Biophysical characterization of K0513, a protein expressed in human invasive glioblastoma

A thesis submitted in fulfilment of the requirements for the degree of

MASTER OF SCIENCE IN BIOCHEMISTRY

SCHOOL OF MATHEMATICS AND NATURAL SCIENCES

UNIVERSITY OF VENDA

By

NEMUKONDENI NDIVHUWO

11640078

Supervisor: Dr. A. Burger

Co-supervisor: Dr. T. Zininga

June 2020

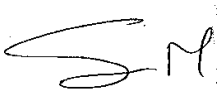
Abstract

Glioblastoma multiforme (GBM) is the deadliest brain tumour. GBM is associated with poor prognosis, with its patients having a very short median survival and poor response to chemotherapy. GBM is a World Health Organization (WHO) grade IV glioma and accounts for up to 78% of all brain tumours. This phenotype harbours a series of mutations that provide cells with selective growth advantages that promote survival and proliferation in a hostile and hypoxic environment. *KIAA0513* is one of the genes identified to be upregulated in GBM phenotype through gene expression profiling. *KIAA0513* gene codes for K0513 protein. *KIAA0513* is ubiquitously expressed and is enriched in the cerebral area of the brain. It is predicted to be involved in signalling pathways including neuroplasticity, cytoskeletal regulation and apoptosis. The main objective of the current study was to perform biophysical characterization of K0513. Characterization of K0513 will allow structure-function annotation which may serve as a basis for establishment of K0513 as a potential biomarker for GBM. Using bioinformatics analysis, a potentially functional SBF2 domain was identified. The three-dimensional homology model shows that K0513 is a globular protein, with the identified SBF2 domain and transmembrane region in proximity to one another. Predicted interacting partners includes members of the Rab GTPase family, membrane proteins, transcription factors and neurotransmitters. The overall *in silico* analysis suggest that K0513 may possess nucleotide exchange factor (NEF) activity for Rab3a. Recombinant proteins, non-codon harmonized ($K0513_W$) and codon harmonized ($K0513_H$) were expressed in *E. coli* XL1-Blue cells. The recombinant $K0513_H$ was successfully purified using nickel-affinity chromatography, this is the first study to report on the recombinant expression and purification of K0513. Using tryptophan fluorescence assay and limited proteolysis, nucleotides were found to have no significant effect on the tertiary structural conformation of K0513. Pull down assay shows promising interactors of K0513 from fibrosarcoma cell line; future studies will identify the interactors using Liquid chromatography-mass spectrometry.

Key words: Glioblastoma; KIAA0513; brain tumour; GTPases; Guanine Exchange Factors; protein-protein interaction; tryptophan fluorescence

Declaration

I, Nemukondeni Ndivhuwo (11640078), declare that this thesis submitted to the University of Venda for the Master of Science degree in Biochemistry under the School of Mathematical and Natural Sciences has not been submitted to any other University. This is my work with exception of referenced material which has been cited in text and acknowledged.

Signature:  Date: 26 August 2020

Dedication

This thesis is dedicated to my late father Nemukondeni Pfuluwani Amon.

Acknowledgements

First and foremost, all thanks to God almighty for giving me all the grace and strength to complete this study.

I also wish to extend my sincere gratitude to my supervisor Dr. A. Burger for giving me an opportunity to do my MSc with her. Most notably, her continuing support, guidance, motivation, enthusiasm and her patience.

I would also like to express my sincere thanks and gratitude to my co-supervisor Dr. T. Zininga, for providing necessary support, guidance and motivation.

I thank the Hair and Skin Research Unit in the Department of Dermatology at the University of Cape Town for giving us the fibrosarcoma cell line.

I wish to further extend my acknowledgement to the following institutions for funding this project;

- The National Research Foundation (NRF) of South Africa
- The University of Venda Research and Publication Committee (RPC) fund

I am extremely grateful for my family for their support, love, prayers and their encouragement which helped me in completion of this project.

Finally, I would like to extend gratitude to my lab mates and the whole Department of Biochemistry for their moral support and motivation throughout this study.

Table of contents

List of symbols and abbreviations

Abbreviations of units	Symbol Interpretation
%	percent
µl	microlitre
A600	absorbance at 600 nanometres
bp	base pair
µM	micromolar
mM	millimolar
nm	nanometres
ºC	degree Celsius
mL	millilitre
L	litre
g/L	gram per litre
mg/mL	milligram per millilitre
µg/mL	microgram per litre
w/v	weight per volume
v/v	volume per volume
mg	milligram
µg	microgram
ng	nanogram
g	gram
β	beta
α	Alpha

List of figures

List of tables

List of outputs

Publications

• Nemukondeni, N., Arowolo, A., Shonhai, A., Zininga, T. and Burger, A., 2020. Biophysical characterization of the human K0513 protein. *Plos One*. Submitted for review. Preprint, doi: https://doi.org/10.1101/2020.06.18.158949

• Nemukondeni, N., Zininga, T. and Burger, A., 2020. Current perspectives in design of small molecule inhibitors targeting glioblastoma. *International Journal of Neuroscience*. Submitted for review

Conferences

• Nemukondeni, N., Zininga, T. and Burger, A., 2018. Characterization of human KIAA0513 protein that is upregulated in invasive glioblastoma multiforme. SASBMB-FASBMB. 8th-11th July 2018, North-West University, Potchefstroom, South Africa. Oral presentation

• Nemukondeni, N., Zininga, T. and Burger, A., 2018. In silico analysis of KIAA0513, a molecule upregulated in glioblastoma multiforme. UNIVEN Research Open Day. 27th February 2018, University of Venda, Thohoyandou. Poster presentation

Chapter 1

Introduction and literature review

Introduction

1.1. Glioblastoma multiforme

Glioblastoma multiforme (GBM) is the most prevalent and lethal brain tumour, accounting for more than 60 % of all adult brain tumours (*Rock et al*., 2012). This highly invasive astrocytic (star-shaped) glioma presents the poorest prognosis with a median survival rate of only 15 months (Thakkar *et al*., 2014). Although various treatment strategies have been developed against GBM, it remains a deadly disease. The highly aggressive and invasive nature of the neoplasm as well as the resistance to chemotherapy results in very few long-term survivors (Adamson *et al*., 2009; Sang, 2016). GBM tumour cells can diffuse and invade through normal parenchyma which results in tumour recurrence in near or distant site (Valasquez *et al*., 2019). The term multiforme describes the tumour heterogeneity which affects GBM cell's morphology, gene expression levels and growth rates resulting in different responses to conventional therapies (Sathonsumetee *et al*., 2007).

GBM tumours harbour a series of mutations which enables the cell to have selective growth advantages promoting its survival and proliferation in a hostile and hypoxic environment (*Furnary et al*., 2007). For example, tumour suppressor genes, such as tumour protein 53 (*TP53)*, cyclin-dependent kinase inhibitor 1 (*CDKN1A/p21*), cyclin-dependent kinase inhibitor 2A (*CDKN2A/p16*), and Phosphatase and tensin homolog (*PTEN*) are commonly mutated in GBMs, which leads to the highly unstable nature of the cells (Chen *et al*., 2012). Primary GBM may include mutations in the telomerase reverse transcriptase (*TERT*) promoter and the *PTEN* tumour suppressor gene and increased gene amplification of proto-oncogenes such as the epidermal growth factor receptor (*EGFR*) (Ohgaki and Kleihues, 2013). GBM has been classified as grade IV glioma by World Health Organization (WHO). The grade is assigned to the cytologically most malignant, mitotically active, necrosis prone neoplasms typically associated with widespread infiltration into the surrounding tissue, microvascular proliferation, rapid pre- and post-operative disease evolution and fatal outcome (Louis *et al*., 2007).

1.2. GBM epidemiology

GBM is presented with annual cases of less than 10 per 100 000 people, with an average survival time of 14 to 15 months after diagnosis for reported cases (Iacob & Dinca, 2009; Thakkar *et al*., 2014). Moreover, the five-year survival rate of GBM is also only 10 %. (American Brain Tumour Association). GBM rarely affects children and only accounts for 8.8 % of childhood brain tumours, peak incidence is between the age of 55 and 60 (Ohgaki and Kleihues, 2005). GBM incidence have been reported to be higher in males than in females (Ohgaki and Kleihues, 2005; Thakkar *et al*., 2014), the incident ratio is attributed to GMB grade. Primary GBM occurs more frequently in males while secondary GBM occurs more frequently in females (Schwartzbaum, 2006). Most incidences are reported in western developed countries as compared with less developed countries (Thakkar *et al*., 2014), this is because diagnostic tools towards GBM are poorly developed or established in these population (Fisher *et al*., 2007; Ohgaki, 2009).

1.3. Molecular classification of GBM

The intricate pathogenesis of GBM involves mutations and alterations in key signalling pathways that facilitate increased cell invasion, proliferation, migration, survival, angiogenesis, and decreased apoptosis. Classically, glioblastomas have been classified as primary or secondary based on natural history and differences in molecular phenotype (Figure 1.1; Maher *et al.,* 2001; Louis *et al.,* 2007). Primary GBM occurs *de novo*, whereas secondary GBM develops from a lower grade astrocytoma; each is characterized by separate genetic events (Louis *et al.,* 2007; Ohgaki and Kleihues, 2007). About 90 % of GBMs are primary. Patients with secondary GBM tends to be younger (mean age = 30 years) than patients with primary GBM (mean age = 62 years).

The isocitrate dehydrogenase gene (*IDH1* or *IDH2*) mutation is a critical biomarker for molecular discrimination between primary and secondary GBM (Figure 1; Aldape *et al.,* 2015). IDH-wild type glioblastoma is classified as primary GBM, whilst the IDH mutants are found in secondary GBM (Louis *et al.,* 2016). Secondary GBM is characterized by the presence of *IDH1/2* mutations, and the amplification of of *EGF/EGFR* genes (Chung *et al.,* 1991; Louis, 1994; Liu *et al.,* 2012). The amplification of chromosome 7 and loss of

heterozygosity on chromosome 10 occurs in both primary GBM and secondary GBM (Colman and Aldape, 2011). Primary GBM may include mutations in the *TP53* gene and the phosphatase and tensin homolog *(PTEN)* tumour suppressor gene, as well as an increase in gene amplification of platelet derived growth factor receptor alpha (*PDGFRα*) (Ohgaki and Kleihues, 2013). Primary GBMs have a high frequency of loss of the long arm of chromosome 10 (Figure 1.1; Zarghooni *et al.,* 2009).

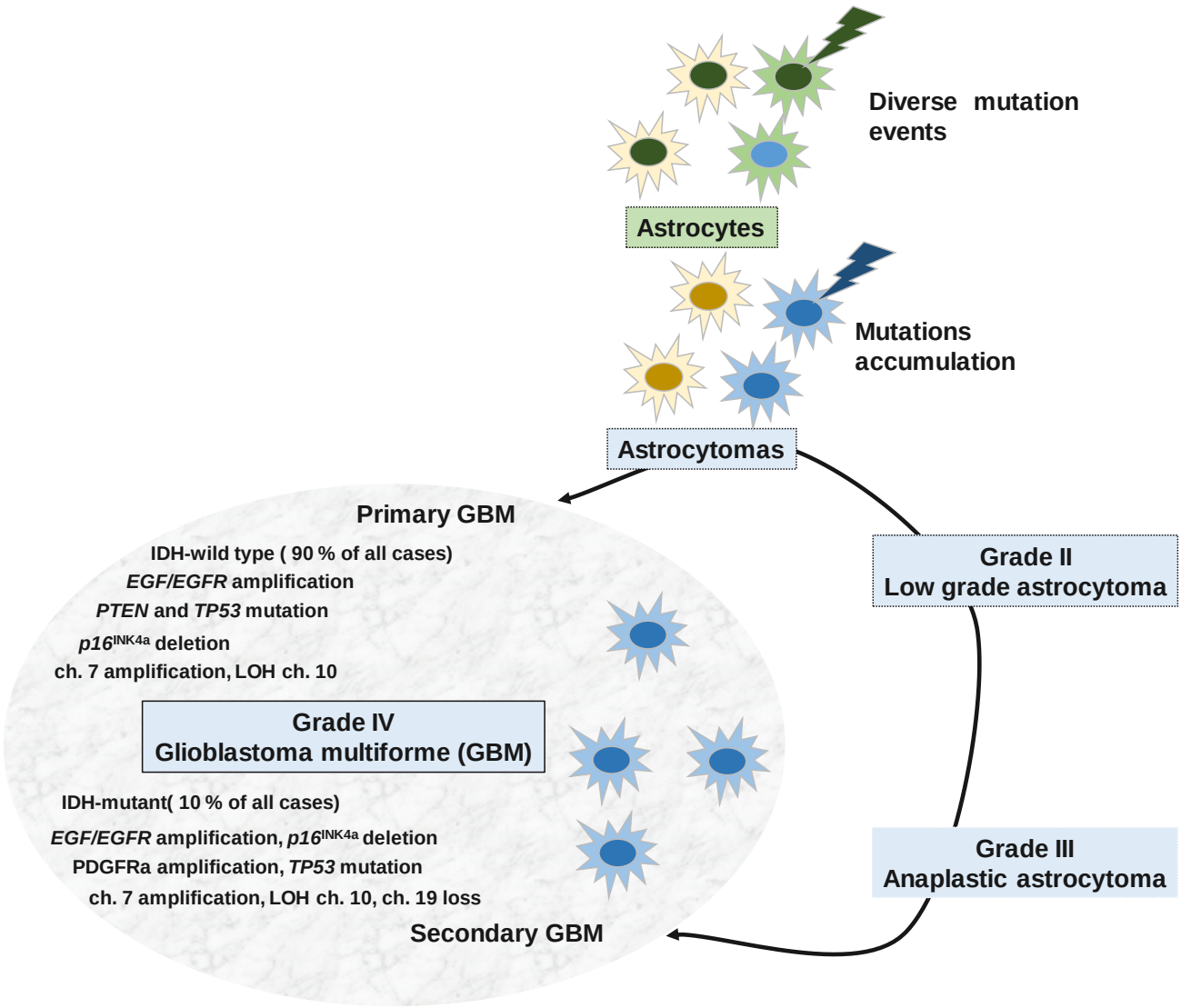


Figure 1.1. Origin of primary and secondary GBM and the genetic changes involved.
Primary GBM (*IDH*-wild type) develops *de novo* from astrocytes. Secondary GBM (*IDH* mutant) arises from lower grade astrocytoma. EGF/EGFR - epidermal growth factor/receptor; CDKN2A - cyclin dependent kinase inhibitor 2A; TP53 - tumour protein p53; PTEN - phosphatase and tensin homolog; PDGFRα - platelet derived growth factor receptor alpha; IDH1/2 - isocitrate dehydrogenase, mitochondrial; LOH - loss of heterozygosity; ch. - chromosome. Adapted from Pawlowska *et al*., 2018.

1.4. Molecular biomarkers for GBM

1.4.1. Isocitrate dehydrogenase 1 and 2 (IDH1/2) genes

The *IDH* catalyses the conversion of isocitrate into α-ketoglutarate within the citric acid cycle. *IDH1* and *IDH2* are involved in several metabolic processes like signal transduction, lipid synthesis, oxidative stress, and oxidative respiration (Riemenschneider *et al*., 2010). Mutation on *IDH1/2* is used as a prognostic marker in WHO grade II and III gliomas (Hartmann *et al*, 2009) and GBMs (Weller *et al*., 2009). *IDH1/2* wild type GBMs shows a pattern of genetic alterations related to primary GBMs, this include gain of chromosome 7, loss of chromosome 10, and *EGFR* amplification (Aldape *et al*., 2015).

1.4.2. p53

The tumour suppressor *TP53* which encodes for the *p53* protein is implicated within the pathogenesis of diffuse low-grade astrocytomas (Colman and Aldape, 2011). *p53* plays a crucial role in several cellular pathways associated with oncogenesis, this includes cellular response to DNA damage, cell cycle regulation, and apoptosis. Previous data had suggested that *p53* alterations were relatively rare in glioblastoma, however, results from the TCGA study reported a higher *p53* mutation in these tumors (Cancer Genome Atlas Research Network, 2008).

1.4.3. Epidermal growth factor receptor (EGFR) and EGFRvIII

EGFR is known to promote a pro-proliferative signal in GBM (Fischer and Aldape, 2010). About 40 % of primary GBMs shows an amplification of the *EGFR* gene (Schlegel *et al*., 1994), and about 50 % of GBMs having *EGFR* amplification also depicts a mutation in *EGFR* variant known as *EGFRvIII* gene (Korshunov *et al*., 2015). Studies show that GBMs harbouring constitutively active *EGFRvIII* receptors show a more invasive phenotype than those with wild type *EGFR* (Fischer and Aldape, 2010).

1.4.4. Platelet-derived growth factor alpha (PDGFRA)

PDGF is a dimeric glycoprotein that binds to the *PDGFR* tyrosine kinase. The expression of *PDGF* and PDGFR occurs within the same cells, resulting in the potential for both autocrine

and paracrine stimulatory activities (Colman and Aldape, 2011). Amplification of *PDGFRA* is reported in approximately 15 % of all tumours and is known to be a hallmark of GBM proneural subtype (Phillips *et al*., 2006). *PDGFR* signalling can be activated upon up-regulation of *PDGF* ligands (A–D) has been observed in about 30 % of glioma (Aldape *et al*., 2015). Amplification of *PDGF* and *PDGFR* has been reported to facilitate aggressive glioma growth (Assanah *et al*., 2006).

1.4.5. Neurofibromatosis type 1 gene (NF1)

NF1 is a tumour suppressor gene which also serves as a Ras-GTpase (Parson *et al*., 2008). Large-scale sequencing analysis by the TCGA has shown that in about 15 % of glioma cases, *NF1* gene is inactivated by genetic loss or mutation (Parson *et al*., 2008). *NF1* alterations are most common on mesenchymal GBM subtype and its loss is associated with tumour aggression (Philips *et al*., 2006; Verhaak *et al*., 2010).

1.4.6. Phosphatase and Tensin Homolog (PTEN)

PTEN is a tumor suppressor gene that negatively regulates the PI3K and AKT pathway and plays a crucial role in the regulation of apoptosis, tumour invasion, and cell proliferation (Simpson and Parsons, 2001). *PTEN* altered in 14 %- 47 % of primary GBM, and its deletion is due to loss of heterozygosity of chromosome 10q in 50 %–70 % of primary cases and 54 %–63 % of secondary GBM (Simpson and Parsons, 2001). *PTEN* is predicted to be a prognostic molecular marker as patients with loss of *PTEN* have decreased survival (Bell *et al*., 2011).

1.5. GBM subtypes

GBM has been further classified into subtypes through gene profiling. In 2006, Phillips *et al* subdivided GBM into three major subclasses; proneural, mesenchymal, and proliferative. Furthermore, another classical mechanism was proposed by Verhaak and colleagues in 2010 which subdivided GBM into four subtypes (Classical, Mesenchymal, Proneural, and Neural) (Figure 1.2).

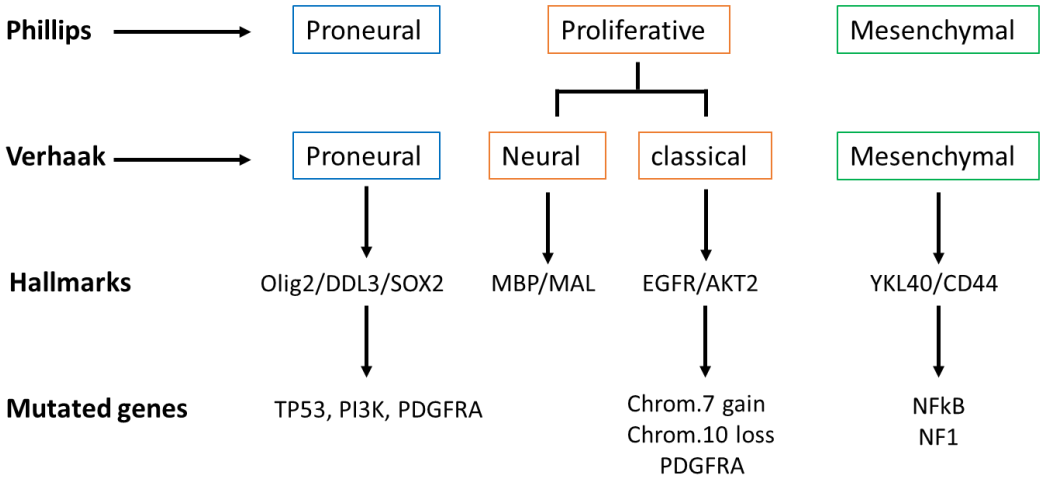


Figure 1.2. Transcriptional subtypes of glioblastomas based on Phillips and Verhaak classification. Molecular classification of GBM based on gene expression profiles. Phillips et al (2006) classified it into proneural, proliferative, and mesenchymal subtypes. Verhaak et al (2010) further subdivided it into proneural, neural, classical, and mesenchymal. Patterns of somatic mutations and DNA copy number alterations has been demonstrated by an Integrated genomic analysis. Alterations in each subtype; classical, mesenchymal, and proneural is represented by alterations in *EGFR*, *NF1*, and *PDGFRA/IDH1*, respectively (Adapted from Aldape *et al*., 2015).

