXS analysis at various concentrations of each construct, were performed and useful structural information was obtained for all constructs. The results will be discussed in this chapter.

4.1.1 The Basics of Small-Angle X-ray scattering

Small-angle scattering by X-ray arises from the secondary wavelets scattered by atoms within a macromolecule in solution and it is a technique performed to give basic low resolution shape information of a protein (Jacques and Trewhella, 2010). The technique, after data collection and appropriate data analysis to confirm data quality, will provide an *ab initio* shape and structure determination of the macromolecules (Svergun and Koch, 2003).

An X-ray beam is used to illuminate the sample in solution (usually a protein of concentration > 1 mg/ml). Radiation is elastically scattered from the atoms in the sample. Electrons in the sample interact with the incident X-rays causing them to

oscillate. Oscillating electrons behave as dipoles becoming sources of spherical waves re-emitting the X-ray energy. The scattering curve is isotropic due to the random orientation of the macromolecules in solution and represents the scattering averaged over all their positions, conformations and orientations. For a macromolecule, the total scatter is a sum of all the scattering amplitudes within the molecule. Total scatter is a function of electron density for each atom and depends on the number of electrons and how the electrons are arranged, thus it depends on the solute concentration and the electron density contrast ($\Delta\rho$) which arises from the difference in the electron density of the solvent and solute. The intensity of scattered waves is recorded on a detector. The direct beam is absorbed by a beam stop which size and position define the minimum angle measured in an experiment (Jacques and Trewhella, 2010).

Figure 4.1 shows a schematic representation of a typical SAXS experiment.

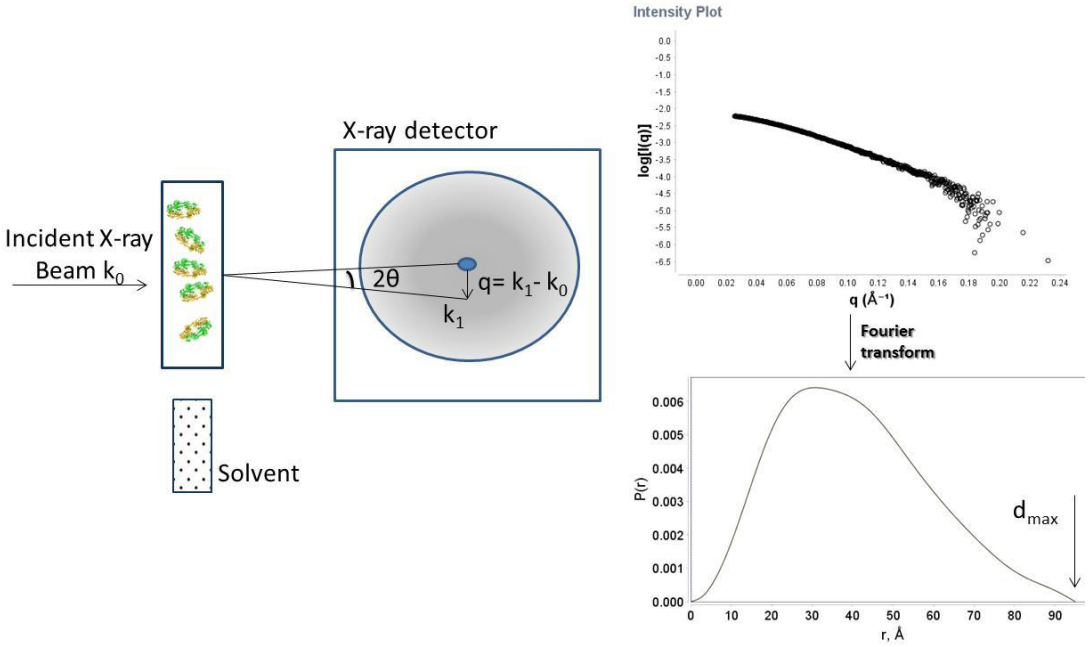


Figure 4.1: Schematic representation of SAXS experiment and the Fourier transform

An incident X-ray beam is directed at a protein sample in buffer. X-ray scattering at angle= 2θ is detected on the X-ray detector. I(q) represents intensity of scatter as a function of momentum transfer, $q= (4\pi \sin\theta)/\lambda$, where λ is the wavelength of the incident X-ray beam and 2θ is the angle between the direct beam and scattered radiation as shown in the figure. The units of q are $Å^{-1}$. The 1D data generated as a curve of log[I(q)/q] can be analysed to obtain information regarding the size, oligomeric state and shape of the molecule. The

Fourier transform of the scattering curve reveals the d_{max}, i.e the linear dimension of the particle. For the calculation of d_{max} the P(r) function is constrained to be zero at r=0 and r=d_{max}. Unfolded proteins are not zero at r=0 whereas r≠0 at d_{max} indicates aggregated protein. Adapted from (Jacques and Trewhella, 2010, Svergun and Koch, 2003, Putnam et al., 2007, Mertens and Svergun, 2010).

The scattering curve (I(q) versus q), is obtained by subtraction of the scattering intensity of the buffer from the scattering intensity of the protein sample plotted against q= $(4\pi \sin\theta)/\lambda$ (Fig. 4.1) where q, the momentum transfer, is the magnitude of the reciprocal space scattering vector. I(q) versus q is basically the intensity as a function of scattering angle because λ is fixed and θ is small, typically less than 3^o (Jacques and Trewhella, 2010). For a specific value of q the largest contribution to the scattering comes from particle dimension $2\pi/q$. Hence, the presence of large protein aggregates can be inferred from scattering intensity rapidly rising very close to the beam stop. The higher resolution data corresponding the small distances is present at larger values of q. From high-quality small-angle scattering data the experimenter can calculate two parameters with accuracy, the zero angle scattered intensity *I*(0) and the radius of gyration Rg, which relate to the shape and size of the particles under study. *I*(0) (scattered intensity from zero angle) cannot be measured directly as it cannot be distinguished from the unscattered radiation, i.e the direct beam. It can be determined though by extrapolation using Guinier's relationship. *I*(0) relates directly to the particle's volume (V) and Δρ scaled by concentration c. From *I*(0) it is possible to calculate the Mr of the particle in solution with the following relationship:

$$I(0) = N(\Delta\rho V)^2 = \frac{C\Delta\rho^2 u^2 Mr}{N_A}, \text{ (Equation 4.1)}$$

Where N is the number of scattering particles per unit volume, Δρ the contrast (the difference in the electron density between the solvent and protein), V the particle volume, C the mass per unit volume, Mr the molecular mass, u the partial specific volume of the particle and N_A is Avogadro's number (Jacques and Trewhella, 2010).

As *I*(0) is proportional to the concentration and Mr of the macromolecule monomer being measured, one can understand that an accurate determination of both

parameters is crucial for the investigation of the oligomeric state of the sample. For flexible systems the electron scattering contrast is not easy to be determined as the particle and solvent scattering are hard to distinguish (Rambo and Tainer, 2011). Thus, the mass estimation will not be accurate based on Guinier analysis. The Rg is the average squared centre of mass distances in the molecule weighted by their scattering densities, providing useful information on the mass distribution within a particle. Rg like the hydrodynamic or Stoke's radius depends on the actual shape of the molecule under study and cannot be used to estimate the molecular mass of the studied molecule (Putnam et al., 2007).

I(0) and Rg, for a monodisperse solution of a globular protein, are estimated using the Guinier equation:

$$I(q) = I(0)e^{\frac{-q^2Rg^2}{q}},$$ (Equation 4.2)

In 1939, Guinier showed that for small values of q (in general for qRg<1.3) the above equation can be used and from a linear fit of ln[I(q)] versus q^2 the slope and y intercept reveal the Rg and *I*(0) respectively. A linear Guinier plot is informative and can be used to show that no significant aggregation and/or no inter-particle interference exist in the protein under study (Jacques and Trewhella, 2010).

Independent from the Guinier analysis, the hydrated particle volume (Vp) can be estimated from Porod's equation assuming uniform electron density inside the particle:

$$Vp = 2\pi^2 \frac{I(0)}{Q}, \quad Q = \int_0^\infty q^2 I(q) dq,$$ (Equation 4.3)

where Q is the Porod-invariant. For macromolecules with a higher Mr of 30kDa, we can consider a uniform electron density by subtracting an appropriate constant from the scattering data. By determining the Porod's equation, errors in concentration measurement will not affect the estimation of the molecular mass. Typically, for a globular protein Vp ($Å^3$) is 1.5-2 times the molecular mass (kDa) (Mertens and Svergun, 2010). In our studies, an average value for an average protein was used and molecular mass was estimated as Mr [kDa]= Vp [$Å^3$]/ 1.7.

Rg and *I*(0) can also be determined from an indirect Fourier transform method which yields the electron distance distribution function, P(r). This is a histogram of

distances of all possible electron pairs. From the shape of this distribution it can be possible to infer the domain arrangement of a protein. Because the scattering data are obtained within a defined limit of q_{min} to q_{max} the indirect Fourier transform depends upon assumptions, such as the d_{max}, the maximum particle diameter. This alternative estimate of Rg makes use of the whole scattering curve and it is much less sensitive to interactions or to the presence of a small fraction of oligomers. To conclude, the possibility to estimate the Rg and $I(0)$ from different mathematical models allows the experimenter to monitor the consistency of the data (Jacques and Trewhella, 2010).

4.1.2 *Ab initio* modelling

The last step in SAXS analysis is to reconstruct a low-resolution molecular envelope of the macromolecule using nothing else than the scattering data. There are a number of available *ab initio* algorithms, but I will only refer to the ones used to generate the 3D models during the present studies, which are the DAMMIN (Svergun, 1999) and DAMMIF (Franke and Svergun, 2009, Jacques and Trewhella, 2010). DAMMIN (Dummy Atom Model Minimisation) represents the molecule as thickly packed beads inside a specific search volume defined by the experimentally determined d_{max}. The algorithm performs the shape reconstruction starting from a random initial approximation by simulated annealing (SA) which after each step creates a new model by a different single bead assignment which will ultimately lead to the creation of a compact model with connected beads (Mertens and Svergun, 2010). The program compares the experimental scattering to the calculated scattering derived from the bead model. The newer version of DAMMIN, the DAMMIF, where F refers to fast, is different in several aspects. First there is no limitation on the search volume, second only the interconnected models are used for the calculation of scattering amplitudes and last each bead contributing to the total scattering at least once is used for the computation of scattering amplitudes (Mertens and Svergun, 2010). However, one must keep in mind that a 3D model from the scattering data using the above algorithms will not necessarily stay consistent when running the programs multiple times. For this reason, DAMAVER (Volkov and Svergun, 2003),

averages all the proposed models to generate a smoothed model containing the most common features among all the 3D models (Mertens and Svergun, 2010).

4.2 Materials and Methods

4.2.1 SAXS sample preparation

SAXS data was collected at the Diamond Light Source, Oxon, UK (DLS) at the B21 BioSAXS beamline on a Pilatus 2M detector with a fixed camera length of 4.014 m and 12.4 keV energy (1 Å wavelength) allowing the collection of the momentum transfer range q between 0.0038–0.42 $Å^{-1}$. All samples were dissolved in Gel Filtration buffer (10mM Tris, HCl, 0.5M NaCl, 5mM $MgCl_2$, 10% glycerol). For N and C domain the buffer is supplemented with 2mM DTT. The samples were centrifuged at 13,000xg at 4^{o}C for 20 mins. Data was collected at 25^{o}C. To check for concentration dependent effect the following concentrations were tested:

1) For full length hHKII: 6.20, 3.10, 1.55, 0.75 mg/ml
2) For N domain hHKII: 5.60, 2.80, 1.40, 0.70, 0.35 mg/ml
3) For C domain hHKII: 2.80, 1.40, 0.70, 0.35 mg/ml

All proteins were stored at 1 mg/ ml and concentrated to the above highest concentration using a Vivaspin with molecular weight cut-off= 30kDa. Twofold serial dilutions were followed and concentrations were measured using UV spectroscopy, all blanked with the sample buffer.

4.2.2 SAXS data analysis and modelling

SAXS data was analysed using the Scatter 3.0H and ATSAS 2.7.2 suite. The primus analysis tool was also used for Guinier analysis. GNOM was used to estimate the d_{max} of each molecule in solution. Low-resolution structure models were constructed by *ab initio* modelling using DAMMIN and DAMMIF.

4.3 Results and Discussion

4.3.1 SAXS analysis of tagged hHKII

SAXS data were collected for tagged hHKII between 0.75 and 6.20 mg/ ml. Guinier analysis was performed for all concentrations. If the oligomerisation state is not altered by the concentration of the protein then we would anticipate the *I*(0)/c and Rg to remain constant. Otherwise, if both parameters increase with increased concentration of protein then it is most likely that protein aggregates or changes its oligomeric state. As seen from Table 4.1 the Rg does not change significantly between all 4 concentrations. The Rg as estimated from Guinier and P(r) distribution are in good agreement with each other with an average Rg at 42.26± 0.96 Å (2.2% error which is negligible). A linear Guinier plot confirms that sample is not aggregated. A slight upturn at low q starts to appear at concentrations 3.10 and 6.20 mg/ml but only for a few data points very close to the beam stop. These points were not included in the analysis. *I*(0) determined from Guinier analysis and concentration estimated from absorbance at 280nm, gave an average particle mass of 76.7 kDa suggesting that the concentration measured by UV was an overestimation (theoretical mass is 102 kDa). The molecular mass calculated from the Porod volume (concentration independent) gave a mass ranging from 101-106 kDa, all of them suggesting that hHKII is a monomer in solution (theoretical mass is 102 kDa). D_{max} was determined including scattering data from 0 to q= 0.34 ($Å^{-1}$) as signal to noise ratio was not optimal for higher q. High q data is most susceptible to slight mismatch between the buffer the sample is dissolved in and the buffer sample used for solvent subtraction.

Table 4.1: SAXS parameters of full length hHKII along a concentration series (0.75-6.20 mg/ ml)

SAXS data were analysed to obtain structural information of hHKII. The Rg of the protein is estimated at 42.3Å from two different methods (Guinier-Real space analysis). The predicted assembly of the molecule is monomer based on the predicted molecular mass from Guinier analysis and Porod volume. The d_{max} ranges from 137-147Å.

hHKII conc. **mg/ml**	Guinier Analysis **Mass (kDa)**	Guinier analysis **Rg (Å)**	Porod volume **Mass (kDa)**	P(r) **Rg (Å)**	Dmax **(Å)**	Predicted assembly
0.75	80	42.1	102	41.0	137	Monomer
1.55	75	43.9	105	41.5	145	Monomer
3.10	74	43	104	41.9	147	Monomer
6.20	78	43.3	107	41.4	145	Monomer

A 3D envelope model of hHKII was constructed from small-angle scattering data (1.55mg/ ml) using the *ab initio* programs DAMMIF and DAMMIN. DAMMIF, which is faster, generated 13 models, which were all very similar to each other. The 13 models were aligned to generate an average model which was used as an input for DAMMIN to develop the final bead model. A molecular envelope was obtained using PyMOL and the known crystal structure of the monomer hHKII was docked into the model manually. Figure 4.2 shows that the 3D model correlates well with the monomer of hHKII shown in cartoon. A comparison of the X-ray solution scattering from SAXS hHKII (blue line) and the best fit by CRYSOL (red line) of the atomic resolution crystallographic data confirms this; χ^2= 0.965 (Svergun et al., 1995).

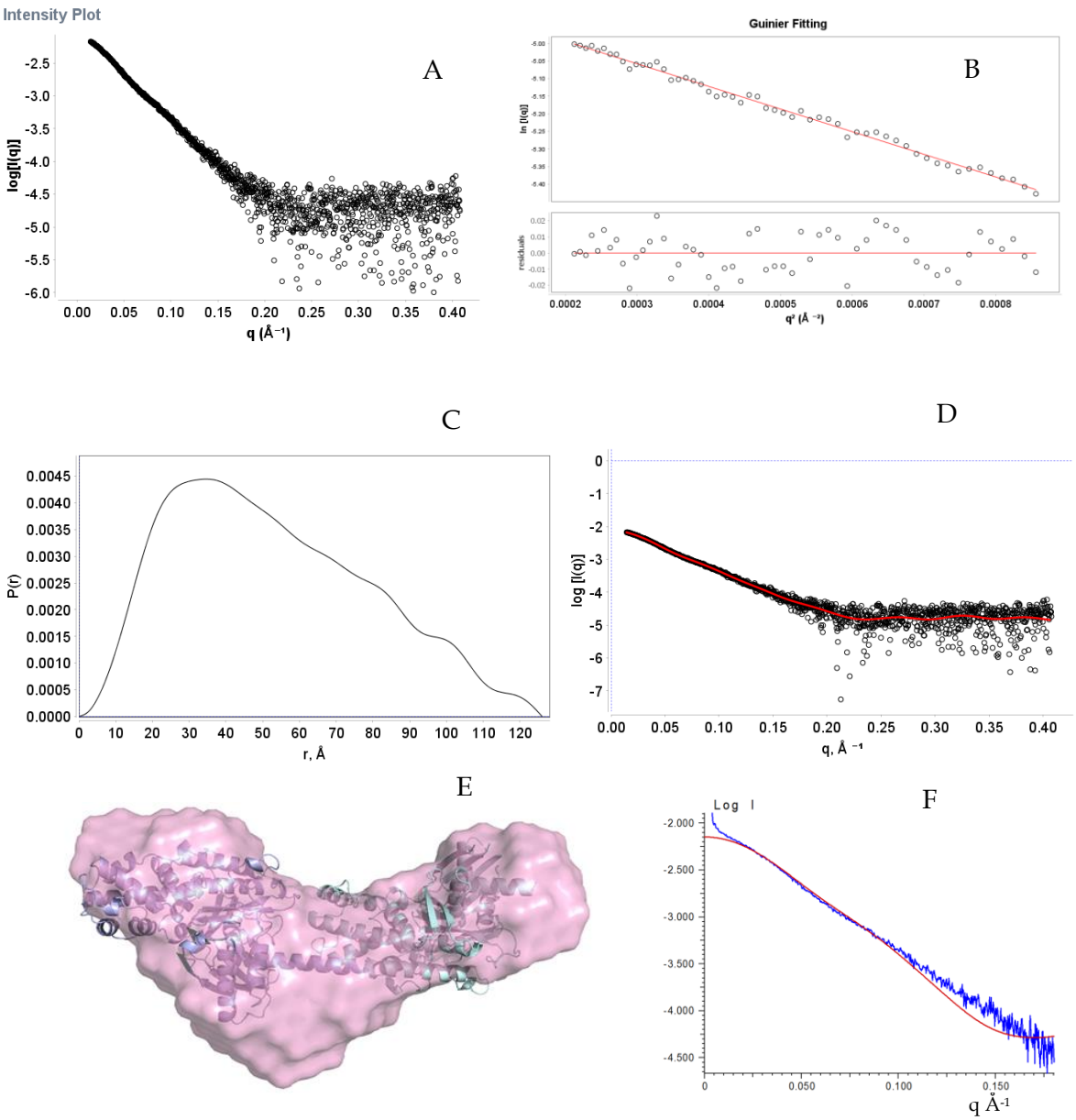


Figure 4.2: The 6His-tag hHKII forms monomeric assembly in solution

A. The intensity plot; log[I(q)] against q ($Å^{-1}$).

B. The linear fit of Guinier analysis is indicative of non-aggregated protein at this concentration. Intensity plot of hHKII (1.55 mg/ ml) does not have a significant upturn or downturn at low q so there is no significant aggregation or inter-particle effects.

C. Real space *I*(0), Rg and estimation of the maximum dimension of the particle, d_{max}, is made from an indirect Fourier-transform of the intensity data into the pair-distance distribution function, P(r). The P(r) distribution contains multiple shoulders after $50Å^{-1}$ until d_{max} which could indicate that protein has more than one domain (Putnam et al., 2007).

D. The experimental scattering data fit well to the calculated scattered curve predicted from the P(r) distribution.

E. 3D envelope of hHKII predicted from DAMMIN modelling program (2nzt.pdb, chain A). There was no significant preference in N and C orientation when fitting the structure to the model.

F. A comparison of the solution scattering from SAXS analysis of hHKII (blue line) and the best fit of the atomic resolution crystallographic data (red line); χ^2= 0.965 from CRYSOL (ATSAS 2.7.2).

4.3.2 SAXS analysis of N domain of hHKII

N domain of hHKII was analysed by DLS and SEC-MALS and predicted to be a monodisperse species of the correct molecular mass. SAXS data were collected for N domain between 0.35 to 5.60 mg/ ml. The signal to noise ratio for 0.35 mg/ ml is not optimal for further analysis so this concentration was omitted. The remaining concentrations were initially analysed to establish that sample was free of aggregation and inter-particle interactions. Guinier analysis was performed for all concentrations of the N domain. The average Rg is estimated at 29.56±1.26Å. The average particle mass derived from Guinier analysis is in range of 37-41 kDa (theoretical mass 53.4 kDa) which again suggests that the concentration measured from UV spectrometry was overestimated. The buffer where N domain was stored contains 2mM DTT; oxidation might interfere with the accurate measurement of the concentration (Jacques and Trewhella, 2010). Interestingly the mass estimated from the Porod's law is quite different (average mass estimated at 64.89±11 kDa) and with a significant error. The excluded volume calculated through the Porod invariant,

$$V = \frac{2\pi^2 I_{exp}^2(0)}{\left(\int_0^\infty I(q)q^2 dq\right)}, \text{ (Equation 4.4)}$$

can be converted to molecular mass from dividing the excluded volume by 1.7. $I_{exp}(0)$ is the experimental intensity at q=0. The integral portion of the above equation is known as the Porod invariant. However, accuracy varies for shape and size. In general, this estimate is true for large globular proteins (>70 kDa) and when it applies to small proteins or proteins with unusual shapes it fails to make a right prediction (Putnam et al., 2007) In our case, due to the fact that N domain is an

elongated protein <70 kDa, using the Porod invariant might mean the estimation of mass is inaccurate. With a closer look to the data, one can see that the predicted mass is larger for SAXS analysis at 0.70mg/ ml (66 kDa) than it is for 5.56mg/ ml (58.29 kDa). The mass estimated from the real-space P(r) distribution is close to the mass predicted from Guinier analysis. The N domain assembly is predicted to be a monomer according to this data. The Rg average (from both Guinier and real-space analysis) is 29.42±1.09 Å, around 30% smaller than the Rg of the full length protein. The highest quality data were used to build the 3D envelope of N domain (2.80mg/ ml). Figure 4.3 depicts the intensity plot and a linear Guinier plot, both of which are indicative that the sample is not aggregated. The d_{max} was estimated at 101Å. As previously the DAMMIF and DAMMIN programs were used to generate the final bead model. DAMMIF, which is faster, generated 13 models, which had similar shapes. The 13 models were aligned to generate an average model which was used as an input for DAMMIN to develop the final bead model. A molecular envelope is shown with PyMOL along with the docked crystal structure of N domain. Figure 4.3 shows that the 3D model correlates well with the monomer of N domain shown in cartoon (manual fit).

Table 4.2: SAXS parameters of N domain along a concentration series (0.70-5.6mg/ ml)

SAXS data were analysed to obtain the Rg of N domain which is estimated at 29.4Å from two different methods (Guinier-Real space analysis). The calculated Rg is roughly 30% smaller than for the full length protein which was expected. The predicted assembly of the molecule is monomer based on the predicted molecular mass from Guinier analysis and Porod volume. However the calculated mass from the UV spectrometry is likely to be overestimation due to oxidation of DTT. The d_{max} ranges from 96-101Å.

Ndomain conc. **(mg/ml)**	Guinier Analysis **Mass (kDa)**	Guinier Analysis **Rg (Å)**	Porod volume **Mass (kDa)**	P(r) **Mass (kDa)**	P(r) **Rg (Å)**	D_{max} **(Å)**	Predicted assembly
0.70	37	28.6	66	41	28.9	101	Monomer
1.40	39	31.6	61	40	28.8	96	Monomer
2.80	39	28.4	60	41	28.8	101	Monomer
5.60	41	30.5	58	45	31.0	99	Monomer

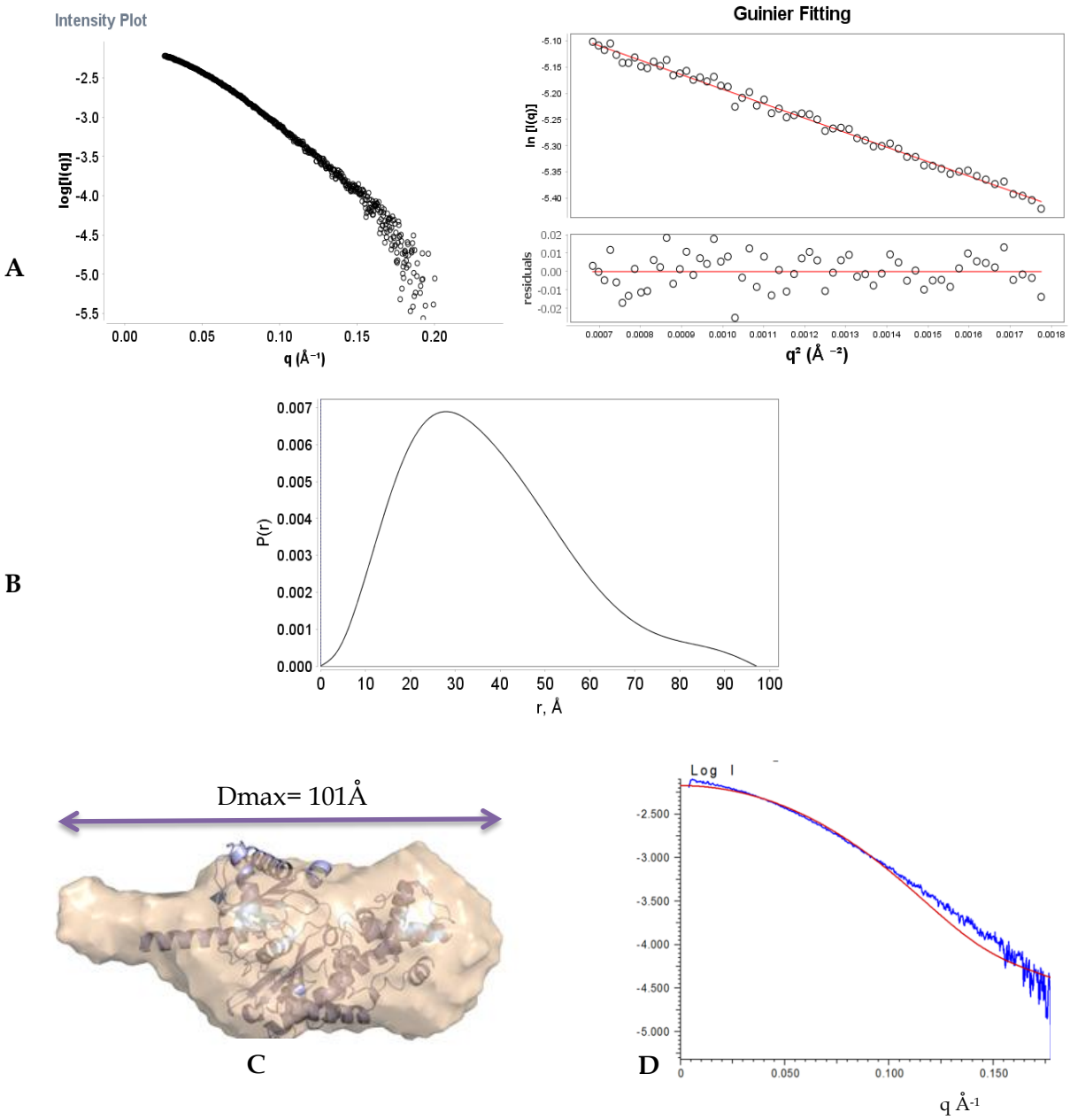


Figure 4.3: The N domain of hHKII forms a compact assembly in solution

A. Intensity plot of N domain (2.80mg/ ml) does not have an upturn or downturn at low q so protein does not have a significant proportion aggregated. The linear fit of Guinier analysis is also indicative of non-aggregated protein.

B. Real space *I(0),* Rg and estimation of the maximum dimension of the particle, d_{max}, is generated from indirect Fourier-transform of the intensity data into the pair-distance distribution function, P(r).

C. 3D envelope of N domain predicted from DAMMIN modelling program. The 3D model has a guitar shape and the neck seems to fit well the long helix that connects N and C domain. PyMOL was used to fit manually the N domain light blue cartoon) to the 3D envelope.

D. A comparison of the solution scattering from SAXS analysis of the N domain (blue line) and the best fit of the atomic resolution crystallographic data (red line); χ^2= 1.7 from CRYSOL.

4.3.3 SAXS analysis of C domain of hHKII

C domain of hHKII was analysed by SAXS in concentration series from 0.35 to 2.7 mg/ ml. As for N domain, the lowest concentration of 0.35mg/ ml was omitted from the analysis as the data were not optimal in such a low concentration. Guinier and real-space analysis were performed to determine the *I(0)* and Rg. The Rg was found to be concentration and method independent with an average of 31.38±0.66 Å. The Rg for both domains is in close proximity which is expected as N domain and C domain have similar size and shape. The average particle mass derived from Guinier analysis is in range of 45-52 kDa (theoretical mass 50.8 kDa). However, the mass estimated from the Porod's law is quite different (average mass estimated at 80.6±6.9 kDa). The same phenomenon is observed for N domain. In contrary to the N domain however, C domain calculated mass does not improve when concentration is higher. At concentration 2.7mg/ ml the mass is estimated at 73.89 kDa which is exactly the average of monomer and dimer of C domain ((100+50)/2). This result seems to agree with SEC-MALS analysis (Chapter 3, section 3.3.2.2) which suggested that C domain is a monomer:dimer (80:20) in equilibrium.

Volume-of-correlation analysis, Vc, overcomes the limitation of Porod analysis for flexible systems. Vc is defined as the ratio of *I(0)* to its total scattered intensity. The total scattered intensity is estimated by integrating the area under the curve of SAXS data transformed as qI(q) versus q. Vc is concentration independent as shown from studies on well-characterized molecules of different mass (Rambo and Tainer, 2011). Vc requires an accurate determination of *I(0)*, preferably by combining both Guinier analysis and real-space P(r) distribution. Vc analysis was performed for C domain

after *I*(0) determination from both methods. The estimated mass according to this method ranges from 58 kDa to 110 kDa.

D_{max} estimations range from 96 to 109 Å. The 3D envelope was not constructed as C domain cannot be treated one rigid body, it is probably an equilibrium of monomer:dimer in solution and the 3D envelope would not be as accurate as for full length and N domain of hHKII.

Table 4.3: SAXS parameters of C domain along a concentration series (0.70-2.80mg/ml)

SAXS data were analysed to obtain structural information for C domain of hHKII. The calculated Rg is 31Å in average. The predicted assembly of the molecule was hard to be answered based on Guinier and Porod analysis. Based on Guinier analysis the predicted assembly is closer to a monomer while from Porod analysis tends to be a dimer. As both have limitations, the volume-of-correlation analysis was performed which enables us to find the molecular mass of the molecule under study in a way that is independent of concentration and of the need for a compact, rigid system. The mass is not consistent between the different concentrations since it ranges from 58 to 110 kDa (monomer-dimer).

Cdomain conc. **mg/ml**	Guinier Analysis **Mass (kDa)**	Guinier analysis **Rg (Å)**	Porod volume **Mass (kDa)**	Vc analysis **Mass (kDa)**	P(r) **Rg (Å)**	Dmax **(Å)**	Predicted assembly
0.70	45	31.1	83	67	30.2	96	Monomer:dimer
1.40	47	30.9	75	110	31.3	109	Monomer:dimer
2.80	50	32.01	74	58	31.1	107	Monomer:dimer

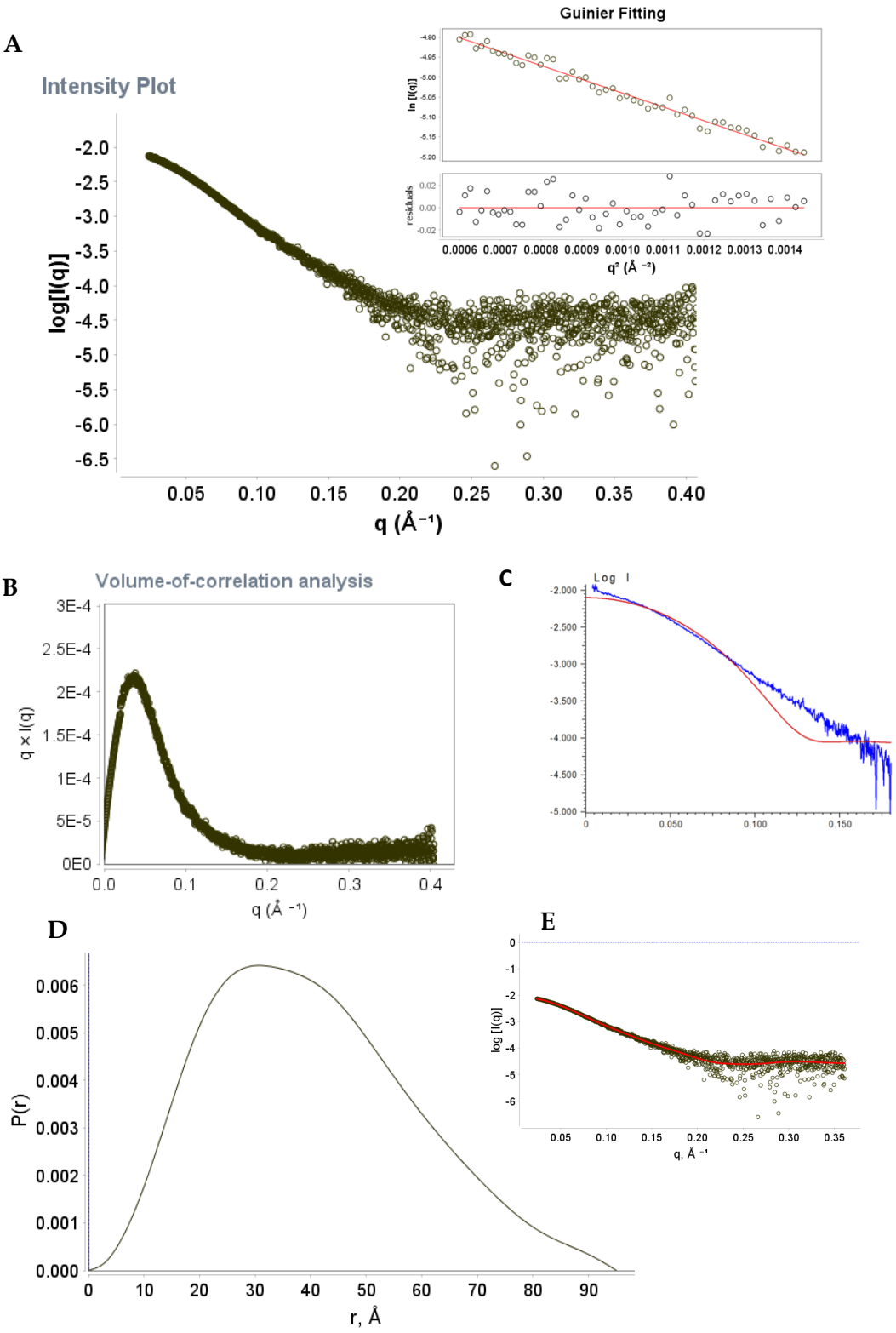


Figure 4.4: The C domain of hHKII forms a multimeric assembly in solution

A. Intensity plot of C domain (2.80 mg/ ml) does not have an upturn or downturn at low q so protein does not have significant aggregation or suffer from inter-particle interactions. The

linear fit of Guinier analysis is also indicative of non-aggregated protein at this concentration.