Classical subtype shows high-level amplification of *EGFR* and a distinct lack of *TP53* mutations (Verhaak *et al*., 2010). Neural precursor and stem cell marker nestin (*NES)*, Notch (*NOTCH3*, jagged 1 (*JAG1)*, and *LFNG O-Fucosylpeptide 3-Beta-N-Acetylglucosaminyltransferase*) and sonic hedgehog (*SMO*, *GAS1*, and *GLI2*) signalling pathways are also highly expressed in this subtype (Verhaak *et al*., 2010).

Mesenchymal subtype is characterized by mutations or alteration in *NF1* gene which codes for Neurofibromin 1. Genes from tumour necrosis factor super family pathway and NF-κB pathway are highly expressed in this subtype, this includes tumour necrosis factor receptor type 1-associated DEATH domain (*TRADD)* and v-rel reticuloendotheliosis viral oncogegene homolog B (*RELB)* which might be a result of higher overall necrosis in the Mesenchymal class (Verhaak *et al*., 2010). Few EGFR alterations are observed in this subtype (Kuehn and Bridget, 2010).

Proneural subtype is well known by alterations of *PDGFRA* (a gene encoding a-type platelet-derived growth factor receptor) and point mutations in *IDH1* gene which encodes for isocitrate dehydrogenase-1 (Figure 1.2). There are also high alterations in *TP53* observed (Verhaak *et al*., 2010).

Neural signature is signified by the expression of neuron markers, such as *NEFL*, Gamma-aminobutyric acid receptor subunit alpha-1 (*GABRA1)*, Synaptotagmin-1 (*SYT1)*, and Solute carrier family 12 member 5 (*SLC12A5)* (Verhaak *et al*., 2010).

1.6. Common pathways disrupted in GBM

Though primary and secondary GBMs seem to be histologically similar, they comprise different genetic alterations and signalling pathways (Kanu *et al*., 2009). High expression of cell surface membrane receptors facilitates tumour growth in GBM by controlling the intracellular signal transduction pathways, thus regulating proliferation and cell cycle abnormalities, such as an increase in DNA repair proteins and abnormal cell death pathways (Furnari *et al*., 2007; Wen and Kasari, 2008). Specific molecular and signalling pathways have been shown to be involved in the development and progression of GBMs, this was from studies done by TCGA research network and other individual labs which revealed several genetic abnormalities. An integrated analysis of several genetic abrasions has led into grouping of these lesions into three main signalling pathways, including receptor tyrosine kinase (RTK) through the RAS/MAPK (mitogen-activated protein kinase) and PI3K/AKT/mTOR which is altered in almost 88% of GBMs, along with the cell cycle-regulating retinoblastoma (RB) tumour suppressor in 78% and p53 pathways altered in approximately 78% of GBMs (Aldape *et al*., 2015).

1.6.1. Growth factor tyrosine kinase receptor/RAS/PI3K pathway

Several signalling pathways important in brain tumour pathogenesis involve specific growth factors and their associated RTK (Colman and Aldape, 2011). Mutations or amplifications of RTK in more than 80 % primary GBM include *EGFR*, *PDGFRA*, insulin-like growth factor receptor (*IGFR-1*), and basic fibroblast growth factor receptor 1 (*FGFR-1*) (Figure 1.3; Ohgaki and Kleihues, 2009). RTK utilizes two pathways; the RAS/RAF/MAPK pathway that leads to cellular proliferation, differentiation, and migration, and the PI3K/ AKT/mTOR pathway that serves to promote cell proliferation and survival by inhibiting apoptosis (Ohgaki and Kleihues, 2009). The most frequent genetic alterations occurring in GBM tumours is the mutation, rearrangements, alternative splicing, and focal amplifications of *EGFR* (Kanu *et al*., 2009). *EGFR* implication in progression and pathogenesis of high-grade astrocycomas has made them to become therapeutic target for brain tumours (Colman and Aldape, 2011).

In addition, 13 % of adults GBM patients shows a high-level amplification of the platelet-derived growth factor receptor alpha gene (*PDGFRA*) (Network, 2013). The expression of platelet derived growth factor receptor (*PDGFR*) is mostly reported in types of gliomas, while that of *EGFR* is mainly in GBM. Similarly, constitutively activating deletion mutants in *PDGFRA* have been reported in receptor amplified GBMs (Ozawa *et al*., 2010). The PI3K/AKT/mTOR, the oncogenic pathway in GBM, can be activated by mutations in either the regulatory (*PIK3R1*) or catalytic (*PIK3CA*) domains of PI3K (Huse and Aldape, 2014). PI3K pathways regulate several malignant phenotypes including cell growth, antiapoptosis, and proliferation. In glioma patients, activation of the PI3K pathway is associated with poor prognosis (Sathornsumetee *et al.*, 2007). Specific mutations affecting Ras are low in GBM; however, high levels of Ras GTPases have been documented in cell lines and primary tumours, which suggest that this signalling pathway is activated by upstream factors such as RTK (Figure 1.3; Ohgaki and Kleihues, 2009).

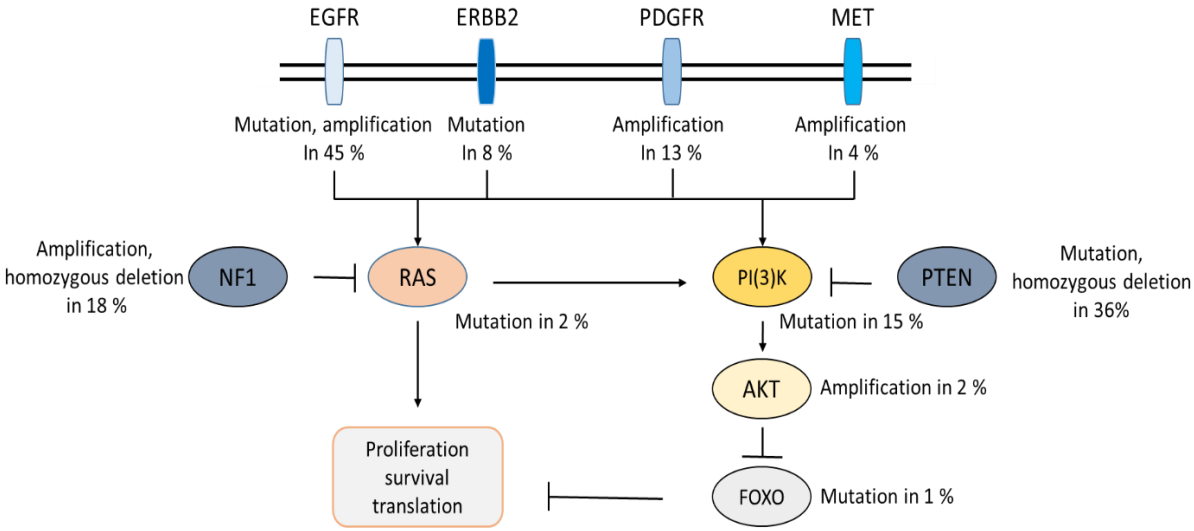


Figure 1.3. Alterations in the RTK/RAS/PI3K signalling pathway in GBM.
Several genes encoding proteins involved in the RTK/RAS/ PI3K signalling pathway are altered in GBM. Frequently amplified genes in this pathway include two membrane receptors; the platelet-derived growth factor receptor α (*PDGFRA*) and epidermal growth factor receptor (*EGFR*) that are known to have tyrosine kinase activity. The most deleted gene in the RTK pathway is the phosphatidylinositol-3 kinase (PI3 K), a cell cycle inhibitor of PARK2 phosphatase, tensin homolog (*PTEN*), retinoblastoma (RB1), and neurofibromin 1 (*NF1*). The most mutated genes in this pathway are *PIK3R1*, *PIK3CA, PTEN*, *NF1*, and lastly the *EGFR*. Adapted from The Cancer Genome Atlas Research Network.

1.6.2. p53 and Rb pathway

TP53 gene is an important tumour suppressor gene located on chromosome 17p13.1 that is centrally involved in multiple pathways regulating DNA integrity, cell cycle, and cell death

(Carson and Lois, 1995). TP53 gene encodes for p53 tumour suppressor, which drives the regulation of genes involved in the cell cycle (Figure 1.4). The *p53* plays a role in DNA repair through cell arrest in G1 phase (Adamson *et al*., 2009), if genetic injuries are irreparable, *p53* induces apoptosis machinery. Following DNA damage, *p53* is activated and induces transcription of genes such as p21Waf1/Cip1 (Sherr and Weber, 2000). GBM demonstrate inactivating mutations in either the *p53* and/or retinoblastoma (RB) pathways (Iwakuma and Lazano, 2003), and their interaction is through p21. *p53* upregulation is triggered by cellular stress such as radiation exposure, DNA strand breaks, and toxins (Adamson *et al*., 2009). Not only the mutation in TP53 gene affect the *p53* pathway, it can also be disrupted by alterations in several other regulatory genes/proteins. Identified antagonists of *p53* are murine double minute 2 homolog (*MDM2)* and *MDM4* genes, they bind to the *p53* promoter and inhibit its transcription. Amplification of both *MDM2* and *MDM4* has been identified in a subset of gliomas with intact *p53* indicating an alternate mechanism for downregulation of this pathway (Iwakuma and Lazano, 2003; TCGA, 2008). Mutations in genes encoding upstream regulators of *RB* (CGAR, 2008; Parsons *et al*., 2008). The *RB* pathway plays a role in the G1/S transition of the cell cycle and encodes for a tumour suppressor (Figure 1.4), *RB1* which negatively regulate the cell cycle (Shapiro, 2006). The *RB* pathway is affected in 80 % of GBM (Brennan *et al*., 2013).

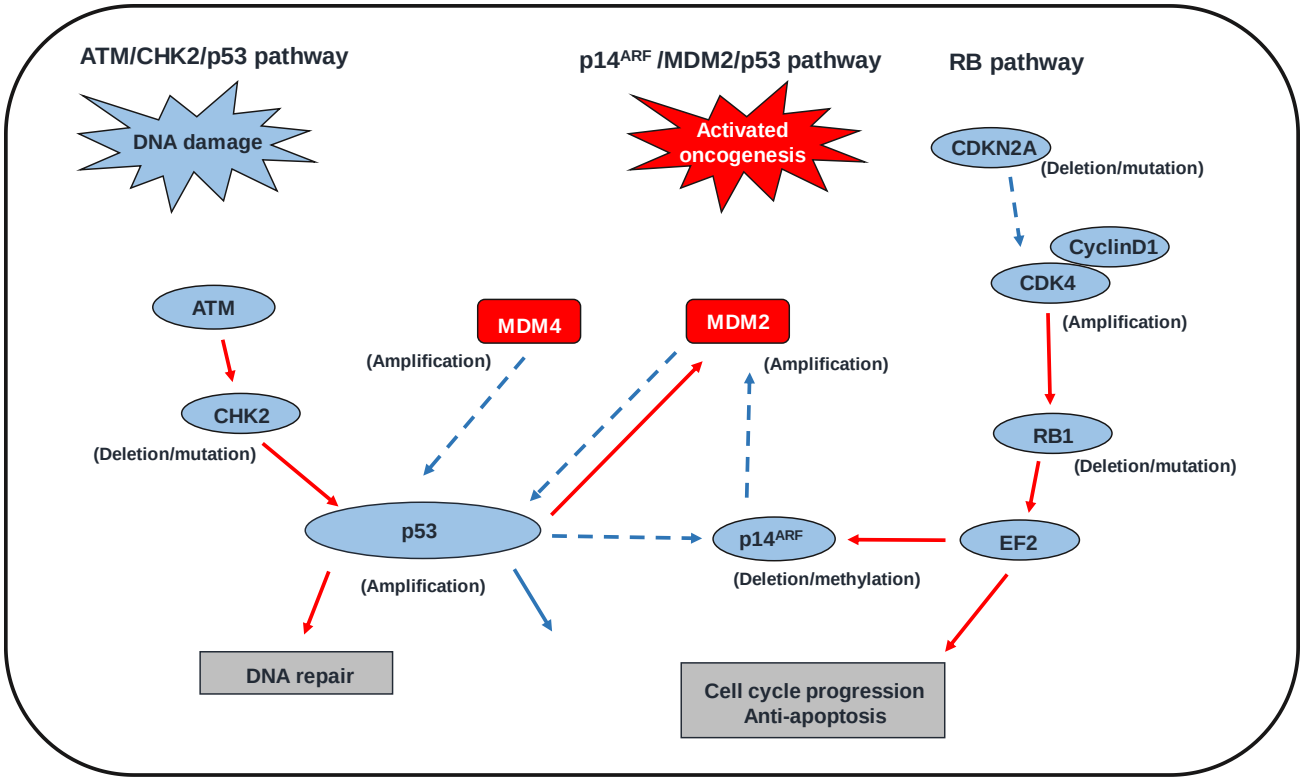


Figure 1.4. p53 and retinoblastoma (RB) pathways in GBM.
DNA damage activates ataxia-telangiectasia mutated (*ATM*) gene which subsequently activates checkpoint kinase (CHK2) and p53. MDM2 and MDM4 acts as negative regulators of p53, MDM2 is inhibited by $p14^{ARF}$ thus promoting p53. In RB pathway, the deletion/mutation of CDKN2 allows the formation of CDK4/CyclinD1 complex, thus allowing cell cycle transition from G1 to S phase. Dashes and solid arrows indicate suppression and activation, respectively. Red cycles represent activating genetic alterations whereas blue cycle represents genetic alterations that leads to loss of function. Adapted from Ohgaki and Kleihuis, 2009.

1.7. Small GTPases

The small GTPase superfamily also known as small G proteins, are monomers similar to the alpha-subunit of the heterotrimeric G protein. They consist of a highly globular structure that includes two flexible motifs known as the switch regions 1 and 2. The human genome encodes 153 members of the small GTPases which are divided into five groups based on function and sequence conservation (Wennerberg *et al*., 2005): RAS (*Rat sarcoma*) which are implicated in cell growth, proliferation and adhesion (Matsumura *et al*. 1998); RAB (*Ras-related proteins in brain*) proteins are involved in membrane trafficking (Stenmark, 2009); RHO (*Ras homologous*) involved in cytoskeletal processes (Aspenstrom *et al*. 2004); AFR (*ADP ribosylation factor*) involved in actin remodelling and vesicle transportation; and RAN (*Ras-like nuclear*) which are involved in nuclear transport and microtubule organization (Li *et al*. 2003).

Upon binding to GTP, small GTPases change conformation and interact with the effector in the cell. Most of the small GTPases are post-translationally modified to allow cell membrane association, this is by lipid anchors such as palmitoyl, acetyl groups, geranylgeranyl or farnesyl. Due to low intrinsic GTP hydrolysis and GDP/GTP exchange in GTPases, their GDP/GTP cycle is facilitated by GEFs and GAPs (Cherfils and Zeghouf, 2013). Some members of small GTPases requires an additional regulatory activity by guanine dissociation inhibitors (GDIs) that mask the lipid group hold the inactive GTPases in the cytosol.

1.7.1. RAB GTPases subfamily

Rab GTPases constitute the largest subfamily in the small GTPases superfamily with more than 70 members in human, and are known for membrane trafficking regulation (Stenmark, 2009; Wandinger-Ness and Zerial, 2014). Like any other small GTPases, Rab GTPases switch molecules by transiting between the active GTP-bound state and the inactive GDP-bound state (Figure 1.5). In their GTP-bound conformation, Rab GTPases are distinctly localized on the surface of vesicle/ membrane and facilitate trafficking through recruitment of the effector proteins to the membrane surface (Fukuda, 2008; Zhen and Stenmark, 2015). To execute their function and localize in the membranes, Rab proteins require prenylation in the C-terminal cysteine residues (Zhang and Casey, 1996). The same way in other small GTPases, the activation and inactivation of Rab GTPases is tightly regulated by GEFs and GAPs respectively (Barr and Lambright, 2010; Cherfils and Zeghouf, 2013). After GTP hydrolysis, Rab GTPases are extracted from the membrane by GDI which subsequently solubilizes the inactive prenylated Rab and keep it ready in the cytosol for next round of vesicle trafficking (Figure 2; Pfeffer and Aivazian, 2004). Rab proteins and their GEFs are key mediators of cytoskeletal function and membrane trafficking (Jean *et al*., 2012).

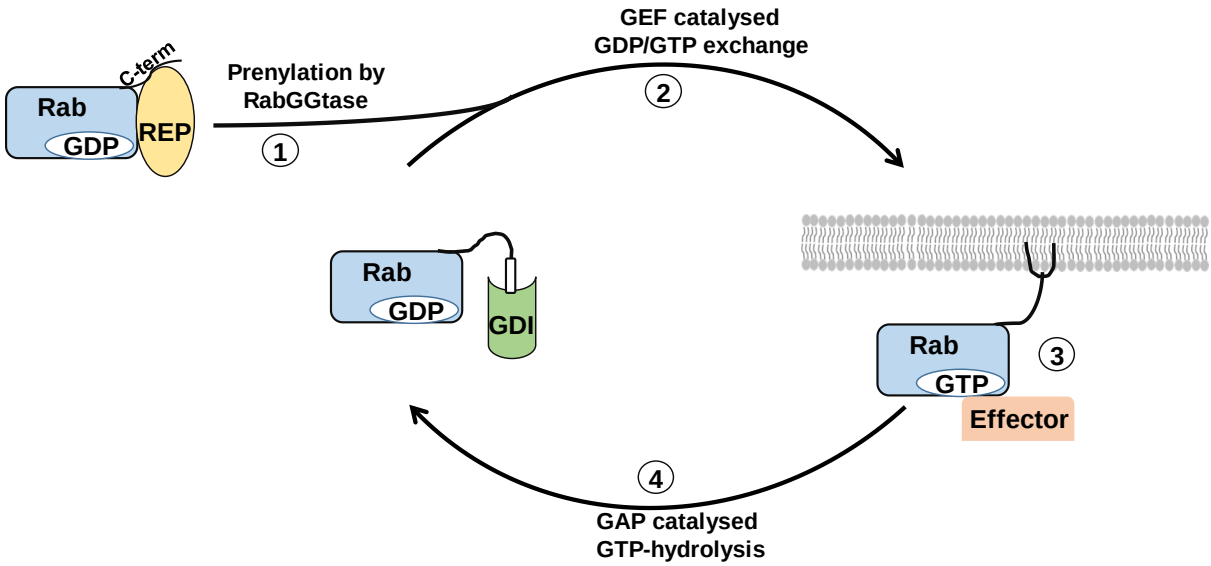


Figure 1.5. Rab GTPases as molecular switches.
Following synthesis, Rab GTPase bind to the Rab escort protein (REP) where they are prenylated the C-terminal cysteine residues by RabGGTase. After prenylation, the protein can be solubilized in the cytosol by the GDP dissociation inhibitor (GDI), which shields the hydrophobic geranylgeranyl groups from the hydrophilic environment (**1**). Guanine nucleotide exchange factors (GEFs) facilitates the exchange of GDP to GTP which subsequently recruits Rabs to the cell membranes (**2**). In their active GTP-bound state, Rabs interact with effectors and facilitate various vesicular trafficking (**3**). Interaction with the GTPase activating proteins (GAPs) hydrolyses GTP resulting into an inactive Gap in a GDP-bound state (**4**). Adapted from Müller et al., 2018.