B. Integrating the area under the curve of SAS data transformed as qI(q)/q gives the total scattered intensity. The plot is more informative on the molecular mass of a flexible particle, where it is not possible to define the area under a Kratky plot.

C. A comparison of the solution scattering from SAXS analysis of the C domain (blue line) and the best fit of the atomic resolution crystallographic data (red line); χ^2= 6.1 from CRYSOL.

D. Real space *I*(0), Rg and estimation of the maximum dimension of the particle, d_{max}, is made from indirect Fourier-transform of the intensity data into the pair-distance distribution function, P(r). The d_{max} is estimated at 95Å. The 3D envelope was not constructed as it would not be accurate because of the likely monomer dimer equilibrium.

E. The experimental scattering data fit well to the calculated scattered curve predicted from the P(r) distribution.

4.3.4 Comparison of flexibility and Kratky plots between full length, N and C domains of hHKII

The Kratky plot is the plot of the scattering pattern as $q^2I(q)$ versus q and it is used to identify unfolded samples. Ideal globular proteins follow the Porod law and the scattering intensity show a decrease as q^{-4} at higher q. This results in a bell shaped curve with a pronounced maximum at low angles (q). Two limitations occur from this approach: 1) the inability to compare molecules with different molecular mass since the scattering intensity will be influenced from different size samples and 2) it is hard to make a decision whether a protein is folded or partially unfolded when the protein contains a high amount of large sized structured regions. To overcome this problem the dimensionless Kratky plot was developed were I(q) is normalized to the *I(0)* and q is normalized to the radius of gyration (plots $I(q)/I(0)\times(q\times Rg)^2$/ qRg. This way the angular scale is independent of the molecular mass and size of the protein. The dimensionless Kratky plot on the other hand has a limitation occurring from the need of an accurate determination of the Guinier region (Rambo and Tainer, 2011). Globular proteins consistently exhibit a maximum value of 1.104 for qRg= √3

(indicated by grey lines on Figure 4.5). On the other hand, for completely unfolded polypeptide chains the plot does not have a bell-shape as the curve keeps rising and plateaus at a region between 1.5 and 2 qRg. In our case all dimensionless Kratky plots (Rg based on Guinier analysis) were overlaid and compared to each other.

As shown in Figure 4.5 C and N domain have the characteristic parabolic shape with a similar maximum around 1.2 at qRg around 2. This indicates that proteins do not scatter very differently from a globular molecule. C domain shows a slight difference in shape when qRg>3 but does not clearly indicate a flexible or a completely unfolded particle. The C domain deviates from ideality slightly more than the N domain. Increases in $I(q)/I(0)\times(q\times Rg)^2$ when qRg>6 is likely due to poor buffer subtraction rather than a property of the particle. The full length protein has a broad peak around the maximum (~1.3) which is suggestive of a multidomain protein with flexible linkers.

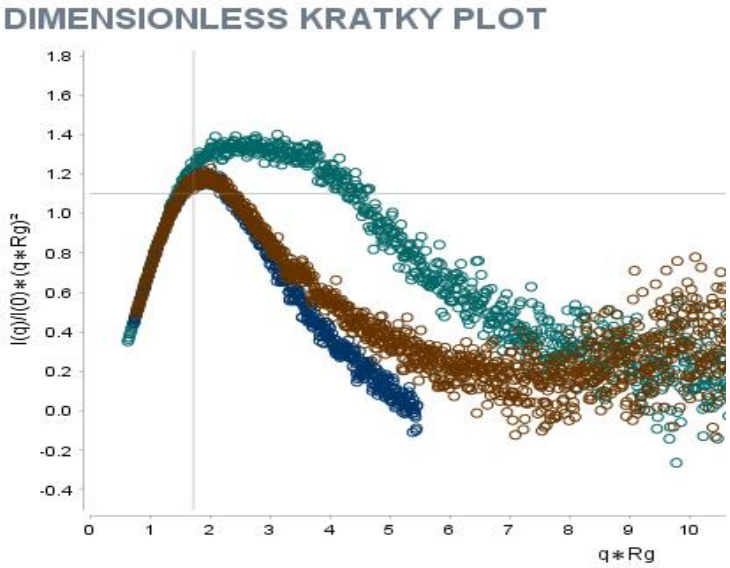


Figure 4.5: Dimensionless Kratky plots between all constructs of hHKII

The dimensionless Kratky plot is based on the Rg as estimated from the Guinier analysis. The grey lines indicate the maxima at which any ideal globular protein should lie. Green plot belongs to full length hHKII, blue to N domain and brown to C domain.

Due the difficulty of definitively determining flexibility from Kratky plot analysis, Porod-Debye plots were examined to identify flexibility. The Porod-Debye law is an approximation that describes a linear relationship between q and I(q) for a small

range of q just larger than the Guinier limit of q. In a Porod-Debye plot $q^4I(q)$ is plotted against q or q^4. Application of the Porod-Debye law should result in a Porod plateau at a value of q just outside the Guinier region for "rigid" particles that have a well-defined electron density contrast between the protein and the solvent. Flexible proteins do not have a well-defined contrast due to the many conformations of the ensemble (Rambo and Tainer, 2011). The Porod plateau is most easily observed in the plot of $q^4I(q)$ vs. q^4 (Figure 4.6). Unstructured particles exhibit a plateau when plotted as $q^2I(q)$ versus q^2 (Kratky-Debye plot) and partially flexible when plotted as $q^3I(q)$ versus q^3 (SIBYLS plot).

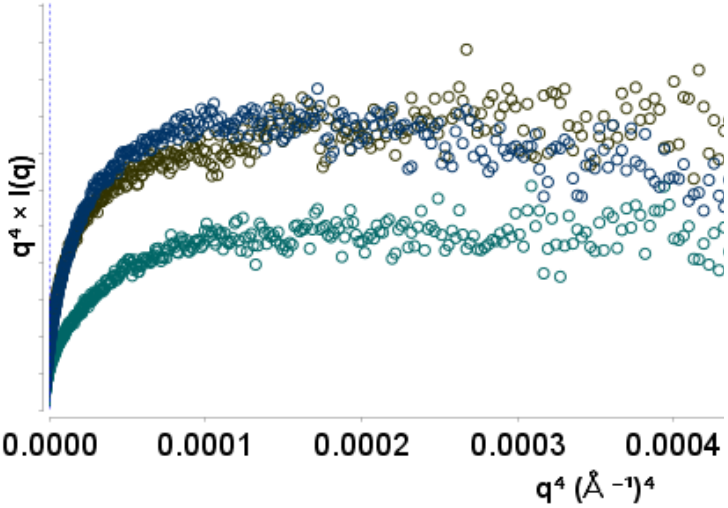


Figure 4.6: Porod-Debye plots for all constructs of hHKII

Porod-Debye plots show characteristic plateau; the intensity decay follows q^{-4}. Green plot corresponds to full length hHKII, blue plot to N domain and brown plot to C domain. This suggests that each construct can be characterized as well-folded proteins in solution.

4.4 Summary and conclusions

The major aim of the present work was to gain structural insights of the full length hHKII and each domain separately (N and C domains). Although full length hHKII is folded, active and homogenous I could not crystallise it and SAXS was sought as an alternative method to obtain structural information on hHKII. SAXS data were collected for a concentration series of all three constructs.

The Rg of the full length hHKII is 42.3 Å, of N domain is 29.4 Å and for C domain is 31 Å. The average Rg from the crystallographic data is estimated at 40.6 Å, 25.9 Å

and 27 Å respectively for monomers. Both methods are essentially in agreement. The Rg for N and C domains is smaller than for the full length protein as anticipated and similar to each other. The average d_{max} of full length hHKII, based on SAXS, is 145 Å, 96 Å for N domain and 109 Å for C domain, compared to 137 Å, 88 Å and 75 Å based on the X-ray structure. Once again, the SAXS results are not far away from those calculated from crystallographic data for the full-length and N-domain constructs. The N domain is estimated to be 96 Å instead of 88 Å but the construct is not identical. The SAXS analysis was performed on a construct that has the 6His tag attached with a TEV cleavable site (15 residues). These residues could account for the longer d_{max}. C domain is the one that shows a significant larger d_{max} possibly because of the complexity of the system.

The molecular mass estimation for full length hHKII was very close to the theoretical mass (102 kDa) of the protein by using the Porod analysis which is concentration independent. Based on Guinier analysis the predicted mass is smaller (75-80 kDa) probably because the measured concentration with UV spectrometry was an overestimation. The mass prediction of N domain using the Guinier analysis and Pr distribution is in agreement, thus smaller than the size of a monomer. The UV concentration measurement was not as accurate. Using the Porod analysis the concentration is closer to a monomer which suggests that the assembly of the N domain in solution is monomeric. The results agree with other biophysical analysis of the N domain.

SAXS data for C domain were analysed as the Guinier plots were indicative of a non-aggregated and soluble sample, well suitable for SAXS analysis. The Rg is in good agreement between Guinier analysis and distance distribution function. On the contrary the mass estimation was not consistent. Based on Guinier analysis it is predicted to be a monomer, based on Porod analysis the mass is an average of monomer and dimer. As C domain was predicted to be an equilibrium of monomer:dimer in solution based on SEC-MALS, the system flexibility might interfere with an accurate mass estimation using the above SAXS methods. For this reason the volume-of-correlation (Vc) was added to the analysis. Vc is expected to estimate a more accurate mass for a flexible/ intrinsically unstructured system and it is also concentration independent. The predicted mass this way was not consistent

across the different concentrations of the C domain. The C domain data could not be exploited to create a 3D envelope as the analysis of the SAXS data were not clear and not as expected thus the generated model would not be correct.

To further compare the three constructs the dimensionless Kratky plots were analysed. As shown in Figure 4.5 C and N domain have the characteristic bell-shaped curve with a similar maximum close to a maximum observed for globular molecules. C domain shows an increase at far qRg which is not likely to indicate a flexible domain. It is more likely to represent a poor subtraction of buffer since the resolution is low at the region qRg>6. The rather flat topped Kratky plot for full-length hHKII is interesting as it is suggestive of a multi-domain protein with some flexibility. However, the Porod-Debye plot for hHKII has a clear Porod plateau which is evidence of a well-defined protein-solvent boundary and a rigid particle. It would be tempting to hypothesize that the hHKII has overall well folded domains with some pivot points that may facilitate enzymatic action.

The 3D envelope was constructed for full length hHKII and N domain hHKII as they are well characterised and are monodisperse species in solution. The crystal structures were manually docked to the envelopes using PyMOL. The manual fits of model-crystal structure were close to crystallographic shape of the proteins. The best fit of experimental data to crystallographic data was seen for full length protein. The 3D-envelope of N domain has a guitar shape and the neck correlates well with the long helix that connects the two domains to each other in the atomic resolution structure.

Taken together, these data would indicate that full length and the N-terminal domain construct of hHKII are monomeric in solution and do not differ much in overall architecture from the crystallographic atomic resolution structures. The C-terminal domain of hHKII is likely to exist in equilibrium between monomeric and dimeric form.

5. Chapter 5: Screening for hits against human hexokinase II

5.1 Introduction

The discovery of a new drug is a long and expensive process. It takes an average of 10-15 years for a drug to reach the market and the cost is estimated between US$800 million up to US$1.8 billion (Macalino et al., 2015). The main technique for the discovery of new lead compounds is the physical screening of large libraries of chemical compounds against a target molecule (High Throughput Screening- HTS) (Shoichet, 2004). In the early 1990s the development of combinatorial chemistry and HTS technologies, which enabled the screening of huge libraries in less time, spread the hope for an accelerated drug discovery process (Lavecchia and Giovanni, 2013). Table 5.1 shows some examples of recently approved drugs, derived from HTS process.

Table 5.1: Examples of recently approved drugs through HTS process (Macarron et al., 2011)

Name	***Target class***	***Disease***	***Year of FDA approval***	***Pharmaceutical company***
Tipranavir	Protease	HIV	2005	Boehringer Ingelheim
Sitagliptin	Protease	Diabetes	2006	MSD
Dasatinib	Tyrosine kinase	Cancer	2006	Bristol-Myers Squibb
Maraviroc	Protease	GPCR	2007	Pfizer
Lapatinib	Tyrosine kinase	Cancer	2007	GlaxoSmithKline

Name	***Target class***	***Disease***	***Year of FDA approval***	***Pharmaceutical company***
Ambrisentan	GPCR	Pulmonary hypertension	2007	Gilead
Etravirine	HIC	Reverse transcriptase	2008	Tibotec Pharmaceuticals
Tolvaptan	GPCR	Hyponatraemia	2009	Otsuka Pharmaceutical
Eltrombopag	Cytokine receptor	Thrombocytopaenia	2008	GlaxoSmithKline

Overall hit rates are often low as many hits fail in the lead optimisation process due to absorption, distribution, metabolism, excretion and toxicity deficiencies (ADMET/Tox). These issues became the reason for an alternative method to be found, which would be cost-efficient and would limit the hits identified with unsuitable properties (Lavecchia and Giovanni, 2013).

Computer-aided drug discovery (CADD) techniques have been used as an alternative and complementary approach to drug discovery. These *in silico* approaches have been developed by a number of research groups and pharmaceutical companies in order to speed up the discovery of potent lead compounds and also minimise the chance of failure in a later stage. It is important to highlight that rational drug design using CADD, studies the interaction of the complex between a protein and a ligand and makes use of that structural knowledge to design more potent lead compounds. HTS on the other hand, requires no *a priori* knowledge of the binding mechanism of the drug on the protein (Macalino et al., 2015).

In general, computer-aided approaches are categorised into ligand-based and structure-based methods.

Ligand and Structure based approaches

Ligand based drug design: In cases where the three-dimensional (3D) structure of a protein is not available, ligand-based approaches utilise structure−activity data from already known actives to discover new candidate compounds with similar properties for experimental evaluation (Scior et al., 2012). It is thought that compounds with similar structure will interact in a similar manner with the target protein (Macalino et al., 2015). Common ligand-based design techniques are quantitative structure-activity relationship (QSARs) and pharmacophore based methods. QSAR modelling aims to predict a correlation between the physicochemical and structural properties of a ligand and its potency. Ligand-based pharmacophores make use of the known biological activities of different (structurally and functionally) ligands, to create a model with the essential atom groups which need to be present for the binding to the target protein (Drwal and Griffith, 2013).

Structure based drug design (SBDD), the use of 3D structural information gathered from biological targets, is a scientific area that has received a lot of attention with many successful applications in recent years. At the beginning of the 1990s the first reviews were published, where the X-ray structure of HIV-1 protease was taken into consideration for the design of inhibitors (Erickson et al., 1990). However, with the completion of the human genome project and the fast development of technology, especially with the development of faster computers, this field has now more opportunities for a successful discovery of drug leads. The development in X-Ray detectors and the exceptionally powerful synchrotron X-ray sources also make possible the determination of many more protein structures (Anderson, 2003). The structure-based drug design process results in the discovery of a lead, i.e a compound that binds specifically to the target protein with at least micromolar affinity. It can often be toxic or unstable, thus it serves as a first step followed by other optimization steps until a low nanomolar drug is discovered (Verlinde and Hol, 1994). A diagram containing the main steps of the SBDD, from the computational screening to the clinical trials, is shown below.

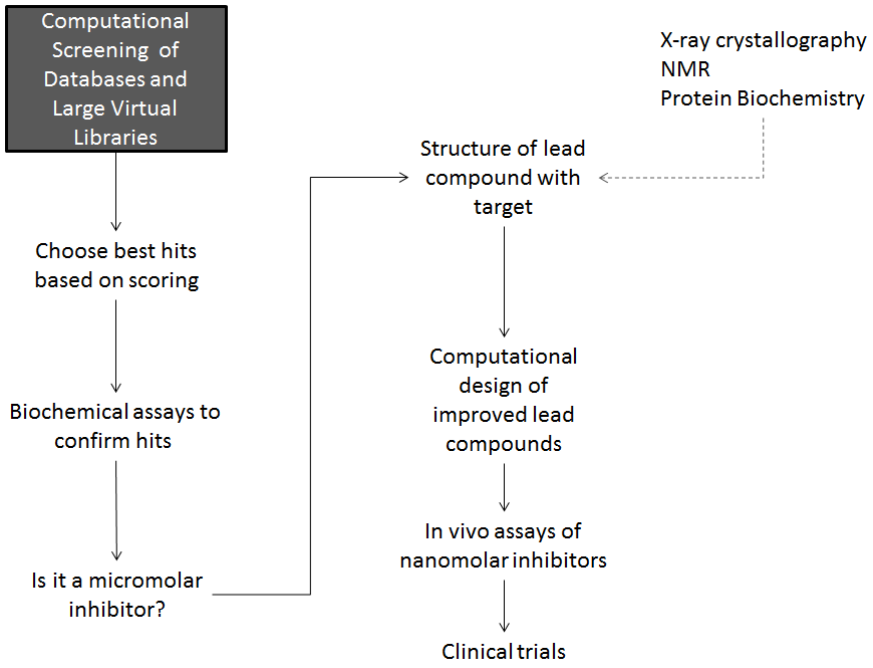


Figure 5.1: Steps performed during a typical SBDD project

Starting from a known target structure virtual screening takes place to identify ligands. Best hits are purchased and tested for affinity and potency. Ideally a structure of the complex receptor-ligand (at least micromolar inhibition shown) should be determined. Analysis of the structure provides helpful insight on the key intermolecular interactions. These are taken into consideration for the computational design of improved lead compounds which are then tested again. In vivo assays are implemented for the highly potent inhibitors and if they are in the nanomolar range clinical trials can be followed.

Structure-based virtual screening (SBVS) utilises the 3D structure of the biological target to dock millions of compounds from a virtual library to the desired site of the macromolecule. Scoring algorithms predict the binding affinity with this site and rank the compounds (Scior et al., 2012). In general SBVS consists of four steps:

i. Molecular target selection/preparation. Accurate structural information is very important. Crystal structures are widely used for structure-based drug discovery but they have to be evaluated for the resolution, reliability, or R factors, coordinate error, temperature factors and chemical “correctness”. The

big advantage of crystal structures is that the water molecules are visible which is useful for the process (Anderson, 2003).

ii. Compound database selection. The selection of a virtual library, among many freely accessible databases of commercial compounds, for the high-throughput virtual screening (Jorge Moura Barbosa and Del Rio, 2012).

iii. Molecular docking. Molecular docking programs aim to identify the most likely binding conformation of a small ligand within a specified binding site in the protein. These programs use specific scoring functions to estimate the binding energetics of the formed complex between the ligand and the receptor. (Ferreira et al., 2015).

iv. Post-docking analysis. The VS process results in a long list of compounds (hundreds of thousands or millions) and a visual analysis is conducted to prioritise hits based on the desired criteria, for example if the ligand makes the predetermined interactions with the target protein (Ferreira et al., 2015).

5.2 Materials and methods for SBVS of hHKII

The structure-based virtual screening was performed by Dr. Douglas Houston. The following diagram summarises the steps used for the identification of virtual hits for hHKII. The structure templates, the compound database, the docking programs and the scoring functions used for this purpose are described here. Moreover, the docking results and the compound selection are explained in detail in this chapter.

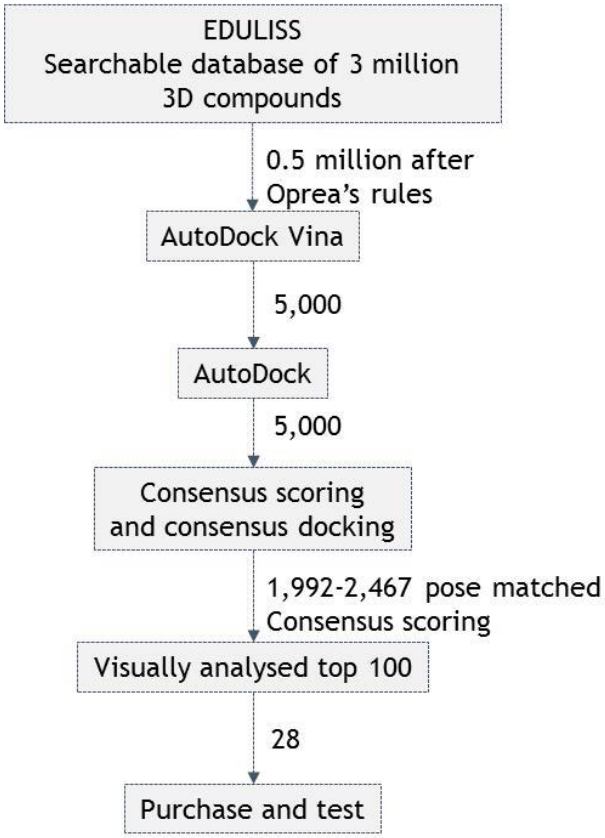


Figure 5.2: Diagram of virtual screening.

EDULISS, the database containing ~3 million compounds, was filtered according to the Oprea's "reduced complexity" rules. 0.5 m compounds were docked with AutoDock Vina and the top 5,000 based on Vina score were docked with AutoDock. The compounds with the agreed or without an agreed binding mode from both programs were then scored applying multiple scoring algorithms. The top hundred compounds were visually analysed to decide which are going to be tested. The number of compounds in a step can be different for different sites docked.

5.3 EDULISS Database

EDULISS (Edinburgh University Ligand Selection System) is a relational database for data mining small molecules. The database comprises of 3 million commercially available compounds from 28 suppliers (Hsin et al., 2011). For each compound a single 3D and 2D coordinates are stored along with over 1600 topological, geometrical, physicochemical and toxicological descriptors per compound. Also, the

vast majority of the compounds fit the Lipinski's rule of five and many compounds fulfill the Oprea lead-like criteria (Hsin et al., 2011). Lipinski's rule of 5 is a rule that is used from the pharmaceutical companies to evaluate a drug. According to Lipinski's rule of 5, (Lipinski et al., 1997) a drug is more likely to have poor absorption or permeation when:

- There are more than 5 hydrogen bond donors (expressed as the sum of OHs and NHs)
- The molecular weight is over 500
- The Log P is over 5
- There are more than 10 hydrogen bond acceptors (expressed as the sum of Ns and Os)

The more stringent Oprea-criteria for a promising lead-like compound (Hann and Oprea, 2004) are the following:

- The molecular weight should not be over 460
- The number of rotatable bonds should be less than 10
- The calculated Log P should be between −4 and 4.2
- The number of hydrogen bond acceptors should be ≤9
- The number of hydrogen bond donors should be ≤5
- The number of rings should be ≤4.

However, one should never forget that a compound with the above characteristics might not necessarily make it to the final stage of drug discovery as it may be toxic, teratogenic, be metabolised quickly, unable to reach the target macromolecule in the right concentration, too difficult to synthesise and/or too expensive. There is no way to estimate what changes it will provoke regarding the metabolic, transport and signaling pathways once administered to the human body (Verlinde and Hol, 1994).

The EDULISS library was initially filtered using the software "Filter-it" (http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/software.html). The number of compounds that proceeded to the docking routine was reduced to 0.5 million. This command-line program filters molecules with unwanted properties and comply with the Oprea's "reduced complexity" rules (Hann and Oprea, 2004).

5.4 AutoDock 4.0 and Vina

Virtual screening, being one of the common strategies for the identification of new lead compounds, relies on a receptor-based computational docking of libraries containing compounds (Shoichet, 2004). The importance of an accurate docking tool for this purpose is clear. In general protein-ligand docking programs consist of two essential components, sampling and scoring. Sampling refers to the generation of putative ligand binding orientations/conformations near a binding site of a protein. The scoring function is used to predict the binding tightness for individual ligand orientations/conformations with a physical or empirical energy function. The lowest energy score indicates the likely best orientation/conformation of a ligand, referred to as the binding mode (Huang and Zou, 2010). Two methods are known for the automated docking: The matching and the docking simulation methods (Rosenfeld et al., 1995). Matching methods try to dock the ligand, as a rigid-body, by matching its geometry to the active site. Docking simulation methods allow flexibility within the ligand combined with more advanced molecular mechanics to calculate the binding energies. (Morris et al., 1998). AutoDock 4 and AutoDock Vina belong in this group (Morris et al., 2009, Trott and Olson, 2010).

AutoDock 4 calculates the interaction energy between a ligand and a macromolecule in a grid-based method, where the target protein is embedded in a grid and different atom types of a ligand are placed at each grid point while computing the interaction energy. This grid of energies is stored and used during the docking simulation. Autogrid is the program that creates the grid points in the interaction map by assigning different atom types around the binding site of the macromolecule.

AutoDock 4 uses the Lamarckian genetic algorithm (LGA) and the semi empirical free energy force field scoring to predict binding free energies of the ligands to the protein (Morris et al., 2009). LGA is a hybrid of the genetic algorithm (GA) method and the local search (LS) method, described in detail from Morris et al, 1998 (Morris et al., 1998). This algorithm overcomes the docking obstacles when more degrees of freedom are involved in the process.

The semi empirical free energy force field uses an improved thermodynamic model for the binding process as it includes the intramolecular energies of the unbound

structure, resulting in a more accurate prediction of the free energy. A full desolvation model that includes both favorable and unfavorable energetics and a model to predict the proper alignment of groups with multiple hydrogen bonds are also included (Huey et al., 2007). AutoDock Vina has an optimised algorithm that takes advantage of the local optimisation method that calculates derivatives to generate a gradient speeding up the optimization significantly. AutoDock Vina also calculates the grid maps internally to further quicken the procedure. Finally, the runs can be performed at the same time by using multithreading. As a result, AutoDock Vina had improved the speed by two orders of magnitude compared to AutoDock 4 while improving the accuracy of the predictive binding modes (Trott and Olson, 2010).

Both programs were compared during a virtual screening for the identification of actives for HIV protease using two different chemical libraries. The library which contained molecules with low molecular weight and few rotatable bonds was screened against both programs and gave similar results with significant level of accuracy. However, Vina was capable of preferentially ranking active compounds in the virtual screen of another library which consisted of larger molecules, with more rotatable bonds, while AutoDock 4 failed to do so (Chang et al., 2010).

5.5 Consensus Docking

For our studies virtual screening was improved using a relatively easy method named consensus docking (Houston and Walkinshaw, 2013). Houston with their work showed that, when combining more than one docking program and using only the docked compounds that are in the same location with the same orientation and conformation (referred to as binding pose), then the accuracy is improved. For this conclusion, a subset of 228 protein-ligand crystal complexes from the PDBbind-CN database, which is supplemented with experimental data, was used. More specifically, combining AutoDock and Vina poses and excluding the ones that were not close to each other, the correctly docked poses (i.e. the RMSD between the docked pose and crystallographic pose was less than 2 Å) were 82%, instead of 64% for the best docking program (55% for the other) (Houston and Walkinshaw, 2013).

In our case Vina was used first, as it is faster than Autodock, and the top 5,000 compounds from this were docked with AutoDock and the binding poses examined.

5.6 Scoring functions and consensus scoring

The scoring function of a protein-ligand docking program predicts the binding energy for individual ligand orientations/conformations aiming to differentiate the preferential binding mode, i.e the one with the lowest energy score. Many scoring functions have been developed which can be divided into three main categories according to their method of derivation: force-field (FF), empirical and knowledge-based scoring functions (Huang and Zou, 2010). For our studies 5 different scoring functions were selected to rank the virtual hits: DrugScore, X-score, NNScore 1.0, NNScore 2.0 and RFscore 4.

DrugScore is a knowledge-based scoring function that consists of distance dependent pair potentials with novel torsion angle potentials and a newly developed potentials for the estimation of solvent accessible surface (Neudert and Klebe, 2011).

X-Score is an empirical consensus scoring function which consists of three different scoring algorithms, each of which has five adjustable terms: atom classification, van der waals interaction, hydrogen bonding, deformation penalty and hydrophobic effect. These 5 adjustable parameters affect the overall free energy change when a protein-ligand complex is formed according to Wang et al. A larger training set (200 protein-ligand complexes) has been used to calibrate these, thus an error of 2 kcal/ mol in the estimated binding free energy was shown (Wang et al., 2002).

RF-Score function performs under a machine learning approach. Compared to scoring functions that generate under a rigid set of parameters which ultimately will fail to conform for all protein complexes, RF-Score does not account for any a priori relationship between the complex components and the binding data. Thus it is more flexible to estimate a more accurate prediction within the big diversity of protein-ligand complexes. RF-Score has been shown a valuable tool especially as a re-scoring function (Ballester and Mitchell, 2010).

NNScore is a neural network based scoring function that can be used to re-score the docked poses of potential hits. A neural network, that is designed to mimic the microscopic organization of the brain, does not need specific formulas describing the relationship that governs the components studied by the network. Knowing that, Durrant & McCammon (2010), while designing NNScore, had only to define the properties of a ligand protein complex that affect the binding affinity and allow for the system itself to find the relationship, analyse and finally characterise the complex. Although the network succeeded to discriminate between well-docked and poorly-docked ligands as well as true ligands from decoy compounds, the designers suggest its use in combination with more traditional scoring functions (Durrant and McCammon, 2010). NNScore 2.0 is an updated version of NNScore neural network, supplemented with a much greater number of binding characterizations (Durrant and McCammon, 2011).

As consensus docking increases the accuracy of the predictive binding mode, consensus scoring, which combines more than one scoring function for the prediction of the binding mode, leads to an improved prediction of hits (Charifson et al., 1999).

5.7 Final choice of the compounds for the bioassays

Consensus docking and consensus scoring were combined to rank the virtual hits. Consensus docking has been shown to be slightly biased towards molecules that have a higher score (Houston and Walkinshaw, 2013). In the present study ~50% of the top hundreds hits are molecules showing a different binding mode but having a lower score for binding affinity. The other ~50% arises from molecules that are predicted to have the same binding mode (RMSD with a cutoff of 2 Å) and the best score.

For the final ranking of all the candidates the "rank-by-rank" strategy was followed (Wang and Wang, 2001), where we determined the rank position of a candidate from all scoring functions and then the average rank was used, i.e if a hit ranks 10^{th} with DrugScore and 20^{th} with X-Score then the average rank will be (10+20)/2=15. The consensus scoring was more sophisticated in our studies as more scoring functions were added to our schemes. As will be explained in detail later we performed several docking studies and different scoring functions were used (or updated versions at the

time of the run). The specific scoring functions for each one will be discussed at the relevant section. The following diagram summarises the steps undertaken, starting from Vina docking until the final ranking of all hits.

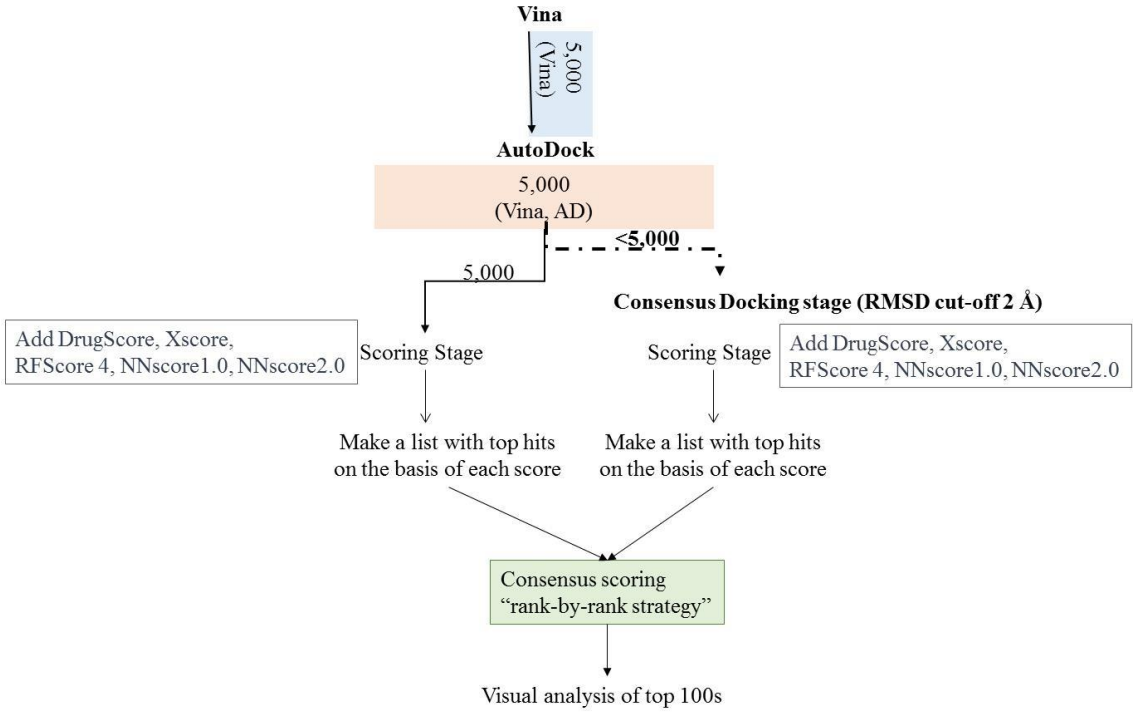


Figure 5.3: Scoring and ranking the virtual hits

The diagram summarises the steps taken from the initiation of the docking studies. Solid lines show the same steps, dashed line shows the difference between the two ranking schemes. 0.5 m of compounds were docked with Vina and the top 5,000 (scored by Vina) proceeded to AutoDock, which added a new score for the predictive binding mode (AD). The 5,000 candidates were further analysed with the scoring functions (X-score 1.2 and 1.3, DrugScore, NNScore1.0, NNScore2.0 and RF-score) to estimate the binding affinity on the target receptor. At the same time only the candidates docked in the same way with both docking software (Vina, AD) were scored using the same scoring algorithms and ranked accordingly. The consensus docking is performed before the scoring stages in order to gain a more accurate prediction of binding affinity due to the fact that the binding mode has more chances to be correct (Houston and Walkinshaw, 2013). The consensus scoring is then performed for each list using the "rank-by-rank strategy" and the top hits are further analysed.

All the virtual hits from *in silico* screening were filtered using multiple techniques, including prioritisation of compounds that met Lipinski rules and Oprea criteria. In addition, the hit list was analysed by an experienced medicinal chemist (Dr Phill Cowley, Head of Chemistry, IOmet Pharma Ltd.) to remove compounds which were potentially reactive, non-specific in terms of their biological action or not drug-like. This allowed the focus to be on compounds that were more likely to be of use as lead compounds in the initiation of a drug discovery program.

5.8 Structure model templates

5.8.1 Structure choice for human hexokinase docking studies hHKII structure

As mentioned in Chapter 1, the crystal structure of hHKII has been determined from SGC in 2006, with Glc (substrate) and G6P (natural inhibitor) present (PDB code: 2NZT). As it was the only available structure at the PDB, the holo form was considered for the SBVS strategy. hHKII contains two domains, the N (1-475) and C (476-917) terminal domain, showing a significant level (~60%) of identity. In contrast to HKI and HKIII, both domains for HKII are functional with comparable catalytic activities (Tsai and Wilson, 1996), (Ardehali et al., 1996). As both catalytic sites are identical choice of domain should not make a difference in the docking process. However, since many studies have been performed on the catalytic C domain and the inactive N domain of isoforms I and III (Arora et al., 1993), (Tsai and Wilson, 1997), (White and Wilson, 1989), the docking was performed at the C domain. Figure 5.4 shows the monomer of hexokinase II consisting of two domains along with a closer-up representation of residues of C domain which interact with Glc and G6P respectively.

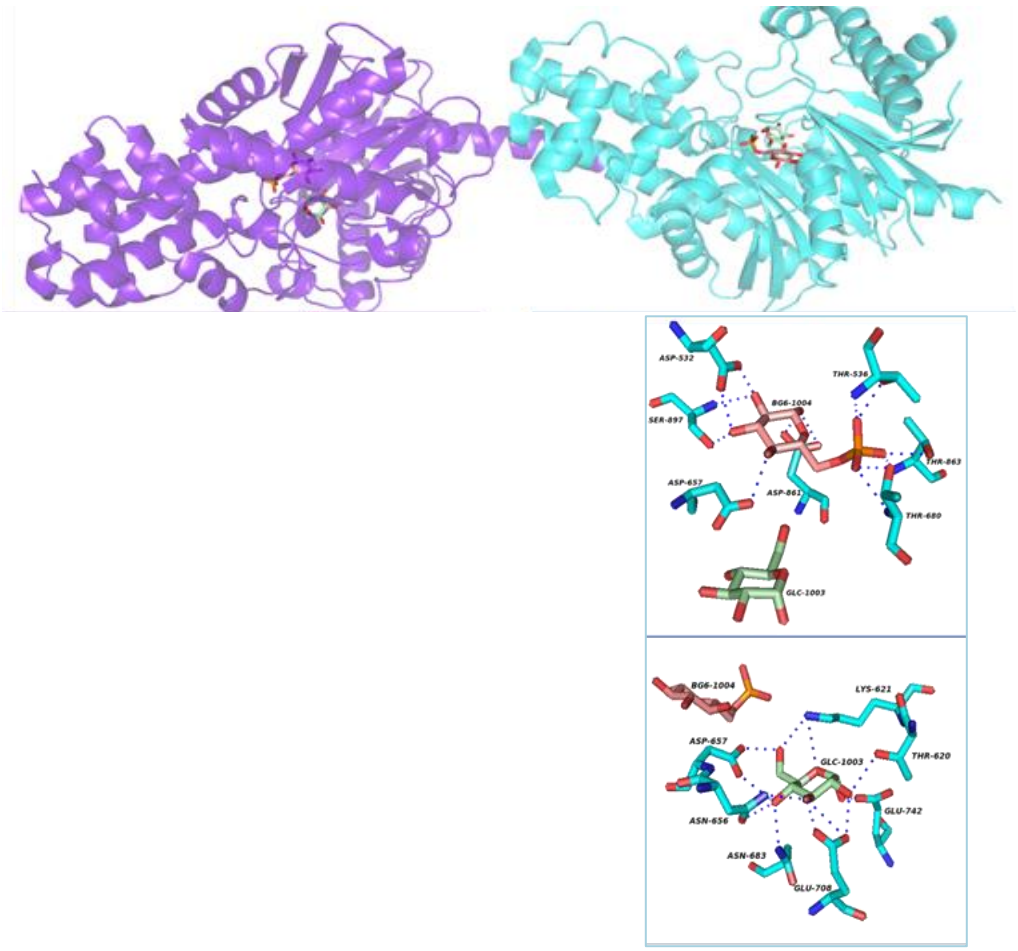


Figure 5.4: Monomer of hHKII and active site occupied by Glc and G6P (Glc-G6P site)

Cartoon representation of monomer hHKII (2NZT) consisting of two domains, the N (purpleblue) and C (cyan) domain. Both domains bind the ligands, Glc and G6P, which are shown as pale green and salmon sticks respectively. The residues interacting with hydrogen bonds with each ligand at the C domain are shown in the box. Residues interacting with Glc: T620, K621, N656, D657, N683, E708, E742. Residues interacting with G6P: D532, T536, D657, T680, D861, T863, S89.

A closer analysis of the C domain from the above structure shows that the active site is not in an optimal conformation for drug discovery. It forms a relatively closed conformation, allowing a restricted docking space which would not allow the accommodation of compounds of lead-like size. An analysis of the available crystal structures for human hexokinase was carried out, searching for another isoform that would serve as a better template while maintaining the same binding site.

hHKI structure

A structure from HKI (PDB code: 1DGK) was found to hold a more open conformation of the active site. The crystal structure of the quadruple mutant (Glu280→Ala, Arg283→Ala, Gly284→Tyr, and Thr536→Ala) is a monomer with one molecule of ADP and Glc bound to the C-terminal half and one molecule of Pi, Glc and ADP to the N terminal half (Aleshin et al., 2000). The monomer of hHKI is shown in cartoon representation in Figure 5.5.

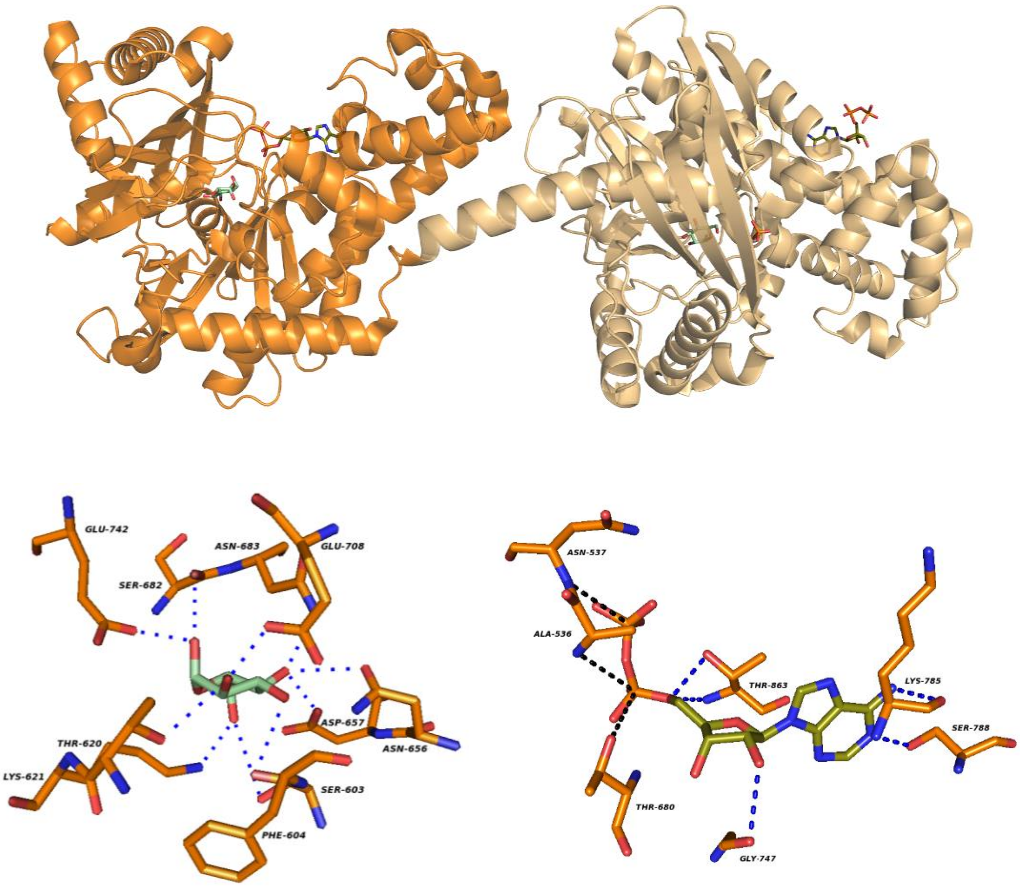


Figure 5.5: Monomer of mutant hHKI with Glc and ADP bound.

Cartoon representation of monomer hHKI (1DGK) consisting of two domains, the N (light orange) and C (orange) domain. Both domains bind the ligands, Glc and ADP, which are shown as pale green and deep olive sticks respectively. Glc interacts with S603, F604, T620, Lys621, N656, D657, S682, N683, E708, E742. Selected interactions from the ADP/Glc monomer complex of the C domain are shown in the box. ADP interacts with S788, Lys785,

G747, T863, T680, A536, N537. Blue dots indicate hydrogen bonds, black dots represent interactions with phosphate groups.

A grid surface representation (Fig. 5.6) of 1DGK and 2NZT reveals a surface cleft where ADP is binding only for 1DGK. 2NZT does not contain the open cavity formed by ADP binding.

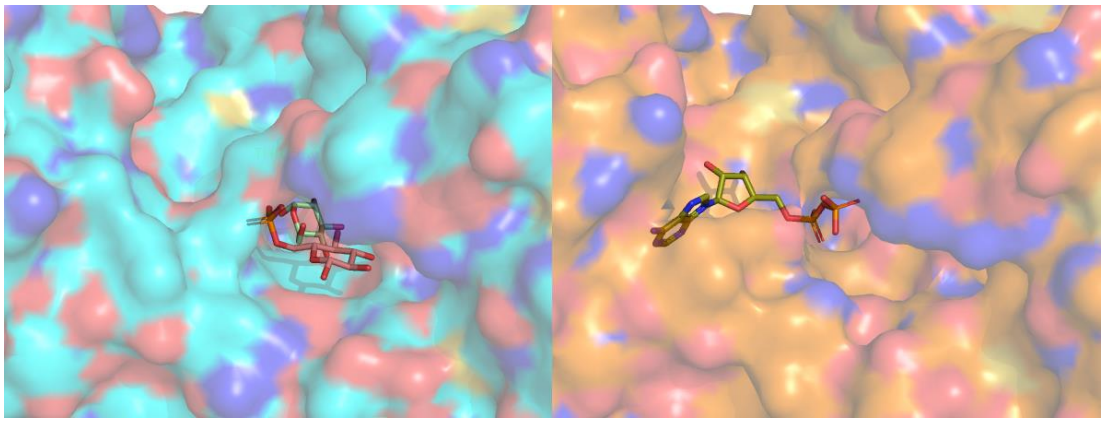


Figure 5.6: Surface grid representation of C domain of 1DGK and 2NZT

Overlay and grid representation of the C domain of 2NZT (cyan) and 1DGK (orange). ADP is bound on the outer surface of the 1DGK while 2NZT has Glc and G6P bound. The difference between the two surfaces can be seen as 2NZT adopts a closed conformation close to the ADP binding site.

An overlay of 1DGK and 2NZT shows that ADP binding on hexokinase I (1DGK) moves a loop and a helix in order to accommodate ADP. Side chains of T784, T536 and G535 from hexokinase II (2NZT) clash on ADP.

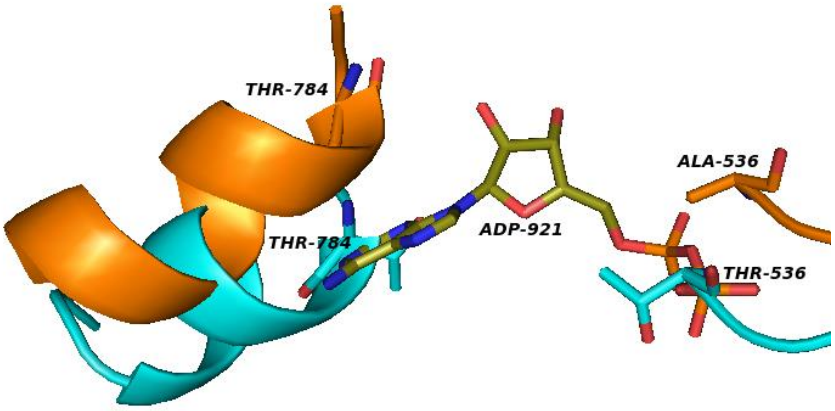


Figure 5.7: Overlay of 1DGK and 2NZT on ADP binding site.

ADP binding on 1DGK moves a helix so that T784 does not clash on ADP. Also a loop is moved so that A536 can interact with the phosphate group from ADP. In comparison with 2NZT, both helix and loop are clashing on ADP, as this structure is only a complex of Glc/G6P, and it adopts a different secondary structure. (2NZT has Thr instead of Ala at 536 as 1DGK is a mutant).

For this reason the use of 1DGK for the docking studies is preferable, since the docking space from this wider pocket would allow the identification of more ligands which will either bind to Glc/G6P binding site or ADP binding site. To make a decision whether the use of 1DGK is feasible we had to evaluate if the residues interacting with the Glc and G6P from 2NZT (0-4.0 Å distance) were the same with the same side chain orientation. By comparing the sequence of both isoforms the binding sites for Glc and G6P are completely conserved as well as the side chain orientation (Fig. 5.8).

```
1DGK:C_domain      AHFHLTKDMLLEVKKRMRAEMELGLRKQTHNNAVVKMLPSFVRRTPDGTENGDFLALDLG
2NZT:C_domain      EHLQLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGDFLALDLG

1DGK:C_domain      GANFRVLLVKIRSGKKRTVEMHNKIYAIPIEIMQGTGEELFDHIVSCISDFLDYMGIKGP
2NZT:C_domain      GTNFRVLLVRVRNGKWGGVEMHNKIYAIPQEVMHGTGDELFDHIVQCIADFLEYMGMKGV

1DGK:C_domain      RMPLGFTFSFPCQQTSLDAGILITWTKGFKATDCVGHDVVTLLRDAIKRREEFDLDVVAV
2NZT:C_domain      SLPLGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRREEFDLDVVAV

1DGK:C_domain      VNDTVGTMMTCAYEEPTCEVGLIVGTGSNACYMEEMKNVEMVEGDQGQMCINMEWGAFGD
2NZT:C_domain      VNDTVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCVNMEWGAFGD

1DGK:C_domain      NGCLDDIRTHYDRLVDEYSLNAGKQRYEKMISGMYLGEIVRNILIDFTKKGFLFRGQISE
2NZT:C_domain      NGCLDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRGLLFRGRISE

1DGK:C_domain      TLKTRGIFETKFLSQIESDRLALLQVRAILQQLGLNSTCDDSILVKTVCGVVSRRAAQLC
2NZT:C_domain      RLKTRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTVVARRAAQLC

1DGK:C_domain      GAGMAAVVDKIRENRGLDRLNVTVGVDGTLYKLHPHFSRIMHQTVKELSPKCNVSFLLSE
2NZT:C_domain      GAGMAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPKCDVSFLQSE

1DGK:C_domain      DGSGKGAALITAVGVRLRTEASS
2NZT:C_domain      DGSGKGAALITAVACRIREAGQ-
```

Figure 5.8: Sequence alignment between 2NZT and 1DGK (C domain).

The residues interacting with Glc and G6P are coloured grey and yellow respectively. Common residues for both ligands are shown as cyan. The active sites are completely conserved apart from one mismatch because 1DGK is a mutant (T536→A536).

An overlay of the structures 1DGK and 2NZT was performed to identify any possible side chain movement between the residues of the active sites. Figure 5.9 shows that all residues from both sites have the same orientation in isoforms I and II. 1DGK crystal structure forms a wider pocket and this template was used for the docking. A536 was mutated to T536 using PyMOL, in order to create an identical binding site for substrate and product as it is for 2NZT, hexokinase isoform II.

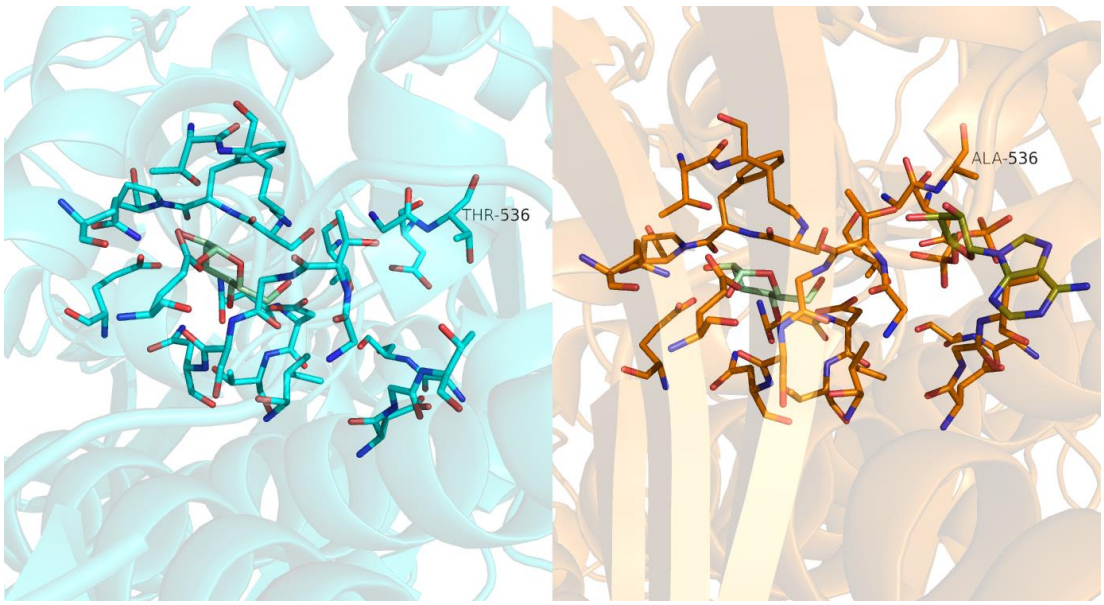


Figure 5.9: Superposition and grid representation of the Glc/G6P site between 2NZT and 1DGK.

The C domains of 2NZT and 1DGK are coloured cyan and orange respectively. Structures are overlaid and shown in a grid mode. The residues of the binding sites of Glc (pale green) and G6P (salmon) are conserved between the two isoforms. Only one different residue differs at position 536, as 1DGK is a mutant of isoform I. ADP, present only in 1DGK, is shown as deep olive stick.

In summary, the computer-simulated docking studies were performed using the widely distributed molecular docking software AutoDock and AutoDock Vina. The protein structure used for this was extracted from the PDB with the code 1DGK. The structure belongs to the hHKI which has an extensive sequence similarity to hHKII. The latter will be used for screening the selected virtual hits. The active site, where the substrate and product are bound in the crystal structure of isoform II, was the selected docking space. Figure 5.10 shows the surface of 1DGK highlighting the docking area. The docking area belongs within the black box.

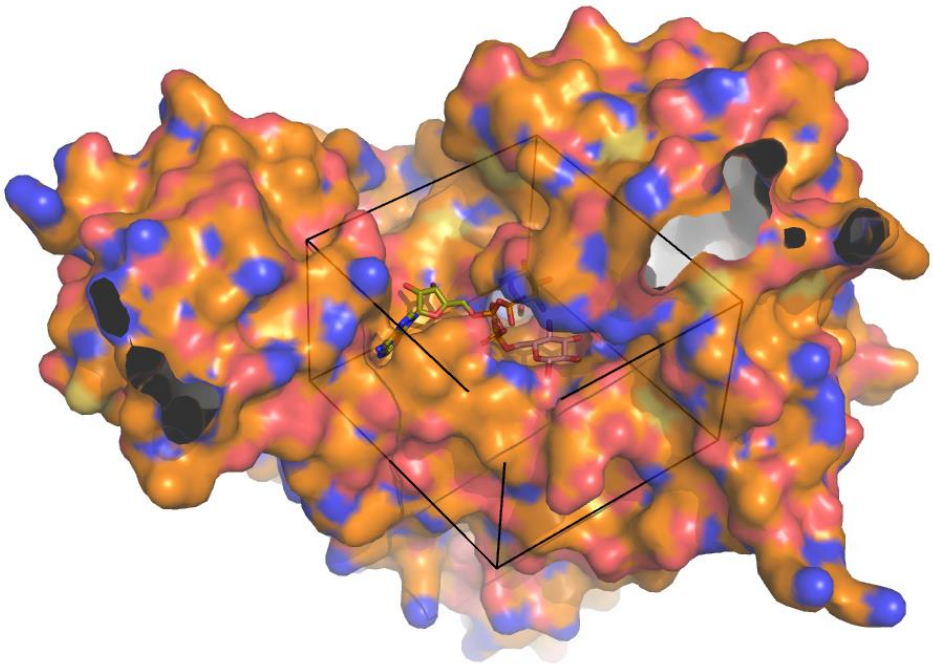


Figure 5.10: Surface of 1DGK highlighting the docking area of the Glc/G6P site.
C domain of 1DGK (orange), Glc (pale green stick) and G6P (salmon stick) are present in the 2NZT structure whereas ADP (deep olive stick) is present in the 1DGK structure. All ligands are added to the overlay of 1DGK-2NZT. The black box represents the limits of the docking area. As can be seen the docking area includes the binding sites of all natural substrates/products.

5.9 Docking studies on the ADP binding site of hHKI

The ADP binding site (1DGK) was also subjected to docking studies. The interactions between the protein and ADP in a distance from 0-4Å were analysed with CCP4 and the following residues were found to form contacts (hydrogen bonds and van der waals):

G535, A536, N537, G679, T680, G747, M748, T784, K785, S788, G862, T863, L864, L867.

An alignment with HKII shows that this site is highly conserved so it is worth docking for potential inhibitors.

```
1DGK:C_domain      AHFHLTKDMLLEVKKRMRAEMELGLRKQTHNNAVVKMLPSFVRRTPDGTENGDFLALDLG
2NZT:C_domain      EHLQLSHDQLLEVKRRMKVEMERGLSKETHASAPVKMLPTYVCATPDGTEKGDFLALDLG

1DGK:C_domain      GANFRVLLVKIRSGKKRTVEMHNKIYAIPIEIMQGTGEELFDHIVSCISDFLDYMGIKGP
2NZT:C_domain      GINFRVLLVRVRNGKWGGVEMHNKIYAIPQEVMHGTGDELFDHIVQCIADFLEYMGMKGV

1DGK:C_domain      RMPLGFTFSFPCQQTSLDAGILITWTKGFKATDCVGHDVVTLLRDAIKRREEFDLDVVAV
2NZT:C_domain      SLPLGFTFSFPCQQNSLDESILLKWTKGFKASGCEGEDVVTLLKEAIHRREEFDLDVVAV

1DGK:C_domain      VNDTVGTMMTCAYEEPTCEVGLIVGTGSNACYMEEMKNVEMVEGDQGQMCINMEWGAFGD
2NZT:C_domain      VNDTVGTMMTCGFEDPHCEVGLIVGTGSNACYMEEMRNVELVEGEEGRMCVNMEWGAFGD

1DGK:C_domain      NGCLDDIRTHYDRLVDEYSLNAGKQRYEKMISGMYLGEIVRNILIDFTKKGFLFRGQISE
2NZT:C_domain      NGCLDDFRTEFDVAVDELSLNPGKQRFEKMISGMYLGEIVRNILIDFTKRGLLFRGRISE

1DGK:C_domain      TLKTRGIFETKFLSQIESDRLALLQVRAILQQLGLNSTCDDSILVKTVCGVVSRRAAQLC
2NZT:C_domain      RLKTRGIFETKFLSQIESDCLALLQVRAILQHLGLESTCDDSIIVKEVCTVVARRAAQLC

1DGK:C_domain      GAGMAAVVDKIRENRGLDRLNVTVGVDGTLYKLHPHFSRIMHQTVKELSPKCNVSFLLSE
2NZT:C_domain      GAGMAAVVDRIRENRGLDALKVTVGVDGTLYKLHPHFAKVMHETVKDLAPKCDVSFLQSE

1DGK:C_domain      DGSGKGAALITAVGVRLRTEASS
2NZT:C_domain      DGSGKGAALITAVACRIREAGQ-
```

Figure 5.11: Alignment of the ADP binding site between 1DGK and 2NZT.

The residues interacting with ADP, as found from 1DGK crystal structure were compared to 2NZT. The results show that the sites are identical since all residues are conserved (one mismatch shown as red, occurs from 1DGK because is a mutant).

The new docking area is restricted to the ADP binding site. Figure 5.12 shows the new docking space. This new docking performance aims to identify compounds that would mimic the ADP binding to the target molecule.

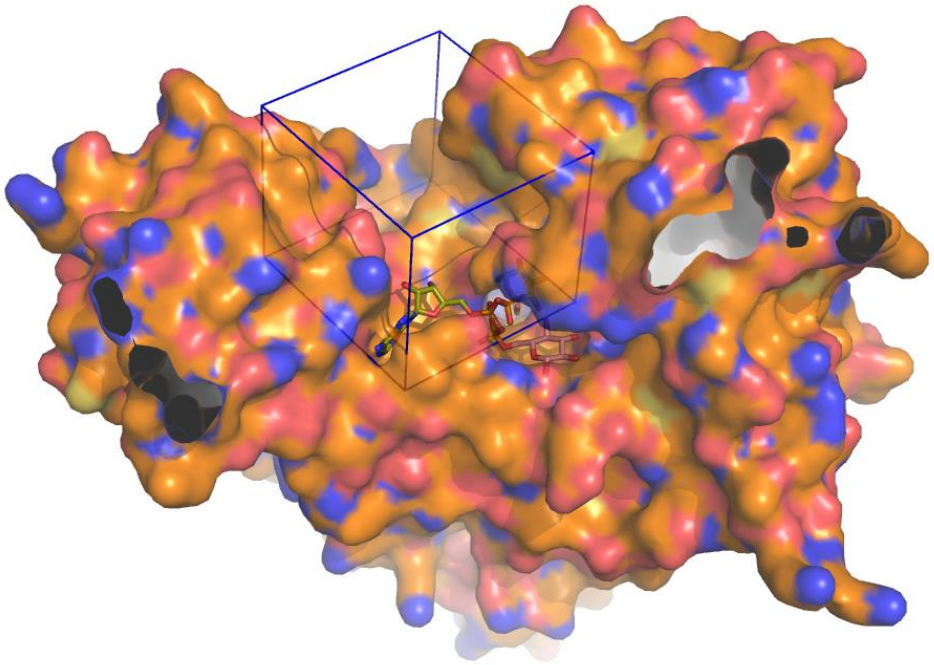


Figure 5.12: Surface of 1DGK highlighting the docking area of the ADP site.
C domain of 1DGK (orange), Glc (pale green stick) and G6P (salmon stick) are present in the 2NZT structure whereas ADP (deep olive stick) is present in the 1DGK structure. All ligands are added to the overlay of 1DGK-2NZT. The new docking study was restricted to the ADP binding site, as the blue box highlighting the docking space excludes the binding of Glc and G6P molecules.

5.10 Docking results for the Glc/G6P site of hHKI

As discussed in section 5.8 the Glc/G6P site of hHKI (Fig. 5.10) was docked using AutoDock Vina and AutoDock. From the initial list of ~3 million compounds, 5,000 were proceeded to the docking process. The consensus docking was performed and 2,467 compounds were found to have the same binding mode predicted from both programs. Two different scoring schemes were performed after the docking which will be analysed below.

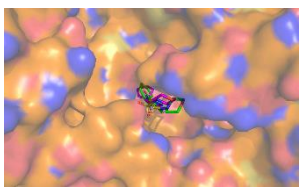
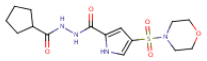
5.10.1 1st scoring scheme for ranking the docked compounds from the active site

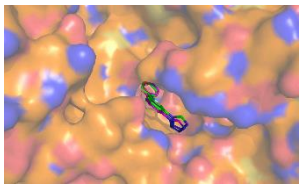
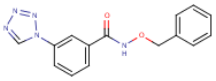
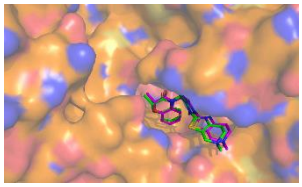
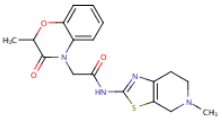
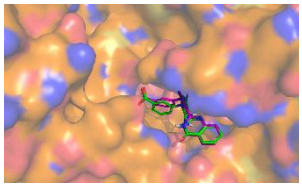
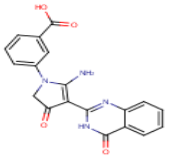
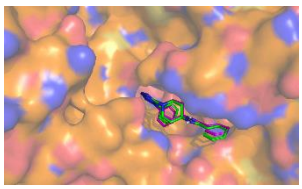
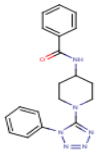
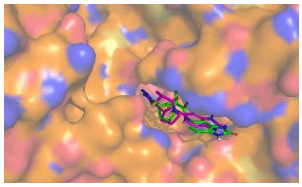
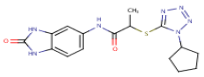
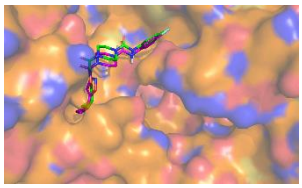
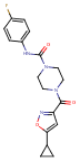
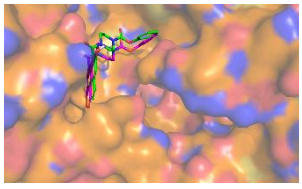
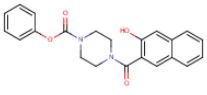
The scoring algorithms from each program were used, i.e DrugScore and X-score 1.2. A complete ranked list of 103 compounds was generated which was visually analysed. The calculated average AutoDock and AutoDock Vina scoring energies of these hits are between -6.12 to -12.1 kcal/mol and -6.7 to -10.4 kcal/mol respectively. Nine compounds were purchased from the first ranking method, named AS1RS**1**, AS1RS**2** etc. The MW for all compounds is <500, with an average of 360 and the average of cLogP is 1.77.

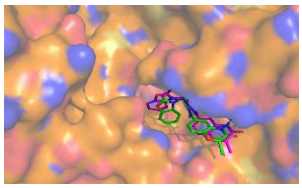
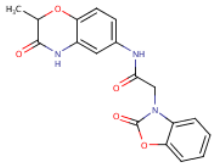
Table 5.2 shows the ligands including their binding modes and their chemical structure.

Table 5.2: Ligands with their docking poses as predicted by AutoDock and Vina in the active site of C domain of hHKI.

The surface of C domain of 1DGK is shown as orange. Ligands coloured in green (Vina) and magenta (Autodock) represent the predicted binding modes by each software. The ranking order is according to the consensus scoring from AD, Vina, DrugScore and X-Score as well as consensus docking. The chemical structure, Vina, AD, DSX and X-score for each compound are shown. The last column shows the percentage of inhibition/ activation at 100μM (final inhibitor concentration) based on the enzymatic assays performed in the present study (see Chapter 7).

Docked Poses from AD and Vina	*Chemical structure*	*Vina (kcal/ mol)*	*AD (kcal/ mol)*	*X-score*	*DSX*	*%I at 100μM*
AS1RS1 		-9.2	-10.15	6.61	-125.822	-4%

AS1RS2 		-9.3	-9.43	6.12	-160.431	-2%
AS1RS3 		-9.7	-8.06	5.99	-132.367	0%
AS1RS4 		-9.9	-7.87	6.28	-120.222	-15%
AS1RS5 		-9.9	-8.09	5.87	131.047	+9%
AS1RS6 		-9.6	-8.68	5.91	-122.967	-13%
AS1RS7 		-8.6	-7.89	6.27	-139.721	+9%
AS1RS8 		-7.8	-8.04	6.76	-127.733	-35%

AS1RS9 		-8.8	-7.43	5.92	-135.320	+1%

5.10.1.1 Predicted interactions between virtual hits and target molecule

The hydrogen bonds of each complex (protein-ligand) were identified using WinCoot. WinCoot cut-off distance measurements for hydrogen bonds were set up to a maximum of 3.3 Å. Table 5.3 summarises all the interactions identified from WinCoot. The two different docking programs show differences even though consensus docking has been performed. However, the number of the different residues is small for all compounds, with zero difference for compound AS1RS2 to a maximum of 4 differences observed for compound AS1RS6. The rest show 1-2 differences so overall agreement between the two programs is observed. The compounds AS1RS7-8 form contacts with different residues than from the other compounds. For these 2 compounds G747, E783, T784 and K785 participate in the interactions, which are not seen for any other compound. Also these 2 compounds are predicted to make fewer contacts with the target. Compound AS1RS5 also shows less contact. The remainder form, in average, six hydrogen bonds.

Table 5.3: List of residues that form hydrogen bonds with virtual hits from Glc-G6P site (1st ranking scheme).

The table shows all the residues that form hydrogen bonds (0- 3.3Å) with the template structure, as identified from WinCoot for the virtual hits of the Glc-G6P site (1st ranking scheme).