1.7.1.1. Prenylation of Rab GTPases

In order to allow Rab proteins to cycle between their cytosolic and membrane bound form, they are irreversibly prenylated at the C-terminal cysteine residues making hydrophobic appendages which serves as membrane anchors (Stenmark and Olkkonen, 2001). This post-translational modification is catalysed by three prenyltransferase enzymes using farnesylpyrophosphate (FPP) or geranylgeranylpyrophosphate (GGPP) to modify target proteins. Farnesyltransferase (FTase) and geranylgeranyltransferase type I (GGTase-I) facilitate the attachment of a single farnesyl or geranylgeranyl-isoprenoid group to the cysteine residue located at the C-terminal CaaX-motifs (Thomas *et al*., 2007). Geranylgeranyltransferase type II (GGTase-II/ Rab geranylgeranyltransferase) on the other side facilitate attachment of two geranylgeranyl groups to two cysteine residues in C-terminal of different motifs such as CC, CXC or close to the C-terminus of Rab proteins (Goody *et al*., 2005).

Following synthesis of Rab proteins, they get escorted by Rab escort protein (REP) to the RabGGTase for prenylation. Most Rab proteins display double prenylation and are released to target membrane when the GGPP molecule binds to RabGGTase active sites, this

subsequently causes the now prenylated Rab in complex with REP to dissociate from RabGGTase.

1.7.1.2. Guanine nucleotide exchange factors

There are at least six identified types of Rab GEFs, this includes the Differentially Expressed in Normal and Neoplastic cells (DENN)-domain-containing proteins with 18 members in human (Marat *et al*., 2011; Table 1.1; Appendix A, Table A1), the Vacuolar Protein Sorting 9 (VSP9)-domain-containing proteins (Carney *et al*., 2006), Sec2-domain-containing proteins, multi-subunit Transport Particle Protein (TRAPP) complexes (Jones *et al*., 2000), and heterodimer GEFs (Barr and Lambright, 2010). Rab GEFs shows a very low homology amongst themselves (Koch *et al*., 2016). Most GEFs have been reported to be controlled by auto-inhibition which is facilitated by regulatory proteins (Cherfils *et al*., 2013). All GEFs follow the same mechanism where they bind to switch I/switch II and Interswitch regions of the nucleotide bound small GTPase, this subsequently induces a conformational change within these regions. The conformations of the two variable regions-switch I and switch II significantly differs depending on the nucleotide-loading state of the GTPase (Kiontke *et al*., 2017), and are remodelled to have a low affinity for the nucleotide. Switch I shows a largest conformational, this involves the interaction between the C-terminal of switch I and the GEF which subsequently open the switch I. This structural rearrangement displaces highly conserved aromatic residues (F or Y) to interact with guanine base thus lowering the nucleotide affinity.

Table 1.1: Human Rab proteins and their known GEFs

Rabs	GEFs	References
Rab3a	MAP kinase-activating death domain protein (MADD)	Yoshimura *et al*., 2010, Ishida *et al*., 2016
Rab35	DennD1A-C	Yoshimura *et al*., 2010; Ishida *et al*., 2016
Rab38	Biogenesis of lysosome-related organelles complex 3 (BLOC-3)	Gerondopoulos *et al*., 2012
Rab27a	MADD (Denn)	Yoshimura *et al*., 2010; Ishida *et al*., 2016
Rab25	Not identified	
Rab23	Not identified	
Rab43	Rabaptin-5-associated exchange factor for Rab5 (Rabex-5)/vacuolar sorting protein 9 (Vps9)	Yoshimura *et al*., 2010; Ishida *et al*., 2016
Rab7a	Monensin sensitivity protein 1 (Mon1(/caffeine, calcium, and zinc 1 (Ccz1)	Yoshimura *et al*., 2010; Ishida *et al*., 2016
Rab8b	Chromosome 9 open reading frame 72 (C9Orf72)	Corbier and Sellier, 2016
Rab10	DennD4	Yoshimura *et al*., 2010; Ishida *et al*., 2016
Rab12	DennD3	Yoshimura *et al*., 2010; Ishida *et al*., 2016
Rab13	DennD1C	Yoshimura *et al*., 2010; Ishida *et al*., 2016
Rab14	DennD6	Yoshimura *et al*., 2010; Ishida *et al*., 2016

1.7.2. Myotubularin related protein 13 (MTMR13), a DENN domain-containing protein

Myotubularin related protein 13 (MTMR13) is a member of myotubularin protein family, this family incorporates catalytically active and inactive enzymes sharing a core of protein domains and is a highly conserved group of ubiquitously expressed phosphatidylinositol 3-phosphatases (Laporte *et al*., 2003). MTMR13 is a known RAB GEF which exchange GTP for GDP on RABs. These GTPases are known to interact with effectors such as kinases, phosphatases and tubular-vesicular cargo (Wandinger-Ness and Zerial, 2014). MTMR13 is an inactive phosphatase, with a monomeric molecular weight of 208 kDa. It is comprised of a DENN domain, glucosyltransferase/Rab-like GTPase activator/myotubularin (GRAM) domain, pseudo-phosphatase domain and a pleckstrin homology (PH) domain, and a pfam assigned SBF2 domain (Figure 1.6). PH domains are involved in membrane association and usually bind to phosphates (Robinson and Dixon, 2005; Berger *et al*., 2006). The GRAM domain is known to mediate membrane attachment by binding to phosphoinositides, (Berger *et al*., 2003; Berger *et al*., 2006). MTMR13 serves as both myotubularin phosphoinositide

phosphatase scaffold and a Rab GTPase GEF (Yoshimura *et al*., 2010; Jean *et al*., 2012). MTMR13 forms part of the 18 known DENN domain-bearing proteins encoded in the human genome (Reviewed in Marat *et al*., 2011). And those that are differentially expressed in neoplastic normal cells were recently shown to exhibit Rab GEF activity (Yoshimura *et al*., 2010).

Figure 1.6. Schematic representation of the domain organisation in MTMR13.
The tripartite DENN module consist of uDENN (light green; 7-172), the cDENN (red; 191-324), and dDENN (dark green; 326-427). Other domains include; the pfam assigned SBF2 domain (531-754) represented by light orange box, glucosyltransferase/Rab-like GTPase activator/myotubularin (GRAM) domain (grey; 871-957), pseudo-phosphatase domain (blue; 1108-1584), and the pleckstrin homology (PH) domain (gold; 1743-1847). Adapted from Marat *et al*., 2011.

1.7.3. Implications of Rab GTPases in GBM

Rab proteins plays an important in cellular processes such as growth and membrane trafficking. Despite their significant role in normal cellular processes, these small GTPases have been reported to play a role in cancer (Stenmark, 2009). The abnormal expression of Rab proteins leads to aberrant tumour behaviour, by promoting invasion though tumour-associated fibroblasts (Ge *et al*., 2017). In glioblastoma, expression of Rab23 facilitates the invasion and proliferation of the tumour cells (Chen *et al*., 2016). Expression of Rab27a and Rab3a has been reported to inhibit apoptosis and promote cell proliferation (Wu *et al*., 2013; Kim *et al*., 2014). Rab38 has been identified a poor prognostic biomarker in GBM and has been associated with the glioma cell progression (Wang and Jiang., 2013). Over expression of Rab43 has also been associated with poor prognosis and epithelial-mesenchymal transition in GBM (Han *et al*., 2016). Rab25 knockout in GBM cells inhibits invasion and reduce cell proliferation through inhibition of PI3K and AKT phosphorylation levels (Ding *et al*., 2017). Reduced expression of Rab35 has been associated with a shorter survival in patients demonstrating pro-neural and mesenchymal GBM (Kulasekaran *et al*., 2020).

1.8. Human KIAA0513

KIAA0513 is a ubiquitously expressed gene enriched in the cerebral area of the brain (Lauriat *et al*., 2006). There have been reports of aberrant implication of this gene in different cancer cells; upregulation in invasive glioblastoma multiforme, prognosis signature gene in

pancreatic cancer (Chen *et al*., 2019). KIAA0513 which codes for K0513 protein has not been characterized as a tumour suppressor gene or oncogene, however, it is associated with tumour progression, cell proliferation and apoptosis (Standfuß, 2019). Low expression leads to unfavourable prognosis in non-small cellular lung cancer through overexpression of Calcium Voltage-Gated Channel Subunit Alpha1 B (CACNA1B) (Zhou *et al*., 2017). Moreover, in Neuroblastoma, the aberrant methylation of K0513 correlates with non-survivors (Olsoon *et al*., 2016). There are three interactors that are known to specifically interact with K0513 through in situ hybridization, these are: kidney and brain expressed protein or WW domain-containing protein 1 (KIBRA/WWC1), a tumour suppressor gene which regulates Hippo signalling pathway (Zhang *et al*., 2014); HS-1associated protein X-1 (HAX-1) involved in apoptosis regulation (Zhao *et al*., 2009); and integrator complex subunit 4 (INTS4) which forms part of transcription regulation (Lauriat *et al*., 2006; Albrecht *et al*., 2018).

1.9. Problem statement

GBM is the deadliest brain tumour associated with poor prognosis. Despite advances in treatment strategies it remains largely incurable. Gene expression profiling has been successfully used to identify genes implicated in GBM for possible therapeutic targets. However, potentially interesting genes of unknown function are identified but often ignored due to the effort required to characterize them. K0513 is one of the proteins identified but with unknown function. K0513 implication has been also reported in pancreatic cancer, neuroblastoma and non-small cellular lung cancer. It is in the interest of this study to characterize K0513 and further elucidate its function. This will enable an understanding of whether it facilitates the progression of tumour cells and possibly be used as a therapeutic target.

1.10. Aim

To characterize the structure-function features of K0513 and elucidate its role in cancer

1.11. Hypothesis

K0513 is a possible guanine nucleotide exchange factor (GEF) which facilitates the exchange of GDP/GTP for Rab GTPases

1.12. Main objectives

1.12.1. To utilise bioinformatics analysis to identify the structure-function features of K0513

1.12.2. To express and purify recombinant K0513 using Nickel-affinity chromatography

1.12.3. To investigate the tertiary structural organisation of K0513 using Tryptophan fluorescence assay, and limited proteolysis

1.12.4. To identify and establish interactors of K0513 using pull-down assay

Chapter 2

Bioinformatics analysis of human K0513

2.1. Introduction

Bioinformatics can be defined as a computational tool used for analysis, visualization and storage of biological data, as to generate and discover a new insight related to biological macromolecules. This is an important tool in protein biochemistry as it paves a way and address problems encountered in wet lab experiments. Computational tools make use of DNA, RNA and protein sequence. These sequences are stored in databases such as The National Centre for Biotechnology Information (NCBI, Brown *et al*., 2015) and UniprotKB (The Uniprot consortium, 2017) to name few. Retrieved sequences are analyzed and compared through multiple alignments to predict molecular functions, conduct evolutionary studies, predict intermolecular interactions and tertiary structure.

Homology determination is important in ascertaining functional features in protein structures especially in the absence of a crystal structure, as well as structural prediction of proteins. The query protein sequence can be aligned with identified homologues for determination of domains and motifs based on the level of conservation between the protein. Human K0513 (NCBI accession number: NP_055547.1) has no reported domain, therefore, identification of its homologues and orthologues would help in identifying key domains possessed by this molecule leading to functional annotation and hypothetical structure prediction. Protein-protein interaction (PPI) prediction is crucial in ascertaining pathways and cellular processes that the protein is involved in. Such a database of PPIs is the Integrated Interaction Database (IID) (Kotlyar *et al*., 2019). IID is the only database with context-specific networks for the common model organisms and domesticated species (Kotlyar *et al*., 2019), it integrates experimentally detected PPIs from nine curated databases (BioGRID, IntAct, I2D, MINT, InnateDB, DIP, HPRD, BIND, BCI). This database is the first to provide tissue-specific protein-protein interaction, this was to counteract limiting factors encountered in other databases such as high false positive rates, and context information (tissue and subcellular location). Predicting the K0513 interactome will help to reveal pathways that K0513 may be involved in and so enable prediction of possible roles.

Based on the gene nomenclature, genes identified by the Kazusa cDNA project without any known information are termed ‘KIAA’ plus a four-digit number (Nagase *et al*., 2006), K0513 is one the discovered cDNAs with no function reported (Braschi *et al*., 2019). *In*

silico analysis of K0513 will be very essential for prediction of its structure-function feature.

The main aim of this study is to utilize bioinformatics for identification of functional features in protein structure of K0513.

Specific objectives:

1. To retrieve and analyze amino acid sequence of human K0513 and its homologues
2. To perform secondary structural analysis of K0513 using Phyre^2 and Biotools
3. To generate a three-dimension homology model using Phyre^2 and BIOVIA Discovery studio
4. To predict possible interactors of K0513 using an Online Integrated Interaction Database (IID)
5. To investigate the evolutionary distribution of K0513 across different forms of life using MEGA X and ITOL

2.2. Methods

2.2.1. Multiple sequence analysis of K0513

The amino acid sequence of human K0513 (NCBI accession number: NP_055547.1) was obtained from National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov; NCBI resource, 2018), together with its homologues from *Pan troglodytes* (XP_003952958.1), *Xenopus tropicalis* (NP_001096454.1), *Canis lupus* (XP_005620695.1), *Bos taurus* (XP_005218484.1), *Mus musculus* (NP_001157232.1), *Rattus norvegicus* (NP_001100906.1), *Gallus gallus* (XP_001235240.2), *Danio rerio* (NP_001002193.1), and from *Macaca mulatta* (XP_001113080.1). The sequence alignments and identity matrix were generated using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/; Edgar, 2004)) and shaded using Boxshade (http://www.ch.embnet.org/software/BOX_form.html) server.

2.2.2. Secondary structural analysis of K0513

Secondary structural analysis of K0513 was carried out from the amino acid sequence. The analysis was carried out using Protein Homology/analogy Recognition Engine Version 2.0 (PHYRE^2) (http://sbg.bio.ic.ac.uk/; Kelley *et al*., 2015) which predicts the percentage composition of alpha-helices, beta-sheets, and loops. Furthermore, the hydrophobicity and ampathicity of K0513 was predicted using Biotools Web-based Hydropathy, Amphipathicity and Topology (WHAT) (http://biotools.tcdb.org/barwhat2.html; da Silva *et al*., 2017). The amino acid composition of K0513 was computed using PEPSTAT module integrated in the EMBOSS software (https://www.ebi.ac.uk/Tools/seqstats/emboss_pepstats/; Madeira *et al*., 2019). Physicochemical characterization of the protein was conducted using the Expasy's ProtParam server (https://web.expasy.org/protparam/; Gasteiger *et al*., 2005).

2.2.3. Prediction of three-dimension model of K0513

The structural prediction of a protein is necessary for domain annotation whereby sequences of the same homology are used to predict the three-dimensional (3D) structure. This will also reveal the secondary structural annotation with regards to alpha-helices, beta-sheets and coils. The homology models were generated using PHYRE^2 a (http://www.sbg.bio.ic.ac.uk/phyre2; Kelley *et al*., 2015). The human K0513 3D model was

generated from several templates, and the GST C-terminal domain-like crystal structure (PDB, *d1z9ha1*) showed the highest probability, *Mus musculus* generated using cnot1 mif4g domain - ddx6 complex crystal structure (*c4ct4C*.pdb; Mathys *et al*., 2014) as a template, *pan troglodytes* was generated using the crystal structure of a transcriptional activator tena from *Staphylococcus epidermidis* (PDB, *c2no6B*; JCSG, 2010). All the generated PDB files were visualized using BIOVIA Discovery Studio (Dassault Systemes Biovia, 2016).

2.2.4. Prediction of K0513 interaction partners

In order to infer functions to a protein, one of the best approaches is through identification of potential interacting partners. An online Integrated Interaction Database (IID) was launched (http://ophid.utoronto.ca/iid; Kotlyar, 2016), K0513 sequence identifier (UNIPROT ID; O60268) was used to generate the predicted interactome. Interaction data in IID comprises three types of PPI networks: experimentally detected PPIs from major databases, orthologous PPIs and high-confidence computationally predicted PPIs.

2.2.5. Evolutionary analysis of K0513

The evolutionary relationship of human K0513 with its homologues and orthologues was investigated by generating a phylogenetic tree. An alignment of K0513 protein sequences (orthologues and homologues) from K0513-expressing organisms was generated and retrieved from EggNOG online database (http://eggnogdb.embl.de/; Huerta-Cepas *et al*., 2019), the alignment was exported in a NEWICK format to MEGA X for evolutionary analysis. The bootstrap value was set for 100 replicates, initial tree(s) for the heuristic search were generated automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a Jones-Taylor-Thornton (JTT) model, and then selecting the topology with superior log likelihood value. The generated evolutionary tree was exported in a NEWICK format to ITOL (https://itol.embl.de/; Letunic and Bork, 2019) for further annotation.

2.3. Results

2.3.1. K0513 is conserved across the homologues

The homologues of human (*Homo sapiens*) K0513 (HsK0513) include chimpanzee (*Pan troglodytes,* PtK0513), Rhesus monkey (*Macaca mulatta*, MamK0513), dog (*Canis lupus,* ClK0513), cow (*Bos taurus*, BtK0513), mouse (*Mus musculus*, MmK0513*)*, rat (*Rattus norvegicus*, RnK0513), chicken (*Gallus gallus,* GgK0513), zebrafish (*Danio rerio,* DrK0513), and frog (*Xenopus tropicalis,* XtK0513). A multisequence alignment of K0513 and its homologues was carried out (Figure 2.1) The N-terminal displayed a partial conservation across the HsK0513, and its homologues compared to the carboxylic end which is highly conserved (Figure 2.1). The alignment matrix shows a very high percentage sequence identity with PtK0513 and a very low identity to DrK0513 (Table 2.1), DrK0513 also showed low N-terminal conservation when aligned with other homologues (Figure 2.1). K0513 possesses a partial myotubularin related/ set binding factor 2 (SBF2) domain as indicated by NCBI (NCBI accession number: NP_055547.1), the 72 amino acids long domain is located towards the carboxyl-terminal and is highly conserved throughout K0513 homologues (Figure 2.1). The secondary structure of K0513 is only comprised of alpha helices and coils, the helices comprising 70 % of the whole structure. The predicted transmembrane region (150-167) shows high level of conservation across all the homologues.

Table 2.1. Percentage identity matrix of HsK0513 and its homologues

	HsK0513	DrK0513	XtK0513	GgK0513	MmK0513	RnK0513	MamK0513	PtK0513	ClK0513	BtK0513
HsK0513	-	68.92	73.35	78.92	89.68	89.43	98.05	99.51	90.00	89.27
DrK0513	68.92	-	64.27	67.17	66.58	66.58	68.92	68.92	67.59	67.59
XtK0513	73.35	64.27	-	74.88	73.33	73.58	72.13	73.11	72.55	72.30
GgK0513	78.92	67.17	74.88	-	78.22	78.22	79.17	78.68	79.36	79.36
MmK0513	89.68	66.58	73.33	78.22	-	99.02	88.94	89.43	87.96	86.45
RnK0513	89.43	66.58	73.58	78.22	99.02	-	88.94	89.43	87.71	86.70
MamK0513	98.05	68.92	72.13	79.17	88.94	88.94	-	98.05	89.27	88.78
PtK0513	99.51	68.92	73.11	78.68	89.43	89.43	98.05	-	89.76	89.27
ClK0513	90.00	67.59	72.55	79.36	87.96	87.71	89.27	89.76	-	90.95
BtK0513	89.27	67.59	72.30	79.36	86.45	86.70	88.78	89.27	90.95	-

HsK0513- *Homo sapiens* K0513, DrK0513- *Danio rerio* K0513, XtK0513- *Xenopus tropicalis, GgK0513- Gallus gallus, MmK0513- Mus musculus, RnK0513- Rattus norvegicus, MamK0513- Macaca mulatta, PtK0513- Pan troglodytes, ClK0513- Canis lupus, BtK0513- Bos taurus*. The numbers represent percentage identity. The matrix was generated using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/).