Active site 1st ranking scheme	AutoDock binding mode	Vina binding mode	Differences between Vina-AutoDock
AS1RS1	A536, N537, K621, D657, G681, T863	A536, N537, **R539**, K621, D657, G681, T863	Yes.
AS1RS2	K621, D657, T661, T680, G681, D861, S897	K621, D657, T661, T680, G681, D861, S897	No.
AS1RS3	**D532**, A536, **N537**, R539, D895	A536, R539, D895	Yes.
AS1RS4	D532, A536, **N537**, R539, T680, T863, K866	D532, A536, R539, T680, **D861**, T863, K866	Yes.
AS1RS5	A536A, **R539**, T680	**D532**, A536, T680, **T863**	Yes.
AS1RS6	**D532**, **N537**, R539, T680, **T863**, D895, **G898**	**A536**, R539, **N557**, T680, D895	Yes.
AS1RS7	T680, G747, T784	**M567**, T680, G747, T784, **K785**	Yes.
AS1RS8	**T680**, E783, **T784**	E783	Yes.
AS1RS9	D532, A536, **N537**, R539, N557, D895	D532, A536, R539, N557, D895	Yes.

On analysing the 1DGK/virtual hits interactions it can be seen that all compounds bind with similar orientation to the docking area apart from compounds 7 and 8. These compounds interact with the polar pocket T784/ E783, whereas the other ligands orient their side chains away from this polar pocket.

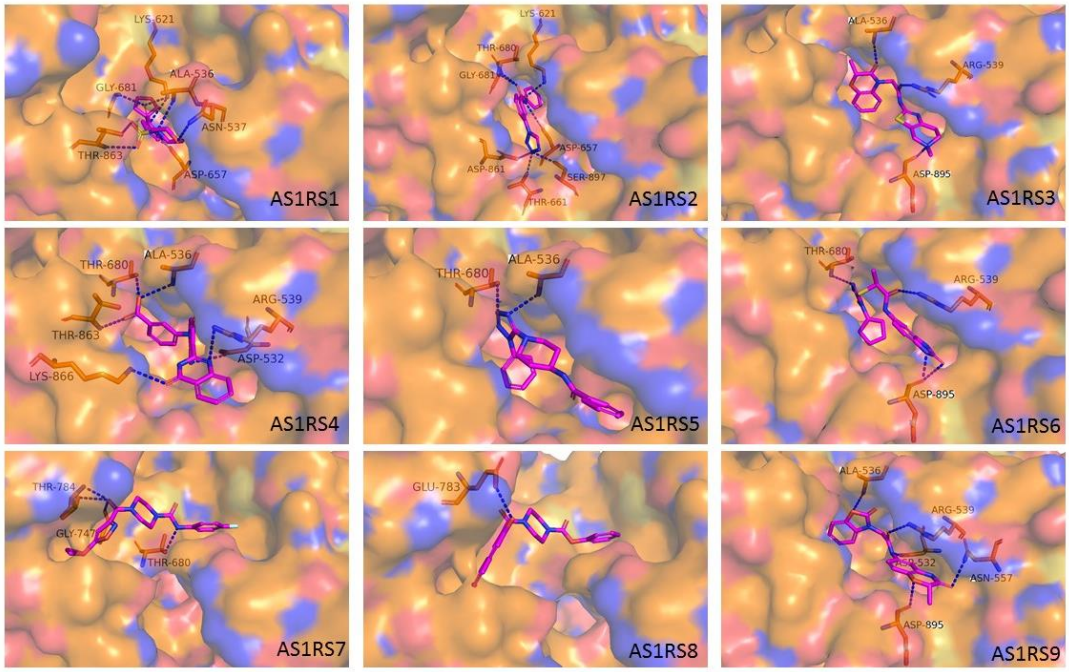


Figure 5.13: The purchased compounds docked in the active site of 1DGK with interacting residues.

WinCoot was used to identify all the hydrogen bonds with the surrounding residues in a maximum distance of 3.3 Å. The common residues (between AD and Vina orientations) are shown. The different contacts are not shown.

The ligands can be further categorised into three groups based on their predictive binding site:

1) Compounds AS1RS1 and AS1RS2 expand in both Glc and G6P binding sites
2) Compounds AS1RS3-6 and AS1RS9 fill the G6P and ADP binding site
3) Compounds AS1RS7-8 bind to the ADP binding site

Figure 5.14 shows the predicted binding site of all the compounds compared to the binding site of the natural ligands of the protein.

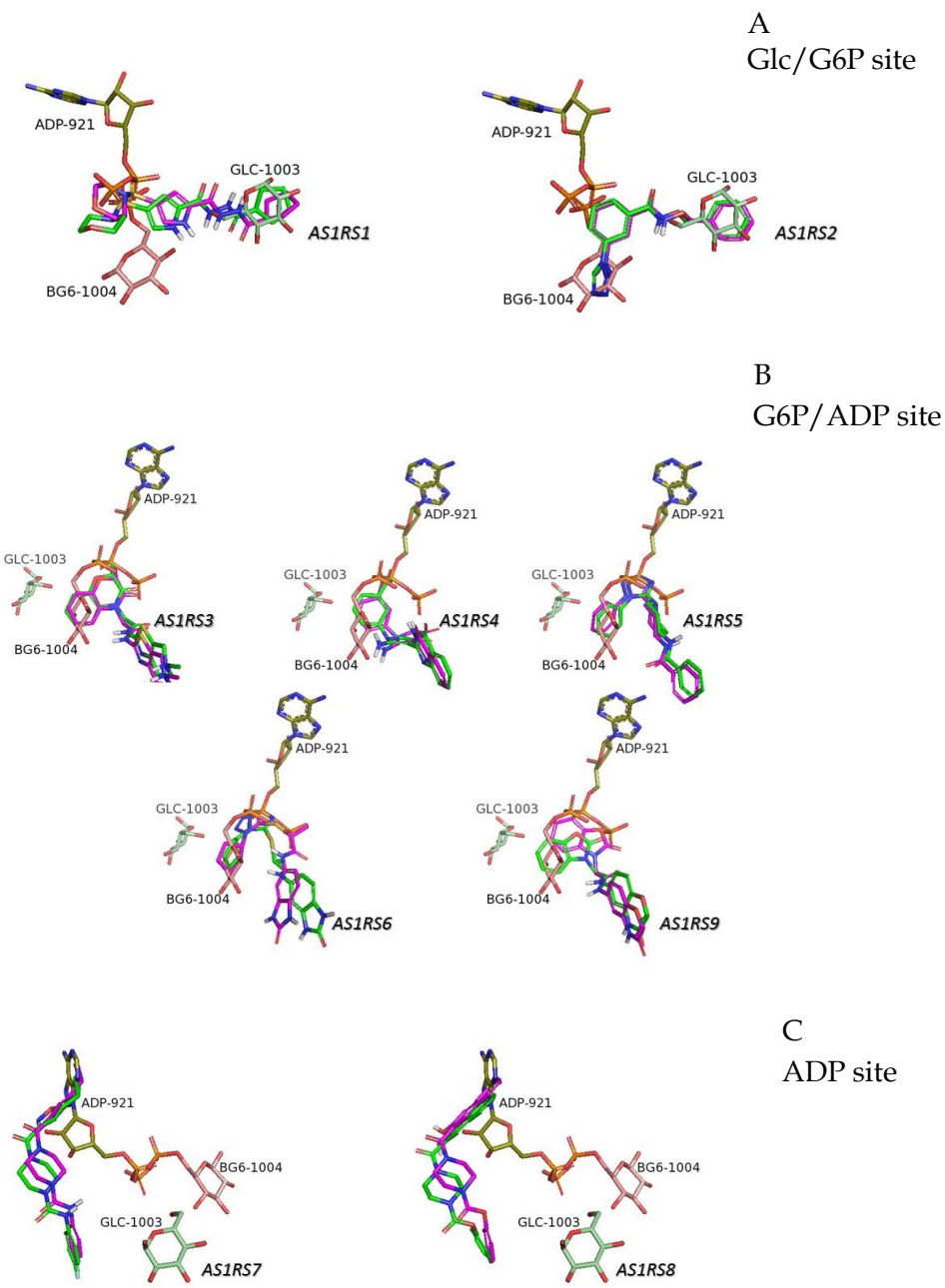


Figure 5.14: Ligands binding site relative to the natural ligands of hHKI and hHKII.
Ligands present in 2NZT, Glc and G6P, are shown as pale green and salmon sticks. ADP present in 1DGK is shown as deep olive stick. The ligand binding poses as predicted from docking programs are shown as green sticks (Vina) and magenta (AutoDock). Panel **A** shows compounds AS1RS1-2 binding to the Glc and G6P binding site. Panel **B** shows

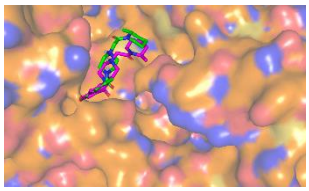
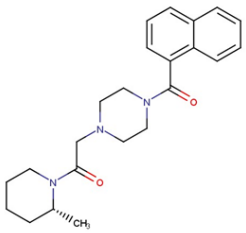
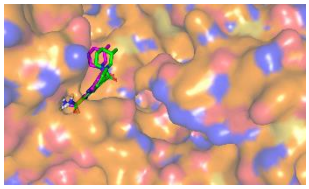
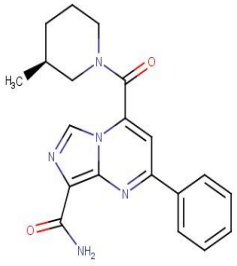
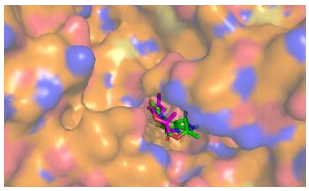
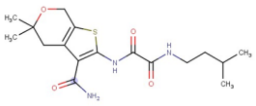
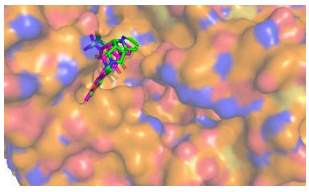
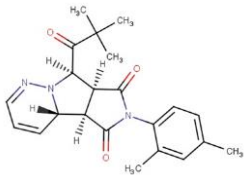
compounds AS1RS3-6 and 9 that bind to both G6P and ADP site and panel **C** shows compounds AS1RS7-8 predicted to bind closer to the ADP binding site.

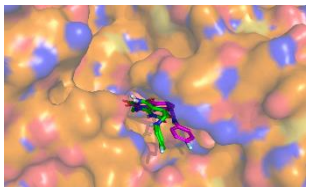
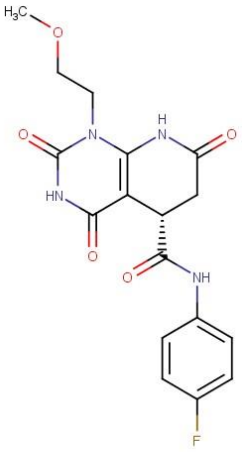
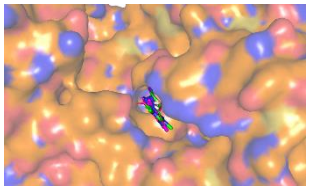
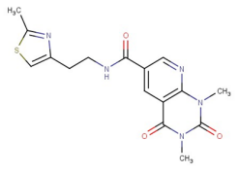
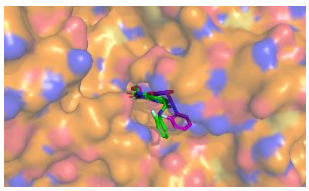
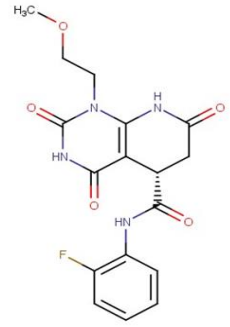
5.10.2 2nd scoring scheme for ranking the top docked compounds from the Glc/G6P site:

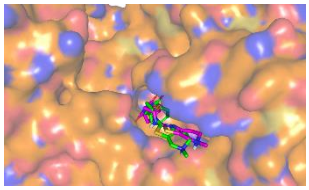
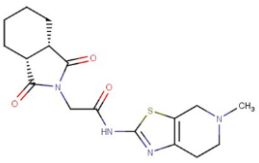
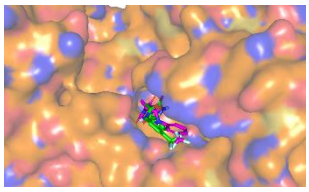
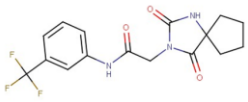
The top 2,467 were re-ranked using the updated version of X-Score 1.3. Moreover, multiple scoring functions were added to the consensus scoring scheme (NNScore 1.0, NNScore 2.0 and RFScore 4). The latter method completely altered the ranked order of the compounds. The calculated average AutoDock and AutoDock Vina scoring energies of 109 top hits are between -5.43 to -12.37 kcal/mol and -6.6 to -10.4 kcal/mol respectively. Nine compounds were chosen for screening from this docking. The MW is less than 380 for all compounds. Table 5.4 shows the ligand predicted binding modes, the chemical structures and the scoring functions.

Table 5.4: Ligands docking poses predicted by AutoDock (AD) and Vina in the active site of C domain of human hexokinase I.

Ligands coloured in green (Vina) and magenta (Autodock) represent the predicted binding modes from each software. The table order is according to the consensus scoring from the total of scoring functions (AD, Vina, DrugScore and X-Score 1.3, NNScore 1.0 and 2.0 and RFScore 4.0). The chemical structure, Vina, AD and NNScore 1.0 for each compound are shown. The last column shows the percentage of inhibition/ activation at 100μM (final inhibitor concentration) based on the enzymatic assays performed (see Chapter 7).

Docked Poses from AutoDock and Vina	*Chemical structure*	*Vina (kcal/ mol)*	*AD (kcal/ mol)*	*NNScore 1.0*	*% I (-)/ A(+) at 100μM*
2RS1 		-8.2	-7.7	-0.7611	+3%
2RS2 		-8.8	-7.7	0.8989	-7%
2RS08628 		-9.0	-9.3	-0.8926	-11%
2RS23615 		-8.4	-8.0	-0.7185	-22%

Docked Poses from AutoDock and Vina	*Chemical structure*	*Vina (kcal/ mol)*	*AD (kcal/ mol)*	*NNScore 1.0*	*% I (-)/ A(+) at 100μM*
2RS11168 		-8.5	-7.8	-0.2358	-12%
2RS4 		-8.1	-8.0	-1.151	+8%
2RS11309 		-8.6	-8.0	-0.7636	+1%

Docked Poses from AutoDock and Vina	*Chemical structure*	*Vina (kcal/ mol)*	*AD (kcal/ mol)*	*NNScore 1.0*	*% I (-)/ A(+) at 100μM*
2RS5 		-8.6	-7.6	-0.3035	-36%
2RS6 		-7.3	-5.9	0.9029	-8%

Interactions between virtual hits and target molecule

Table 5.5 shows all the residues that form hydrogen bonds (0- 3.3 Å) with the template structure, as identified from WinCoot. Between the two different docking programs, different interactions are found because of small differences in compound orientation between them. The number of the different residues is small for all compounds. Another interesting observation is that 2RS1 in AutoDock binding mode forms no hydrogen bond while only one interaction is formed with Vina binding mode. 2RS23615 forms no hydrogen bonds according to Vina orientation but 3 bonds according to AutoDock. The latter also forms interactions with the residues that take part for ADP binding, i.e 747G. The same is observed for 2RS2.

Table 5.5: List of residues that form hydrogen bonds with virtual hits for Glc/G6P site (2nd ranking scheme).

Active site 2nd ranking scheme	AutoDock binding mode	Vina binding mode	Differences between Vina-Auto
2RS1	*No hydrogen bond*	**T680**	Yes.
2RS2	**G747**, T784, S788, **T863**	T784, S788	Yes.
2RS08628	K621, D657, T680, G681, **T863**	**D532**, **G535**, K621, D657, T680, G681	Yes
2RS23615	**G747, T784, 7K85**	*No hydrogen bond*	Yes
2RS11168	**D532**, N537, R539, T680, G681, **D895**, T863	N537, R539, T680, G681, T863	Yes
2RS4	**A536**, K621, D657, **G679**, T680, **G681**, **E708**	K621, D657, T680, **T863**	Yes.
2RS11309	**D532**, N537, R539, T680, G681, T863	N537, R539, T680, G681, T863	Yes
2RS5	A536, N537, R539, **G862**, T863, D895	A536, N537, R539, T863, D895	Yes.
2RS6	A536, N537, R539, **D861**, **G862**, **T863**, K866, E894	A536, N537, R539, K866, G894, **G896**	Yes.

Figure 5.15 depicts the hydrogens formed as predicted from the docking programs. As shown in Table 5.5 there are differences between the two programs but the common interactions are shown here. Compound 2RS1, 2RS2 and 2RS23615 bind to the ADP binding site. Compound 2RS1 and 2RS23615 are not in close proximity to form hydrogen bonds according to AutoDock and Vina respectively thus they are predicted to interact with the other program. For these compounds both binding modes are shown, whereas for the remainder only the AutoDock binding mode is shown.

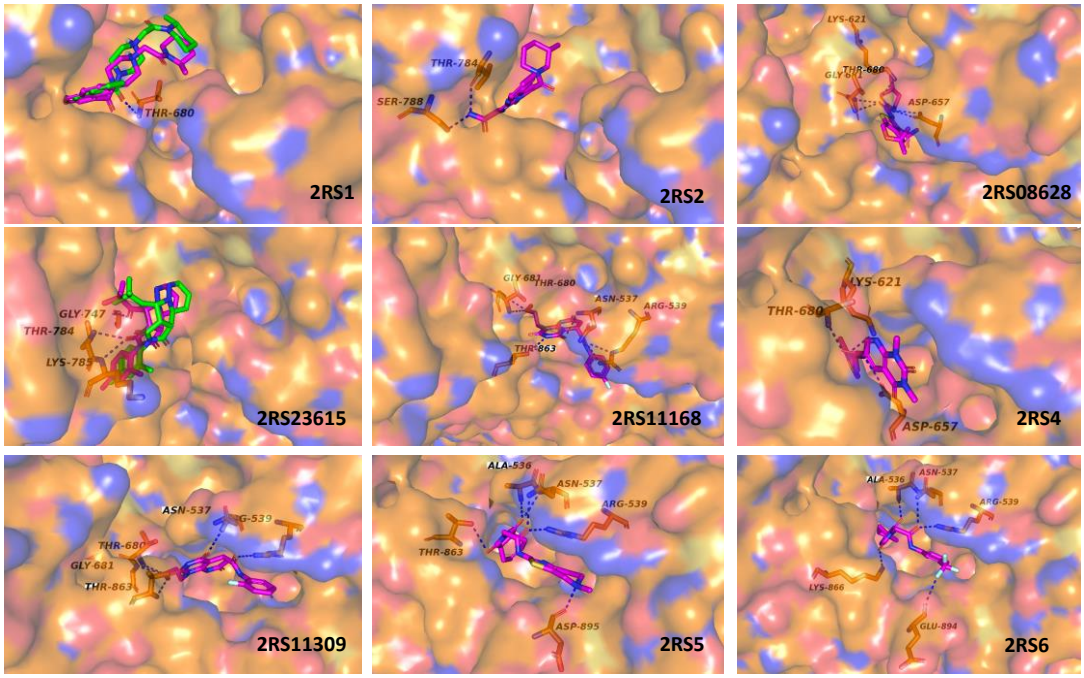


Figure 5.15: Docked compounds in the Glc/G6P of 1DGK (from re-ranking the top hits) with interacting residues.

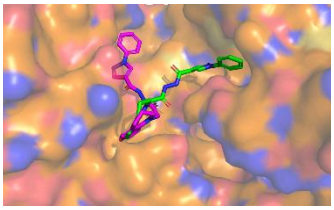
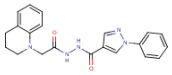
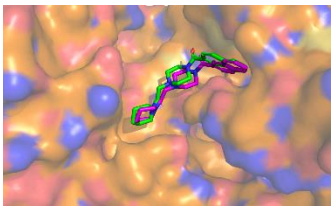
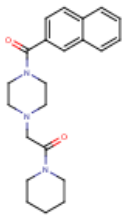
WinCoot was used to find the hydrogen bonds within a distance of 0-3.3 Å from the ligand.

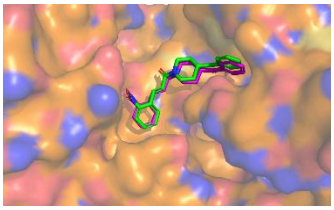
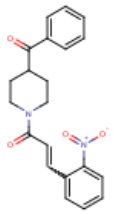
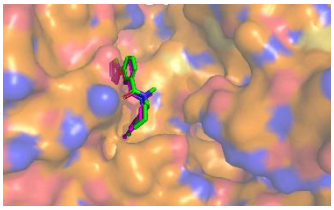
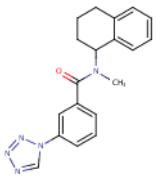
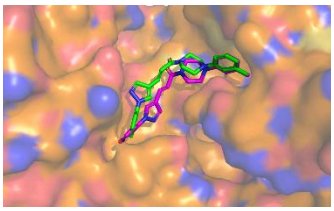
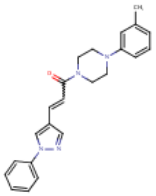
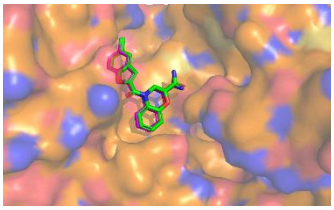
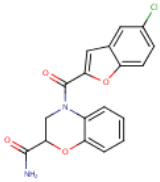
5.11 Docking results for ADP binding site

The filtering process of the library for the ADP site resulted in 5,000 compounds which entered the docking process. Then scoring algorithms DSX and X-Score 1.2 were applied to rank the compounds based on their binding affinity. The top 93 compounds were analysed and six compounds were purchased to test their inhibitory potency against hHKII. All the compounds had MW<380 and cLogP from 2.3-4.3. The surface of the 1DGK is shown. Table 5.6 shows the ligands predicted binding modes, the chemical structures and the scoring functions.

Table 5.6: Ligands docking poses predicted by AutoDock (AD) and Vina in the ADP pocket of C domain of human hexokinase I

Ligands coloured in green (Vina) and magenta (Autodock) represent the predicted binding mode from each software. The table order is according to the consensus scoring (AD, Vina, DrugScore and X-Score). The chemical structure, Vina, AD, DSX and X-score for each compound are shown. The last column shows the percentage of inhibition/ activation at 100µM (final inhibitor concentration) based on the enzymatic assays performed (see Chapter 7).

Docked Poses from AutoDock and Vina	*Chemical structure*	*Vina (kcal/ mol)*	*AD (kcal/ mol)*	*X-score*	*DSX*	*% I/A at 100 µM*
NP1 		-9.1	-9.2	7.28	-140.17	+32
NP2 		-8.9	-9.3	7.14	-143.23	+12

Docked Poses from AutoDock and Vina	*Chemical structure*	*Vina (kcal/ mol)*	*AD (kcal/ mol)*	*X-score*	*DSX*	*% I/A at 100 μM*
NP3 		-8.9	-9.1	6.80	-149.30	-3
NP4 		-9.2	-9.9	6.56	-135.68	-3
NP5 		-8.8	-8.8	7.39	-161.80	+12
NP6 		-9.8	-9.2	6.43	-139.50	-14

Interactions between virtual hits and target molecule

Table 5.7 shows all the residues that form hydrogen bonds (0- 3.3 Å) with the template structure, as identified from WinCoot. Only NP4 forms more than two hydrogen bonds (with both AutoDock and Vina). The remaining compounds are limited to 2-3 bonds. The compounds from the ADP pocket could be characterised as less likely to inhibit the protein in low micromolar concentration. NP5 according to Vina is not in a hydrogen bond distance from any residue.

Table 5.7: List of residues that form hydrogen bonds with virtual hits for ADP pocket.

Novel pocket	**AutoDock binding mode**	**Vina binding mode**	**Differences between Vina-Auto**
NP1	**M748, T784**	**T680**	Yes.
NP2	T680, M748	T680T, M748	No.
NP3	G747, T784	G747, T784, **K785**	Yes.
NP4	G747, M748, **Y749**, **L750**, G751, E752, **T784**, L785	G747, M748, G751, E752, L785	Yes.
NP5	*No hydrogen bond*	**T784**	Yes.
NP6	T680, T784	T680, **G747**, T784	Yes.

Figure 5.16 depicts the hydrogens formed as predicted from the docking programs. As shown in Table 5.7 there are differences between the two docking programs but the common ones are shown in this figure. NP1 is the only compound that does not

show a posematch. For this reason both binding modes from AD (magenta) and Vina (green) are shown. In addition, both binding modes are shown for compound 5. These are in agreement (posematch) but the predicted interactions are poor, resulting in no hydrogen bond for AutoDock and just one hydrogen bond with Thr784 for Vina.

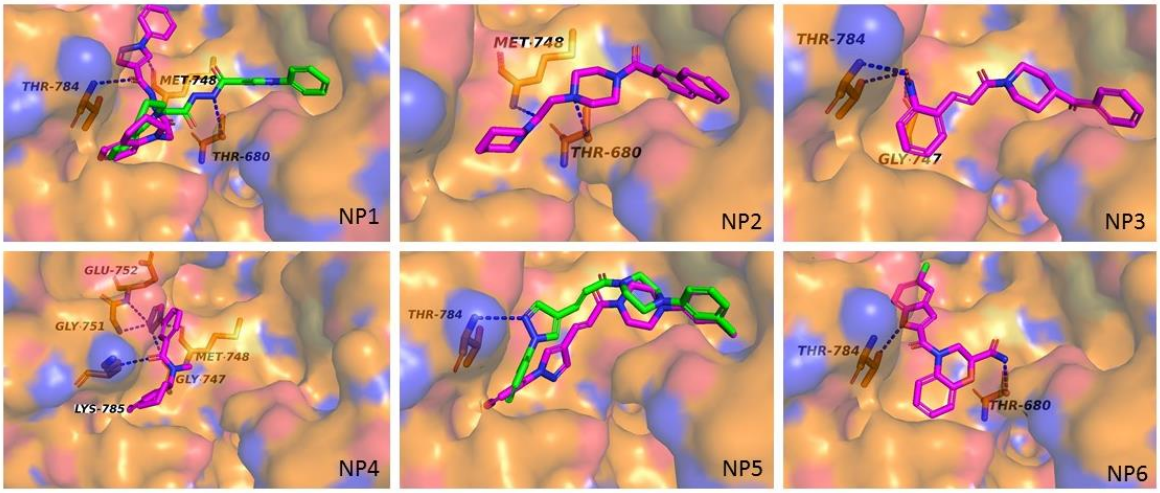


Figure 5.16: Docked compounds in the ADP site of 1DGK with interacting residues.
Six compounds were purchased after analysis of the top hits from docking the ADP binding site. WinCoot was used to find the hydrogen bonds within a distance of 0-3.3 Å from the ligand.

5.12 Discussion and summary

In this chapter the docking studies, which were performed with the aim to discover novel inhibitors for human HKII, are explained in detail. The SBVS has been a very common method for the generation of potential lead-compounds. The Glc/G6P site and the ADP site of hHKI were docked using two different, widely known docking softwares, AutoDock and Vina. Hexokinase isoform I has identical sites with isoform II and the structure of the former (1DGK) was used as it was found to be in a better conformation. For both sites the following scoring functions were used to predict the binding affinity of the predicted binding modes: DrugScore and X-score 1.2. The consensus scoring method ("rank-by-rank") was used as it has been shown to improve the accuracy of the scoring algorithms. For the Glc/G6P site only, in order to be "time and money" efficient, the docked compounds were re-ranked using an additional list of scoring algorithms such as: NNScore 1.0 and NNScore 2.0, RFScore 4.0 and an updated version of X-Score (1.3 instead of 1.2). This completely altered the ranking order of the compounds and it was anticipated that the compounds purchased from these results would show improved inhibition efficiency.

Surprisingly one compound from the ADP site is an activator. NP1 was found to activate significantly the enzyme (by 32%) by binding to the ADP binding site. It is difficult to make a hypothesis on the activation of the enzyme. There are not any known synthetic activators of hHKII to date.

The most potent inhibitors are the AS1RS8 and 2RS5 which bind to the ADP and the Glc/G6P site respectively. The former compound was predicted to make fewer contacts based on the binding mode of the docking software (Fig. 5.13). However, the hydrophobic rings of AS1RS8 at each end of the molecule participate in hydrophobic interactions whereas the other inhibitors lack any of those interactions, as they bind to the polar Glc/G6P site. Figure 5.17 shows the inhibitor binding mode and the residues that could interact with it. The hydrophobic double ring at one end could participate in hydrophobic interactions (yellow dash) with T863, L864, L867 which are 4.4, 3.4 and 3.7 Å away respectively. While the phenyl ring at the other end is in close distance with M567 and F623 (3.6 and 3.8Å respectively). These hydrophobic interactions could stabilise the ligand as they are stronger than hydrogen bonds (blue dash) and this could explain the higher inhibition observed despite the fact that the other inhibitors form more hydrogen bonds.

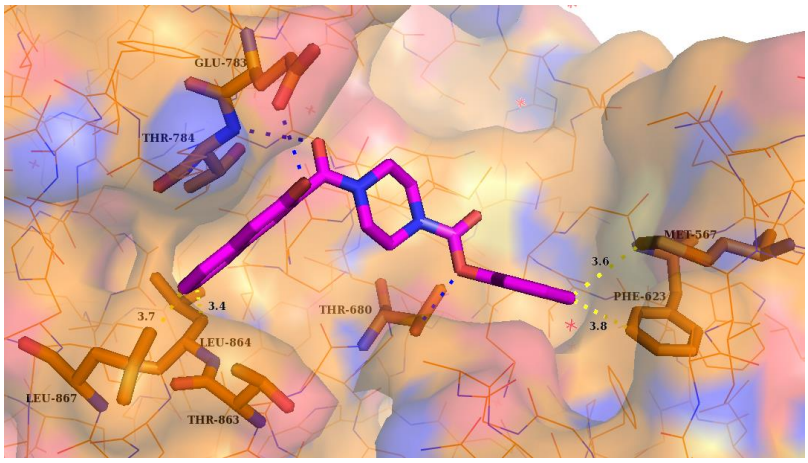


Figure 5.17: Hydrophobic interactions of AS1RS8 with 1DGK

The predicted binding mode of AS1RS8 is shown in magenta sticks. This inhibitor is predicted to bind to the ADP binding site rather than the polar Glc/G6P site.

2RS5, shows 36% inhibition while it is predicted to bind to the Glc/G6P site. It is the most potent inhibitor compared to the others that are predicted to bind in the same pocket (AS1RS4, AS1RS6). However, according to the predictive binding mode the compound makes similar interactions with the latter two. Figure 5.18 shows the residues that are predicted to interact with 2RS5 via hydrogen bonds proposing a modification that could improve the compound potency.

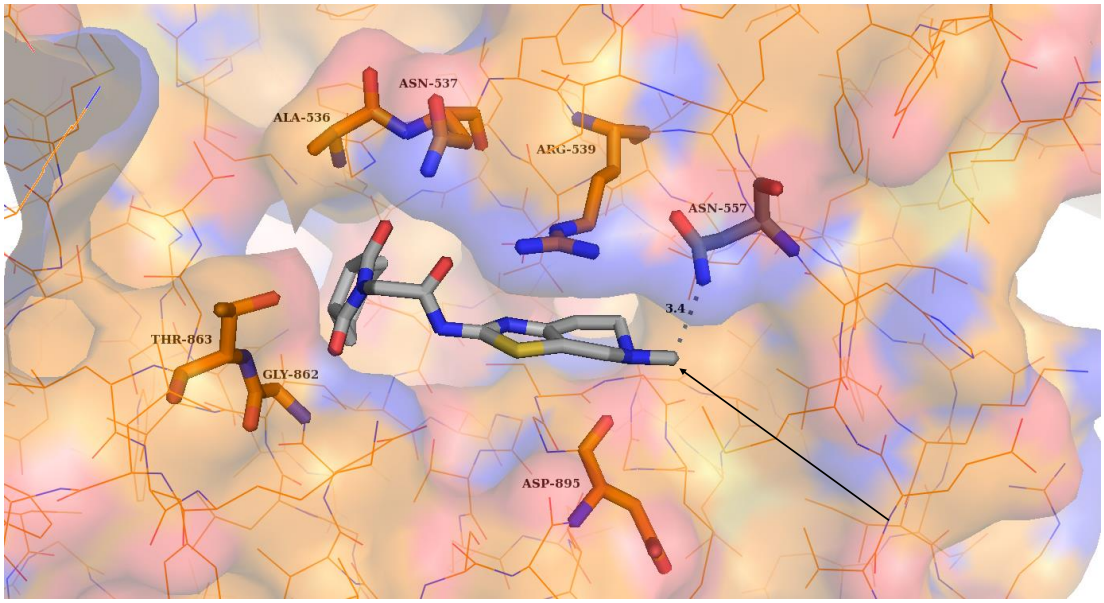


Figure 5.18: SAR of 2RS5 in the active site of 1DGK

The grey sticks show the modelling of the compound 2RS5 in the active site of hHKI. 2RS5 was purchased after the second ranking method of the docked compounds in the Glc/G6P site. It forms hydrogen bonds with the highlighted residues (orange sticks). An extra hydrogen bond could be formed with N557 by changing the part shown with the black arrow.

From the first ranking scheme, of the Glc/G6P site AS1RS8, AS1RS4 and AS1RS6 show 35%, 15% and 13% inhibition at ~100 μM. To improve the latter two inhibitors, we have identified a number of substitutions that could be made to the rings of the molecules in order to increase the number of hydrogen bonds and/or hydrophobic interactions with the target. It is extremely difficult to make a sophisticated SAR analysis in the absence of a crystal structure showing the actual binding site of the compound and the key interactions with its target. However, based on the modelling of the compounds to the Glc/G6P site of the enzyme we could propose a number of key changes to the molecules aiming to optimize the activity.

AS1RS4 is predicted to bind to the polar Glc/G6P site of the enzyme participating in six hydrogen bonds. In Figure 5.19 the black arrow shows the benzene ring on the one end of the molecule that could be modified with the addition of a methyl which could interact with the hydrophobic side chains of M555 and L541. The $–CH_3$ could improve the lipophilicity as well. Otherwise a MeOH group could be added that would act as a hydrogen bond donor (HBD) and could potentially make an extra hydrogen bond with N557. The ring on the other end of the molecule (shown by the blue arrow) could be modified with a methyl ($-CH_3$) substituent to shorten the distance between the hydrophobic side chain of I677 and the hydrophobic ring in order to interact. The yellow dashes show the distance between the lead candidate and the residues (shown as grey) that are not predicted to interact with the molecule based on the modelling.

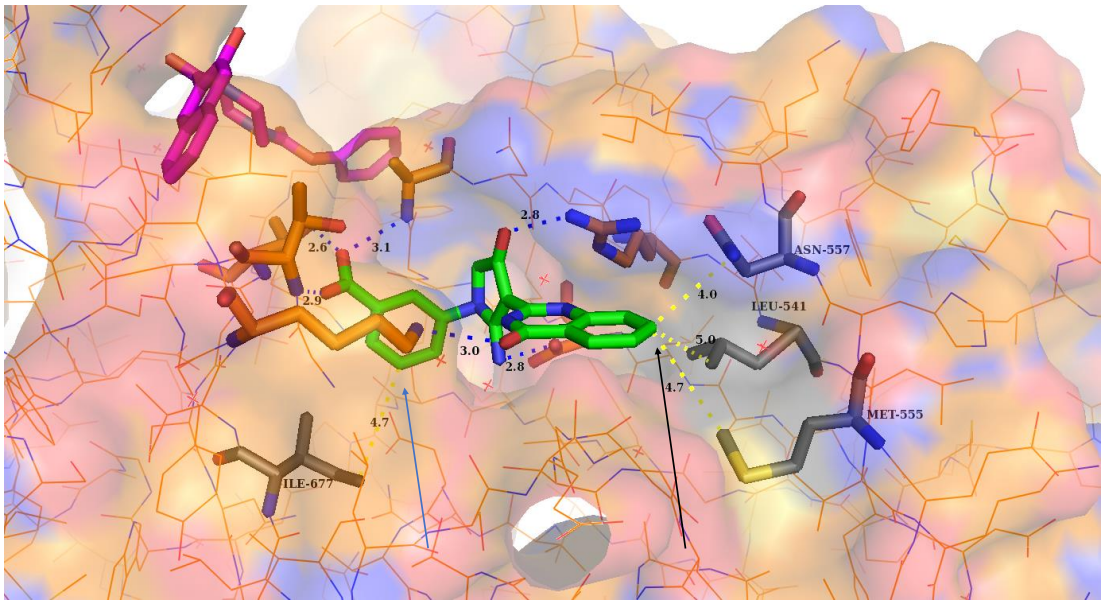


Figure 5.19: SAR of AS1RS4 in the active site of 1DGK

The green sticks show the modelling of the compound AS1RS4. Magenta sticks show the inhibitor AS1RS8. AS1RS4 is predicted to participate in six hydrogen bonds shown as blue dashes. The benzene ring on the one end of the molecule could be modified to make stronger hydrophobic interactions. Otherwise a HBD could be added to make an extra hydrogen bond. The ring on the other end of the molecule could be modified with a methyl ($-CH_3$) substituent to increase the hydrophobic interactions as well.

AS1RS6 contains a polar ring on the end that could be modified with the addition of an extra HBD linker that could interact with an extra residue (Fig. 5.20). The black arrow shows the polar ring on the end of the molecule that could interact with A536. The grey dashes show the distance between the polar ring and A536 (4 Å).

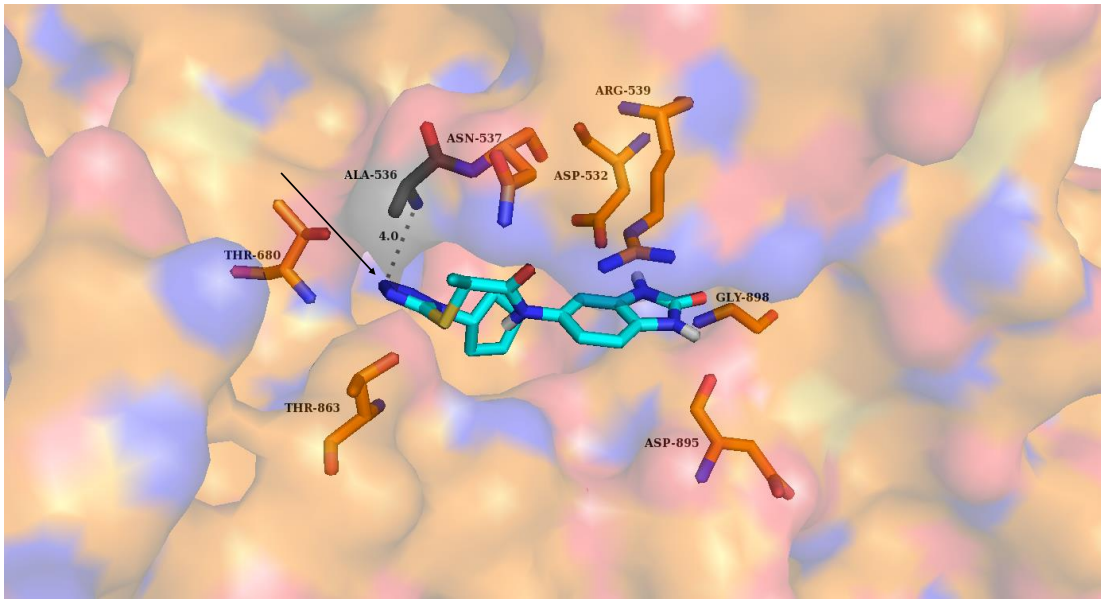


Figure 5.20: SAR of AS1RS6 in the active site of 1DGK

The cyan sticks show the modelling of the compound AS1RS6. AS1RS6 participates in hydrogen bonds with the residues shown as orange sticks.

However, these modifications were not applied to the compounds for further studies. The lack of a crystal structure in complex with at least one of the inhibitors from each site makes it very hard to be confident about the key elements of the compounds for the binding and the inhibition of the target protein.

From the compounds purchased, after re-ranking as mentioned above, only two compounds show considerable inhibition compared to three compounds from the first scoring scheme. The new strategy did not improve the number of inhibitors as anticipated. Compounds 2RS23615 and 2RS5 show considerable inhibition of the enzyme (22% and 36% inhibition respectively). Compound 2RS23615 is predicted to bind to the ADP site.

6. Chapter 6: Screening hits for *Trypanosoma brucei* hexokinase I (TbHKI)

6.1 Introduction

This chapter will discuss the methods performed to discover novel compounds that could potentially inhibit the TbHKI *in vitro*. The approach followed here involves the Structure Based Virtual Screening (SBVS), as for hHKII, with the difference that for the specific target (TbHKI), an experimental X-ray structure is unavailable.

Protein modelling can however provide a reasonably reliable template to allow structure-based drug design studies. Among the three major approaches to three-dimensional structure prediction, homology modelling is the easiest one (Krieger et al., 2003) and is based on the observation that homologous proteins with high percent of sequence identity adopt practically identical structures. When the sequence homology rises up to 50%, a good protein model is expected to be predicted, whereas in cases that sequence identity drops to 20% the structural differences are impossible to predict (Chothia and Lesk, 1986). More recently, Rost, analysed more than a million sequence alignments showing that it is possible to differentiate true from false positives for sequences which share a low level of similarity (Rost, 1999).

In our case, a homology model of TbHKI was generated to discover ligands of the active site. In general, the following steps are performed during a homology model generation: 1) fold assignment and template selection, 2) target-template alignment, 3) model building and 4) model evaluation (Martí-Renom et al., 2000). The generated model is then used as a docking template to identify virtual hits which can then be tested against TbHKI. Since the docking will only take place in the active site of the protein, which is expected to be conserved among proteins of the same family, we will not evaluate the whole model.

6.2 Materials and methods

6.2.1 Template selection for TbHKI homology modeling

We compared the TbHKI sequence to all the sequences of known structures stored in the PDB, using BLAST, and the highest sequence identity was found to be 38% for Arabidopsis hexokinase 1 (athxk1), followed by human hexokinase isoform I (37% identity).

We decided to model the amino acid sequence using the X-ray structure of human hexokinase isoform I (PDB code: 1DGK). TbHK1 shares 37% sequence identity with the full length of hexokinase isoform I. However, the active site is conserved, and it may provide a good model regardless that the overall sequence homology is not so high. The active sites of related proteins will have similar geometries because of the necessity to maintain a functional binding site (Lesk and Chothia, 1980). Figure 6.1 shows the pairwise comparison of TbHK1 with human hexokinase isoform I, using Clustal Omega (Larkin et al., 2007). The generated model should be in a "ligand bound conformation", so the sidechain of the residues interacting with Glc and G6P will be compared between 1DGK and generated model.

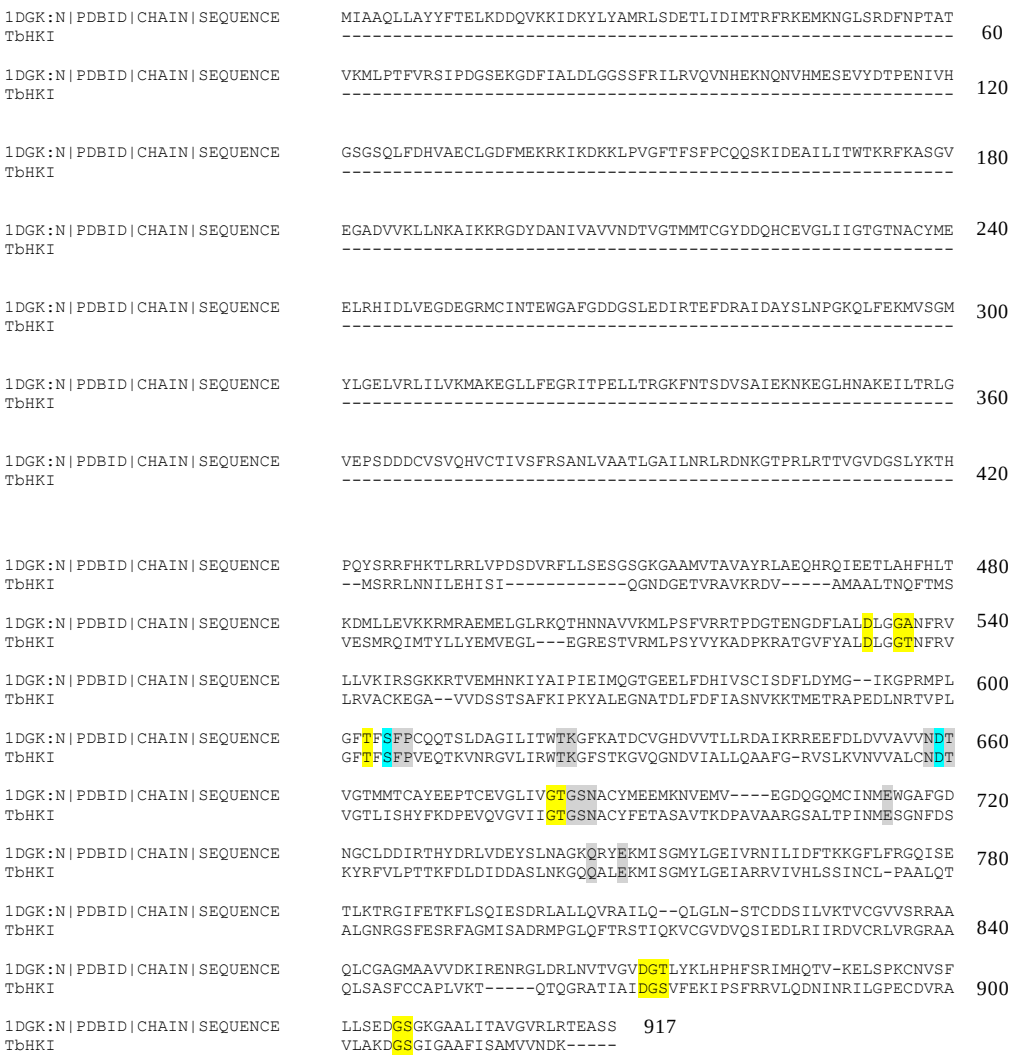


Figure 6.1: Amino acid alignment of human hexokinase isoform 1 (1DGK) and TbHK1.

Protein sequence for TbHK1 was found from http://tritrypdb.org/tritrypdb/ (Aslett et al., 2009). Clustal Omega was used to align the two sequences. Grey residues interact with Glc, yellow residues interact with G6P and cyan residues interact with both ligands. Glc binding site is 100% conserved, while G6P site is not completely identical between the two different species. Apart from the mismatch at residue 536 there is another true mismatch. Residue 863 is different and not because 1DGK is a mutant. Residue at 863 is threonine for human hexokinase I while it is serine for TbHK1 (S421).

6.2.2 Model building

Two different softwares were employed for the protein structure prediction, Phyre2 (Kelley et al., 2015) and I-TASSER (Zhang, 2008). Phyre2 has two primary modes for model building, the normal and intensive. For 'normal' mode, Phyre2 builds a hidden Markov model (HMM) of the user sequence and compares this to a library of HMMs built for a representative set of known protein structures. Regions of the user sequence with no significant match to known structures are left unmodelled. The 'intensive mode' will create a complete model even in the absence of known homologs structures, using a simplified protein-folding simulator (Kelley et al., 2015). In our case the 'normal' mode was chosen with one-to-one threading using 1DGK as the template.

I-TASSER builds 3D models based on multiple-threading alignments by LOMETS and iterative TASSER assembly simulations. Yang, 2015, discusses the operation of the I-TASSER suite in detail (Yang et al., 2015). One-to-one threading was performed, with 1DGK as the template. 3 models were generated from I-TASSER that differ in the C-score. C-score is a confidence score for estimating the quality of predicted models by I-TASSER. It is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of higher value signifies a model with a high confidence and vice-versa. The model with the higher C-score (1.45 in our case) was chosen as our model template.

A superposition of the models from Phyre2 and I-TASSER with 1DGK was performed and analysed.

6.3 Results

6.3.1 Homology model choice

Phyre model was chosen as the template for our docking studies as the modeling seems to have worked exceptionally well, maintaining the orientation of all the residues. The different residues have been also modelled keeping the ligand bound conformation.

I-TASSER, did not work very well in our case, as almost all the residues have a different orientation in the pocket with N213 for TbHK1 to have a rotated side chain up to 3.2Å away from the template. N656 in 1DGK makes a hydrogen bond with Glc (N from Asn with O from Glc), so it is essential to keep the same orientation for the docking studies, for a glucose competitive inhibitor.

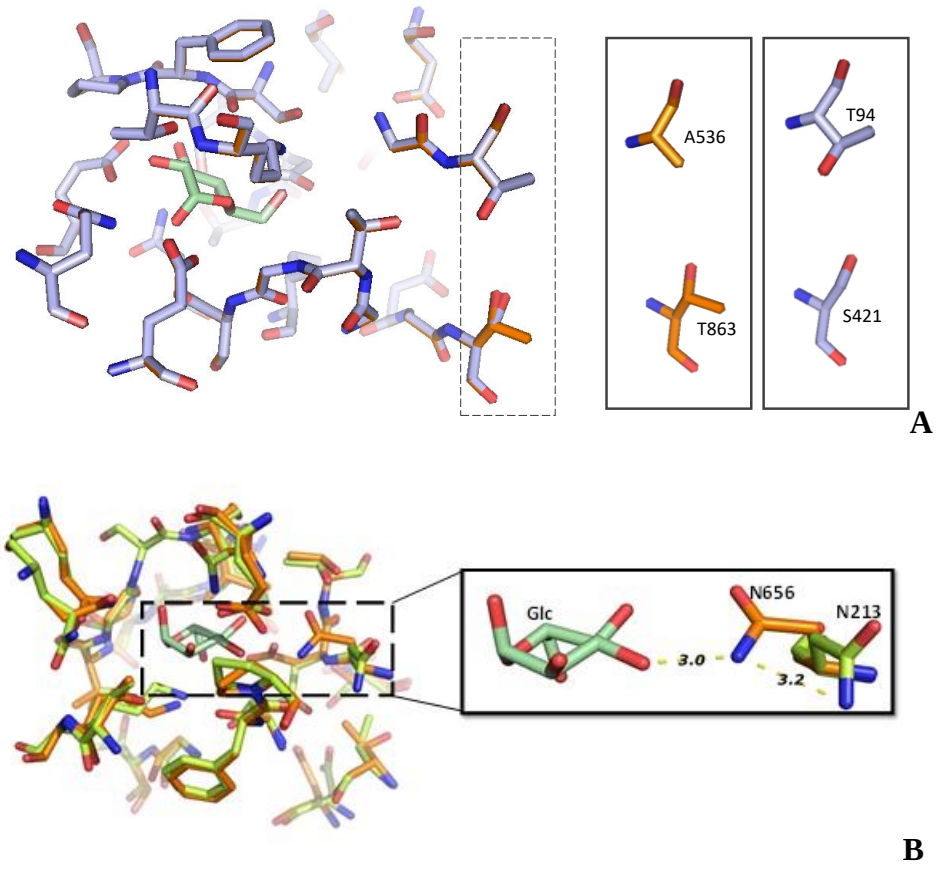


Figure 6.2: Superposition of models generated from Phyre-2 and I-TASSER against template 1DGK.

Panel **A** shows the overlay of Phyre-2 model (light blue) with 1DGK (orange). The residues of Glc and G6P binding site (stick representation) have the same ligand bound orientation as desired. Only two residues are different (T94 and S421 for *T. brucei*) in the protein sequence, as highlighted on the right. The modeling maintained the same atom orientation of template.

Panel **B** shows the alignment between 1DGK (orange) and I-TASSER model (limon). The residues do not seem to have maintained the same orientation as template and especially N213 side chain, as shown on the right, has a big rotation of 3.2Å away from the ligand. This interrupts the formation of hydrogen bond with oxygen of glucose.

6.3.2 Docking space of TbHKI

The Glc/G6P and ADP sites of TbHKI were used for the docking studies. The docking space is shown by the black box. A closer look into the box identifies two pockets that could accommodate lead-size compounds. The residues present in the pocket are identified and shown in a stick representation.

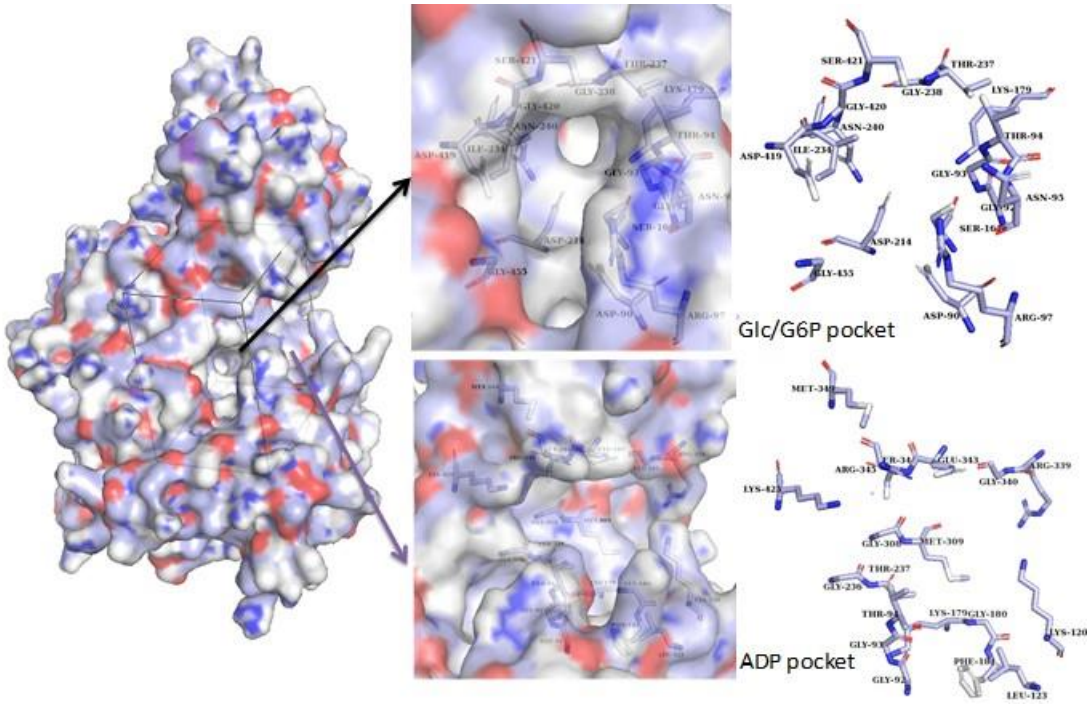


Figure 6.3: Docking space of TbHKI. Two pockets are present in the docking space.
The surface of TbHKI is shown as light blue. The black box shows the limits of the docking area. Inside the box, two pockets very close to each other, are present. The black arrow shows a closer view of Glc/G6P_1 with the residues shown as sticks.
The residues which form this pocket are: D90, G92, G93, T94, N95, R97, S161, K179, D214, I234, T237, G238, N240, D419, G420, S421, G455.
Purple arrow shows ADP pocket which consists of the following residues: G92, G93, T94, K120, L123, K179, G180, F181, G236, T237, G308, M309, R339, G340, E343, S344, R345.

Vina and AutoDock were the docking softwares used for the docking studies. From the initial list of ~3 million compounds, 5,000 were proceeded to the docking process. The consensus docking was performed and 1,595 compounds were found to have the same binding mode predicted from both programs. The compounds were

ranked following the “rank-by-rank” strategy as mentioned in Chapter 5, section 5.7. The scoring algorithms used for the final ranking are the following: X-score 1.3, Drugscore, NNScore 1.0 and 2.0 and RF-score. The final choice of compounds was made using multiple techniques, including prioritisation of compounds that met Lipinski rules and Oprea criteria. In addition, the hit list was analysed by an experienced medicinal chemist (Dr Phill Cowley, Head of Chemistry, IOmet Pharma Ltd.) to remove compounds which were potentially reactive, non-specific in terms of their biological action or not drug-like. Following the prioritisation of the compounds, 4 compounds were purchased.

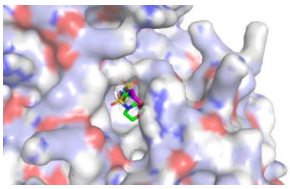
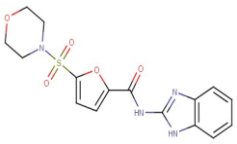
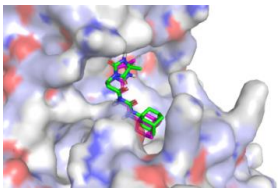
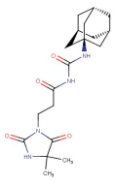
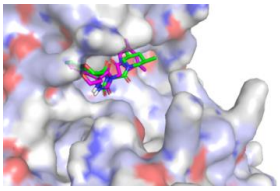
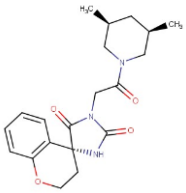
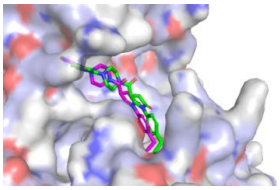
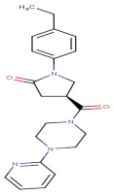
6.3.3 Virtual hits obtained from SBVS

Table 6.1 shows the structure, the binding mode and the scores of the purchased compounds from the *in silico* studies against TbHKI. It also shows the % of inhibition at 100μM based on the enzymatic assays performed. The assay and conditions used for the inhibition studies are described in detail at Chapter 7.

It is interesting that 2 out of 4 compounds activate the enzyme by a significant amount (over 20%). One compound shows a small, non-significant inhibition and final one decreases the enzymatic activity around 65% at 100μM.

Table 6.1: Ligand docking poses predicted by AD and Vina in the active site of *Trypanosoma brucei* hexokinase I.

Ligands coloured in green (Vina) and magenta (AD) represent the predicted binding modes from each software. The ranking order is according to the consensus scoring from AD, Vina, DrugScore and NNScore1.0 as well as consensus docking. The chemical structure, Vina, AD, and NNScore and the % of inhibition or activation are shown for each compound. The cLogP values are in the range of 1.4-4.3 and the molecular weight is <380 for all. All the compounds show consensus docking. The surface of the TbHKI model active site is shown as light blue.

Docked Poses from AD and Vina	*Chemical structure*	*Vina (kcal/ mol)*	*AD (kcal/ mol)*	*NNScore 1.0*	*%I (-)/ A (+) at 100μM*
TbHK1.1 		-8.7	-11.0	0.4824	+45
TbHK1.2 		-8.0	-7.8	0.9668	+20
TbHK1.3 		-9.0	-8.5	0.6498	-9
TbHK1.4 		-8.8	-7.6	0.2039	**-65**

According to the binding mode only TbHK1.1 is predicted to bind to Glc/G6P site while the rest to the ADP site (Figure 6.3). The predicted hydrogen bonds based on the binding mode of each program (AD, Vina) are calculated using WinCoot (3.3 Å cut-off distance). Table 6.2 shows the residues that participate in these interactions. TbHK1.4 is predicted to have the least interactions thus it is the compound that shows the best inhibition (65%) of the enzymatic activity. TbHK1.1 and TbHK1.2 activate the enzyme while they bind on a different pocket.

Table 6.2: Predicted interactions with TbHKI based on binding mode of the docking Fsoftware.

WinCoot was used to calculate the hydrogen bonds formed between each compound and each predicted mode with the target molecule. Residues in bold are the residues which are not in agreement between the programs.

Compound ID	**AD predicted interactions**	**Vina predicted interactions**	**Differences Between AD and Vina**
TbHK1.1	**T94**, N95, R97, K179, **D214**, T**237**, G238, S421	N95, R97, S161, K179, G238, **D419, G420**, S421	Yes
TbHK1.2	T94, T237, **G308,** S344, R345	T94, T237, S344, R345	Yes
TbHK1.3	T237, **G308**, S344, R345, K425	**T94**, T237, S344, R345, K425	Yes
TbHK1.4	S344, **M349**	S344	Yes

The compounds obtained from the hHKII *in silico* studies (Chapter 5) were screened against the TbHKI since the proteins share such a high percentage of sequence identity in the active site. The enzymatic biochemical assays showed that most of compounds activate the enzyme rather than inhibit it. More specifically, 17

compounds out of 24 were found to activate the enzymatic activity (from 6.25% up to 77.5%, Table 6.3).

Table 6.3: Inhibitions studies of purchased compounds on TbHKI

The compounds purchased from docking studies on hHKII were screened against TbHKI. 17 compounds were found to activate the enzyme. Only 4 compounds inhibited the enzyme in a significant percentage (>20%) while the remaining 3 compounds had no significant effect on the enzymatic activity.

Compound ID	**% Inhibition (-)/ Activation (+) at 100µM**	Compound ID	**% Inhibition (-)/ Activation (+) at 100µM**
AS1RS1	+28.75	2RS5	+77.5
AS1RS2	+27.5	2RS6	+30
AS1RS3	+30%	2RS08628	**-66.25**
AS1RS4	+10	2RS23615	+32.5
AS1RS5	+17.5	2RS11309	-5
AS1RS6	+32.5	2RS11168	+6.25
AS1RS7	**-28.5**	NP1	+58.75
AS1RS8	-7.5	NP2	+27.5
AS1RS9	+12.5	NP3	**-56.25**
2RS1	+7.5	NP4	**-31.25**
2RS2	+21.25	NP5	0
2RS4	+22.5	NP6	+11.25

6.4 Discussion and summary

The compound screening, yielded inhibitors and activators, suggesting that there might be an activator-binding site. There are no known activators for TbHKI in the literature. The finding of novel activators for TbHKI is very interesting. To further investigate this we searched the PDB and identified a number of synthetic activators for human glucokinase. These activators serve as a potential therapeutic approach for type 2 diabetes mellitus therapy. All activator compounds are found to bind to an allosteric site, close to the glucose-binding site with the following residues to interact with the activator: V452, V455, r63, Y215, V62, M210, I211, Y214 and M235 (Kamata et al., 2004). A sequence alignment was performed to investigate whether this is a conserved pocket, also present in TbHKI. However, the alignment showed that the only three residues are conserved while the remaining 6 are different. So we cannot argue that the present activators bind to the same pocket.

```
human_glucokinase       -------ML--------DDRARMEAAK-KEKVEQILAEFQLQEEDLKKVMRRMQKEMDRGL
TbHKI                   MSRRLNNILEHISIQGNDGETVRAVKRDVAMAALTNQFTMSVESMRQIMTYLLYEMVEGL

human glucokinase       RLETHEEASVKMLPTYVRSTPEGSEVGDFLSLDLGGTNFRVMLVKVGEGEEGQWSVKTKH
TbHKI                   E---GRESTVRMLPSYVYKADPKRATGVFYALDLGGTNFRVLRVACKEGAV----VDSST

human_glucokinase       QMYSIPEDAMTGTAEMLFDYISECISDFLDKHQ--MKHKKLPLGFTFSFPVRHEDIDKGI
TbHKI                   SAFKIPKYALEGNATDLFDFIASNVKKTMETRAPEDLNRTVPLGFTFSFPVEQTKVNRGV

human glucokinase       LLNWTKGFKASGAEGNNVVGLLRDAIKRRGDFEMDVVAMVNDTVATMISCYYEDHQCEVG
TbHKI                   LIRWTKGFSTKGVQGNDVIALLQAAFGR-VSLKVNVVALCNDTVGTLISHYFKDPEVQVG

human_glucokinase       MIVGTGCNACYMEEMQNVELVEGDE----GRMCVNTEWGAFGDSGELDEFLLEYDRLVDE
TbHKI                   VIIGTGSNACYFETASAVTKDPAVAARGSALTPINMESGNFDSKYRFVLPTTKFDLDIDD

human_glucokinase       SSANPGQQLYEKLIGGKYMGELVRLVLLRLVDENLLFHGEASEQLRTRGAFETRFVSQVE
TbHKI                   ASLNKGQQALEKMISGMYLGEIARRVIVHLSSINCLP-AALQTALGNRGSFESRFAGMIS

human_glucokinase       SDTGDRKQIYN--ILSTLGLR-PSTTDCDIVRRACESVSTRAAHM----CSAGLAGVINR
TbHKI                   ADRMPGLQFTRSTIQKVCGVDVQSIEDLRIIRDVCRLVRGRAAQLSASFCCAPLV-----

human_glucokinase       MRESRSEDVMRITVGVDGSVYKLHPSFKERFHASVR-RLTPSCEITFIESEEGSGRGAAL
TbHKI                   ----KTQTQGRATIAIDGSVFEKIPSFRRVLQDNINRILGPECDVRAVLAKDGSGIGAAF

human glucokinase       VSAVACKKACMLGQ
TbHKI                   ISAMVVNDK-----
```

Figure 6.4: Sequence alignment of human glucokinase with TbHKI

The sequence alignment between human glucokinase and TbHKI was performed with Clustal Omega (Larkin et al., 2007). The highlighted residues are the residues that form the allosteric pocket found in glucokinase. Binding of ligands in that pocket has been found to activate the enzyme. The identical residues are highlighted with green colour.

TbPFK (*Trypanosoma brucei* phosphofructokinase) is an allosteric protein with respect to its substrate, fructose-6-phosphate (Nwagwu and Opperdoes, 1982). Also trypanosomatid pyruvate kinase is allosteric regulated by fructose 2,6-bisphosphate, which promotes an active conformation of the enzyme (R-state) (Verlinde et al., 2001). TbHKI is not proven to be an allosteric enzyme. There are no studies that prove any cooperativity on substrate binding by TbHKI and no allosteric activators/ inhibitors are known.

Interesting kinetics have been observed for the enzyme of *T. cruzi* HK. *T. cruzi* HK has a hysteretic behaviour and shows transition from a less active to a more active enzyme and the opposite way while conformational changes are taking place (Acosta et al., 2014). This could also be the case for TbHKI, but no studies have been performed to prove that. However, if this is the case then the activators upon binding to TbHKI could affect the enzyme's activity by bringing the enzyme into a more active conformation.

Based on the present studies, we cannot propose an activation mechanism. The novel finding of the docking and screening studies against the TbHKI should be further examined in future work to establish the mechanism behind this. It would be interesting to test whether the activators increase the affinity for glucose or ATP. Or do they only increase Vmax? Also kinetic measurements with different concentrations of enzyme varying the Glc or ATP concentration could show if there is a cooperative behaviour depending on the concentration of TbHKI. Nevertheless, we could suggest that TbHKI could behave like *T. cruzi* hexokinase showing a hysteretic behaviour.

Regarding the inhibitors, TbHK1.4 inhibits the enzyme around 65% and is predicted to bind to the ADP site. When the compounds from the hHKII docking were tested against TbHKI, four were identified as inhibitors AS1RS7, NP4, NP3 and 2RS08628. The latter is a more potent inhibitor for TbHKI since the percentage of enzymatic inhibition is higher for TbHKI (66% versus 11% for hHKII). The remaining three had almost no effect on hHKII so they could be characterised as specific for TbHKI.

7. Chapter 7: Inhibition studies on hHKII and TbHKI

7.1 Introduction

The characterisation of the interactions of a compound with its target, and how the natural substrates and physiological conditions will influence this activity, is an important piece of information when performing inhibition assays. There are three main types of inhibition to describe the binding of a compound to its target molecule:

a) Competitive inhibition: a competitive inhibitor will bind to the free enzyme. The binding of the substrate to the active site will exclude binding of the inhibitor in the same site and vice-versa. As a result, the binding of a competitive inhibitor will result in an increase of the apparent K_m for its substrate without changing the V_{max}.
b) Noncompetitive inhibition: in this case the inhibitor binds to a different site than the substrate, thus the inhibitor can bind equally well to both free enzyme and the enzyme bound to its substrate. The binding of a noncompetitive inhibitor will lower the V_{max} without changing the apparent K_m for the substrate.
c) Uncompetitive inhibition: This type of inhibitor binds exclusively to the enzyme-substrate complex resulting in a lower V_{max} and K_m respectively.

An allosteric inhibitor binds to a different site than the active site (known as the allosteric site) followed by a conformational change which is necessary for the inhibition to take place. This conformational change can lead to a different conformation of the active site, or reduce the ability to lower the activation energy of catalysis. An allosteric inhibitor can be all the three above categories, competitive, noncompetitive or uncompetitive (Strelow et al., 2012).