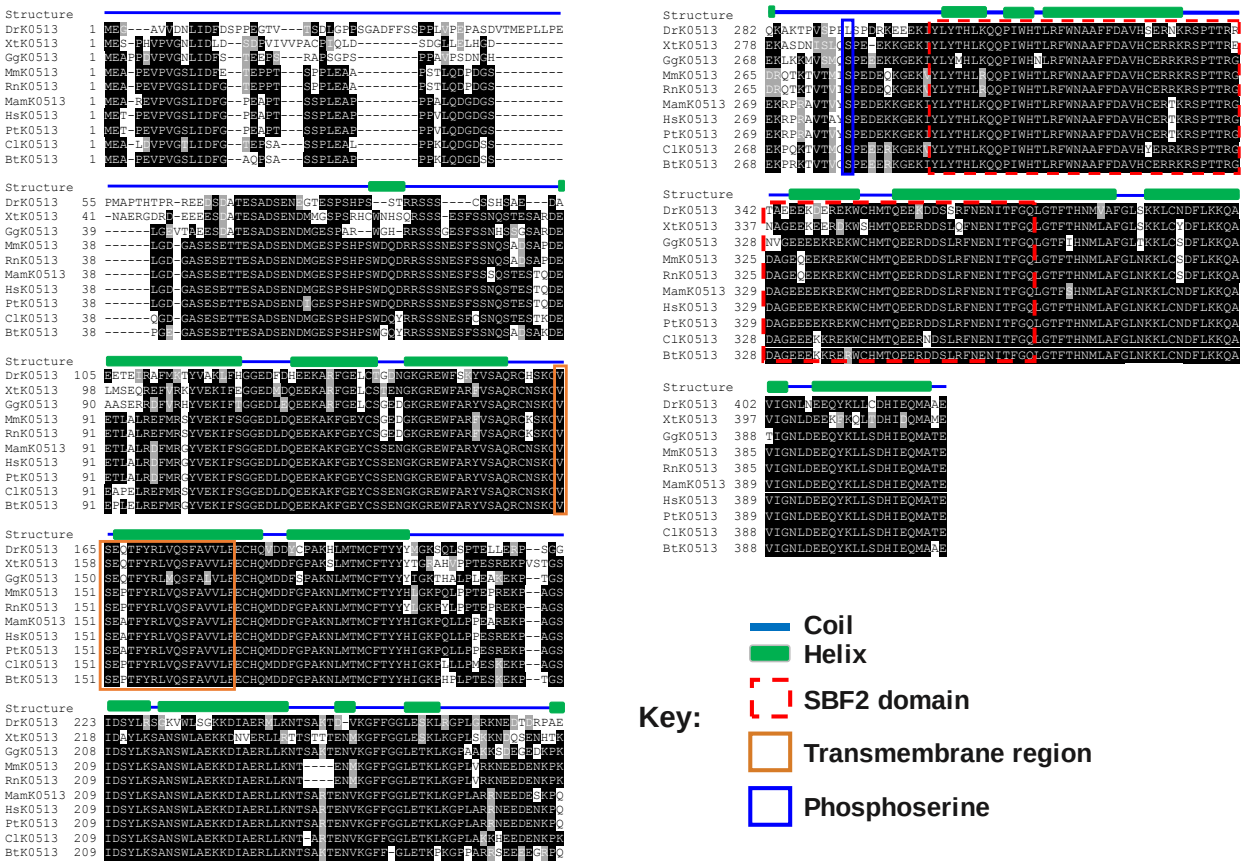


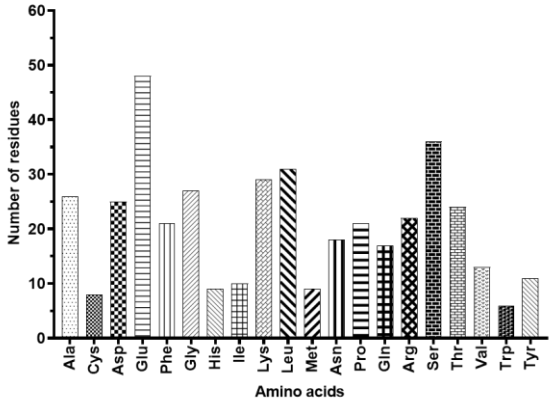
Figure 2.1. Multiple sequence alignment of human K0513 and its homologues.
Multiple sequence alignment of K0513 from *Homo sapiens* (HsK0513, NP_055547.1), with its homologues from *Pan troglodytes* (PtK0513, XP_003952958.1), *Xenopus tropicalis* (XtK0513, NP_001096454.1), *Canis lupus* (ClK0513, XP_005620695.1), *Bos taurus* (BtK0513, XP_005218484.1), *Mus musculus* (MmK0513, NP_001157232.1), *Rattus norvegicus* (RnK0513, NP_001100906.1), *Gallus gallus* (GgK0513, XP_001235240.2), *Danio rerio* (DrK0513, NP_001002193.1), and from *Macaca mulatta* (MamK0513, XP_001113080.1). Identical residues are represented by white font against a black background; similar residues are represented by white font against a grey background whereas the black against white background represents non-conserved residues. The alignment was generated using MUSCLE (https://www.ebi.ac.uk/Tools/msa/muscle/). The numbers on the left-hand side represent the position of the residues. The secondary structure was generated using Phyre[2] and GST C-terminal domain-like crystal structure was used as a template. Coils are represented by a thin blue line and the helices with green solid box.

2.3.2. Structural analysis and physiochemical properties of K0513

The amino acid composition and physiochemical properties determine the fundamental properties of a protein. The amino acid composition of HsK0513 was calculated using PEPSTAT and represented in a form of a bar graph (Figure 2.2A). The primary structure analysis suggests that HsK0513 has a high percentage of polar residues (55.474 %) making it more hydrophilic in nature (Figure 2.2A and 2.2B). The presence of the Cysteine residues suggests the possible presence of disulphide bridges in HsK0513 for tertiary structure stability. The computed isoelectric point (pI) of HsK0513 is 4.6994, this suggest that the molecule is acidic in nature. This is due to high number of an acidic glutamic acid (Glu)

residues (Figure 2.3). K0513 homologues shows similar pI (Appendix A1). The acidity may suggest K0513 involvement in cytoskeleton processes.

A



B

Property	Residues	Number	Mole%
Tiny	(A+C+G+S+T)	121	29.440
Small	(A+B+C+D+G+N+P+S+T+V)	198	48.175
Aliphatic	(A+I+L+V)	80	19.465
Aromatic	(F+H+W+Y)	47	11.436
Non-polar	(A+C+F+G+I+L+M+P+V+W+Y)	183	44.526
Polar	(D+E+H+K+N+Q+R+S+T+Z)	228	55.474
Charged	(B+D+E+H+K+R+Z)	133	32.360
Basic	(H+K+R)	60	14.599
Acidic	(B+D+E+Z)	73	17.762

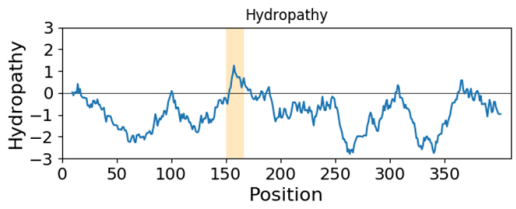
C

Molecular weight	46638.57 Da
Theoretical pI	4.6994
Residues	411
Net charge	-17.5
Extinction coefficient	49390

Figure 2.2. Amino acid composition and physicochemical properties of K0513.
The amino acid composition of K0513 was computed using PEPSTAT module integrated in the EMBOSS software (**A**). The properties of present amino acids are summarized in (**B**). physicochemical characterization of the protein was computed using the Expasy's ProtParam server (**C**).

The secondary structure of human K0513 was further analyzed for hydropathy, amphipacity and topology. K0513 molecule has both hydrophilic, hydrophobic and amphipathic residues with hydrophilic residues dominating most of the protein surface (Figure 2.3A). Amino acid residue 150 to 167 were predicted to be hydrophobic and the region is suggested to form part of the transmembrane anchor (Figure 2.3A and 2.3B). Both the N-terminal and the C-terminal are cytosolic and anchored to the membrane by the transmembrane region (Figure 2.3B).

A



B

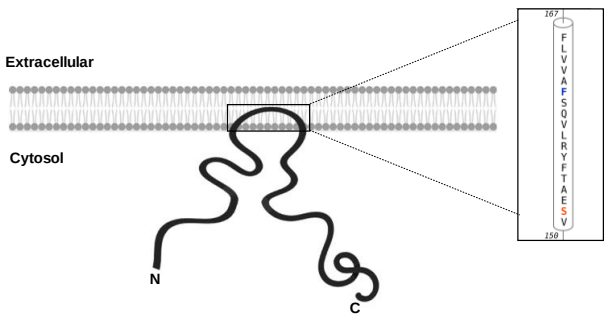


Figure 2.3. K0513 secondary structural analysis.
Hydropathy plot showing the degree of hydrophobicity and hydrophilicity, the hydrophobic transmembrane region is highlighted in brown (**A**). The Hidden Markov Model for Topology Prediction (HMMTOP) showing helical-transmembrane residues represented by mustard colour and the position of the carboxyl-terminal and amino-terminal relative to the cell cytosol (**B**). The models were generated using Biotools Web-based Hydropathy, Amphipathicity and Topology (WHAT) (http://biotools.tcdb.org/barwhat2.html).

2.3.3. K0513 homology modelling

A three-dimension model of K0513 was generated using Phyre[2]. The human K0513 exhibit a globular 3D homology model with the N-terminal forming a loop which is in proximity with the C-terminal end (Figure 2.4A). The N-terminal and the C-terminal end of HsK0513 are more hydrophobic compared to some parts of the middle domains (Figure 2.4B). The 3D models for HsK0513 (Figure 2.4A), PtK0513, MmK0513, and DrK0513 (Figure 2.4) shows structural differences despite high amino acid sequence identity (Table 2.1). This can be accounted to different template structure used for modelling. The 3D model K0513 and the homologues consist of alpha helices and loops (Figure 2.4).

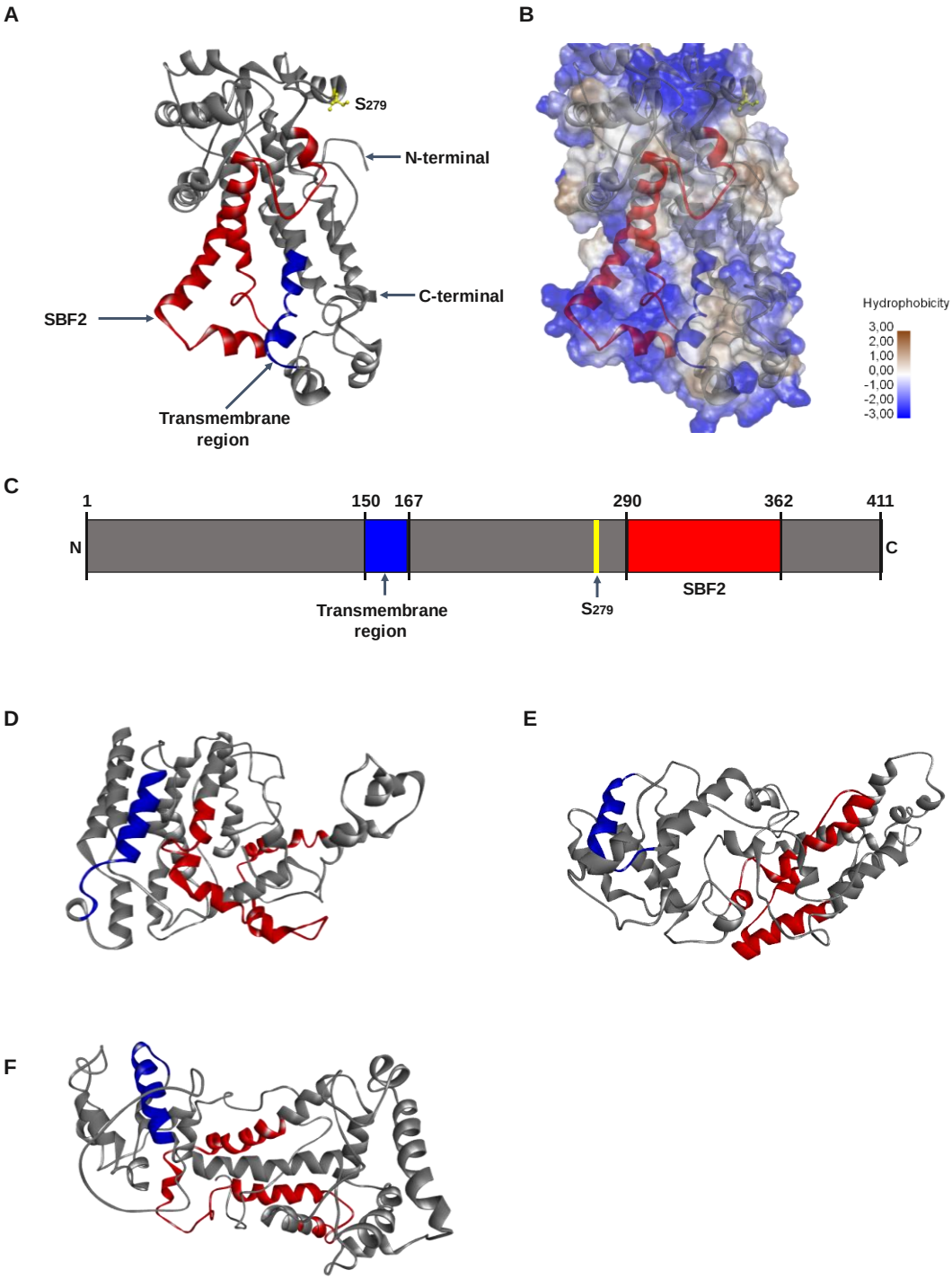


Figure 2.4. Three-dimension homology model of K0513 and its homologues.
The 3D model of K0513 showing the SBF2 domain (red) and the phosphoserine residue represented in balls and sticks (yellow), and the transmembrane region (blue) (**A**). The hydrophobicity surface plot showing hydrophobic (brown), amphipathic (grey) and hydrophilic residues (blue) (**B**). Schematic presentation of KO513 structure (**C**). 3D homology models of K0513 from Chimpanzee PtK0513 (**D**), Mouse MmK0513 (**E**) and Zebrafish DrK0513 (**F**) were also generated. All the 3D models were generated by Phyre[2] (http://sbg.bio.ic.ac.uk/phyre2) and visualized using BIOVIA Discovery studio (Dassault Systemes Biovia, 2016).

2.3.4. Predicted interactome of human K0513

The PPIs study revealed 71 possible interactors of K0513 (Table 2.2; Appendix A, Table A3), 67 interaction partners were generated through prediction. Six of predicted interactors; INTS4, HAX1, WWC1, forkhead box K1 (FOXK1), forkhead box K2 (FOXK2), and amyloid precursor protein (APP) have been identified and validated using yeast 2-hybrid system (Lauriat *et al*., 2006). The predicted interactome include proteins involved in GTPase activities [ras related protein in brain 3a (Rab3a), rho GDP-dissociation inhibitor 3 (ARHGDIG), IQ motif and sec7 domain ArfGEF 3 (IQSEC3) and syntaxin-binding protein 1 (STXBP1)], vesicle trafficking/membrane processes [AP2-associated protein kinase 1(AAK1), guanine nucleotide-binding protein G (o) subunit alpha (GNAO1), calcium channel gamma-3 subunit (CACNG3), Synaptophysin (SYP), Reticulon-1 (RTN1), AP2-interacting clathrin-endocytosis protein (KIAA1107), Ephrin type-B receptor 6 (EPHB6) and Vesicle-fusing ATPase (NSF)], neurotransmission/synaptic activity [Synaptosomal-associated protein 25 (SNAP25), GABRA1, Alpha-synuclein (SNCA), RUN domain-containing protein 3A (RUNDC3A), zinc finger protein 365 (ZNF365), SYT1, Protein bassoon (BSN), Synaptic vesicle glycoprotein 2A (SV2A), and Neuronal pentraxin-1 (NPTX1)], transcription regulation [FOXK1, FOXK2, Chromodomain-helicase-DNA-binding protein 5 (CHD5), and INTS4], kinases [Phosphatidylinositol 4-kinase alpha (PI4KA), Microtubule-associated serine/threonine-protein kinase 3 (MAST3), and Protein kinase C beta type (PRKCB)], signal transduction [WWC1, Modulator of apoptosis 1 (MOAP1), C-Jun-amino-terminal kinase-interacting protein 2 (MAPK8IP2), and HAX1], as well as solute carriers (SLC12A5, Sodium/hydrogen exchanger 6 (SLC9A6) and Sodium/calcium exchanger 2 (SLC8A2)].

Table 2.2: The predicted interactors of human K0513

UniProt ID	Name	Description	Localization	PMIDs
Q2M2I8	AAK1	AP2-associated protein kinase 1, Regulator of the clathrin-mediated endocytosis	Cell membrane	25402006
P42356	PI4KA	Phosphatidylinositol 4-kinase alpha, Acts on the production of the second messenger inositol-1,4,5, -trisphosphate as phosphatidylinositol (PtdIns)	Cell membrane	25402006
P61764	STXBP1	Syntaxin-binding protein 1, Interacts with GTP-binding proteins Participates in the regulation of synaptic vesicle docking and fusion	Cytosol	25402006
P20336	RAB3A	Ras-related protein in brain 3A, Small GTPase protein that serves central role in vesicle trafficking, exocytosis and secretion	Cytosol, Cell membrane	25402006
Q9UPP2	IQSEC3	IQ motif and SEC7 domain-containing protein 3, A guanine nucleotide exchange factor (GEF) for small GTPase ARF1	Cytosol	25402006
Q13387	MAPK8IP2	C-Jun-amino-terminal kinase-interacting protein 2, Group of scaffold proteins selectively mediates JNK signalling	Cytosol	25402006
Q8IX03	WWC1	Protein KIBRA, Plays an important role in tumour suppression through proliferation restriction and apoptosis promotion	Nucleus, cell membrane	17010949; 21836163
Q99819	ARHGDIG	Rho GDP-dissociation inhibitor 3, Inhibits the exchange of GDP/GTP in RhoB	Cytoplasm	25402006
Q9H2X9	SLC12A5	Solute carrier family 12 member 5, required for neuronal Cl- homeostasis, and facilitates electroneutral potassium-chloride cotransport in mature neurons	Cell membrane	25402006
O00165	HAX1	HCLS1-associated protein X-1, and regulates reorganization of the cortical actin cytoskeleton through recruitment of the Arp2/3 complex to the cell cortex	Nucleus membrane	17010949; 21836163

2.3.5. Domain conservation analysis between K0513 and MTMR13

To further elucidate the link between K0513 and SBF2 domain, a well characterized MTMR13 was aligned with K0513 for similarity (Figure 2.5). The pairwise alignment shows that the N-terminal of K0513 is partially conserved the cDENN and dDENN modules of the tripartite DENN domain. The predicted SBF2 domain (290-362) within K0513 is partially conserved with the MTMR13's SBF2 region (531-754).