The classical steady-state experimental conditions involve the measurement of the initial velocity where the substrate conversion is less than 10% or the product formation is also less than 10%. In this case the initial velocity depends on the enzyme and substrate concentrations and it fits to the linear portion where slope does not change with time (V_o stays constant). V_o for each reaction progress curve is equivalent to product formation divided by difference in time which essentially is the slope of the initial region of the curve, V_o.= $\Delta Y/\Delta X$= slope (Brooks et al., 2012).

Concentration-response plots are studied for the effect of a ligand on the enzymatic activity. The enzyme concentration and the substrate concentration respectively, are kept constant and the inhibitor is titrated. The inhibition at each concentration is measured. It is important to use an adequate concentration of inhibitor to saturate the reaction and provide well-defined top and bottom plateau values. The concentration of the compound that depletes the enzymatic activity by 50% is termed the IC_{50}. For an enzymatic assay to detect competitive inhibitors the reaction conditions should meet the following criteria: the reaction should run under initial velocity conditions and the substrate concentration should be equal or below the K_m value (Brooks et al., 2012).

Cell based assays are used after the initial discovery of a promising molecule. If the molecule has a significant effect in the cell based assay then it goes to the next phase of the process.

In this work we describe the inhibitions studies performed to identify the inhibitors potency against hHKII and TbHKI. More specifically, the biophysical methods, the biochemical enzymatic assay and the cell based assay for TbHKI which produced interesting results that will be discussed below.

7.2 Materials and methods

7.2.1 SPR analysis of ligand binding to hHKII

SPR experiments were performed on a BIAcore T200 (GE Healthcare). His-tagged hHKII was immobilised and covalently stabilised on an NTA sensor chip essentially as described previously with minor modifications (Chapter 3, section 3.2.4). Briefly, the sensor surface was primed with a 60 sec injection of 500 μM $NiSO_4$ at 10 μl/min. The surface was then minimally activated with a 240 sec injection at 5 μl/min, of a mixture of 0.2 M EDC and 50 mM NHS. 500 nM hHKII in 10 mM Hepes (pH 7.5), 500 mM NaCl, 1% DMSO, 5 mM $MgCl_2$, 62.5 μM EDTA and 0.05% P20, was passed over the sensor surface at a flow rate of 30 μl/min. After attainment of the required level of immobilisation, this was followed by a 240 sec injection (at 5 μl/min) of 1 M ethanolamine (pH 8.5) to quench any remaining active succinamide

esters. The final amount of hHKII covalently immobilised on the surface was around 8,000 RU. A two-fold dilution series of glucose ranging from 2 mM to 0.0625 mM was passed over the sensor surface. The binding curves were analysed with a one-to-one binding model using the analysis software provided with the instrument (v2.02, GE Healthcare). 15 compounds from active site and ADP binding site of hHKII (Chapter 5) were screened at 100 μM at final DMSO concentration of 1%. Compounds were initially screened at 100 μM on a surface of 40-200 RU of hHKII (flowcell-2) in 1% DMSO at 30 μl/min with a contact time of 30 sec and dissociation time of 30 sec. Solvent correction, carry-over assessment and a 25 % DMSO wash between samples were performed as standard. hHKII specific hits were further analysed with a two-fold concentration series from 100 μM to 3.12 μM in 1 % DMSO, at 30 μl/min with a 30 sec contact and dissociation time.

7.2.3 hHKII inhibitions studies using resazurin coupled assay

Glucose-6-phosphate-dehydrogenase from *Leuconostoc mesenteroides* (G6PDH) was purchased from Merck Millipore (code: 346774). NAD^+, ATP, D-glucose, diaphorase from *Clostridium kluyveri* (code: D5540) and resazurin (code: R7017) were all purchased from Sigma-Aldrich. hHKII assays were also an adaptation of the coupled enzymatic assay containing G6PDH and resazurin (Chapter 3, section 3.2.5). Briefly, test compounds (10 mM in 1 μL volume) were added to a 96 well black, opaque plate for a final compound test concentration of 100 μM in an 89 μL mixture containing 10 nM of hHKII, 18.8 mM NAD^+, 0.01 U G6PDH, 0.02 U diaphorase, 150 mM NaCl, 5 mM $MgCl_2$ and Tris (50 mM, pH 7.5). The plate was incubated for 15 min at 4^{o}C and a further 15 min at RT (21^{o}C). To initiate the reaction, 10 μL of 1 mM Glc, 5 mM ATP and 1 mM resazurin was added to the wells and fluorescence monitored at 530 nm excitation and 590 nm emission wavelength every 25 sec. A control reaction was supplemented with 1% DMSO and blank controls were made in the absence of hHKII. To account for possible inhibition of the coupled enzyme (G6PDH), all the compounds were screened to assess their activity against G6PDH in the absence of hHKII. Once more, test compounds (10 mM in 1 μL volume) were added to a 96 well black, opaque plate for a final compound test concentration of 100

μM in a 99 μL mixture containing 17 mM NAD^+, 0.01 U G6PDH, 100 μM resazurin, and 0.02 U diaphorase. The plate was incubated for 15 min at 4^oC and 15 min at RT (21^oC). To initiate the reaction 12.5 μM of G6P was added to each well. The change in fluorescence was monitored as above.

7.2.4 TbHKI inhibition studies using resazurin coupled assay

TbHKI assays were also an adaptation of the coupled enzymatic assay containing G6PDH and resazurin. Briefly, test compounds (10 mM in 1 μL volume) were added to a 96 well black, opaque plate for a final compound test concentration of 100 μM in an 89 μL mixture containing 0.6 μg of TbHKI, 18.8 mM NAD^+, 11.1 μM ATP, 0.01 U G6PDH, 0.02 U diaphorase, 5 mM $MgCl_2$ and TEA (50 mM, pH 8.0). The plate was incubated for 15 min at 4^oC and a further 15 min at RT (21^oC). To initiate the reaction, 10 μL of 1 mM glucose was added to the wells and fluorescence monitored at 530 nm excitation and 590 nm emission wavelength every 25 sec. The control reaction was supplemented with 1% DMSO and blank controls were made in the absence of TbHKI.

7.2.5 IC_{50} equation

The titration curve was fitted with log [inhibitor] vs response –variable slope (four parameters) using GraphPad 5 (GraphPad software, San Diego California USA).

The equation used is shown below:

$$Y = Bottom + (Top - Bottom)/(1 + 10^{((LogIC_{50} - LogX) \times Hill\ Slope)}) \quad \text{(Equation 7.1)}$$

Where Y= the response
X= compound concentration
Bottom= initial velocity with compounds
Top= Initial velocity of free enzyme
Hill Slope= Slope factor

7.2.6 *T. brucei* viability assay

Dr. Li-Hsuan Yen performed the growth inhibition tests of bloodstream-form *Trypanosoma brucei brucei* (strain 427) using the LILT (long incubation low inoculation test) method (Brun and Lun, 1994, Räz et al., 1997, Hoet et al., 2004). The parasites were cultured in complete HMI-9 medium (Invitrogen) containing 10% FBS (Invitrogen) at trypanosome density in each well of 2500/well. Three-fold dilution series (from 45 µM – 0.020 µM) of the indicated compounds were prepared in the plate; the maximal concentration of DMSO was 0.45%. After 72 hours of incubation at 37°C and 5% CO_2, 50 µl Alamar blue (0.2% diluted with DPBS 1x) was added to each well and plates were incubated at 37°C and 5% CO_2 for an additional 4 hours. Fluorescence was measured on the multiplate reader. The cells were counted using the Beckman Coulter cell counter. Two known anti-trypanosome drugs, Suramin and Fexinidazole, were used as controls in the assay. The EC_{50} values obtained for these controls (excel data analysis) were reproducible and consistent with the published values (Chauviere et al., 2003, Sokolova et al., 2010).

7.3 Results

7.3.1 Ligand identification from SPR

SPR is a biophysical label-free detection method of small-molecule binding affinity and kinetics. To identify the binding of small ligands to the target protein it is crucial to maintain the correct conformation, structure and binding-site accessibility. It is necessary to achieve the correct immobilisation conditions that will not affect the ligand biding characteristics. After protein immobilisation in a sensor chip (CM5 and NTA most commonly used) it is essential to study a positive control to prove that the protein of interest is in ligand binding conformation and shows the expected binding affinity (K_D). Subsequently, the ligand of interest is injected over the sensor surface with a continuous flow and its association on the target protein is monitored in real time, followed by running buffer which causes the ligand to dissociate and the dissociation rate is also monitored in real time (Cusack et al., 2015).

In our studies, glucose (Glc) was chosen as the positive control. Glc was used in a two-fold dilution series from 2 mM to 0.0165 mM and was successfully bound to hHKII as shown in Figure 7.1.

ATP was also tested on the same surface (CM5), having hHKII covalently attached. When the solution containing ATP in running buffer passed over the surface a negative binding response occurred. A negative signal could imply that the analyte binds more strongly to the reference channel. However by analysing the raw data we could see that this was not the case for our results. Another reason for a negative signal could be due to a buffer mismatch. The ATP solution was made from a 100 mM stock in weak buffer system (10 mM Hepes). Such a high concentration of ATP results in an acidic solution (pH around 3.0) thus the addition of concentrated NaOH was necessary. This could result in a buffer mismatch and different refractive index can be seen as a negative binding curve. Apart from the above scenarios, the high concentration of ATP and the running time (0.9 mM, 30 sec) in combination to the millimolar affinity for the target might result in a fast dissociation of the analyte which remains associated on the matrix longer than anticipated. This phenomenon is not matched by the reference cell where there is no specific binding and thus a negative binding response occurs. Finally, some negative binding responses originate from the different behaviour of reference and ligand channels to the injected analyte solution because of the difference in ligand density between the two channels. When the running buffer is replaced by the analyte buffer (in this case ATP containing buffer) the analyte interacts disproportionally on both surfaces. Because the volume of the ligand (protein) is larger on the ligand channel the reference channel will produce a larger bulk solvent response than the ligand surface resulting in an inaccurate reference subtraction. For these reasons the study of the ATP binding needs further optimization. However, since the K_D for Glc was as predicted (298 μM) and is close to the K_m value, we can safely assume that the Glc active site conformation maintains its functionality and can proceed to the inhibition studies.

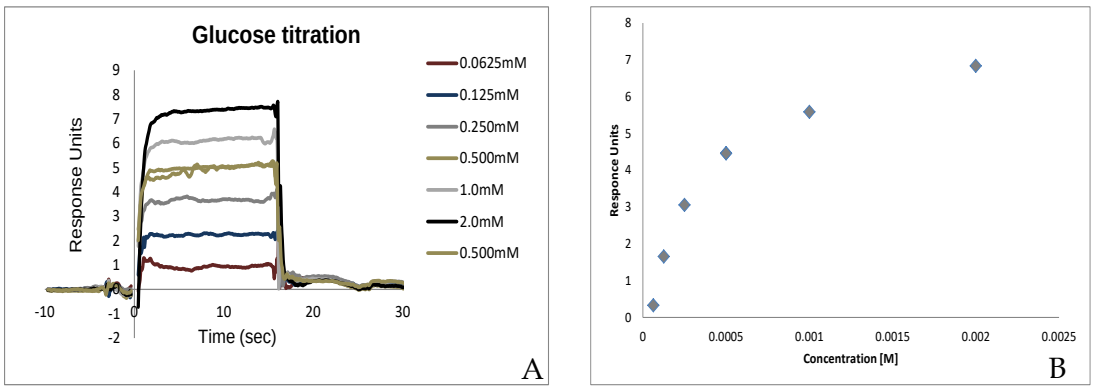


Figure 7.1: Glucose serves as a positive control for the SPR inhibition studies

Glucose, the natural substrate of hHKII, was run as a positive control with SPR (2mM to 0.0625 mM in two-fold serial dilutions). The molecule shows a non-linear concentration dependent binding to hHKII. (A). The lines are in accordance with the increasing concentration of Glc, the higher the concentration the higher the reported RU. 0.5 mM of Glc was repeated to monitor the consistency of the binding. The binding curves were analysed and fit to a 1:1 Lagmuir binding model. K_D was determined from this experiment at 298 μM.

15 compounds were analysed with SPR for binding to hHKII. The compounds tested are the virtual hits from the 1st ranking scheme of the Glc/G6P site of hHKII (AS1RS1-9, Chapter 5) and the purchased compounds from the docking of the ADP site respectively (NP1-6, Chapter 5). Following single concentration screen assessment of binding, three were found to specifically inhibit the ligand on the active surface (Figure 7.2).

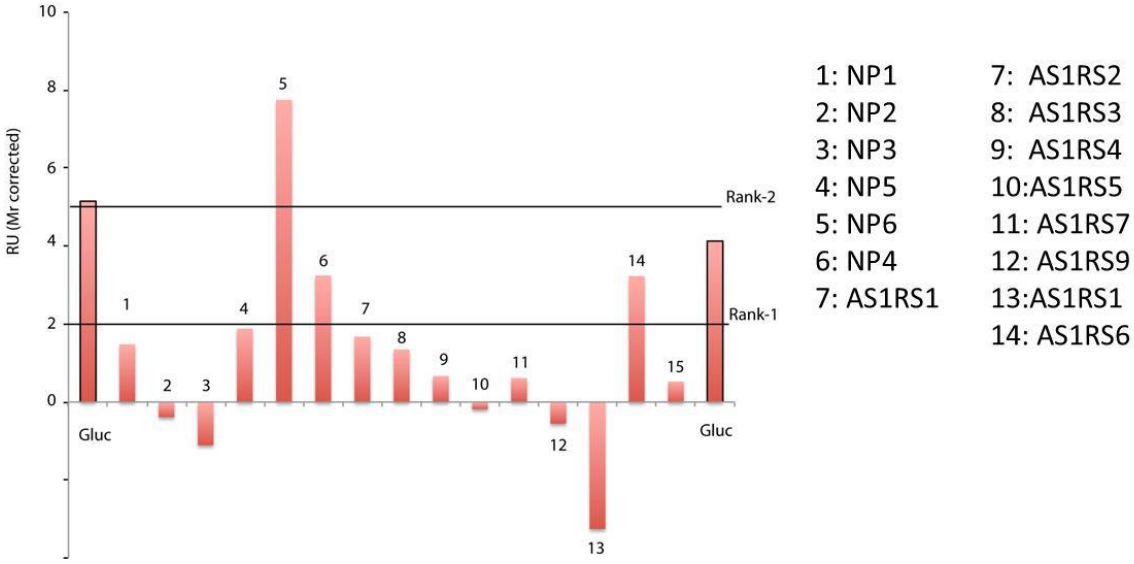


Figure 7.2: Assessment of 15 ligands tested on single concentration

The ligands were tested at single concentration of 100 μM. Each compound was measured once and is represented above by a single number. The Response Unit (RU) is corrected for the MW of each ligand. The two glucose controls are marked. The ranks (Rank-1, 2) are the theoretical boundaries for K_D between 500 μM and 100 μM. Three ligands that show a maximum K_D of 500 μM were selected for the dilution series. Compound 5 which corresponds to NP6, compound 6 to NP4 and finally compound 14 to AS1RS6.

The three compounds with a predicted K_D lower than 500 μM were screened with serial dilutions (from 100 μM to 3.1μM) to determine a more accurate K_D value(Fig. 7.3). NP4 shows a concentration dependent response which does not saturate under the present concentrations tested. There is a big jump of the RU from 50 μM to 100 μM (from ~2RU to ~5RU) which possibly indicates that the inhibitor is weak and higher concentrations should be tested in order to saturate the binding of the protein. However, the data follow a 1:1 Lagmuir model and K_D is calculated at 600 μM. NP6 binding to the surface is complicated showing non-specific binding along with possible solubility issues and the K_D was not determined. AS1RS6, is the only compound from the Glc/G6P site of hHKII. NP4 and NP6 are predicted to bind to the ADP binding site. Once more, for this compound a K_D could not be determined. The reason for this is that the ligand is very weak and no saturation was reached, or the binding of the ligand is non-specific to the protein.

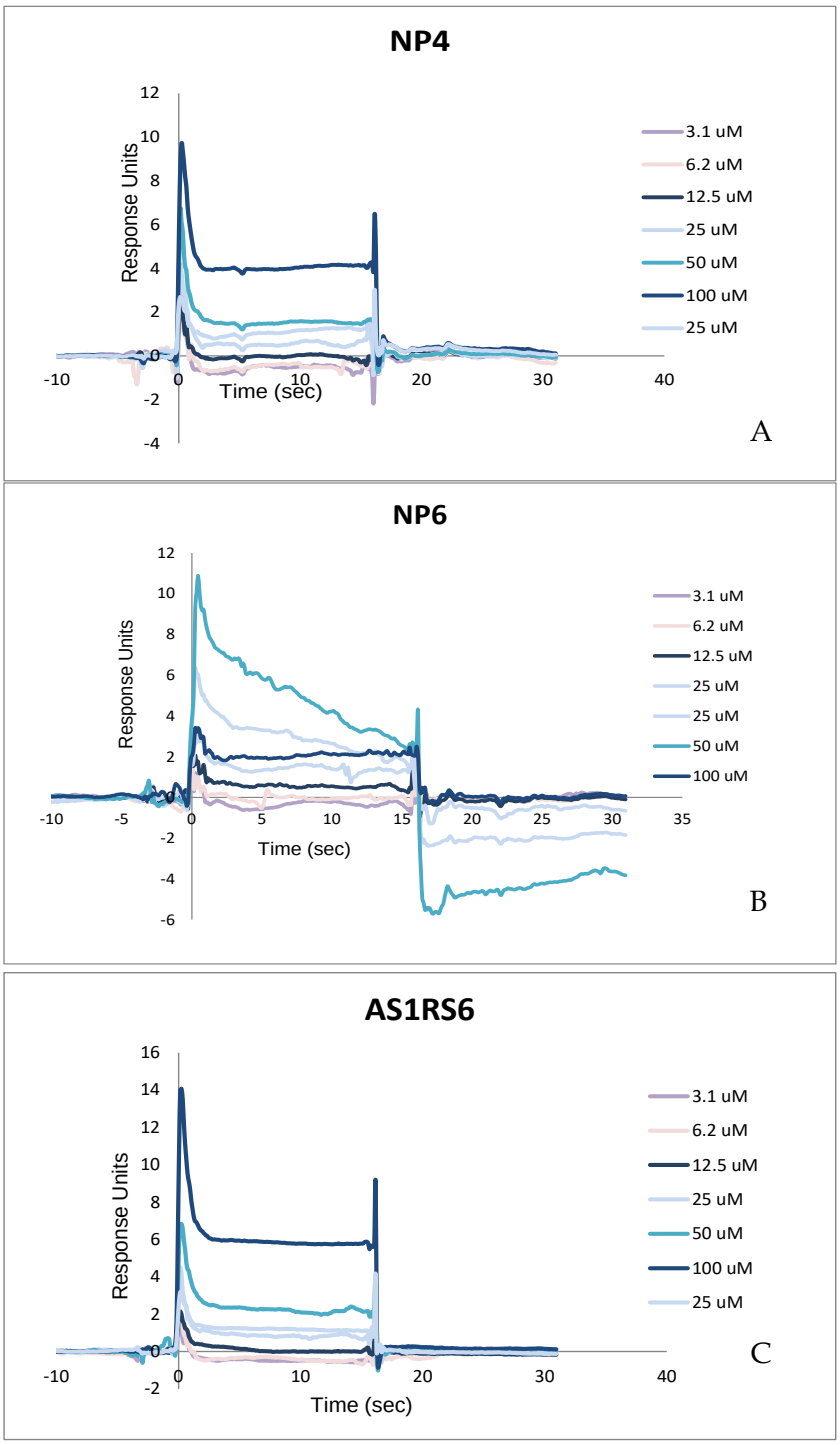


Figure 7.3: Binding sensor grams of six concentrations tested for ligands NP4, NP6 and AS1RS6

(A) NP4 predicted K_D is calculated at 600 µM. (B) NP6 binding to the surface is complicated showing non-specific binding along with possible solubility issues. (C) AS1RS6, is presumambly very weak or nor specific to the protein.

7.3.2 Assay development and optimisation for HK inhibition studies

The most common assay for the study of hexokinase activity is the enzymatic assay which couples the G6P formed via the hexokinase reaction to its catalysis by G6PDH and consequent reduction of NAD^+. The reaction is monitored via absorbance at 340 nm while using the extinction coefficient of NADH. The lower detection limit of this assay is estimated between 1-5 μM NADH (Batchelor and Zhou, 2002). It is also known that most of compounds absorb at the same wavelength. A different approach was sought to decrease the potential background signal of compounds and to improve the limit of detection of the assay to be more appropriate for HTS. An alternative enzymatic assay has been described previously where the highly fluorescent molecule resorufin is produced from the oxidation of the Amplex Rex reagent and submicromolar concentrations of NADH are detected by measuring excitation at 530 and emission detection at 590 nm (Batchelor and Zhou, 2002). More recently the same group (Batchelor and Zhou, 2004) developed a coupled enzymatic assay to quantify cell death based upon the measurement of G6PDH which is known to exit the cytosol when plasma membrane integrity is compromised. In this assay G6PDH is coupled to the reduction of resazurin to the highly fluorescent molecule resorufin.

For our studies we propose the use of a coupled-enzymatic assay where hexokinase is coupled to both G6PDH and diaphorase (the enzyme that reduces the resazurin to resorufin) in order to monitor the kinetics of hexokinase measuring increase in fluorescent intensity. As mentioned previously in Chapter 3 (Fig. 3.1), the assay relies on the reduction of the non-fluorescent resazurin dye (blue colour) to the highly fluorescent molecule resorufin (pink-red colour) from diaphorase which simultaneously oxidises NADH to NAD^+.

Regarding the choice of G6PDH, it is reported that the signal-to-noise ratio for *Leuconostoc mesenteroides* G6PDH is 1.6 times better than for baker's yeast (Zhu et al., 2009) thus the former was purchased for our studies. It is also reported that 0.1 U G6PDH/ml of reaction is both economical and sensitive so we kept the same guideline (Zhu et al., 2009).The enzyme G6PDH can use both NAD^+ and $NADP^+$ as an electron acceptor. The reaction velocity however is approximately 1.8 times

greater when NAD^+ is used so the latter was chosen for our studies. Finally, Zhu and his colleagues adopt the addition of 0.2 U diaphorase/ ml of reaction mix since the detection capability is not improved by the addition of a higher amount (Zhu et al., 2009).We also adopted the same concentration.

To validate the specificity of the reaction we performed a series of dropout experiments in which only one of the reagents was omitted from the reaction mixture each time. This would enable us to monitor if the increase in fluorescence is significant only when the complete reaction is performed. Figure 7.4 shows that there is a significant increase of the fluoresence for the complete reaction which is not affected by the concentration of the G6PDH. However, the reaction without the addition of NAD^+ (or NADH) shows considerable enzymatic activity. The fluorescence units are higher compared to the remaining four dropout experiments. In addition, the activity is significantly higher when 0.1 U of G6PDH is used. This could indicate that the enzyme G6PDH is not pure. It is possible that NAD^+ or/and $NADP^+$ is also present in the protein powder that allow the coupled reaction to take place. The signal-to-noise ratio is better when 0.01 U G6PDH is added in the reaction (2.77:1) rather than after the addition of 0.1 U G6PDH (1.86:1). The addition of 0.01 U G6PDH in the reaction mix was kept constant for all subsequent experiments.

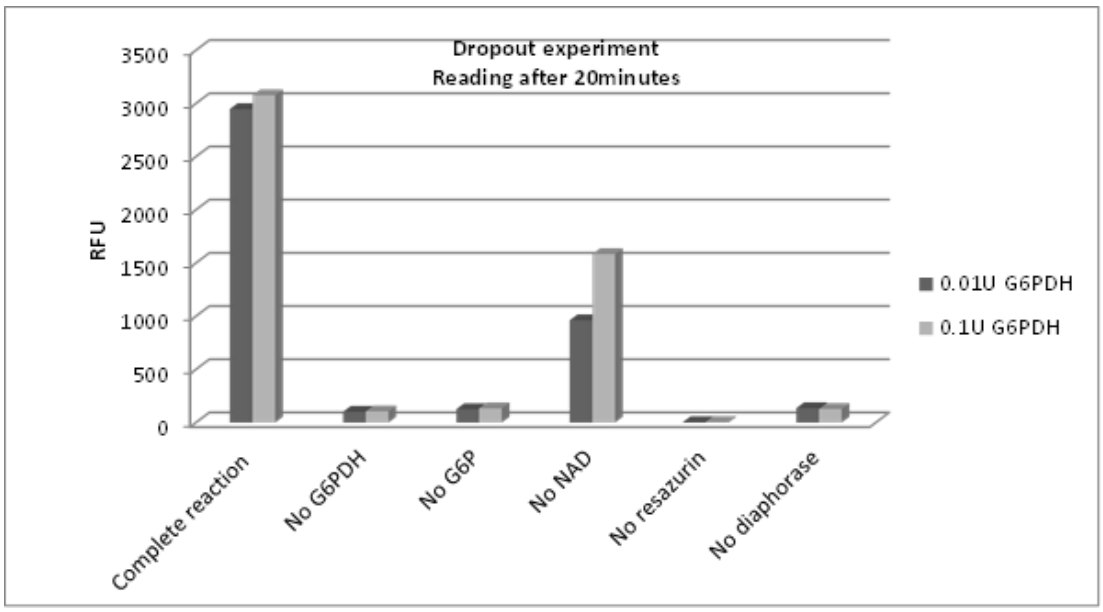


Figure 7.4: Dropout experiment for the coupled enzymatic assay

A dropout experiment was performed at two different amounts of G6PDH, 0.01 U and 0.1 U (final concentration in 0.1 ml of reaction mixture). Results are listed as follows: 1) complete reaction (0.01 or 0.1 U G6PDH, 100 μM G6P, 100 μM NAD+, 10 μM resazurin, 0.02 U diaphorase), 2) No G6PDH, 3) No G6P, 4) No NAD+, 5) No resazurin, 6) No diaphorase.

The generation of a standard curve is essential in order to be able to convert the fluorescence units into moles of product (G6P). For this reason we monitored the increase of fluorescence units as a function of G6P concentration. The concentration of G6P ranges from 25 μM to 1.56 μM. Data from three experiments are plotted in the graph. The points fit the line with a high correlation coefficient.

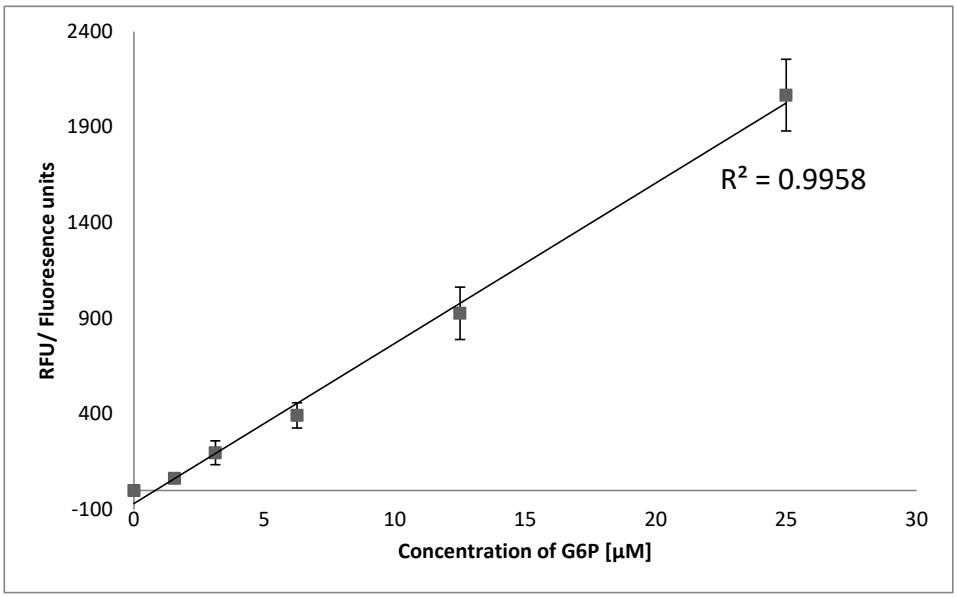


Figure 7.5: Standard curve for the coupled enzymatic assay

Background (no G6P) has been subtracted for these points. The standard curve covers a range of G6P from 25 to 0 µM G6P. The R^2 is 0.996.

To continue with the assay development we had to determine the K_m of the substrates of the coupled enzyme system (i.e G6P, NAD^+ and resazurin) under the specific conditions (buffer, pH, temperature). The assay buffer is made up with 50 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$ and 150 mM or 0mM NaCl and the assays are performed at RT (21^{o}C). For measuring the kinetics of hexokinase the reaction catalysed by this enzyme needs to be the rate-limiting step, i.e the slowest step. The remaining components of the system need to be in saturating conditions so that their concentrations do not change while measurable concentrations of the G6P (product of hexokinase reaction) are formed.

For this reason sequential reactions were performed to identify the K_m of each substrate. The initial rate of the reaction was taken from the linear slope and plotted against the substrate concentration and the data least squares fit to equation 3.4 using Kaleidagraph v4.0 software. Table 7.1 contains the calculated K_m values for each intermediate of the assay system.

Table 7.1: K_m values for the components of the enzymatic assay

The K_m values were calculated for each intermediate of the coupled enzymatic assay. The table shows the corresponding values under two different anionic strength environments. For G6P and NAD^+ the K_m values increase when the ionic strength increases.

Substrate	K_m (M) in 0 mM NaCl	K_m (M) in 150 mM NaCl
G6P	$289 \pm 17\times10^{-6}$	$680 \pm 36\times10^{-6}$
NAD^+	$348 \pm 10 \times10^{-6}$	$1{,}771 \pm 81\times10^{-6}$
NADH	$1.3 \pm 0.4\times10^{-6}$	$4 \pm 1.4\times10^{-6}$
Resazurin	$83 \pm 9.9\times10^{-6}$	$59 \pm 8.2\times10^{-6}$

To have an excess of the NAD^+ in the reaction mix, a concentration 10-fold higher than the K_m in 150 mM NaCl was used (17 mM). The K_m of resazurin was calculated using 10-fold lower concentration of diaphorase (0.002 U) since a higher amount saturated the reaction very rapidly and it was impossible to measure the initial rate. However, in the final assay to maintain the saturating conditions 10 times more enzyme is added (0.02 U). The signal to noise ratio is decreasing with increasing concentration of resazurin thus a similar value to the K_m was initially added to the reaction mix (100 μM) followed by measuring the K_m values of HK (Glc and ATP) in order to monitor the accuracy of the assay.

The predicted K_m values for hHKII are in agreement with the literature (Chapter 3). This suggests that the assay is ready to use for screening of inhibitors. As with the majority of enzymatic screens, the resazurin coupled assay is designed to identify inhibitors regardless of their mode of action. For this reason the assays are performed at the estimated K_m values for both substrates. For ATP only, a 10-fold highest concentration was tested to identify any competitive inhibitors. After the initial screen of the compounds, the actives (the ones with the best inhibition potency) were followed-up by retesting them for an IC_{50} determination in a concentration response curve (Strelow et al., 2012).

7.3.3 3-bromopyruvate as a control compound for the enzymatic assay

3-bromopyruvate (3BrPa) is a known inhibitor of the glycolytic pathway. 3BrPa is a halogenated analogue of pyruvic acid with alkylating properties. Among the other glycolytic enzymes, 3-BrPa is believed to inhibit HKII through a covalent modification of cysteine residues or by dissociating the HKII from the mitochondrial apoptosis-inducing factor (AIF) leading to cell death of the rapid proliferating cancer cells (Cardaci et al., 2012). Among a great number of studies implemented in cancer cells regarding the ability of 3BrPa to inhibit their growth (See chapter 1 for references), two reviews were found to estimate an IC_{50} and a Ki value for 3-bromopyruvate with biochemical methods. Ko and his colleagues performed studies on hexokinase activity from VX2 tumour tissues and they determined a Ki value for 3BrPa at 2.4mM (Ko et al., 2001). Wang and his colleagues developed a capillary electrophoresis-based method for screening inhibitors of hexokinase by monitoring the ADP formation. The IC_{50} against hexokinase from *Saccharomyces cerevisiae* was estimated at 3.5 mM (Wang and Kang, 2009) which is in agreement with that previously reported.

During our studies, two different IC_{50} values were obtained since a 10 min pre-incubation of the enzyme with 3BrPa leads to a lower IC_{50} than when the reaction is started immediately after the addition of 3BrPa. In more detail, when the assay was performed without incubation of 3BrPa-HK, the compound was titrated from 1000 μM to 0.97 μM (2-fold serial dilution). The final Glc concentration was 200 μM and for ATP was 700 μM. The IC_{50} was determined at 128 μM. The reaction is not saturated since the enzyme retains almost 40% of the activity at the highest concentration of the inhibitor. The IC_{50} was also calculated after pre-incubation of the enzyme with 3BrPa (10 min). 50 μM of 3BrPa were incubated with 50 nM HK and the reaction was initiated after addition of the reaction mix while diluting the complex 3BrPa-HK by 5-fold. The initial velocity versus log [3BrPa] was plotted and the IC_{50} was calculated at ~2 μM. It is obvious that the IC50 is much lower when the pre-incubation period of 10 min was performed. Apart from the dramatic drop of the IC_{50} (65 times) the enzymatic activity was completely depleted now leading to a more accurate IC_{50} determination. We could argue that the incubation period is presumably allowing time for the covalent modification to take place.

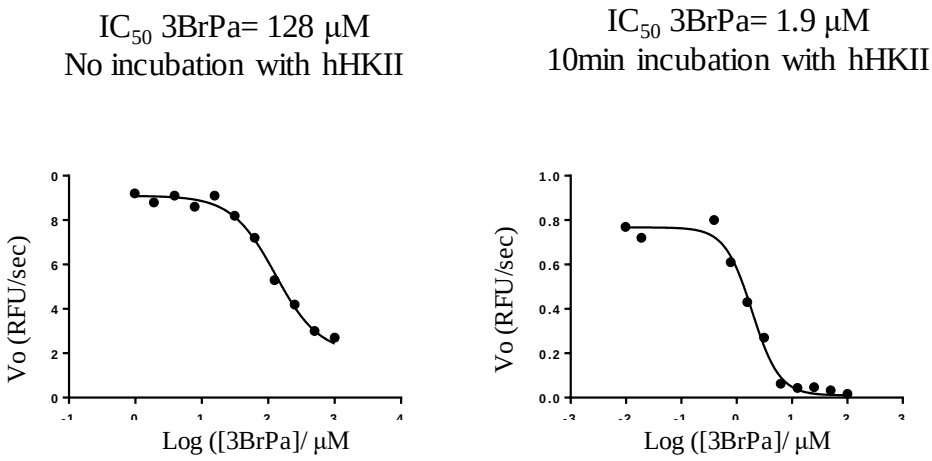


Figure 7.6: IC_{50} determination for 3BrPa under different incubation periods

3BrPa was purchased from Sigma Aldrich (code:16940). 3BrPa in powder was solubilised in Assay buffer (50 mM Tris, pH7.5, 150 mM NaCl, 5 mM $MgCl_2$). Each inhibitor concentration was measured in triplicate and the average value was used to construct the inhibition plots.

7.3.4 Enzymatic screen of 28 virtual screens against hHKII

28 compounds were tested against hHKII at a single concentration (100 μM) for an initial screen. The compounds tested are all weak inhibitors since none of them completely inhibited the enzymatic activity. The highest percentage of inhibition observed is approximately 36%.

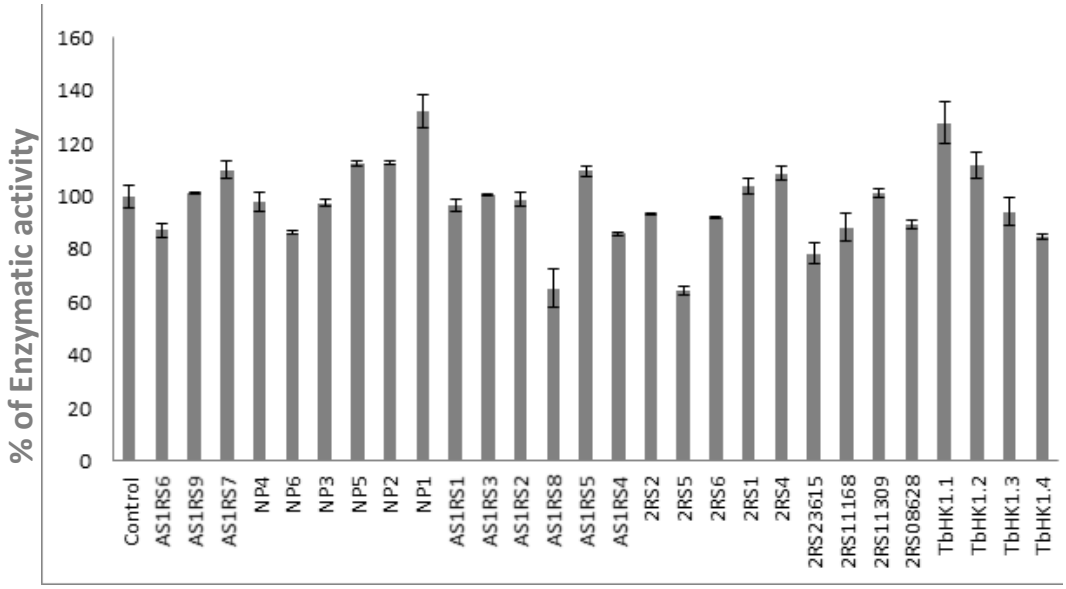


Figure 7.7: Enzymatic screen of 28 hits at 100 μM against purified hHKII

The total number of compounds in house was tested against pure purified hHKII. The protein was incubated with the compounds for 30min before addition of the substrates which initiated the reaction. Four compounds show a significant inhibition >20%, compounds 2RS23615 (22% inhibition), AS1RS8 (35% inhibition), AS1RS6 (13% inhibition) and 2RS5 (36% inhibition). These 4 compounds were further titrated to obtain a concentration response curve.

Four compounds which showed inhibition above 13% and which were dissolved in a high DMSO stock concentration were further studied in a dilution series to determine the IC_{50} of each one. As all four are predicted to be weak inhibitors the higher concentration tested was increased. More specifically for compound 2RS23615 the higher concentration tested was 200 μM, for compound AS1RS8 130 μM, for compound AS1RS6 500 μM and finally for compound 2RS5 200 μM. The IC_{50} of 2RS23615 was not determined as the concentrations tested were not high enough to

produce a sigmoidal curve. The IC_{50} for the remaining inhibitors were calculated at 447 μM for AS1RS6, 62.3 μM for AS1RS8 and finally 127 μM for 2RS5.

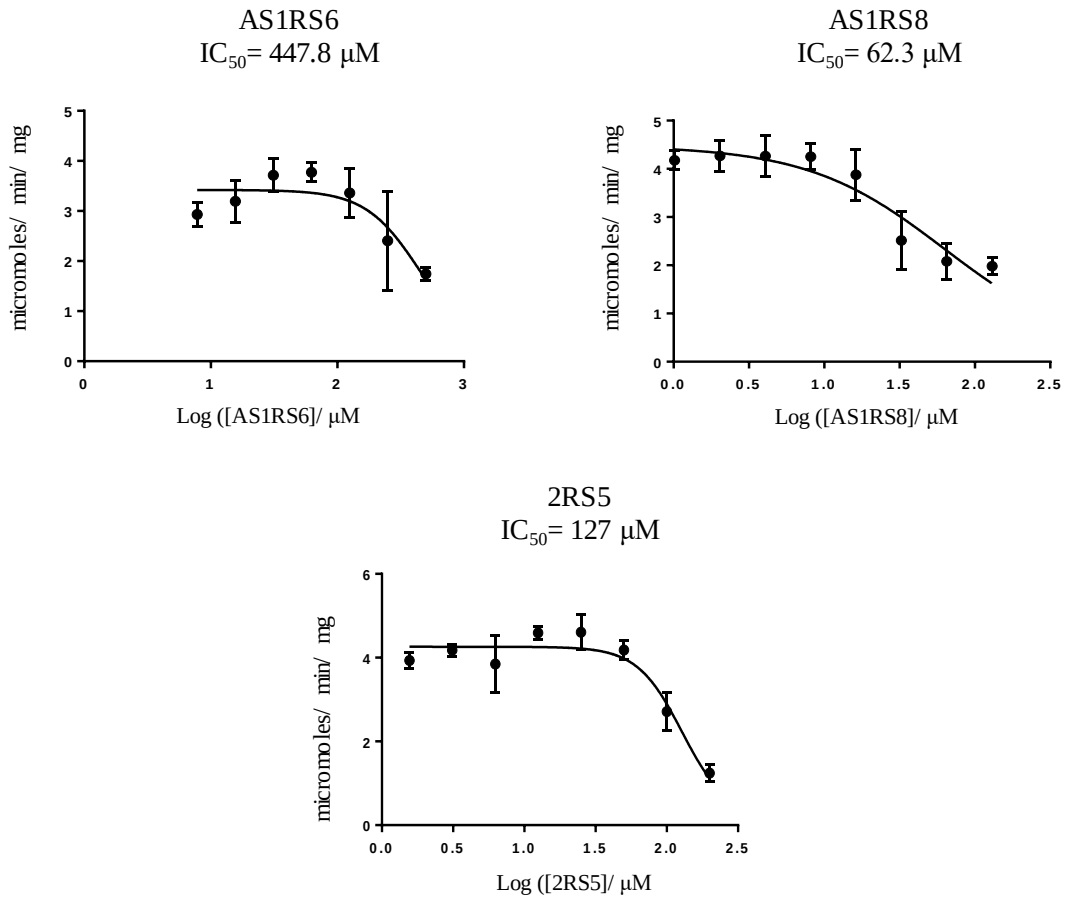


Figure 7.8: Concentration response curves for 3 inhibitors of hHKII

Concentration-response plots were constructed to determine the effect of the inhibitors on the enzymatic activity. Initial rate of reaction (*Vo* in micromoles/min/mg) is plotted against log [inhibitor concentration] at constant enzyme and substrate concentration. The reaction did not reach complete inhibition. The maximum inhibition observed is around ~60% thus the IC_{50} is an estimate. The IC_{50} was calculated by setting manually the bottom restraint at a value between 0 and 0.2 (in the same range as the non-origin starting values for the assay out-put signal). The data are a mean average of 3 measurements. The titration curve was fitted with log [Inhibitor] vs response – variable slope (four parameters) using GraphPad Prism 5.

7.3.5 Enzymatic screen of 28 virtual screens against TbHKI

28 compounds were tested against TbHKI at a single concentration (100 μM) for an initial screen. The virtual hits obtained after docking the active site of the TbHKI homology model are 4 (TbHK1-4). However all of the purchased compounds were tested against TbHKI since both enzymes are highly conserved thus it is likely that they are also active against the parasites isoform. When the assay was performed with the addition of 1×Km of ATP the initial velocity was calculated and the activity of the control reaction (no inhibitor) was compared versus the activity after the addition of the inhibitors. The results are shown in Figure 7.9.

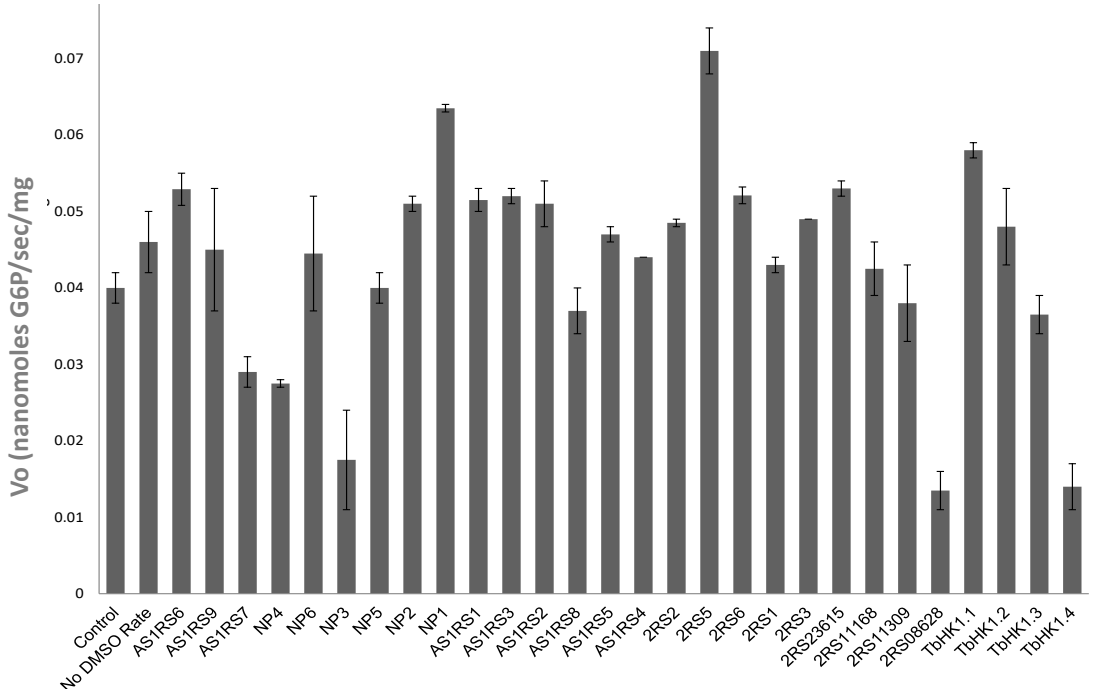


Figure 7.9: Enzymatic screen of 28 hits at 100μM against purified TbHKI

The total number of purchased compounds was tested against purified TbHKI. The protein was incubated with the compounds for 30 min before addition of the substrates (at K_m values) which initiated the reaction. The initial rates were calculated from the slope of the line which is defined as the change in the product formation divided by the change in time. Two compounds showed a significant inhibition (65%) of the initial rate at 100 μM (2RS08628 and TbHK1.4). These were further titrated to obtain an IC_{50}. The majority of the compounds were found to activate the enzyme. They are novel activators of TbHKI (see chapter 6).

Among the virtual hits for TbHKI only one inhibits the enzyme, TbHK1.4. This compound and the compound 2RS08628 (docking of the Glc/G6P site of hHKII) were titrated from 200 μM to 6.25 μM and an inhibition curve was obtained. The inhibitors did not inhibit completely the reaction. They both inhibit the reaction by 65% thus the IC_{50} is an estimate. In order to predict an accurate value of IC_{50}, higher concentrations of the compounds should be tested. However, the IC_{50} here is predicted at 80 μM and 65 μM for TbHK 1.4 and 2RS08628 respectively.

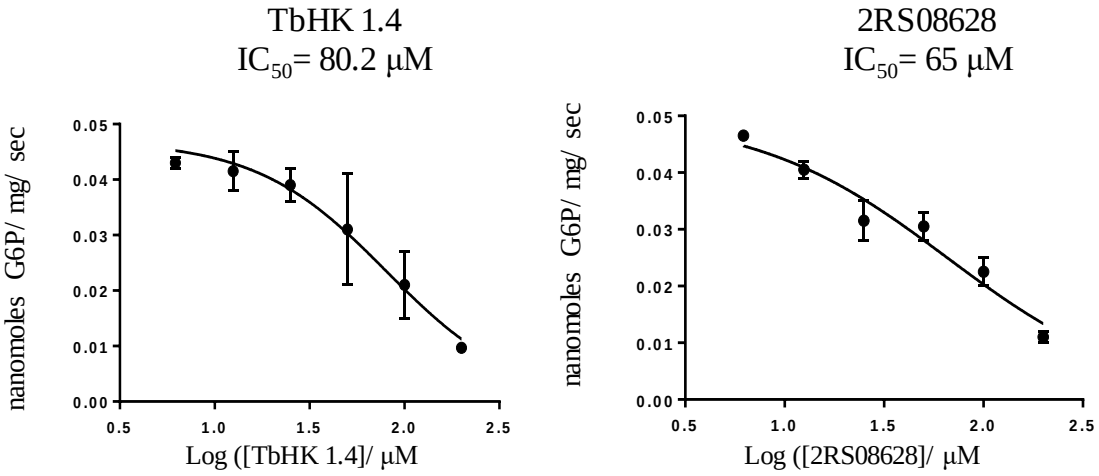


Figure 7.10: Concentration response curves for 2 inhibitors of TbHKI

Concentration-response plots were constructed to determine the concentration of the inhibitor that is required for a 50% inhibition *in vitro*. Initial rate of reaction (*Vo* in nanomoles/mg/sec) is plotted against log [inhibitor concentration] at constant enzyme and substrate concentration. The IC_{50} was calculated by setting manually the bottom restraint at a value between 0 and 0.01 (in the same range as the non-origin starting values for the assay out-put signal). The data are a mean average of 3 measurements. The titration curve was fitted with log [Inhibitor] vs response – variable slope (four parameters) using GraphPad Prism 5.

7.3.6 *T. brucei* viability assay results

All of the purchased compounds were tested with the viability assay under the conditions described previously. A list showing the % of inhibition at 45 μM with the calculated EC_{50} is shown at Table 7.2. Not all compounds have an EC_{50} value as they

did not saturate the inhibition of the parasite and the EC_{50} could not be determined. The compounds are ranked based on their potency on killing the parasites at 45 μM. The first compound on the list, 2RS2 has an EC_{50} in low μM but the same compound was found to activate the enzyme on the enzymatic assay so the results here may arrive from inhibition of another enzyme of the glycolytic pathway. In fact only 1 compound seems to be in agreement between the two assays. TbHK1.4 has an IC50 potency of 80 μM and the % of killing is >55% at 45 μM. The other compound that inhibited the enzymatic assay (2RS08628 with IC_{50} calculated at 65 μM), shows a small percent of killing the parasites (29%) which is too low to estimate an accurate $EC_{50.}$ Both compounds however could be considered as specific for hexokinase of *T. brucei*.

Figure 7.11 shows the inhibition of growth curves at different concentrations of compounds the two most potent inhibitors (2RS2 and AS1RS8). The highest concentration of compound 2RS2 tested kills all cells and the EC_{50} is calculated at 5.40μM. For compound AS1RS8 the same concentration does not lead to a complete saturation since 97% of cells died. To be more accurate the experiment should be repeated starting with a higher concentration, although the EC_{50} is estimated around ~20μM. However these results are not in agreement with the enzymatic assay as 2RS2 is an activator for TbHKI and AS1RS8 inhibits the enzyme just by 7.5% at 100μM.

Table 7.2: Cell viability assay results. The % killing of parasite cells and the EC_{50} values are shown

The compounds were assayed with the cell viability assay to monitor their potency against the parasites in BSF. The virtual hits obtained from hHKII and TbHKI docking were tested. 9 of them kill more than 55% of parasite cells and an EC_{50} was calculated. Among these 9 compounds only one of them has been found to inhibit the enzyme in the enzymatic assay (TbHK1.4). This compound was discovered after docking the active site of TbHKI. It can be considered as a true hit specifically inhibiting the target enzyme. The remaining 3 compounds purchased after the same docking (TbHK1.2, 1.3 and 1.4) seem to kill a significant percentage (>48%) of the parasites. However, the enzymatic assay results do not correlate; therefore their target could be another enzyme. (For chemical structures of the above compounds please refer to Chapter 5).

Compound ID	% Killing at 45 μM	EC_{50} (μM)	Compound ID	% Killing at 45 μM	EC_{50} (μM)
2RS2	100%	5.39	AS1RS7	41%	n.a
AS1RS8	97%	17.47	2RS4	38%	n.a
NP1	87%	n.a	AS1RS9	37%	n.a
NP6	85%	n.a	AS1RS2	36%	n.a
NP3	82%	n.a	2RS6	33%	n.a
TbHK1.1	66%	n.a	AS1RS5	33%	n.a
2RS23615	64%	n.a	AS1RS3	31%	n.a
TbHK1.3	58%	n.a	2RS08628	29%	n.a
TbHK1.4	55%	n.a	AS1RS1	29%	n.a
NP5	55%	n.a	2RS11309	25%	n.a
NP4	51%	n.a	NP2	23%	n.a
AS1RS6	50%	n.a	2RS5	21%	n.a
TbHK1.2	48%	n.a	2RS11168	20%	n.a
2RS1	48%	n.a	AS1RS4	19%	n.a

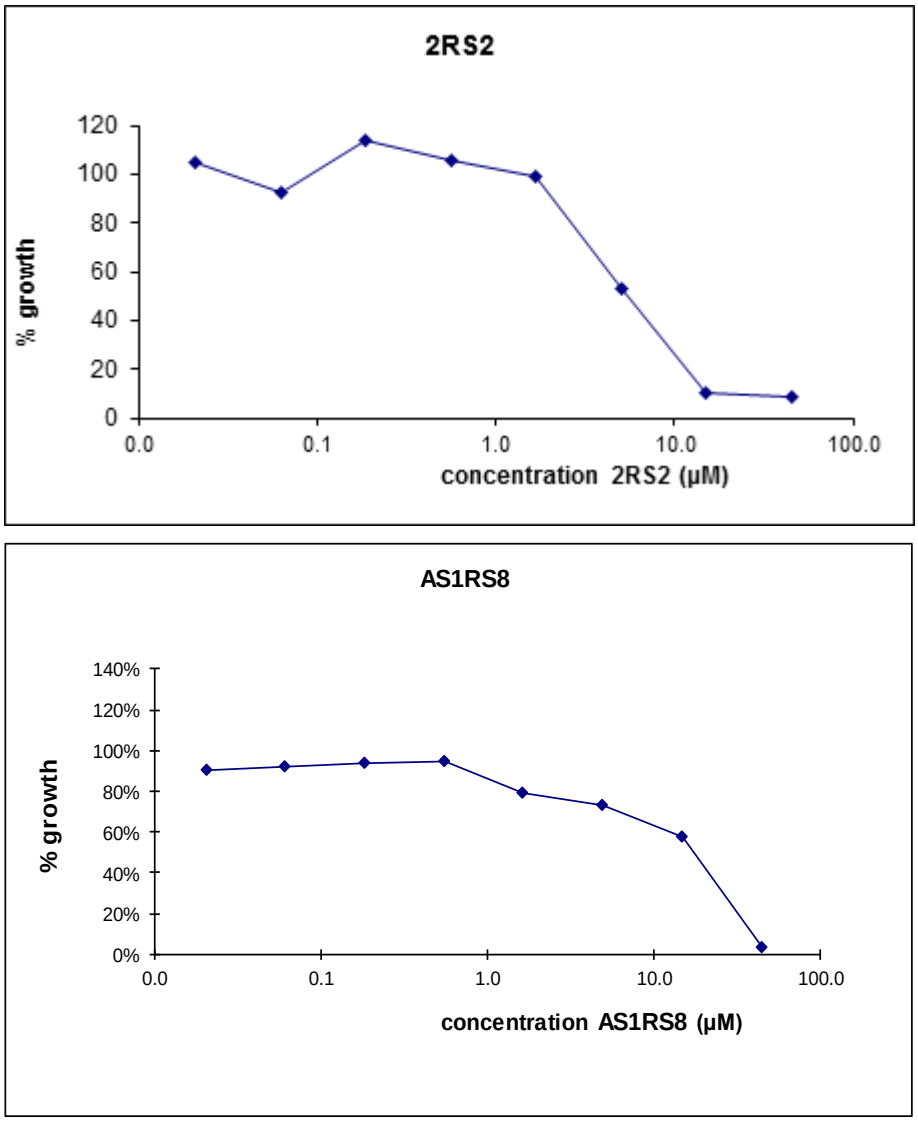


Figure 7.11: Inhibition of growth curves at different concentrations of compounds 2RS2 and AS1RS8

The highest concentration of compound 2RS2 tested kills all cells and the EC_{50} is calculated at 5.40 µM. For compound AS1RS8 the EC_{50} is estimated around ~20 µM.

7.4 Discussion and Summary

The aim of this study is the development of a novel coupled assay which will be used for inhibitor screening. The most common assay for the study of hexokinase activity is the enzymatic assay which couples the hexokinase reaction with G6PDH while measuring the absorbance at 340 nm. However, in order to decrease the potential background signal of compounds and to improve the limit of detection a different assay was developed. The novel assay couples the hexokinase reaction to G6PDH along with diaphorase, an enzyme that converts resazurin to resorufin which is a fluorescent molecule. The concentrations of the intermediates were 10-fold higher than the K_m values determined here under the specific enzymatic conditions. This way the rate limiting step belongs to the hexokinase reaction while the remaining rates are saturated. The K_m values for Glc and ATP of hexokinase were determined with the new assay and are in excellent agreement to the literature (Chapter 3).

After the K_m determination a control inhibitory assay was performed. 3BrPa was used as the control inhibitor. The IC_{50} calculated using the assay developed here is much lower than the reported values. One can bear in mind that the IC_{50} depends on the enzyme concentration. The reported values elsewhere (Refer to 7.3.3) have been determined following a completely different strategy making it hard to compare the results.

In addition, the pre-incubation time of the enzyme with the ligand decreased the IC_{50} by more than 65 times. This dramatic drop was anticipated as the incubation time gives extra time for the ligand to saturate the binding sites before addition of substrate. For 3BrPa, which is believed to covalently modify the hHKII, a pre-incubation time could lead to a lower IC_{50} value since the ligand has modified the enzyme prior to the substrate addition which initiates the reaction. However, no comparison between the two methods could be made because different concentrations are used for the two experiments. The final hHKII concentration is the same for both assays but the incubation period was carried out with a 5-fold higher concentration. The concentrations of 3BrPa tested are also different since the activity was completely inhibited when the same concentration of compound was used for

incubation. Thus 10-fold lower compound concentration was used for incubation with the enzyme.

Screening of inhibitors was performed. Firstly, the majority of purchased compounds was screened at constant concentration and the primary hits were further studied with a concentration-response curve. Comparing the results to the previously performed SPR studies, two ligands were found to inhibit hHKII based on the enzymatic assay, AS1RS6 and NP6. The remaining compound, NP4, was not detected as a hit based on the enzymatic assay so the two methods do not agree. It is likely that the binding of the ligand on the surface was not specific or that SPR is a more powerful method and is able to identify the interaction with the protein and those compounds.

The most potent inhibitor based on the IC_{50} is the ligand AS1RS8 which was discovered from docking of the hHKII active site with an estimated IC_{50} at 62.3 μM. The predicted binding site and interactions are discussed in Chapter 5. The other compounds range from 127 μM (2RS5) to 448 μM (AS1RS6). The compounds tested did not inhibit completely the reaction. They were found to inhibit the reaction with a maximum of ~60% thus the IC_{50} is an estimate. In order to have an accurate IC_{50} the experiments should be repeated with a higher concentration of ligand. However it is not known if the compounds will be soluble at higher concentrations.

The same strategy was followed for screening against TbHKI. The compounds were initially screened at a constant concentration of Glc equal to the K_m (100μM) and different concentrations of ATP, one that was equal to the K_m (10μM) and one 10 times higher than the K_m (100 μM) to distinguish a competitive inhibitor from a noncompetitive. When the assay was performed with the addition of $10\times K_m$ of ATP the reactions were completely depleted, which suggests that all of the inhibitors are competitive against ATP. Among the 28 compounds, two were further studied with a response concentration curve and two IC_{50} values at micromolar ranges were estimated. TbHK1.4 (80 μM) was specifically purchased after docking of the model for TbHKI. The other inhibitor, 2RS08628 (65 μM), was obtained from the docking of hHKII. Dr. Li-Hsuan Yen performed the cell viability results to show whether these compounds were effective in killing *T.brucei* parasites *in vitro*. The compounds were tested at 45 μM. TbHK1.4 shows 55% of killing and 2RS08628 only 29%.

These do not show strong inhibition potency and an EC_{50} was not determined. In contrast, the top two compounds that show almost 100% killing with an estimated EC_{50} at low micromolar were not found to inhibit the enzymatic reaction. This could probably be due to the fact that the real target of the ligands is a different enzyme.

8. Chapter 8: Conclusion and future outlook

8.1 Overall conclusions of the work

As explained in detail in the Introduction the aim of the present study is the discovery of novel potent and specific inhibitors against hexokinase from human and *Trypanosoma brucei.* Hexokinase, the first enzyme of the glycolytic pathway, is a validated target for therapeutic development of cancer and Human African Trypanosomiasis respectively. A structure-based strategy was used to identify and test novel inhibitor molecules. Despite testing a variety of constructs and testing a large number of crystallisation conditions it was not possible to obtain suitable crystals for X-ray studies. However published X-ray structures were used and provided suitable templates for modelling and virtual screening studies. Large virtual chemical libraries were used in docking studies and top hits (best binders) were selected and tested for enzyme inhibition against the target protein. The steps performed here in this Structure-Based study are summarised below:

Protein purification and characterisation: Regarding the hHKII, different constructs were produced and purified (Chapter 2 Fig. 2.1); with the 6His-tag protein used for the inhibition studies. The protein elutes as a monomer of the correct molecular mass (based on GF) and is extremely pure (>95%). SEC-MALS further supported this notion as the predicted molecular mass is 105kDa (theoretical MW is 102kDa) with an excellent monodispersity. The protein was extensively characterised by DLS, CD and SPR and it was proven to be folded, non-aggregated and active with the K_d for Glc to be determined for the very first time using the SPR method (217μM). It was very interesting to discover by performing the SEC-MALS method that the protein aggregates immediately when is exchanged to buffer with low anionic strength. A series of NaCl concentrations were tested (500, 250, 100, 0mM) and it was found that protein aggregates in lower concentrations of NaCl without losing its activity. The activity is actually improved when the assay is performed with 0mM NaCl. In addition, the protein seems to increase its hydrodynamic radius in a significant way when exchanged to 0mM NaCl which could indicate a mass conformation. One hypothesis behind this is that the long helix holding the N and C domains together might be flexible resulting in an elongated

form of the protein. The electrostatic surface potential of this long helix was studied with PyMOL and we did not find any possible interactions as the helix itself and the surrounding environment are mostly negatively charged. On the contrary, the domains are known to be flexible since they adopt a closed conformation upon the substrate biding (Kuser et al., 2008) thus they could adopt different conformations when the anionic strength lowers/ increases and this might result in a different hydrodynamic radius of the protein which can be seen with SEC-MALS. The protein did not crystallise even though many different conditions were tested (Appendix A). For this reason the N and C terminal domains were purified separately and a low resolution SAXS structure was determined for the N- terminal domain. The C-terminal domain is likely to exist in equilibrium of monomeric and dimeric forms making the SAXS analysis unreliable for the development of a 3D envelope. On the contrary, the 6His-tag hHKII and the N-terminal domain constructs of hHKII are monomeric in solution and a 3D envelope was successfully constructed from SAXS data. The two macromolecules do not differ much in overall architecture from the crystallographic atomic resolution structures (PDB code: 2NZT).

The untagged protein was found to be inactive while not very pure after a 3 step-purification explained in detail in Chapter 2, section 2.2.4. Appendix B shows a number of different protocols investigated with unfavourable results. The purification of the protein needs further optimisation in order to obtain a pure and active form.

Regarding TbHKI, the expression trials were unsuccessful to yield enough soluble protein. Most of TbHKI is expressed in IBs. For this reason, two different refolding protocols were tested; the second trial using the Optimised Buffer (Chapter 2, section 2.2.10) resulted in an improved activity. The latter purified protein was unstable in solution and it precipitated during concentration and/or buffer exchange so we could not continue the studies in order to investigate the oligomerisation state of the protein; instead this protein was used to screen the potential inhibitors. The protein was active with K_m values of 65μM for Glc and 6μM for ATP. The affinity for ATP is increased compared to the reported values (Morris et al., 2006). It is shown that TbHKI can form heterohexameric assemblies (Chambers et al., 2008b) albeit the monomeric assembly (based on the GF) was studied here.

Virtual Screening: The Glc/G6P site and the ADP site of hHKI were used as templates for docking studies using two different softwares, AutoDock and Vina. hHKI has very similar sites with isoform II and the structure of the former (PDB code: 1DGK) was used as it was found to be in a better conformation which is expected to accommodate lead-size compounds. For TbHKI, since the crystal structure is not resolved to date, two homology models using the on-line tools Phyre2 and i-TASSER were generated. The structure of hHKI was used as a template (PDB code: 1DGK, 37% identity based on BlastP) for the 1:1 threading. Between the two homology models obtained the model from Phyre2 was used for our docking studies since the amino acid residues of the active site are in a ligand bound conformation (as shown from structure superposition with 1DGK). 28 compounds in total were purchased from the docking studies which were further tested to determine their inhibitory potency against hHKII and TbHKI.

The computational docking studies in the present work, led to the discovery of a number of inhibitors for both targets. More specifically for hHKII the successful ligands reach a percentage of 30% (8 out of 28) and 25% respectively (1 out of 4) for TbHKI. We could argue that these results are a very good starting point that could be followed by ligand modifications in order to build more efficient actives. It is widely known that many drugs in the market have been discovered through the process of SBDD (i.e Tamiflu, Gleevec, Exanta, Nesina). Based on our results we are confident that the *in silico* methods are of great importance in drug discovery and can result in the discovery of actives in a cost-effective and time-saving manner. However, even though the molecular docking was successful in identifying a number of inhibitors the predicted binding modes could not be confirmed; this unfortunately makes it impossible to fully evaluate the docking programs outcome.

Inhibition studies on hHKII and TbHKI: The purchased compounds after the virtual screening were assayed in order to determine their inhibitory potency and determine an IC50 value if possible. Regarding the hHKII, 8 compounds were found to inhibit (>12%) the enzymatic reaction at 100 μM. The SAR analysis is explained at Chapter 5, section 5.12. The most potent inhibitor, AS1RS8, is predicted to bind to

the ADP site and not to the Glc/G6P site. It is predicted to form fewer hydrogen bonds, compared to the total of virtual hits obtained. However, the higher potency could be attributed to the hydrophobic interactions seen by studying the predicted binding mode. The IC_{50} is estimated at ~62 μM. 2RS5 which is predicted to bind to the Glc/G6P site has an IC_{50} of 127 μM while AS1RS6 (same site) ranks in the final position with IC_{50} at ~448 μM. Unfortunately, no crystal structure was obtained with a ligand bound to the protein, thus future work should include crystallisation trials of at least the most potent ligand, AS1RS8.

Table 8.1: Summary of 8 compounds that show >12% inhibition of the enzymatic assay at 100 μM

The IC50 was successfully determined for three inhibitors. The remaining five inhibitors are predicted to be quite weak and the higher stock concentration available to test would not be high enough for an IC_{50} determination.

Compound ID	*% Inhibition at 100 μM*	*IC_{50}*
AS1RS8	-35%	62 μM
2RS5	-36%	127 μM
AS1RS6	-13%	448 μM
TbHK1.4	-16	Could not be determined
NP6	-14	Could not be determined
2RS11168	-12%	Could not be determined
2RS23615	-22%	Could not be determined
AS1RS4	-15%	Could not be determined

In respect to TbHKI, four compounds were purchased after the performance of the docking studies (Chapter 6, section 6.3.3 for chemical structure) which are predicted to bind to different pockets in close proximity (Fig. 6.3). Because the enzyme from *T. brucei* is expected to have maintained a conserved binding site all compounds (hits against hHKII as well) were screened against the parasite isoform. The screening of the compounds with an enzymatic assay was followed by testing them directly against the parasites (Chapter 7, section 7.2.5). Among the 28 compounds an IC_{50} value at micromolar range was estimated for TbHK1.4 (80 μM) and 2RS08628 (65 μM). The cell viability assay was performed to show whether these compounds were effective in killing *T. brucei* parasites *in vitro*. The compounds were tested at 45 μM and TbHK1.4 shows 55% of killing and 2RS08628 only 29%. The potency is not high enough for an accurate EC_{50} to be determined. On the contrary, a low EC_{50} was estimated for 2RS2 and AS1RS8 (5 μM and 17 μM respectively) which show almost 100% killing of parasites. However, these compounds did not inhibit the enzymatic reaction suggesting that the real target of the ligands is a different enzyme.

Table 8.2: Summary of 4 compounds that inhibit either the enzymatic activity in vitro or kill the parasites in vitro

Among the 28 compounds tested against purified TbHKI two were found to inhibit the enzyme in low micromolar. Two more were found to completely kill the parasites and an EC_{50} was determined. However the target is not the hexokinase enzyme.