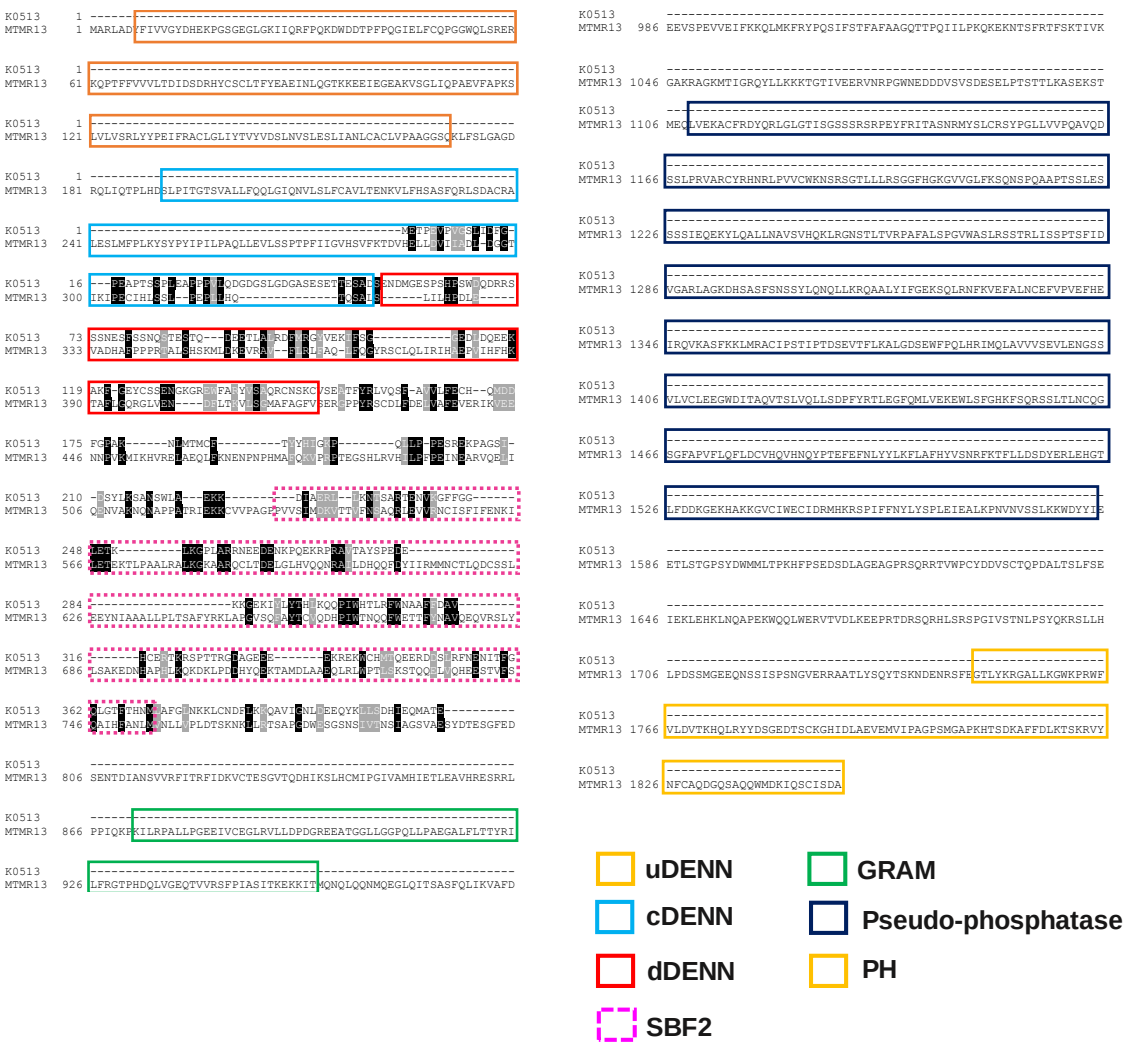


Figure 2.5. Pairwise alignment of K0513 with Denn domain containing MTMR13
K0513 was aligned with the known DENN domain containing GEF from myotubularin family. Coloured figures represent domains found in MTMR13 and their respective positions. The uDENN (orange, 7-172) is situated upstream of the cDENN (blue, 191-324) followed by the dDENN (red, 326-427) domain. The SBF2 domain is represented in magenta (531-754), the GRAM (Glucosyltransferase/Rab-like GTPase activator/myotubularin domain) domain is represented in green (871-957) followed by the pseudo-phosphatase (purple, 1108-1584), and lastly the PH (Plenkin homolog) represented in gold (1743-1847) situated towards the C-terminal. The alignment was generated using EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/).

The SBF2 domain in K0513 is consists of alpha helices and loop, this observation is different compared to the MTMR13's SBF2 consisting of beta sheets, alpha helices and loops. However, high percentage of the models is made of helices which may suggest some structural similarities between the N-terminal of MTMR13 and K0513 (Figure 2.6). The secondary structural difference can be a result of low homology in which K0513 3D model was generated and the diversity among the DENN domains (Wu *et al*., 2011). The MTMR13's SBF2 domain is found in upstream the tripartite DENN (uDENN, cDENN and dDENN) domain, this may suggest functional relationship within these domains.

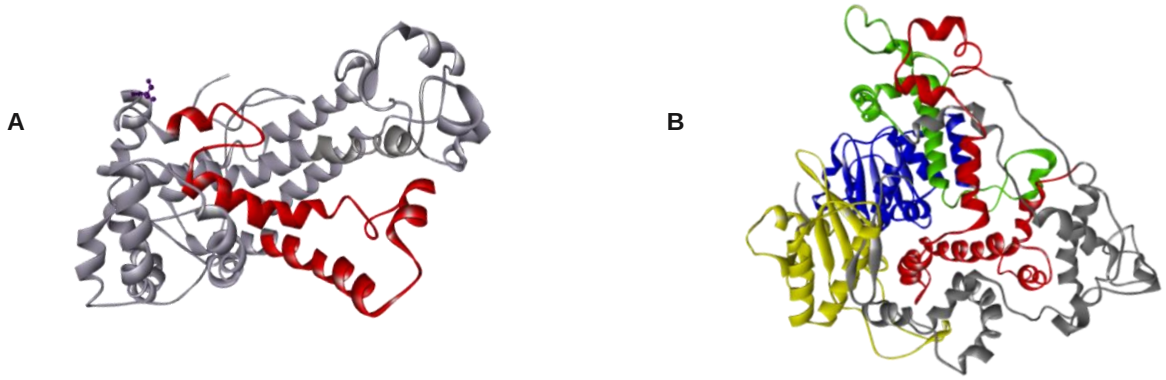


Figure 2.6. structural comparison of K0513 and MTMR13 three-dimensional structures.
(**A**) K0513 3D model shows the secondary structural composition of the SBF2 domain (red). (**B**) shows the structural and domain organisation in MTMR13 from residue 1 to 755; uDENN (yellow, 7-172), the cDENN (blue, 191-324), dDENN (light green, 326-427), SBF2 (red, 531-754). All the 3D models were generated by Phyre[2] (http://sbg.bio.ic.ac.uk/phyre2) and visualized using BIOVIA Discovery studio (Dassault Systemes Biovia, 2016).

2.3.6. Evolutionary analysis of K0513

A phylogenetic tree was generated using MEGA to investigate the evolutionary relationships of K0513 in different species (Figure 2.7). The analysis revealed 65 species expressing K0513 molecule (Figure 2.6). Of all the 65 protein sequences, there were nine homologues and 56 orthologues of human K0513. Most of the species are members of endopterygotes with 14 species, this is a superorder of insects that undergoes metamorphosis between their larval and adult stage. Human K0513 clusters with *Gorilla gorilla* K0513 in a monophyletic primate's clade close to *Pan troglodytes* (identified homologue). Human K0513 was also found to cluster closely with K0513 from rodents signifying its importance in regulating signal pathways in different species.

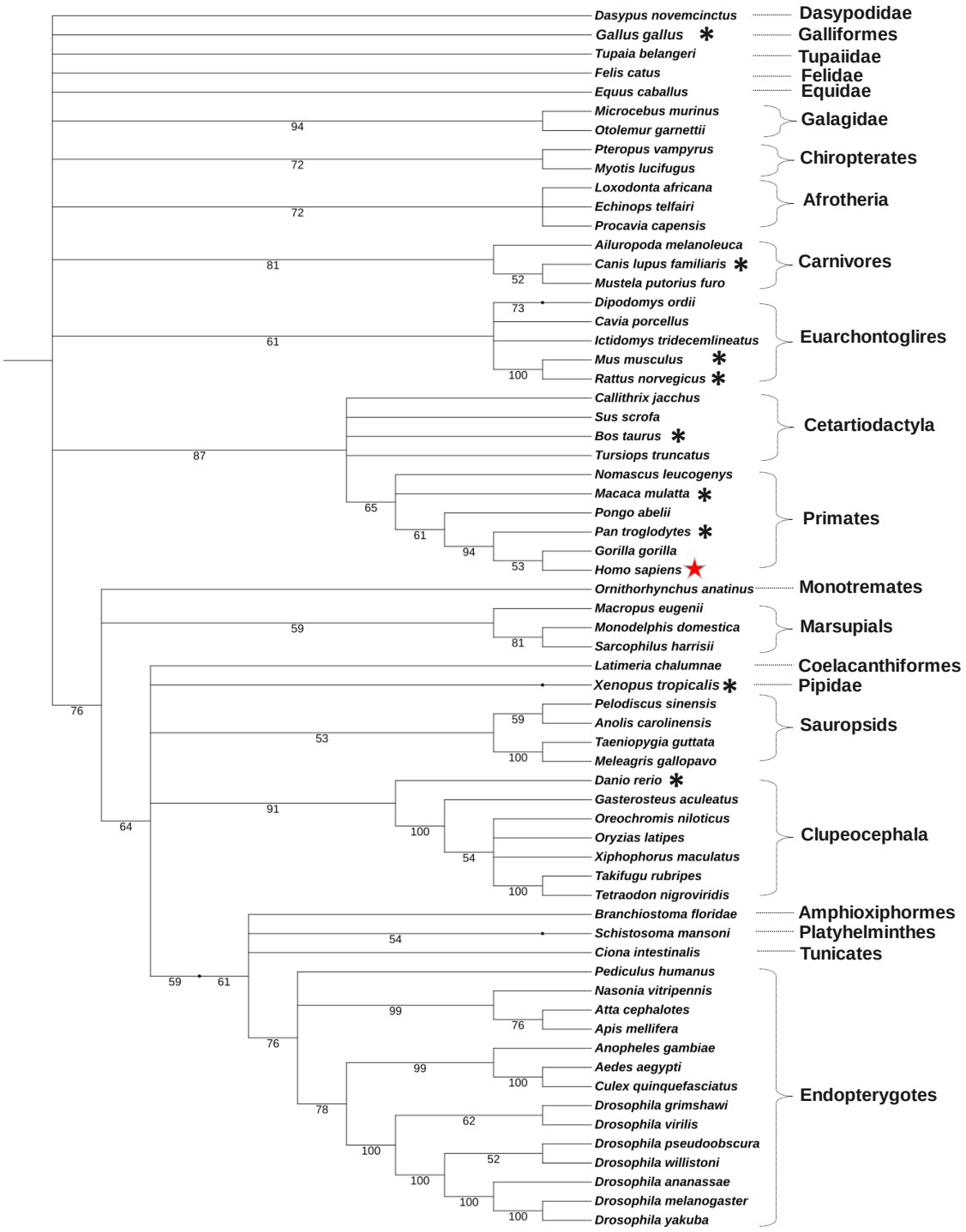


Figure 2.7. Evolutionary analysis of K0513 by Maximum Likelihood method
The evolutionary history of K0513 across different species was inferred by using the Maximum Likelihood method and JTT matrix-based model (Jones *et al.,* 1992). The bootstrap consensus tree inferred from 100 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. This analysis involved 65 amino acid sequences. Red allosteric indicates Homo sapiens K0513, and all homologues are indicated with a star. Evolutionary analysis were conducted in MEGA X (Kumar *et al.,* 2018) and further annotated using ITOL (https://itol.embl.de/, Leturnic and Bork, 2019).

2.4. Discussion

Protein domain can be defined as a structural unit of a protein and provides an insight into the function of the protein or its interaction. Sequence retrieved from NCBI (NCBI accession number: NP_055547.1) revealed the presence of 72 residues long Set Binding Factor 2 (SBF2) domain located at position 290 to 362 of K0513. SBF2 domain is also known as the myotubularin domain as stated by the Pfam database (El-Gebali *et al*., 2019). Multiple sequence alignment shows a high conservation of SBF2 domain across all K0513 homologues, this suggest that the SBF2 domain may be a functional feature that defines and identifies K0513 protein molecule across different species. The SBF2 domain is found and forms part of the middle region of a key myotubularin family member known as myotubularin related protein 13 (MTMR13) (Senderek *et al*., 2003). The SBF2 domain in MTMR13 is 223 amino acid in length which is 151 residues longer that the 72 amino acid long K0513's SBF2 domain. It is interesting to find that despite K0513's SBF2 domain being 151 amino acid shorter than the MTMR13's SBF2, it has been identified through BLAST as an SBF2 domain. This suggest that K0513 contains the truncated form of the SBF2 domain which may serves the same function as the full length SBF2 domain. The SBF2 domain of MTMR13 is found in association with the Differentially Expressed in Neoplastic versus Normal cells (DENN) domain (Levivier *et al*., 2001). The DENN domain binds to switches I and II of Rab21 GTPase resulting in conformational change in the region to allow GDP to GTP exchange (Wu *et al*., 2011). The tripartite DENN module is comprised of the upstream DENN (uDENN), followed by central DENN (cDENN), and the downstream DENN (dDENN). The uDENN may occur independently of the DENN module and interact with various GTPases (Schlenker *et al*., 2006). The central cDENN and dDENN occur together, the d-DENN is reported to make fewer contact with the GTPase thus it always associates with cDENN to maintain the GEF function (Wu *et al*., 2011). When K0513 is aligned with MTMR13 using pairwise alignment, its N-terminal shows a partial conservation with the cDENN and dDENN. In addition to that, the predicted SBF2 domain of K0513 shows a partial conservation with MTMR13's SBF2. This taken together suggest that K0513 possess nucleotide exchange property and may serve as a GEF for Rab GTPases. There is no evidence on how SBF2 domain facilitate the nucleotide exchange activity of the DENN domain, however, their co-existence suggest they have functional relationship.

Physicochemical analysis shows that K0513 possess a transmembrane region (Figure 2.3). This region consists of 17 residues forming up an alpha-helix. Transmembrane proteins can be integrally anchored to the lipid bilayer for interaction with membrane proteins. For example, general receptor for phosphoinositides 1 (Grp1), a known GEF interact with PI3-kinase in the inner membrane through its pleckstrin homology (PH) domain (*Albers, 2012*). When anchored to the membrane, Grp1 catalyses the nucleotide exchange and activation of ADP-ribosylation factor-6 (Arf6) (Venkataraman *et al*., 2012). The presence of the transmembrane may suggest that K0513 performs its nucleotide exchange function on membrane localized Rab GTPases when it is anchored to the lipid bilayer. Furthermore, it may directly and indirectly facilitate cellular processes such as vesicle trafficking or serve in transducing signal in pathways such as neuroplasticity, cytoskeletal regulation and apoptosis (Lauriat *et al*., 2006). The predicted interaction with the transmembrane proteins, most notably the Na^{+}/Ca^{2+} Solute Carrier Family 8 Member A2 (SLC8A2) further substantiate this finding (Table 2.2). Most importantly, Rab3a, a GTPase involved in vesicle trafficking has been predicted to interact with K0513 supporting the role of K0513 along the membrane.

Another important finding is the identification of the possible phosphorylation in Ser-279 (Figure 2.1). Protein phosphorylation serves as an initial step crucial for facilitating cellular and organic functions such as regulation of metabolism, subcellular trafficking, and signal transduction (Ardito *et al*., 2017). Phosphoresidues can be activated or deactivated by kinases or phosphatases, this modification induces conformational change when the protein interacts with other proteins. Some GEFs require phosphorylation to activate their GEF activity, p115, the smallest member of the RH-Rho GEF subfamily gets phosphorylated at Tyr-738 by Janus kinase 2 (JAK2) to positively regulate its GEF activity (Hodgson, 2014). The presence of the phosphoserine residue suggest that K0513 is required to undergo phosphorylation to perform its GEF function or other cellular processes.

The three-dimension homology model of K0513 was successfully generated using $PHYRE^{2}$. This is the first study to report on the hypothetical structure of K0513

molecule (Figure 2.4). The three-dimension model reveal the globular nature of K0513 implying water solubility with the N-terminal end in proximity with the C-terminal. The 72 residues long SBF2 domain forms a helix that coils towards the transmembrane region, this further strengthen our hypothesis that K0513 interact with Rab GTPases along the membrane through its N-terminal and the SBF2 domain. However, due to low homology of the templates used, we acknowledge that K0513 may resemble a more complex structure than the one presented. The K0513 structure undergo conformational change when phosphorylated at Ser-279.

Interaction prediction study shows diverse interactors that forms part of important cellular processes (Table 2.2; Table A3). This includes prediction of membrane proteins, amongst this are vesicle trafficking proteins such as VAMP2. VAMP2 mediates trafficking of α5β1 along the plasma membrane (Hasan *et al*., 2010). Prediction of membrane proteins indicates that K0513 may be involved in regulating membrane processes. Notably, the prediction shows that K0513 possibly interacts with small GTPases, most importantly Rab3a (Table 2), this GDP/GTP binding protein is involved in vesicle trafficking (Van Weering *et al*., 2007). This finding further substantiates the role of K0513 as a potential GEF (Wang *et al.,* 2017). The predicted kinases (PI4KA, MAST3, and PRKCB) indicates a possible phosphorylation of K0513 on Ser-279 to execute its functions. Proteins involved in transcription regulation were also identified as part of the interactome (Table 2.2, Table A3). This amongst others include INTS4 which has been experimentally validated to interact through yeast two-hybrid system (Lauriat *et al*., 2006). INTS4 is a transcriptional regulatory complex which is associated with RNA polymerase II (Albrecht *et al*., 2018).

Chapter 3

Overexpression and purification of recombinant K0513

3.1. Introduction

Production and purification of recombinant protein serves as a basis for structural and functional characterization of proteins. The production requires an appropriate expression system, *E. coli* is used as one of the systems in recombinant protein expression (Ferrer-Miralles, 2015). The preference is based on; amongst other, the rapid expression of proteins (Ferrer-Miralles, 2015), high culture cell density (Shiloach and Fass, 2005), and easy transformation of exogenous DNA (Pope and Kent, 1996). However, the *E. coli* system lacks most of posttranslational modifications found in eukaryotes such as glycosylation, palmitation, hydroxylation and sulfation (Sahdev *et al*., 2008; Brondyk, 2009).

The low expression yields of eukaryotic gene in *E. coli* has been reported as one of the major draw backs in using this expression system. Several strategies have been developed to counteract this problem; this includes optimization of the expression temperature (Song *et al*., 2012), use of improved host system such as *E. coli* BL21 (DE3) (Hortsch and Weuster-Botz, 2011), and of interest to this study the harmonization of codon usage (Angov *et al*., 2008). When a foreign eukaryotic protein is expressed in a different host such as the prokaryotic *E. coli* system, there may be a significant difference in frequency of occurrence in synonymous codons relative to the host's DNA. This phenomenon is referred to as codon bias (Rosano and Ceccarelli, 2014). This discrepancy leads to low abundance of tRNAs which results in either incorporation of amino acids, or truncation of the protein (Gustafsson *et al*., 2004; Rosano and Ceccarelli, 2014).

In order to trace the expression and purification scheme of the recombinant proteins, proteins are expressed in tandem with a stretch of amino acids (fusion tag) which allows production of a chimeric protein (Nilsson *et al*., 1997). The presence of the fusion tag may also maximise the solubility and allow for easy purification from the *E. coli* system, and their presence has been reported to be less likely interfere with the function of the fused protein (Khan *et al*., 2012). Available vectors allow these small tags to be positioned at either the N-terminal or the C-terminal of the fused protein, therefore it is wise to choose a vector wherein the tag will be at the solvent accessible end of your protein of interest (Rosano and Ceccarelli, 2014). The most commonly

used tags include poly-His, poly-Arg, c-Myc, FLAG-, and Strep II- tags (Terpe, 2003). Tags allow for easy purification by using resins that specifically bind to the tag such as nickel-affinity resin, the purification as well as expression can also be confirmed using antibodies designed against the specific tag (Bornhorst and Falke, 2000).

In this study, an *E. coli* expression system is used to produce recombinant His-tagged wildtype ($K0513_W$) and the codon harmonized ($K0513_H$) human K0513. Recombinant expression and purification of K0513 has never been reported. K0513 is cytosolic and ubiquitously expressed, whether it has functions related to cellular membrane which may affect its solubility has never been reported. Therefore, this study seeks to heterologously express and purify recombinant K0513 for further structural and functional analysis.