Compound ID	*% Inhibition (-)/ Activation (+)*	*IC_{50}*	*% killing*	*EC_{50}*
2RS08628	-66.25	65 μM	29	-
TbHK1.4	-65	80 μM	55	-
2RS2	+21.25	-	100%	5 μM
AS1RS8	-7.5	-	100%	17 μM

Future work: Further work is required in order to characterise the TbHKI regarding its oligomeric state, secondary structure as a function of the storage conditions of the recombinant enzyme. All results presented here were obtained using protein directly from the purification process. Moreover, since the enzymatic assay revealed that most of the ligands are activators it would worth studying the mechanism behind the activation of the protein. It would be interesting to test whether the activators increase the affinity for Glc or ATP, or whether they only increase the V_{max}. Kinetic measurements with different concentrations of enzyme varying the Glc or ATP concentration could show if there is a cooperative behaviour depending on the concentration of TbHKI in order to suggest that TbHKI behaves like *T. cruzi* hexokinase showing a hysteretic behaviour (Acosta et al., 2014).

For hHKII it is essential to obtain a crystal structure of the protein with the present inhibitors to study the interactions formed in order to perform a more sophisticated SAR analysis. It would be also very interesting to obtain the structure of the protein under different anionic environment to monitor the transition mode of the protein to an elongated form as SEC-MALS predicts. Finally, to propose whether the inhibitors are Glc or ATP competitive, more enzymatic assays need to be performed with excess of each substrate at each time.

References

2017. UniProt: the universal protein knowledgebase. *Nucleic Acids Research,* 45**,** D158-D169.

ACOSTA, H., CÁCERES, A., GONZÁLEZ-MARCANO, E., QUIÑONES, W., AVILÁN, L., DUBOURDIEU, M. & CONCEPCIÓN, J. L. 2014. Hysteresis and positive cooperativity as possible regulatory mechanisms of Trypanosoma cruzi hexokinase activity. *Molecular and Biochemical Parasitology,* 198**,** 82-91.

ALBERT, M.-A., HAANSTRA, J. R., HANNAERT, V., VAN ROY, J., OPPERDOES, F. R., BAKKER, B. M. & MICHELS, P. A. 2005. Experimental and in silico analyses of glycolytic flux control in bloodstream form Trypanosoma brucei. *Journal of Biological Chemistry,* 280**,** 28306-28315.

ALESHIN, A. E., KIRBY, C., LIU, X., BOURENKOV, G. P., BARTUNIK, H. D., FROMM, H. J. & HONZATKO, R. B. 2000. Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites and conformational changes relevant to allosteric regulation1. *Journal of Molecular Biology,* 296**,** 1001-1015.

ANDERSON, A. C. 2003. The Process of Structure-Based Drug Design. *Chemistry & Biology,* 10**,** 787-797.

ARDEHALI, H., YANO, Y., PRINTZ, R. L., KOCH, S., WHITESELL, R. R., MAY, J. M. & GRANNER, D. K. 1996. Functional Organization of Mammalian Hexokinase II: RETENTION OF CATALYTIC AND REGULATORY FUNCTIONS IN BOTH THE NH- AND COOH-TERMINAL HALVES. *Journal of Biological Chemistry,* 271**,** 1849-1852.

ARORA, K. A. & PEDERSEN, P. 1988. Functional significance of mitochondrial bound hexokinase in tumor cell metabolism. Evidence for preferential phosphorylation of glucose by intramitochondrially generated ATP. *Journal of Biological Chemistry,* 263**,** 17422-17428.

ARORA, K. K., FILBURN, C. R. & PEDERSEN, P. L. 1993. Structure/function relationships in hexokinase. Site-directed mutational analyses and characterization of overexpressed fragments implicate different functions for the N- and C-terminal halves of the enzyme. *Journal of Biological Chemistry,* 268**,** 18259-18266.

ASLETT, M., AURRECOECHEA, C., BERRIMAN, M., BRESTELLI, J., BRUNK, B. P., CARRINGTON, M., DEPLEDGE, D. P., FISCHER, S., GAJRIA, B. & GAO, X. 2009. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic acids research,* 38**,** D457-D462.

BALLESTER, P. J. & MITCHELL, J. B. 2010. A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. *Bioinformatics,* 26**,** 1169-1175.

BATCHELOR, R. H. & ZHOU, M. 2002. A Resorufin-Based Fluorescent Assay for Quantifying NADH. *Analytical Biochemistry,* 305**,** 118-119.

BATCHELOR, R. H. & ZHOU, M. 2004. Use of cellular glucose-6-phosphate dehydrogenase for cell quantitation: applications in cytotoxicity and apoptosis assays. *Anal Biochem,* 329**,** 35-42.

BONDOS, S. E. & BICKNELL, A. 2003. Detection and prevention of protein aggregation before, during, and after purification. *Analytical Biochemistry,* 316**,** 223-231.

BONNIER, F., KEATING, M., WROBEL, T., MAJZNER, K., BARANSKA, M., GARCIA, A., BLANCO, A. & BYRNE, H. J. 2014. Cell viability assessment using the Alamar blue assay: A comparison of 2D and 3D cell culture models. *Toxicol In Vitro,* 29**,** 124-31.

BROOKS, H. B., GEEGANAGE, S., KAHL, S. D., MONTROSE, C., SITTAMPALAM, S., SMITH, M. C. & WEIDNER, J. R. 2012. Basics of enzymatic assays for HTS.

BRUN, R. & LUN, Z.-R. 1994. Drug sensitivity of Chinese Trypanosoma evansi and Trypanosoma equiperdum isolates. *Veterinary Parasitology,* 52**,** 37-46.

CARDACI, S., DESIDERI, E. & CIRIOLO, M. R. 2012. Targeting aerobic glycolysis: 3-bromopyruvate as a promising anticancer drug. *J Bioenerg Biomembr,* 44**,** 17-29.

CÁRDENAS, M. A. L., CORNISH-BOWDEN, A. & URETA, T. 1998. Evolution and regulatory role of the hexokinases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1401**,** 242-264.

CHAMBERS, J. W., FOWLER, M. L., MORRIS, M. T. & MORRIS, J. C. 2008a. The anti-trypanosomal agent lonidamine inhibits Trypanosoma brucei hexokinase 1. *Mol Biochem Parasitol,* 158**,** 202-7.

CHAMBERS, J. W., KEARNS, M. T., MORRIS, M. T. & MORRIS, J. C. 2008b. Assembly of Heterohexameric Trypanosome Hexokinases Reveals That Hexokinase 2 Is a Regulable Enzyme. *Journal of Biological Chemistry,* 283**,** 14963-14970.

CHANG, M. W., AYENI, C., BREUER, S. & TORBETT, B. E. 2010. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PloS one,* 5**,** e11955.

CHARIFSON, P. S., CORKERY, J. J., MURCKO, M. A. & WALTERS, W. P. 1999. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of medicinal chemistry,* 42**,** 5100-5109.

CHAUVIERE, G., BOUTEILLE, B., ENANGA, B., DE ALBUQUERQUE, C., CROFT, S. L., DUMAS, M. & PERIE, J. 2003. Synthesis and biological activity of nitro heterocycles analogous to megazol, a trypanocidal lead. *J Med Chem,* 46**,** 427-40.

CHEN, Z., ZHANG, H., LU, W. & HUANG, P. 2009. Role of mitochondria-associated hexokinase II in cancer cell death induced by 3-bromopyruvate. *Biochim Biophys Acta,* 1787**,** 553-60.

CHOTHIA, C. & LESK, A. M. 1986. The relation between the divergence of sequence and structure in proteins. *The EMBO Journal,* 5**,** 823-826.

CUSACK, K. P., WANG, Y., HOEMANN, M. Z., MARJANOVIC, J., HEYM, R. G. & VASUDEVAN, A. 2015. Design strategies to address kinetics of drug binding and residence time. *Bioorganic & medicinal chemistry letters,* 25**,** 2019-2027.

DODSON, H. C., LYDA, T. A., CHAMBERS, J. W., MORRIS, M. T., CHRISTENSEN, K. A. & MORRIS, J. C. 2011. Quercetin, a fluorescent bioflavanoid, inhibits Trypanosoma brucei hexokinase 1. *Experimental parasitology,* 127**,** 423-428.

DRWAL, M. N. & GRIFFITH, R. 2013. Combination of ligand- and structure-based methods in virtual screening. *Drug Discovery Today: Technologies,* 10**,** e395-e401.

DURRANT, J. D. & MCCAMMON, J. A. 2010. NNScore: A neural-network-based scoring function for the characterization of protein− ligand complexes. *Journal of chemical information and modeling,* 50**,** 1865-1871.

DURRANT, J. D. & MCCAMMON, J. A. 2011. NNScore 2.0: a neural-network receptor–ligand scoring function. *Journal of chemical information and modeling,* 51**,** 2897-2903.

ERICKSON, J., NEIDHART, D. J., VANDRIE, J., KEMPF, D. J. & PAUL, D. A. 1990. Design, Activity, and 2.8 (angstrom) Crystal Structure of a C (2) Symmetric Inhibitor Complexed to HIV-1 Protease. *Science,* 249**,** 527.

FERREIRA, L., DOS SANTOS, R., OLIVA, G. & ANDRICOPULO, A. 2015. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules,* 20**,** 13384.

FLORIDI, A., PAGGI, M. G., D'ATRI, S., DE MARTINO, C., MARCANTE, M. L., SILVESTRINI, B. & CAPUTO, A. 1981. Effect of Lonidamine on the Energy Metabolism of Ehrlich Ascites Tumor Cells. *Cancer Research,* 41**,** 4661-4666.

FRANKE, D. & SVERGUN, D. I. 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of applied crystallography,* 42**,** 342-346.

GOLDIN, N., ARZOINE, L., HEYFETS, A., ISRAELSON, A., ZASLAVSKY, Z., BRAVMAN, T., BRONNER, V., NOTCOVICH, A., SHOSHAN-BARMATZ, V. & FLESCHER, E. 2008. Methyl jasmonate binds to and detaches mitochondria-bound hexokinase. *Oncogene,* 27**,** 4636-43.

GOLDSTEIN, L. 1972. Microenvironmental effects on enzyme catalysis. Kinetic study of polyanionic and polycationic derivatives of chymotrypsin. *Biochemistry,* 11**,** 4072-4084.

GROSSBARD, L. & SCHIMKE, R. T. 1966. Multiple Hexokinases of Rat Tissues: PURIFICATION AND COMPARISON OF SOLUBLE FORMS. *Journal of Biological Chemistry,* 241**,** 3546-3560.

GUILBAULT, G. G. 1975. [11] Fluorometric determination of dehydrogenase activity using resorufin. *Methods in enzymology,* 41**,** 53-56.

HAMADA, H., ARAKAWA, T. & SHIRAKI, K. 2009. Effect of additives on protein aggregation. *Current pharmaceutical biotechnology,* 10**,** 400-407.

HANN, M. M. & OPREA, T. I. 2004. Pursuing the leadlikeness concept in pharmaceutical research. *Current Opinion in Chemical Biology,* 8**,** 255-263.

HARTL, F. U., BRACHER, A. & HAYER-HARTL, M. 2011. Molecular chaperones in protein folding and proteostasis. *Nature,* 475**,** 324-332.

HOET, S., OPPERDOES, F., BRUN, R., ADJAKIDJÉ, V. & QUETIN-LECLERCQ, J. 2004. In vitro antitrypanosomal activity of ethnopharmacologically selected Beninese plants. *Journal of Ethnopharmacology,* 91**,** 37-42.

HOUSTON, D. R. & WALKINSHAW, M. D. 2013. Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context. *Journal of Chemical Information and Modeling,* 53**,** 384-390.

HSIN, K.-Y., MORGAN, H. P., SHAVE, S. R., HINTON, A. C., TAYLOR, P. & WALKINSHAW, M. D. 2011. EDULISS: a small-molecule database with data-mining and pharmacophore searching capabilities. *Nucleic acids research,* 39**,** D1042-D1048.

HUANG, S.-Y. & ZOU, X. 2010. Advances and challenges in protein-ligand docking. *International journal of molecular sciences,* 11**,** 3016-3034.

HUEY, R., MORRIS, G. M., OLSON, A. J. & GOODSELL, D. S. 2007. A semiempirical free energy force field with charge-based desolvation. *J Comput Chem,* 28**,** 1145-52.

IRWIN, D. M. & TAN, H. 2008. Molecular evolution of the vertebrate hexokinase gene family: Identification of a conserved fifth vertebrate hexokinase gene. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics,* 3**,** 96-107.

JACQUES, D. A. & TREWHELLA, J. 2010. Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls. *Protein Science,* 19**,** 642-657.

JOICE, A. C., LYDA, T. L., SAYCE, A. C., VERPLAETSE, E., MORRIS, M. T., MICHELS, P. A. M., ROBINSON, D. R. & MORRIS, J. C. 2012. Extra glycosomal localization of Trypanosoma brucei hexokinase 2. *International journal for parasitology,* 42**,** 401-409.

JORGE MOURA BARBOSA, A. & DEL RIO, A. 2012. Freely accessible databases of commercial compounds for high-throughput virtual screenings. *Current topics in medicinal chemistry,* 12**,** 866-877.

KAMATA, K., MITSUYA, M., NISHIMURA, T., EIKI, J.-I. & NAGATA, Y. 2004. Structural Basis for Allosteric Regulation of the Monomeric Allosteric Enzyme Human Glucokinase. *Structure,* 12**,** 429-438.

KATZEN, H. M. & SCHIMKE, R. T. 1965. Multiple forms of hexokinase in the rat: tissue distribution, age dependency, and properties. *Proceedings of the National Academy of Sciences of the United States of America,* 54**,** 1218-1225.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols,* 10**,** 845-858.

KELLY, S. M., JESS, T. J. & PRICE, N. C. 2005. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* 1751**,** 119-139.

KO, Y. H., PEDERSEN, P. L. & GESCHWIND, J. F. 2001. Glucose catabolism in the rabbit VX2 tumor model for liver cancer: characterization and targeting hexokinase. *Cancer Letters,* 173**,** 83-91.

KO, Y. H., SMITH, B. L., WANG, Y., POMPER, M. G., RINI, D. A., TORBENSON, M. S., HULLIHEN, J. & PEDERSEN, P. L. 2004. Advanced cancers: eradication in all cases using 3-bromopyruvate therapy to deplete ATP. *Biochem Biophys Res Commun,* 324**,** 269-75.

KOSOW, D. P. & ROSE, I. A. 1970. Product inhibition of the hexokinases. *Journal of Biological Chemistry,* 245**,** 198-204.

KRIEGER, E., NABUURS, S. B. & VRIEND, G. 2003. Homology modeling. *Methods of biochemical analysis,* 44**,** 509-524.

KUSER, P., CUPRI, F., BLEICHER, L. & POLIKARPOV, I. 2008. Crystal structure of yeast hexokinase PI in complex with glucose: A classical "induced fit" example revised. *Proteins,* 72**,** 731-40.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-2948.

LAVECCHIA, A. & GIOVANNI, C. D. 2013. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Current Medicinal Chemistry,* 20**,** 2839-2860.

LESK, A. M. & CHOTHIA, C. 1980. How Different Amino-Acid-Sequences Determine Similar Protein Structures - Structure and Evolutionary Dynamics of the Globins. *Journal of Molecular Biology,* 136**,** 225-&.

LIN, H., ZENG, J., XIE, R., SCHULZ, M. J., TEDESCO, R., QU, J., ERHARD, K. F., MACK, J. F., RAHA, K., RENDINA, A. R., SZEWCZUK, L. M., KRATZ, P. M., JUREWICZ, A. J., CECCONIE, T., MARTENS, S., MCDEVITT, P. J., MARTIN, J. D., CHEN, S. B., JIANG, Y., NICKELS, L., SCHWARTZ, B. J., SMALLWOOD, A., ZHAO, B., CAMPOBASSO, N., QIAN, Y., BRIAND, J., ROMINGER, C. M., OLEYKOWSKI, C., HARDWICKE, M. A. & LUENGO, J. I. 2016. Discovery of a Novel 2,6-Disubstituted Glucosamine Series of Potent and Selective Hexokinase 2 Inhibitors. *ACS Med Chem Lett,* 7**,** 217-22.

LIPINSKI, C. A., LOMBARDO, F., DOMINY, B. W. & FEENEY, P. J. 1997. In Vitro Models for Selection of Development CandidatesExperimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews,* 23**,** 3-25.

LIS, P., DYLĄG, M., NIEDŹWIECKA, K., KO, Y. H., PEDERSEN, P. L., GOFFEAU, A. & UŁASZEWSKI, S. 2016. The HK2 Dependent “Warburg Effect” and Mitochondrial Oxidative Phosphorylation in Cancer: Targets for Effective Therapy with 3-Bromopyruvate. *Molecules,* 21**,** 1730.

MACALINO, S. J. Y., GOSU, V., HONG, S. & CHOI, S. 2015. Role of computer-aided drug design in modern drug discovery. *Archives of Pharmacal Research,* 38**,** 1686-1701.

MACARRON, R., BANKS, M. N., BOJANIC, D., BURNS, D. J., CIROVIC, D. A., GARYANTES, T., GREEN, D. V. S., HERTZBERG, R. P., JANZEN, W. P., PASLAY, J. W., SCHOPFER, U. &

SITTAMPALAM, G. S. 2011. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov,* 10**,** 188-195.

MALIK, S. & SHRIVASTAVA, T. 2013. Protein characterization using modern biophysical techniques. *Advances in Protein Chemistry, ed by Ashraf GM and Sheikh IA. OMICS Group eBooks, Foster City***,** 4-18.

MARIE, S. K. N. & SHINJO, S. M. O. 2011. Metabolism and brain cancer. *Clinics,* 66**,** 33-43.

MARTÍ-RENOM, M. A., STUART, A. C., FISER, A., SÁNCHEZ, R., AND, F. M. & ŠALI, A. 2000. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure,* 29**,** 291-325.

MATHUPALA, S. P., KO, Y. H. & PEDERSEN, P. L. 2009. Hexokinase-2 bound to mitochondria: cancer's stygian link to the "Warburg Effect" and a pivotal target for effective therapy. *Semin Cancer Biol,* 19**,** 17-24.

MCDONNELL, J. M. 2001. Surface plasmon resonance: towards an understanding of the mechanisms of biological molecular recognition. *Current Opinion in Chemical Biology,* 5**,** 572-577.

MERTENS, H. D. & SVERGUN, D. I. 2010. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *Journal of structural biology,* 172**,** 128-141.

MORRIS, G. M., GOODSELL, D. S., HALLIDAY, R. S., HUEY, R., HART, W. E., BELEW, R. K. & OLSON, A. J. 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry,* 19**,** 1639-1662.

MORRIS, G. M., HUEY, R., LINDSTROM, W., SANNER, M. F., BELEW, R. K., GOODSELL, D. S. & OLSON, A. J. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal Of Computational Chemistry,* 30**,** 2785-2791.

MORRIS, J. C., WANG, Z., DREW, M. E. & ENGLUND, P. T. 2002. Glycolysis modulates trypanosome glycoprotein expression as revealed by an RNAi library. *The EMBO Journal,* 21**,** 4429-4438.

MORRIS, M. T., DEBRUIN, C., YANG, Z., CHAMBERS, J. W., SMITH, K. S. & MORRIS, J. C. 2006. Activity of a second Trypanosoma brucei hexokinase is controlled by an 18-amino-acid C-terminal tail. *Eukaryotic cell,* 5**,** 2014-2023.

NAKASHIMA, R. A., MANGAN, P. S., COLOMBINI, M. & PEDERSEN, P. L. 1986. Hexokinase receptor complex in hepatoma mitochondria: evidence from N, N'-dicyclohexlycarbodiimide-labeling studies for the involvement of the pore-forming protein VDAC. *Biochemistry,* 25**,** 1015-1021.

NEDELKOV, D., NELSON, R. W. & FOLTA-STOGNIEW, E. 2006. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *New and Emerging Proteomic Techniques***,** 97-112.

NEUDERT, G. & KLEBE, G. 2011. DSX: a knowledge-based scoring function for the assessment of protein–ligand complexes. *Journal of chemical information and modeling,* 51**,** 2731-2745.

NØRBY, J. G. & ESMANN, M. 1997. The effect of ionic strength and specific anions on substrate binding and hydrolytic activities of Na, K-ATPase. *The Journal of general physiology,* 109**,** 555-570.

NWAGWU, M. & OPPERDOES, F. 1982. Regulation of glycolysis in Trypanosoma brucei: hexokinase and phosphofructokinase activity. *Acta tropica,* 39**,** 61-72.

PASTORINO, J. G., SHULGA, N. & HOEK, J. B. 2002. Mitochondrial binding of hexokinase II inhibits Bax-induced cytochrome c release and apoptosis. *J Biol Chem,* 277**,** 7610-8.

PEDERSEN, P. L., MATHUPALA, S., REMPEL, A., GESCHWIND, J. F. & KO, Y. H. 2002. Mitochondrial bound type II hexokinase: a key player in the growth and survival of many cancers and an ideal prospect for therapeutic intervention. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1555**,** 14-20.

PELICANO, H., MARTIN, D. S., XU, R. H. & HUANG, P. 2006. Glycolysis inhibition for anticancer treatment. *Oncogene,* 25**,** 4633-46.

PORPORATO, P. E., DHUP, S., DADHICH, R. K., COPETTI, T. & SONVEAUX, P. 2011. Anticancer targets in the glycolytic metabolism of tumors: a comprehensive review. *Front Pharmacol,* 2**,** 49.

PUTNAM, C. D., HAMMEL, M., HURA, G. L. & TAINER, J. A. 2007. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics,* 40**,** 191-285.

RAMBO, R. P. & TAINER, J. A. 2011. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers,* 95**,** 559-571.

RÄZ, B., ITEN, M., GRETHER-BÜHLER, Y., KAMINSKY, R. & BRUN, R. 1997. The Alamar Blue® assay to determine drug sensitivity of African trypanosomes (Tb rhodesiense and Tb gambiense) in vitro. *Acta tropica,* 68**,** 139-147.

REMPEL, A., MATHUPALA, S. P., GRIFFIN, C. A., HAWKINS, A. L. & PEDERSEN, P. L. 1996. Glucose catabolism in cancer cells: amplification of the gene encoding type II hexokinase. *Cancer research,* 56**,** 2468-2471.

ROBEY, B. 2005. Mitochondrial hexokinases, Guardiand of mitochondria.

ROSENFELD, R., VAJDA, S. & DELISI, C. 1995. Flexible docking and design. *Annual review of biophysics and biomolecular structure,* 24**,** 677-700.

ROST, B. 1999. Twilight zone of protein sequence alignments. *Protein engineering,* 12**,** 85-94.

RUSSELL, S., RAHMANI, R., JONES, A. J., NEWSON, H. L., NEILDE, K., COTILLO, I., RAHMANI KHAJOUEI, M., FERRINS, L., QUREISHI, S., NGUYEN, N., MARTINEZ-MARTINEZ, M. S., WEAVER, D. F., KAISER, M., RILEY, J., THOMAS, J., DE RYCKER, M., READ, K. D., FLEMATTI, G. R., RYAN, E., TANGHE, S., RODRIGUEZ, A., CHARMAN, S. A., KESSLER, A., AVERY, V. M., BAELL, J. B. & PIGGOTT, M. J. 2016. Hit-to-Lead Optimization of a Novel Class of Potent, Broad-Spectrum Trypanosomacides. *Journal of Medicinal Chemistry,* 59**,** 9686-9720.

SCHEIN, C. H. 1989. Production of soluble recombinant proteins in bacteria. *Nature Biotechnology,* 7**,** 1141-1149.

SCIOR, T., BENDER, A., TRESADERN, G., MEDINA-FRANCO, J. L., MARTINEZ-MAYORGA, K., LANGER, T., CUANALO-CONTRERAS, K. & AGRAFIOTIS, D. K. 2012. Recognizing pitfalls in virtual screening: a critical review. *J Chem Inf Model,* 52**,** 867-81.

SHARLOW, E., GOLDEN, J. E., DODSON, H., MORRIS, M., HESSER, M., LYDA, T., LEIMGRUBER, S., SCHROEDER, C. E., FLAHERTY, D. P. & WEINER, W. S. 2011. Identification of inhibitors of Trypanosoma brucei hexokinases.

SHARLOW, E. R., LYDA, T. A., DODSON, H. C., MUSTATA, G., MORRIS, M. T., LEIMGRUBER, S. S., LEE, K. H., KASHIWADA, Y., CLOSE, D., LAZO, J. S. & MORRIS, J. C. 2010. A target-based high throughput screen yields Trypanosoma brucei hexokinase small molecule inhibitors with antiparasitic activity. *PLoS Negl Trop Dis,* 4**,** e659.

SHOICHET, B. K. 2004. Virtual screening of chemical libraries. *Nature,* 432**,** 862-865.

SOKOLOVA, A. Y., WYLLIE, S., PATTERSON, S., OZA, S. L., READ, K. D. & FAIRLAMB, A. H. 2010. Cross-resistance to nitro drugs and implications for treatment of human African trypanosomiasis. *Antimicrobial agents and chemotherapy,* 54**,** 2893-2900.

STEITZ, T. A., ANDERSON, W., FLETTERICK, R. J. & ANDERSON, C. 1977. High resolution crystal structures of yeast hexokinase complexes with substrates, activators, and inhibitors. Evidence for an allosteric control site. *Journal of Biological Chemistry,* 252**,** 4494-4500.

STRELOW, J., DEWE, W., IVERSEN, P. W., BROOKS, H. B., RADDING, J. A., MCGEE, J. & WEIDNER, J. 2012. Mechanism of Action assays for Enzymes.

SUBRAMANIAN, S. & ROSS, P. D. 1984. Dye-ligand affinity chromatography: The interaction of Cibacron Blue F3GA® with proteins and enzyme. *CRC critical reviews in biochemistry,* 16**,** 169-205.

SVERGUN, D. 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal,* 76**,** 2879-2886.

SVERGUN, D., BARBERATO, C. & KOCH, M. 1995. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of applied crystallography,* 28**,** 768-773.

SVERGUN, D. I. & KOCH, M. H. 2003. Small-angle scattering studies of biological macromolecules in solution. *Reports on Progress in Physics,* 66**,** 1735.

TARAZONA, M. P. & SAIZ, E. 2003. Combination of SEC/MALS experimental procedures and theoretical analysis for studying the solution properties of macromolecules. *Journal of Biochemical and Biophysical Methods,* 56**,** 95-116.

TROTT, O. & OLSON, A. J. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem,* 31**,** 455-61.

TSAI, H. J. 1999. Functional organization and evolution of mammalian hexokinases: mutations that caused the loss of catalytic activity in N-terminal halves of type I and type III isozymes. *Archives of biochemistry and biophysics,* 369**,** 149-156.

TSAI, H. J. & WILSON, J. E. 1996. Functional Organization of Mammalian Hexokinases: Both N- and C-Terminal Halves of the Rat Type II Isozyme Possess Catalytic Sites. *Archives of Biochemistry and Biophysics,* 329**,** 17-23.

TSAI, H. J. & WILSON, J. E. 1997. Functional Organization of Mammalian Hexokinases: Characterization of the Rat Type III Isozyme and Its Chimeric Forms, Constructed with the N- and C-Terminal Halves of the Type I and Type II Isozymes. *Archives of Biochemistry and Biophysics,* 338**,** 183-192.

TSUMOTO, K., EJIMA, D., KUMAGAI, I. & ARAKAWA, T. 2003. Practical considerations in refolding proteins from inclusion bodies. *Protein Expression and Purification,* 28**,** 1-8.

VAGENENDE, V., YAP, M. G. & TROUT, B. L. 2009. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry,* 48**,** 11084-11096.

VERLINDE, C. L. & HOL, W. G. 1994. Structure-based drug design: progress, results and challenges. *Structure,* 2**,** 577-587.

VERLINDE, C. L. M. J., HANNAERT, V., BLONSKI, C., WILLSON, M., PÉRIÉ, J. J., FOTHERGILL-GILMORE, L. A., OPPERDOES, F. R., GELB, M. H., HOL, W. G. J. & MICHELS, P. A. M. 2001. Glycolysis as a target for the design of new anti-trypanosome drugs. *Drug Resistance Updates,* 4**,** 50-65.

VOLKOV, V. V. & SVERGUN, D. I. 2003. Uniqueness of ab initio shape determination in small-angle scattering. *Journal of applied crystallography,* 36**,** 860-864.

WANG, R., LAI, L. & WANG, S. 2002. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *Journal of computer-aided molecular design,* 16**,** 11-26.

WANG, R. & WANG, S. 2001. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *Journal of Chemical Information and Computer Sciences,* 41**,** 1422-1426.

WANG, T. & KANG, J. 2009. Hexokinase inhibitor screening based on adenosine 5'-diphosphate determination by electrophoretically mediated microanalysis. *Electrophoresis,* 30**,** 1349-1354.

WEAR, M. A., PATTERSON, A., MALONE, K., DUNSMORE, C., TURNER, N. J. & WALKINSHAW, M. D. 2005. A surface plasmon resonance-based assay for small molecule inhibitors of human cyclophilin A. *Anal Biochem,* 345**,** 214-26.

WEAR, M. A., PATTERSON, A. & WALKINSHAW, M. D. 2007. A kinetically trapped intermediate of FK506 binding protein forms in vitro: chaperone machinery dominates protein folding in vivo. *Protein Expr Purif,* 51**,** 80-95.

WHITE, T. K. & WILSON, J. E. 1989. Isolation and characterization of the discrete N- and C-terminal halves of rat brain hexokinase: Retention of full catalytic activity in the isolated C-terminal half. *Archives of Biochemistry and Biophysics,* 274**,** 375-393.

WHITMORE, L. & WALLACE, B. 2004. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic acids research,* 32**,** W668-W673.

WILLSON, M., SANEJOUAND, Y.-H., PERIE, J., HANNAERT, V. & OPPERDOES, F. 2002. Sequencing, Modeling, and Selective Inhibition of Trypanosoma brucei Hexokinase. *Chemistry & Biology,* 9**,** 839-847.

WILSON, J. E. 2003. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *Journal of Experimental Biology,* 206**,** 2049-2057.

WOLF, A., AGNIHOTRI, S., MICALLEF, J., MUKHERJEE, J., SABHA, N., CAIRNS, R., HAWKINS, C. & GUHA, A. 2011. Hexokinase 2 is a key mediator of aerobic glycolysis and promotes tumor growth in human glioblastoma multiforme. *J Exp Med,* 208**,** 313-26.

YANG, J., YAN, R., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods,* 12**,** 7-8.

ZENG, C., ALESHIN, A. E., HARDIE, J. B., HARRISON, R. W. & FROMM, H. J. 1996. ATP-binding site of human brain hexokinase as studied by molecular modeling and site-directed mutagenesis. *Biochemistry,* 35**,** 13157-13164.

ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics,* 9**,** 40-40.

ZHU, A., ROMERO, R. & PETTY, H. R. 2009. An enzymatic fluorimetric assay for glucose-6-phosphate: application in an in vitro Warburg-like effect. *Anal Biochem,* 388**,** 97-101.

Appendix A

Crystallisation trials for hHKII

Construct	Method	Conditions screened	Reservoir	Comments
Tagged hHKII, 30mg/ml, 10mM Glc, 10mM G6P, 20mM $MgCl_2$	Sitting drop 1:1	**0.5-3M ammonium sulphate** (0.5 intervals), **pH 7.4-8.0** (0.2 intervals), 16^{o}C	1mM DTT, 0.1M Bis-Tris propane	Precipitates after 2M ammonium sulphate/ No crystal growth
Tagged hHKII, 30mg/ml, 10mM Glc, 10mM G6P, 20mM $MgCl_2$	Sitting drop 1:1	**5-30% PEG3350** (5% intervals), **pH 7.4-8.0** (0.2 intervals), 16^{o}C	1mM DTT, 0.1M Bis-Tris propane	Precipitates after 15% PEG3350 immediately/ No crystal growth
Tagged hHKII, 27mg/ml, 10mM Glc, 10mM G6P	Hanging drop 1:1, 1:2 in reservoir	16% **PEG screen (PEG 1450, 3350, 800) pH screen 7.4-8.5** (0.2 intervals)	1mM DTT, 0.5M NaCl, 0.1M Tris, 0.02M MgCl2,	Protein precipitates immediately at PEG 3350 16% and PEG 8000 16%. No precipitation at PEG1450/ No crystal growth either. PEG 3350 should not exceed 16%.
Tagged hHKII, 37.8mg/ ml, 10mM Glc, 10mM G6P	Hanging drop, 1:1, 1:2 in reservoir	**pH 5.7, 7, 8.0, 8.5), ammonium sulphate 1.2M-2.2** (0.2 intervals)	0.1M Tris, 20mM MgCl2, 5mM DTT, 0.5M NaCl,	No crystal growth.

Construct	Method	Conditions screened	Reservoir	Comments
Tagged hHKII, 7mg/ml	Hanging drop, 1:1 in reservoir	**PEG 3350 (6, 7, 10, 14, 16 20%)** pH 7.0 - 8.5 (0.5 intervals)	10% ethylene glycol, 0.2M sodium malonate, 1mM DTT, 0.1M Bus-Tris propane	No bad precipitants, 14-22% PEG 3350, pH 7.0-8.0 gave the best conditions close to crystallization/ No crystal growth
Tagged hHKII, 22mg/ml	Robot	Morpheus, Structure Screen 1 & 2 from Molecular Dimensions		No crystal growth
N domain hHKII, 20mg/ml, 10mM Glc, 10mM G6P, 20mM MgCl2	Hanging drop	**PEG 3350 12-22%** (2% intervals) **PEG 1450 14-24% Ammonium sulphate 0.5-2.5M** (0.5M intervals)	0.1M Tris, pH 7.5, 10% ethylene glycol, 0.2M sodium malonate, 1mM DTT	Precipitates after 18%PEG 3350 Precipitates after 20%PEG 1450 Precipitates after 1M ammonium sulphate No crystal growth

Appendix B

Purification trials of Untagged hHKII

Cell pellet treatment	Purification steps	Comments
Ammonium sulphate precipitation (40%)	HIC Column screening	Butyl FF gave the best separation of target protein, followed by Octyl FF, Phenyl LS, Butyl S, Phenyl HS
Ammonium sulphate precipitation (40%)	Butyl FF, Gel Filtration Superdex 26/60 200pg 2.5ml/ min	Not active compared to 6His tagged 90-95% pure
Lysate	AIEX protocols: HiTrap QFF 1ml Buffer A: 0mM NaCl, Buffer B: 1M NaCl	
	A: Linear gradient 100% 20CV	Broad peak-not clean HK
	B: Step gradient: Step 1: 30%B for 20CV Step 2: 50% B for 20CV Step 3: 100% B for 20CV	Not clean HK