Specific objectives

1. To confirm the integrity pQE30-$K0513_W$ and pQE60-$K0513_H$ using restriction digestion analysis
2. To express recombinant $K0513_W$ and $K0513_H$ in *E. coli* XL1-Blue cells
3. To purify $K0513_W$ and $K0513_H$ using Nickel-affinity chromatography

3.2. Materials and methods

3.2.1. Materials

Reagents used were purchased from Merck (Germany), Sigma-Aldrich (United States), lford (United Kingdom), Thermo Fisher (United States). The following antibodies were used to validate the presence of both K0513 constructs. Mouse raised monoclonal anti-polyhistidine antibody, goat raised horseradish peroxidase (HRP) conjugated anti-mouse IgG antibody (Thermo Fisher Scientific, USA), rabbit raised polyclonal anti-KIAA0513 antibody and goat raised HRP conjugated anti-rabbit IgG antibody (Abcam, UK). The special reagents used in this study are listed in the table B4. The *E. coli* strains, and plasmid constructs used are listed in the Table 3.1.

Table 3.1. Description of *E. coli* strains and plasmids constructs used

***E. coli* Strains**	**Description**		**Source/reference**
JM109 DE3	e14– (McrA–) recA1 endA1 *gyr*A96 thi-1 hsdR17 (rK – mK+) supE44 *rel*A1 Δ(lac-proAB) (F´ traD36 proAB lacI^q ZΔM15)		Thermo Fisher Scientific
XL1-Blue	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 supE44 *rel*A1 lac [F´ proAB laclq ZΔM15 Tn10 (Tetr)].		Bullock *et al*., (1987)
Plasmids	**Description**	**Antibiotic resistance**	
pQE30-K0513$_W$	pQE30 encoding non-codon harmonized K0513	Ampicillin resistance (Amp^R)	GeneScript This study
pQE60-K0513$_H$	pQE60 encoding codon harmonized K0513	Amp^R	GeneScript This study

3.2.2. Construction of plasmid vector encoding human K0513

The constructs expressing the human K0513 were synthesized by GeneScript (United States). The wild type K0513 was cloned into a pQE30 plasmid using *Bam* HI and *Hind* III restriction enzymes to generate pQE30-K0513$_W$ harbouring the N-terminal histidine tag and an ampicillin resistance gene. The pQE60 expression vector was used to clone the codon harmonized K0513 using *Nco* I and *Bgl* II restriction enzymes

to generate pQE60-K0513$_H$, this plasmid possesses, most importantly the C-terminal histidine tag and ampicillin resistance gene. Restriction analysis was used to confirm the identity of the constructs. *Bam* HI and *Hind* III were used to digest pQE30-K0513$_W$, *Nco* I and *Bgl* II for pQE60-K0513$_H$ (Appendix B1, Table B1). Agarose gel electrophoresis was used to visualize the resulting DNA fragments (Appendix B1).

3.2.3. Purification of His-tagged constructs

3.2.3.1. Recombinant expression of pQE30-K0513$_W$ and pQE60-K0513$_H$ constructs

Chemically prepared *E. coli* XL1-Blue competent cells (Appendix B2) were transformed with pQE30-K0513$_W$ or pQE60-K0513$_H$ (Appendix B3). A colony was inoculated into 25 mL of 2 x YT media (1.6 % tryptone, 1 % yeast extract and 0.5 % NaCl) supplemented with 100 µg/mL ampicillin and incubated overnight with shaking at 37 °C. The overnight culture was diluted 10 times into a freshly prepared 2 x YT media supplemented with 100 µg/mL ampicillin and incubated to mid exponential growth at OD_{600}. The expression of both the constructs was induced with 1mM IPTG and samples were collected every hour for 5 hours. The cells were harvested at the 5th hour by spinning at 4000 x g for 20 min at 4 °C. The pellets were resuspended in lysis buffer (0, 1 M Tris, [pH 7.5]; 300 mM NaCl, 10 mM Imidazole, 1 mg/mL lysozyme; 1 mM PMSF) and frozen at -80 °C. samples were prepared and analyzed using Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Western blotting (Appendix B4).

3.2.3.2. Solubility study of recombinant pQE30-K0513$_W$ and pQE60-K0513$_H$ constructs

To determine the solubility of K0513, cell pellets from -80 °C freezer were allowed to thaw on ice. The cells were reconstituted with native buffer and were resuspended in lysis buffer containing either 8 M urea, 1 % Triton X-100 or the combination of two. Collected samples were prepared and analyzed using SDS-PAGE and Western blotting (Appendix B4).

3.2.4.3. Purification of recombinant pQE30-K0513$_W$ and pQE60-K0513$_H$ constructs

Recombinant protein was purified using nickel-affinity chromatography under denaturing conditions for the N-terminal His-tagged K0513$_W$ and under native conditions for C-terminal His-tagged K0513$_H$. The native purification followed a protocol described by Zininga *et al*., (2015a and b) with some modifications. Briefly, cell lysates from -80 ºC were allowed to thaw on ice and subsequently sonicated at 30 % amplitude with pulses of 15 seconds. The homogenate was centrifuged at 5000 x g for 20 minutes at 4 ºC, the soluble fraction was kept on ice for further purification and the pellet was resuspended with 2.5 mL PBS (137 mM NaCl, 27 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4). The soluble fraction was incubated with 2 mL (50 % slurry) HisPur Ni-NTA Resin (Thermo Fisher Scientific) for four hours on ice. A column was prepared and washed with 70 % ethanol followed by equilibration with the lysis buffer. The lysate was added to the column, beads resins were allowed to sediment before collecting the flow through. The column was washed three times with two bed volume wash buffer (0, 1 M Tris, [pH 7.5]; 300 mM NaCl, 25 mM Imidazole) and samples were collected. The protein was eluted with 2 mL of elution buffer (0, 1 M Tris, [pH 7.5]; 300 mM NaCl, 500 mM Imidazole, 1 mM PMSF). The recombinant proteins were dialyzed overnight in dialysis buffer (300 mM NaCl, 10 mM Imidazole, 10 mM Tris-HCl, pH 7.5, 10 % (v/v) glycerol, containing 1 mM PMSF) and quantified using Bradford's assay (Appendix B5, Table 5). SDS-PAGE analysis was used to validate the purity of K0513$_W$ and K0513$_H$ and Western blot analysis was used to verify the identity and integrity of the protein (Appendix B4).

3.3. Results

3.3.1. Constructs

3.3.1.1. Confirmation of pQE30-K0513-w$_T$ plasmid integrity

The integrity of pQE30-K0513$_W$ was successfully validated using restriction digestion analysis (Figure 1B). The undigested pQE30-K0513$_W$ shows a single fragment of 2500 bp, this suggests the supercoiled form of the plasmid DNA (Figure 3.1B, lane 2). A single digest using either *Bam* HI or *Hind* III resulted in a linearized form of the construct and shows a fragment at 4661 bp (Figure 3.1B, lane 3 and 4). This is in line with the hypothetical size of pQE30-K0513$_W$ construct as generated on the plasmid map (Figure 3.1A). The double digest using both *Bam* HI and *Hind* III resulted in two fragments, pQE30 at 3425 bp, and K0513$_W$ at the size of 1236 bp (Figure 3.1, lane 5). The supercoiled form of the construct shows a fragment at 3000 bp (Figure 3.1D, lane 1). Single digest of pQE60-K0513$_H$ with *Nco* I and *Bgl* II resulted in 4653 bp linearized plasmid as resolved by agarose electrophoresis (Figure 3.1D, lane 3 and 4). The double digest with both *Nco* I and *Bgl* II resulted in two fragments, pQE60 resolved at 3417 bp and K0513$_H$ at 1236 bp (Figure 3.1D, lane 5). The resulting sizes correspond with the generated plasmid map (Figure 3.1C).

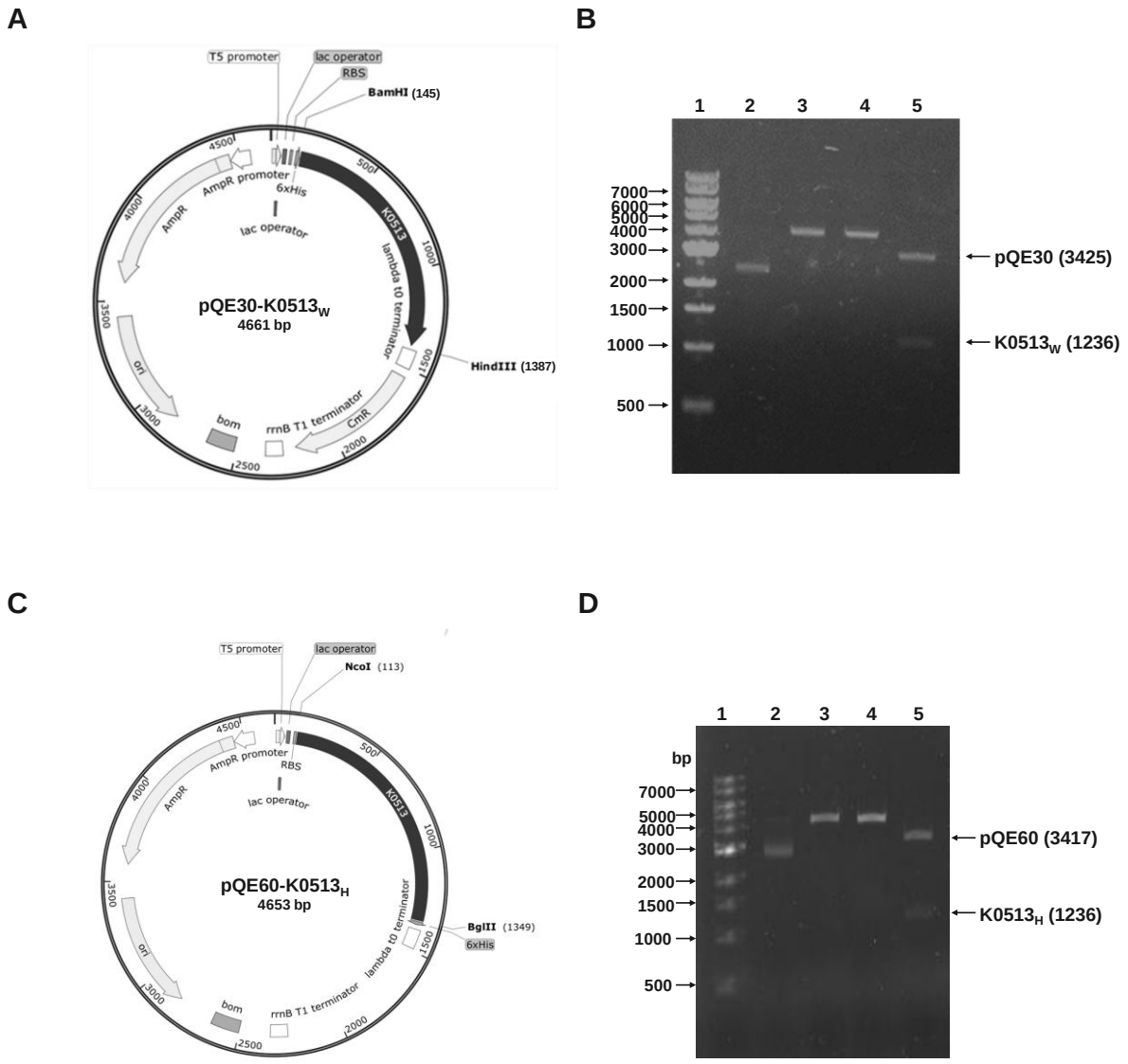


Figure 3.1: Restriction analysis of pQE30-K0513$_W$ and pQE60-K0513$_H$.
Restriction digest analysis of the constructs. (**A**) Plasmid map of pQE30-K0513$_W$ showing restriction sites for *Bam* HI and *Hind* III, the size of K0513$_W$ and the overall construct size. (**B**) Agarose gel electrophoresis of pQE30-K0513$_W$: samples were loaded as follows; Lane 1 - 1 Kb DNA ladder, Lane 2 - uncut pQE30-K0513$_W$, Lane 3 - pQE30-K0513$_W$ restricted with *Bam* HI, Lane 4 - pQE30-K0513$_W$ restricted with *Hind* III, Lane 5 - pQE30-K0513$_W$ double digested with *Bam* HI and *Hind* III. (**C**) Plasmid map of pQE60-K0513$_H$ showing restriction sites for *Nco* I and *Bgl* II, the size of K0513$_H$ and the overall construct size. (**D**) Restriction analysis of pQE60-K0513$_H$ plasmid: samples were loaded as follows; Lane 1 - 1 Kb DNA ladder, Lane 2 - uncut pQE60-K0513$_H$, Lane 3 - pQE60-K0513$_H$ restricted with *Nco* I, Lane 4 - pQE60-K0513$_H$ restricted with *Bgl* II, Lane 5 - pQE60-K0513$_H$ double digested with *Nco* I and *Bgl* II. The plasmid maps were generated using SnapGene.

3.3.2. Recombinant protein production of recombinant K0513w and K0513H

3.3.2.1. Expression of the pQE30-K0513w

The K0513w was expressed in *E. coli* XL1-Blue cells at 37 °C. The expression was assessed using SDS-PAGE and Western blot analysis (Figure 3.2A). The SDS-PAGE results did not depict a distinct K0513w band before and after IPTG induction, however, the expression was confirmed by Western blot analysis which confirms the presence of a band at 55 kDa. The Western blot shows that K0513w is expressed prior to induction by IPTG, a phenomenon described as leaky expression. The pre-induction sample also shows a presence of a distinct band at approximately 38 kDa which was also picked úp by the antibody. The 38 kDa band may suggest the presence of a truncated form of K0513w as a result of protease activity within the *E. coli* system. The induction study shows a similar K0513w yield from the first hour to the fifth hour post-induction (Figure 3.2A). To counteract the leaky expression, K0513w truncation and to improve the expression yield, the temperature was reduced to 20 °C (Figure 3.2B and C). Heterologous expression of K0513w at 20 °C resulted in no truncation prior to and after induction (Figure 3.2B). The high expression yield was observed at 24 hours post induction (Figure 3.2B). Addition of 1 % glucose resulted in a significant reduction of the leaky expression before induction (Figure 3.2C); however, this led to a more prominent truncation of the K0513w (Figure 3.2C).

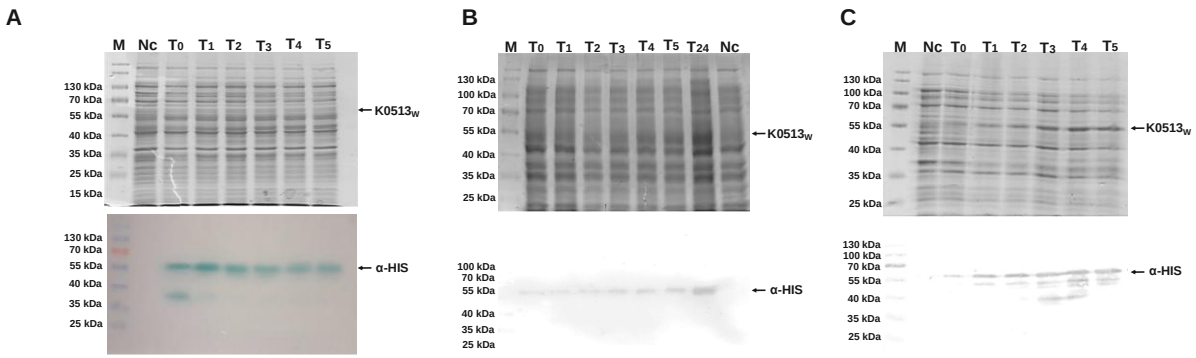


Figure 3.2. Heterologous expression of K0513w.
SDS-PAGE (12 %) analysis of K0513w expression in *E. coli* XL1-Blue cells at 37 °C (**A**), 20 °C (**B**), and in the presence of 1 % glucose (**C**). Samples are as follows; lane M - Prestained protein marker, Nc- cells transformed with pQE30, T0 - uninduced cells transformed with pQE30- K0513w, T1 – T5 (1, 2, 3, 4 and 5-hour post induction samples respectively). Western blot was done using anti-Histidine.

3.3.2.2. Solubility study of pQE30-K0513w

The solubility of recombinant pQE30-K0513w was assessed in the presence and absence of denaturants. K0513w was not soluble in lysis buffer, depicted by a 55 kDa band in the pellet fraction (Figure 3.3B, Lane Pf). This was further confirmed by Western blot analysis which detected a band on sample after sonication and in the pellet fraction, but a very faint band at 55 kDa was detected in the soluble fraction (Figure 3.3B, untreated) which suggest K0513w may be expressed in inclusion bodies or possible hydrophobic interaction with the membrane making it insoluble. The K0513w was then treated by denaturants, urea and Triton X-100. The use of high concentration of urea (8 M) has been shown to completely denature proteins by disrupting the existing secondary structures (Singh *et al*., 2015), the same has been reported when using 1 % of the nonionic detergent Triton X-100 (Noda *et al*., 2017).The denaturants showed a significant improvement in the solubilization of K0513w. The solubilization with 1 % Triton X-100 known to unfold proteins (Noda *et al*., 2017) was partial (Figure 3.3A, 1 % Triton X-100), this is shown by a pronounced 55 kDa band detected in the soluble fraction, however, there was still significant amount of protein detected in the pellet fraction as confirmed by Western blot (Figure 3.3A, 1 % Triton X-100). Highest concentration of 8 M urea known to unfold proteins (Singh *et al*., 2015) showed a significant improvement, this is shown by a pronounced 55 kDa band in the soluble fraction and a faint band in the pellet fraction (Figure 3.3B, 8 M urea). The denaturants (Triton X-100 and urea) were further used in combination to improve the solubilization of K0513w (Figure 3.3A). The combination treatment shows a significant improvement compared to a single treatment using either urea or Triton X-100. This is depicted by a thick band detected using anti-His antibody in the soluble fraction (Figure 3.3A). The negative control shows the absence K0513w.

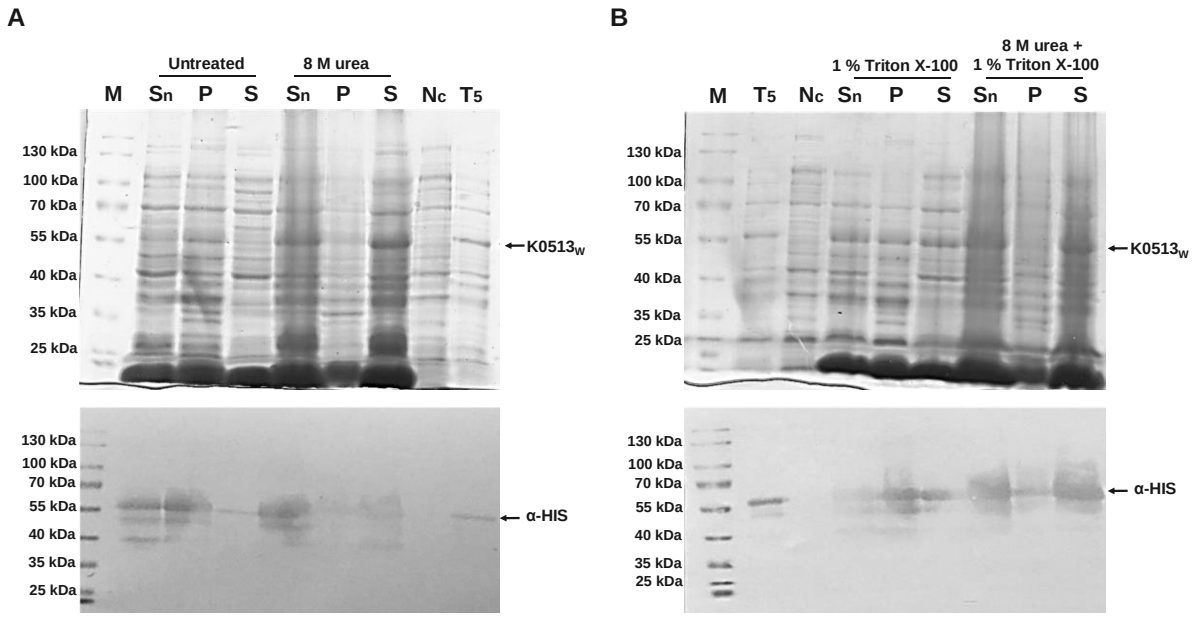


Figure 3.3. Solubility study of K0513$_W$.
K0513$_W$ was solubilized in the presents and absence of detergents. (**A**) Shows untreated and 8 M treated lysates. (B) Shows lysates treated with 1 % Triton X-100 and combination treatment with 8 M urea and 1 % Triton X-100. Lane M – Prestained molecular marker, Sn – cells after sonication, P – pellet fraction, S – soluble fraction, Nc - cells transformed with pQE30, T$_5$ – time of harvest (5 hour) sample. Samples were resolved using 12 % SDS-PAGE and the presence of K0513$_W$ was validated using anti-His (Lower panel).

3.3.2.3. Purification of recombinant K0513$_W$

The recombinant K0513$_W$ was purified under denaturing conditions using Nickel-affinity chromatography. K0513$_W$ was detected in the whole lysate as a 55 kDa species (Figure 3.4, lane Sn), this was also observed in the soluble fraction to show that K0513w was successfully solubilized (Figure 3.4, lane S). There was some protein detected in the flow through, which suggests that K0513$_W$ did not fully bind to the Nickel-charged beads (Figure 3.4, lane Ft). K0513$_W$ was absent from the wash sample (Figure 3.4, lane W1 and W2). The purification of the protein was not successful due to the high level of other contaminating species coming out prominently at 100 kDa, 45 kDa and 25 kDa (Figure 3.4, lane E1, E2 and E3). There is a notable breakdown of K0513$_W$ detected in all the samples, this might be a results of prolonged incubation time as supported by the Western blot (Figure 3.4). To ensure stability of the protein, a shorter incubation time of four hours was used in further purifications. To obtain a pure protein, several modifications to the purification protocol were made.

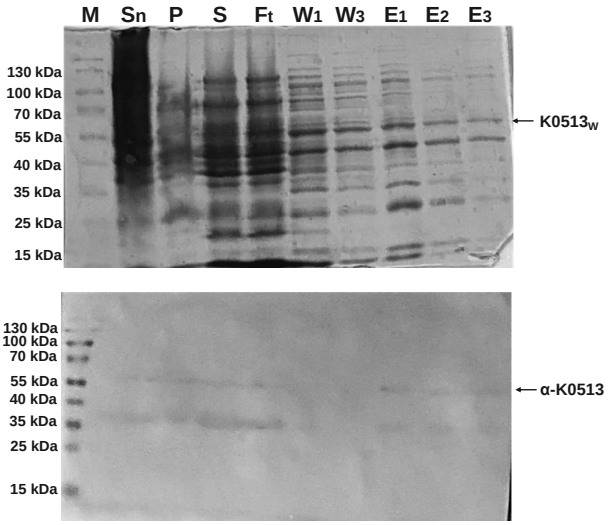


Figure 3.4. Purification of recombinant K0513$_W$.
Purification of K0513$_W$ using Ni-affinity chromatography, SDS-PAGE (12%) analysis (top panel) and Western blot analysis (lower panel). Upper panel: Lanes, M – Protein marker; Nc - *E. coli* XL1-Blue cells transformed with pQE30; Pf – pellet fraction; Sf – supernatant fraction; FT – flow through; W$_1$ and W$_2$ – wash 1 and wash 3, respectively; E$_1$ – E$_3$ – elution 1, elution 2 and elution 3, respectively. Lower panel: detection of K0513$_W$ at 55 kDa using anti-Histidine antibody and visualized using TMB.

To reduce the presence of contaminating protein species, the wash volume was increased and the imidazole concentration for the first wash was increased from 25 mM to 50 mM, and gradient elution was performed from 150 mM to 500 mM imidazole (Figure 3.5A). There was some of K0513$_W$ protein lost in the first wash which is expected because of an increased imidazole concentration (Figure 3.5A, W1). However, there was no significant improvement on the purity of K0513$_W$ since the contaminants were still observed. The contaminating proteins observed during the expression led to an assumption that they might be chaperones from the *E. coli* system, therefore, the column was washed with a wash buffer supplemented with 10 mM ATP and 5 % glycerol (Figure 3.5B). Addition of ATP and glycerol did not show a significant reduction of the contaminants (Figure 3.5B, lane W3). There was a prominent band at 35 kDa suggesting a possible breakdown or truncation of K0513$_W$ detected in all elution's as shown by the Western blot using anti-His antibody.

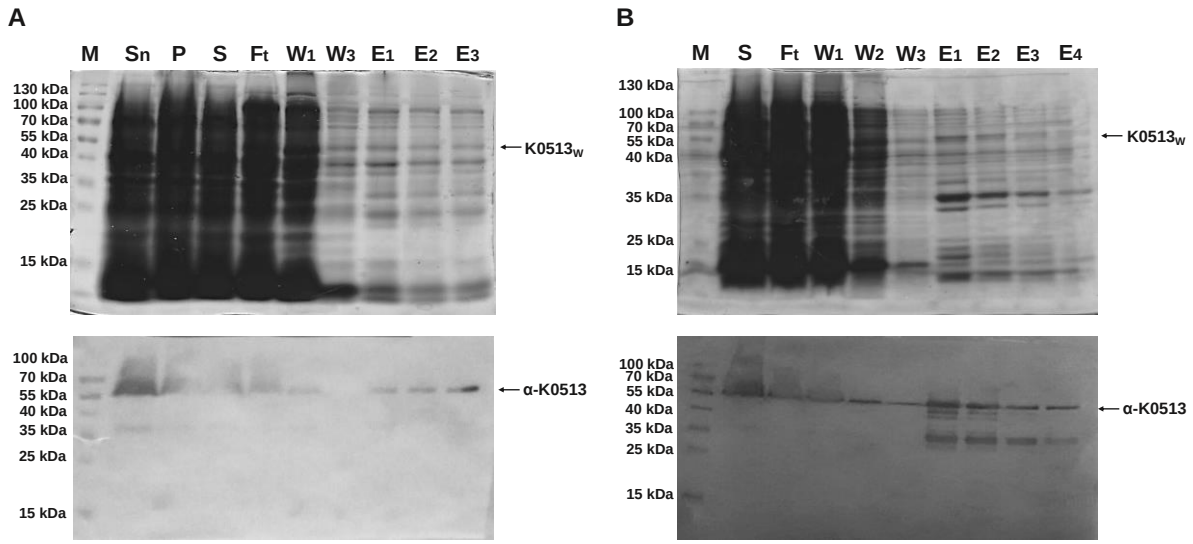


Figure 3.5. Purification optimization of recombinant pQE30-K0513$_W$.
Purification of K0513$_W$ using Ni-affinity chromatography, SDS-PAGE (12%) analysis (top panel) and Western blot analysis (lower panel). Purification was optimized using gradient elution (**A**) and washing the column with ATP and glycerol (**B**). samples were loaded as follows; Lanes, M – Protein marker; Nc - *E. coli* XL1-Blue cells transformed with pQE30; Pf – pellet fraction; Sf – supernatant fraction; FT – flow through; W$_1$ and W$_3$ – wash 1 and wash 3, respectively; E$_1$ – E$_3$ – elution 1, elution 2 and elution 3, respectively. Lower panel: detection of K0513$_W$ at 55 kDa using anti-His and visualized using TMB.

To purify K0513$_W$, alterations made include the use enriched media (terrific broth) (Figure 3.6A). This was an attempt to increase the expression of K0513$_W$ and to counteract the toxicity which could have resulted to truncation, the induction OD was not altered since the increase might promote formation of inclusion bodies. The truncation was not observed, there was no significant improvement in band intensity (Figure 3.6, lane S). The target protein was still detected with a lot of other protein species (Figure 3.6, lane E1, E$_2$, and E$_3$). In order determine if contaminating proteins *E. coli* XL1-Blue strain specific, JM109 cells were used to produce recombinant K0513$_W$ (Figure 3.6B). The contaminating protein species prominently at 35 kDa, 30 kDa were detected in the elution (Figure 3.6B; E1, E2 and E3). The Western blot analysis shows no breakdown product. There was no significant improvement detected in lowering the amount of contaminating proteins even after a strain switch (Figure 3.6B, lane E1, E$_2$, and E$_3$).

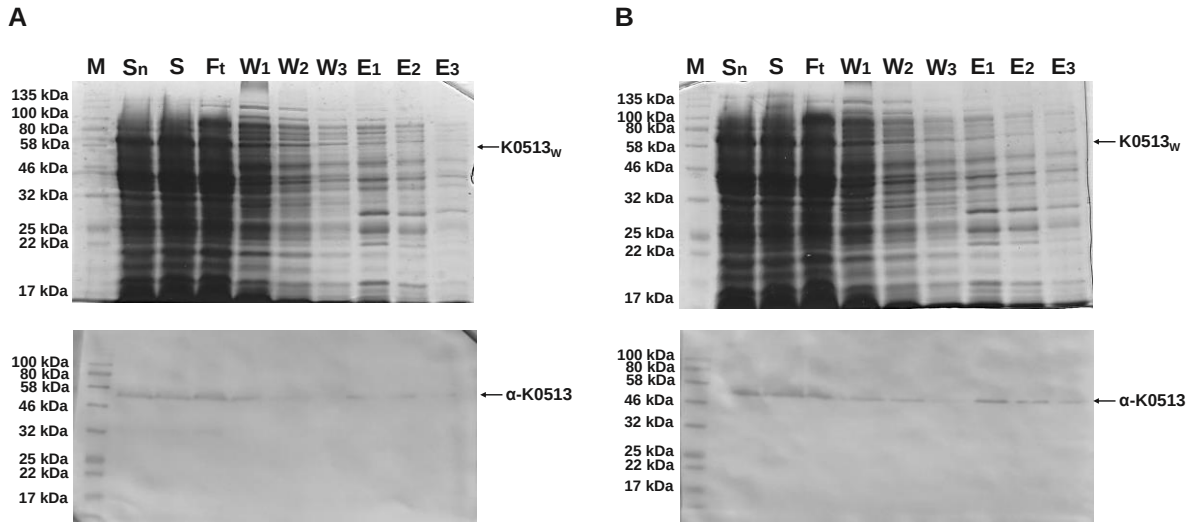


Figure 3.6. Purification of recombinant K0513$_W$ expressed in JM109 cell line and in TB media. Purification of K0513$_W$ using Ni-affinity chromatography, SDS-PAGE (12%) analysis (top panel) and Western blot analysis (lower panel). Purification of K0513$_W$ expressed in *E. coli* JM109 cell line (**A**) and expressed in terrific both (TB) enriched media (**B**). samples were loaded as follows; Lanes, M – Protein marker; Nc - *E. coli* XL1-Blue cells transformed with pQE30; Pf – pellet fraction; Sf – supernatant fraction; FT – flow through; W$_1$ and W$_3$ – wash 1 and wash 3, respectively; E$_1$ – E$_3$ – elution 1, elution 2 and elution 3, respectively. Lower panel: detection of K0513$_W$ at 55 kDa using anti-His and visualized using TMB.

3.3.2.4. Expression of recombinant K0513$_H$

Wild type K0513 was not purified, so the codon harmonized K0513 was explored through expression and purification in the *E. coli* system. The recombinant K0513$_H$ was expressed in the *E. coli* XL1-Blue system. The induction study was assessed by SDS-PAGE and confirmed by Western blot using anti-K0513 and anti-Histidine antibodies (Figure 3.7). *E. coli* XL1-Blue cells transformed with pQE60 was used as a negative control for the expression study as it does not express K0513$_H$. The pre-induction sample exhibits the heterologous expression of the 47 kDa K0513$_H$ protein (Figure 3.7, lane T$_0$), this might show that the promoter is not tightly regulated leading to leaky expression. The post-induction samples show an increase in the expression K0513$_H$, this however does not show a significant increase from the first and last hour (Figure 3.7, T$_1$ – T$_5$). The induction and expression of the 47 kDa K0513$_H$ was successfully confirmed using Western blot analysis using anti-K0513 and anti-Histidine antibodies.

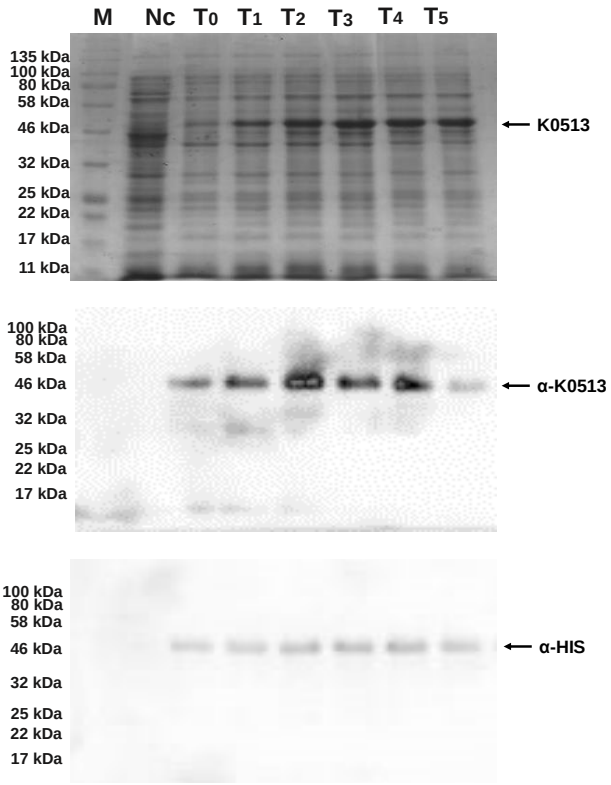


Figure 3.7. Heterologous expression of recombinant His-tagged K0513$_H$.
SDS-PAGE (12 %) analysis of K0513$_H$ expression in *E. coli* XL1-Blue cells. Samples are as follows; lane M - Prestained protein marker, Nc - cells transformed with pQE60, T0 - uninduced cells transformed with pQE60-K0513$_H$, T1 - T5 (1, 2, 3, 4 and 5-hour post induction samples respectively). Cells were grown at 37 ºC. Western blot was done using anti-K0513 and Anti-Histidine.

3.3.2.5. Solubility study of recombinant K0513$_H$

The solubility study was carried out to determine the solubility profile of K0513$_H$. To increase the solubility, 8 M urea, 1 % Triton X-100, and the combination of the two detergents was used (Figure 3.8, A and B). The untreated sample shows good solubilization wherein approximately 80 % of K0513$_H$ is in the soluble fraction (Figure 3.8A, untreated). There is no significant improvement in samples treated with 8 M urea and 1 % Triton X-100 (Figure 3.8B), however, the combination of the two detergents shows a slight improvement as depicted by a slight band in the pellet fraction (Figure 3.8B). Although the treatment with Triton X-100 and urea and the combination thereafter yielded best results, the untreated was sufficient to purify from since it maintains K0513$_H$ in its native form.

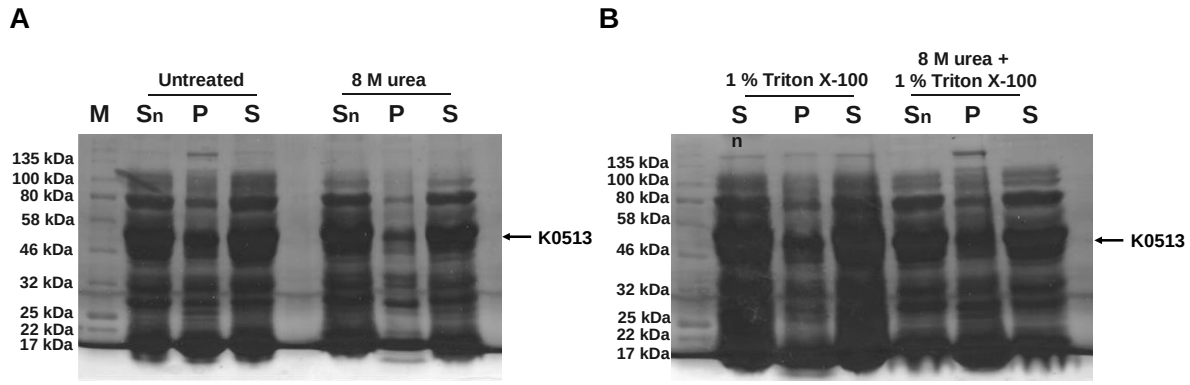


Figure 3.8. Solubility study of K0513$_H$.
K0513$_H$ was solubilized in the presents and absence of detergents. (**A**) Shows untreated and 8 M treated lysates. (**B**) Shows lysates treated with 1 % Triton X-100 and combination treatment with 8 M urea and 1 % Triton X-100. Samples were loaded as follows; Lane M – Prestained molecular marker, Sn – cells after sonication, P – pellet fraction, S – soluble fraction, Nc - cells transformed with pQE30, T$_5$ – time of harvest (5 hour) sample. Samples were resolved using 12 % SDS-PAGE and the presence of K0513$_H$ was validated using anti-His (Lower panel).

3.3.2.6. Purification of recombinant pQE60-K0513$_H$

Recombinant K0513$_H$ was successfully purified from the soluble fraction in its native form using Nickel-affinity chromatography (Figure 3.9). This was from the lysate treated with lysis buffer without denaturants in order to maintain the native state of the protein (Figure 3.9A). Some of the protein did not fully bind to the beads as indicated by a distinct 47 kDa band in the flow through (Figure 3.9, lane Ft), some protein was lost as unbound protein which did not bind to the nickel beads. A little but insignificant amount of protein was lost in the washes (Figure 3.9, lane W$_3$). However, the recombinant K0513$_H$ was successfully purified as depicted by a band of approximately 47 kDa (Figure 3.9, lane E$_1$ and E$_2$). The overall purification yielded 5 mg/mL from 250 mL total culture volume and shows about 80 % purity. The purified K0513$_H$ was further validated by Western blot using anti-K0513 and anti-Histidine antibodies.

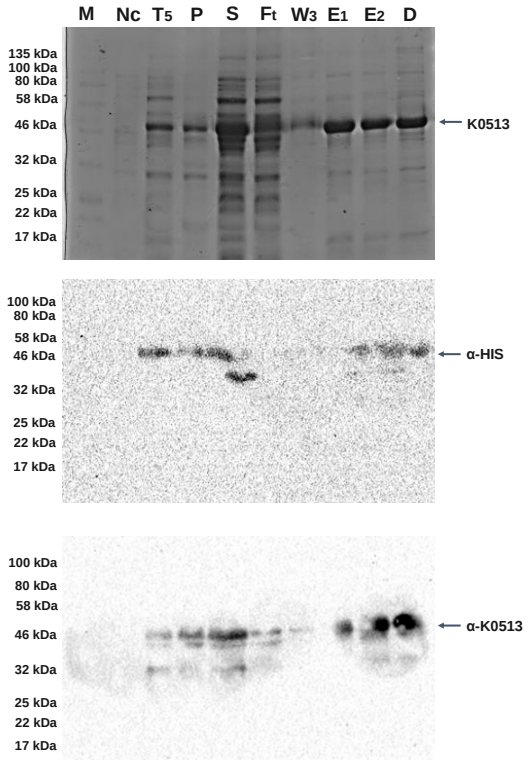


Figure 3.9. Purification of recombinant K0513$_H$ using nickel affinity chromatography.
SDS-PAGE (12%) analysis for K0513$_H$ purification and confirmation using Western blot. Samples are as follows; lane M – Prestained protein marker, Nc - cells transformed with pQE60, T$_5$ – time of harvest (5 hour) sample, P - pellet fraction, S - soluble fraction; Ft - flow through; W$_3$ - third wash, E$_1$ -– elution 1, E$_2$ - elution 2, D - dialyzed protein. Western blot was done using anti-K0513 and anti-histidine antibodies and visualized using ECL.

3.4. Discussion

The main objective of this study was to express and purify recombinant K0513 from the wild type and codon harmonized constructs pQE30-K0513$_W$ and pQE60-K0513$_H$, respectively) for further structural analysis. Both K0513$_W$ and K0513H were successfully expressed in the *E. coli* system, however only K0513$_H$ was successfully purified for analysis. This is the first study to report on the expression and purification of full length human K0513 from a prokaryotic system.

Recombinant K0513$_W$ was shown to resolve at 55 kDa, a different size to its theoretical molecular weight of 47 kDa, this molecule was expressed in its insoluble form in the *E. coli* system (Figure 3.2). This taken together suggests the possible formation of aggregates or inclusion bodies within the *E. coli* system (Gopal and Kumar, 2013; Ferrer-Miralles, 2015). This can further explain the truncation observed which may be a result of the host's protease machinery (Spiess *et al*., 1999), and the induction of chaperone complexes leading to a high-level of contaminants encountered during purification of K0513$_W$. The truncation and low expression could have been an indication of the harmful effects that the heterologous protein has on the cell, this could be a result of toxicity or codon bias (Dumon-seignovert *et al*., 2004). The expression of toxic protein also activates the stress response of the host cell which may include the induction of chaperone machinery (Akiyama, 2009), molecular chaperones will then execute their holdase and foldase function forming a complex with K0513$_W$. The observations suggest that some of the contaminating proteins that were observed to co-purify with K0513$_W$ at approximately 70 kDa, 40 kDa, 35 kDa, and 15 kDa are molecular chaperones (Figure 3.4; Figure 3.5). It has been reported that supplementing the growth media with 1 % glucose can reduce truncation as well as leaky expression, as the glucose tightens the weakly regulated promoter system (Strudier, 2005). Observations from this study was in support of the hypothesis since the leaky expression was reduced, however, this subsequently increased the intensity of the truncation (Figure 3.2).

Most of these factors translates to challenges that are faced in expressing a non-codon harmonized eukaryotic gene in a prokaryotic system such as *E. coli* system. Several optimization strategies were used to improve expression and solubility of K0513$_W$. This

includes lowering the expression temperature (Song *et al*., 2012), and use of different media and changing the *E. coli* strain (Hortsch and Weuster-Botz, 2011). The expression of K0513$_W$ was improved at low temperature (Figure 3.2), this is because lowering the temperature slows the transcription, translation and refolding rate which may allow proper folding of the protein (Vera *et al*., 2006). Although there was a significant improvement on the expression levels, the solubility of K0513$_W$ remained challenging. Purification of this construct can be tackled using a more improved approaches such as size exclusion chromatography.

We therefore assumed that all these challenges are conferred by codon bias between K0513$_W$ as an exogenous protein and the host system machinery (Rosano and Ceccarelli, 2014). We resorted to generating a codon-harmonized gene (K0513$_H$) cloned into a pQE60 plasmid which encodes a C-terminal histidine tag to allow a proper compatibility with the *E. coli* system (Figure 3.1).

The leaky expression of K0513$_H$ in *E. coli* XL1-Blue cells was observed prior induction by IPTG, this phenomenon may indicate weakly regulated promoter (Figure 3.7). This can be improved by addition of 1 % glucose to the media (Strudier, 2005), or the use of the inducible T7 phage lysozyme system into strains such as the BL21(DE3) pLysS (Stano and Patel, 2004; Rosano and Ceccarelli, 2014; Briand *et al*., 2016). Recombinant K0513$_H$ was successfully expressed at 37 °C which is the optimum culture temperature for *E. coli* cells (Figure 3.7). K0513$_H$ is a soluble protein, which is expected to the localization study which reported KIAA0513 to be cytosolic (Lauriat *et al*., 2006). Furthermore, the observed partial solubility might indicate that K0513$_H$ has some hydrophobic properties suggesting a possible functional activity along the cell membrane (Figure 3.8; Rawlings, 2016). This includes in cellular processes such as cytoskeletal regulation and neuroplasticity (Lauriat *et al*., 2006). The partial solubility has been observed in membrane proteins found in magnetosomes that exist as soluble assemblies before they are recruited to the membranes, their overall solubility prevents them from being incorporated to the inner membrane and only the hydrophobic rich domain interacts with the membrane (Tanaka *et al*., 2006). Recombinant K0513$_H$ was successfully purified using Nickel-affinity chromatography (Figure 3.9).

Chapter 4

Investigation of K0513 tertiary structure conformation

4.1. Introduction

The protein structure determines their function, the tertiary structure is a functional form of a protein and exist in three dimensions. The study of tertiary structural organization can reveal and lead to ascertaining and annotation of functions. Several, approaches have been developed to study this structure, this includes amongst others, monitoring the fluorescence of aromatic residues and the use of a controlled proteolytic digest through limited proteolysis.

Protein conformation can be determined by measuring the intrinsic fluorescence from aromatic amino acids (Ghisaidoobe and Chung, 2014). Three aromatic amino acids are known to fluoresce when exposed to ultraviolet (UV) light; tryptophan (W), tyrosine (Y) and phenylalanine (F). F has a very low quantum yield as well as low absorptivity, therefore its intrinsic fluorescence is negligible. on the other hand, Y has a similar quantum yield to W and its emission is often quenched through energy transfer to W or interaction with the peptide chain (Chen *et al*., 1998). Of the three, tryptophan is the most dominant with a concentration of 1 mol % and 3 mol % for cytosolic and membrane proteins respectively (Swaminathan *et al*., 1994; Lakowicz, 2006).

Tryptophan is uniquely sensitive to collisional quenching; this is due to its propensity to donate electrons when it is in its excited-state indole (Guo *et al*., 2014; Xu *et al*., 2014). This quenching can either be a result of external quenchers or from the groups around the W residue within the protein (Van de Weert and Stella, 2011). It has been suggested that the amide group can also lead to fluorescence quenching under appropriate conditions (Ghisaidoobe and Chung, 2014). Moreover, in addition to ligand binding, an inner-filter effect which is an absorbance of light at the excitation or emission wavelength may lead to fluorescence quenching (Ghisaidoobe and Chung, 2014).

A change in the solvent environment of the tryptophan residue due to denaturation of the protein, subunit association or ligand binding induces conformational changes which lead to fluorescence quenching and shift of the fluorescence maxima (Vivian and Callis, 2001; Akbar *et al*., 2016). The primary structure of K0513 possesses a total of six tryptophan residues at positions W65, W134, W218, W300, W306, and W339.

Therefore, tryptophan fluorescence spectrophotometry was used to assess the tertiary structural conformation of K0513.

Limited proteolysis is based on the controlled exposure of a protein to the proteolytic enzyme. This technique is used to determine the structural conformation features of the protein. This is because cleavage takes place when the protein is in its tertiary and quaternary structure, the first cleavage is suggested to be on the cleavage site located on the surface of the protein and accessible to the enzyme (Petrotchenko *et al*., 2014). The presence of the bound ligand can affect the susceptibility of a protein to proteolytic digest (Zeinoddini *et al*., 2013). Several proteases have been explored in the study of protein, this amongst others include trypsin, chymotrypsin and proteinase k. Trypsin is known to specifically cleave the c-terminal of arginine and lysine (Olsen *et al*., 2014). K0513 has a total of 47 trypsin cleavage sites.

Specific objectives

1. To investigate the effect of nucleotides on K0513 tertiary structure using tryptophan fluorescence spectroscopy
2. To investigate the effect of nucleotides on K0513 tertiary structure using limited proteolysis
3. To determine direct interactors of K0513 using pulldown assay

4.2. Materials and methods

4.2.1. Tryptophan based fluorescence spectroscopy

The tertiary structure of K0513 was analyzed using intrinsic tryptophan fluorescence spectroscopy. The assay was conducted as precisely described in (Zininga *et al*., 2016). Spectra was generated and monitored after initial excitation at 295 nm. The emission spectra were monitored between 280 to 500 nm, 1000 nm/min scan speed, 3 accumulations and 10 nm emission bandwidth. Effect of nucleotide on K0513 tertiary structure was determined by incubating the protein with various concentrations ranging from 1 mM to 5 mM ATP/ADP and 0.1 mM to 1 mM GDP/GTP for 10 min at 37 °C before fluorescence reading. The effect of chemical denaturants on the conformation of K0513 was assessed by incubating the protein with 1 M to 6 M guanidine-HCL and 1 M to 8 M Urea concentration range prior the fluorescence reading.

4.2.2. Limited proteolysis

Nucleotide dependent conformational alterations of recombinant his-tagged K0513 were investigated by limited trypsin proteolysis. The assay was conducted as precisely described in (Zininga *et al*., 2015; da Silveira Tome *et al*., 2018) with some modifications. Briefly, limited proteolysis was performed in TBS buffer (250 mM NaCl, 50 mM Tris pH 7.5) using an enzyme to substrate ratio of 1:2000 (3. 21 nM trypsin:6.42 µM K0513). The assay was carried out after incubation of K0513 for 15 minutes at 37 °C in the absence of nucleotides or in the presence of 5 mM ATP/ADP or 1 mM GTP/GDP. Aliquots were collected at time intervals and the reaction was stopped by the addition of 4X SDS loading buffer [0.5 M Tris-HCl, pH 6.8, 10 % (v/v) glycerol, 10 % (w/v) SDS, 5 % (v/v) β-mercaptoethanol, 1 % (w/v) Bromophenol Blue]. The digestion pattern was analyzed using SDS-SPAGE on 12 % acrylamide gels stained with Coomassie brilliant blue R-250 solution. The gel was scanned using ChemiDoc to allow densitometric analysis.

4.2.3. Pull-down assay

Pull down assay was used to identify possible interactors of K0513 from fibrosarcoma cell lysates. The recombinant K0513 (bait) immobilized on HisPur Ni-NTA Resin to facilitate pull down of protein interactors from fibrosarcoma lysate. This was extensively washed with wash buffer (0, 1 M Tris, [pH 7.5]; 300 mM NaCl, 25 mM Imidazole). The volume of 200 µL Fibrosarcoma cell lysate (prey) containing 1 mg/mL of proteins was added incubated for 4 hours to allow possible binding with K0513 (bait). The interactions were cross-linked by incubating with 1 % formaldehyde for 7 minutes at room temperature and quenched with PBS. The column was washed 3 times with two times bed volume of the wash buffer. The interactome was eluted with 2 mL of the elution buffer (0, 1 M Tris, [pH 7.5]; 300 mM NaCl, 500 mM Imidazole, 1 mM PMSF). The pull-down eluates were analyzed using 12 % SDS-PAGE and visualized using silver stain.

4.3. Results

4.3.1. Tertiary structural organization of K0513

The tertiary structure of K0513 was assessed using intrinsic tryptophan fluorescence analysis in the presence of varying concentration of guanidine-HCl and urea (Figure 4.1). The native K0513 registered its maximum fluorescence intensity at a wavelength of 332 nm suggesting a folded state. An increase in guanidine-HCl concentration resulted in reduced fluorescence intensity, this suggest fluorescence quenching (Figure 4.1A1). Analysis of fluorescence intensity maxima shows a significant wavelength shift with respect to increasing guanidine-HCl concentration, the change followed a red shift indicating that K0513 now in an unfolded state (Figure 4.1A2). To validate this, the tertiary structural analysis was again performed in the presence of varying concentration of urea (Figure 4.1B1). An increase in urea concentration resulted in fluorescence quenching in a concentration-dependent manner (Figure 4.1B1). The shift in fluorescence in response to urea concentration displays a significant red shift as expected (Figure 4.1B2). The fluorescence spectra show the presence of two distinct peaks at approximately 332 nm and 365 nm respectively. The 365 nm peak appears small in the absence of denaturants and become more pronounced as the denaturant concentration is increased. This may indicate that the protein was not completely native.

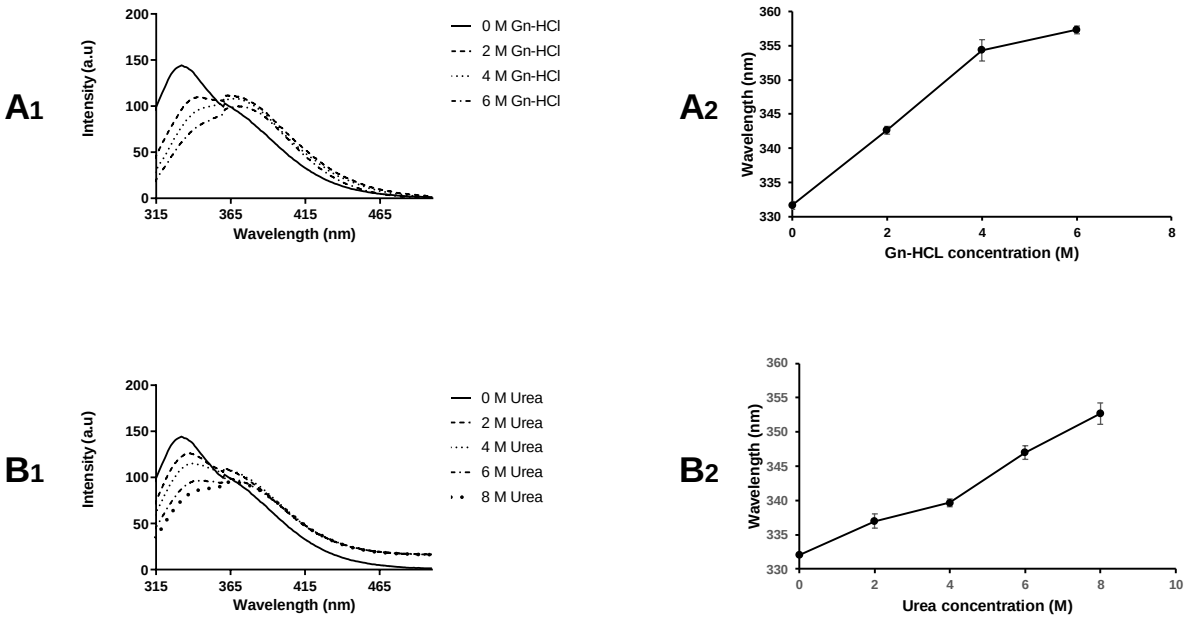


Figure 4.1. Conformation of K0513 in the presence of guanidine-HCl and urea.
K0513 conformation was assessed using tryptophan fluorescence spectrophotometry in the presence of various guanidine-HCl concentration ranging from 1 M to 6 M (**A1**), and the wavelength shift was plotted against the concentration of urea (**A2**). The same procedure was followed in the presence of urea at concentration ranging from 1 M to 8 M (**B1**) and the wavelength shift was plotted against its concentration (**B2**). Initial excitation was at 295 nm and the spectra were monitored between 280 t0 500 nm, 1000 nm/min scan speed, 3 accumulations and 10 nm emission bandwidth. The 332 and 365 nm peak represents the native and the denatured K0513, respectively.

4.3.2. Effects of nucleotides on the tertiary structure of K0513

The ability of K0513 to change conformation in the presence of nucleotides was assessed using tryptophan fluorescence (Figure 4.2). The fluorescence analysis in the presence of ATP shows a quenching effect which is proportional to an increasing ATP concentration (Figure 4.2A1). there was no significant red or blue shift observed on the intensity maxima of the 332 nm peak (Figure 4.2A2). Similarly, the fluorescence assessment in the presence of ADP shows a decrease in fluorescence intensity with an increasing ADP concentration (Figure 4.2B1). The wavelength shift based on fluorescence maxima remained insignificant (Figure 4.2B2). Furthermore, fluorescence analysis was done in the presence of GTP (Figure 4.2C1). Similar to what was observed with ATP and ADP, the fluorescence intensity was decreased with increasing concentration of GTP. Notably, the second peak coming out at 365 nm become more pronounced as the GTP concentration increases. These changes were more prominent in 0.8 mM and 1 mM concentration (Figure 4.2C1). The wavelength shift of the intensity maxima shows no significant change (Figure 4.2C2). Tryptophan

fluorescence intensity was decreased with an increasing concentration of GDP (Figure 4.2D1). Higher concentration of GDP at 1 mM shows a slight disappearance of the 365 nm peak. There was no significant wavelength shift based on fluorescence intensity maxima observed (Figure 4.2D2). This suggests that the presence of GDP and GTP affected the surrounding environment, favouring the fluorescence of either the native or denatured population of K0513.

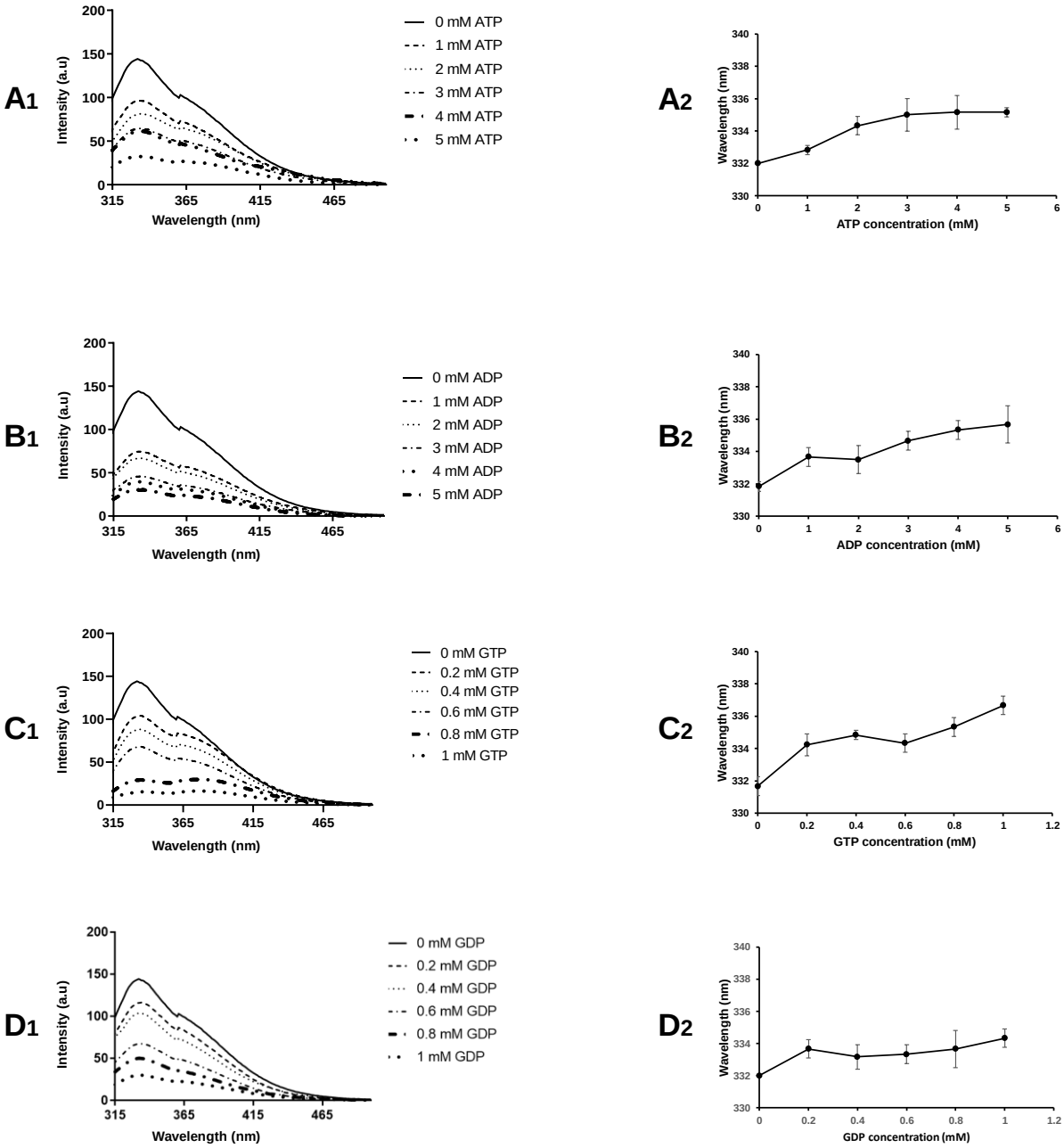


Figure 4.2. Effect of nucleotides on the tertiary structure of K0513.
The conformational change of K0513 was assessed in the presence of various concentration of ATP (**A1**) and the shift in wavelength was plotted against concentration (**A2**), ADP (**B1**) with its wavelength shift (**B2**). The same was followed for GTP (**C1**) and GDP (**D1**), and the wavelength shift was plotted in (**C2**) and (**D2**) respectively. Initial excitation was at 295 nm and the spectra were monitored between 280 t0 500 nm, 1000 nm/min scan speed, 3 accumulations and 10 nm emission bandwidth.

To elucidate if the pronouncement of the 365 nm peak was a not a result of GTP hydrolysis, K0513 was incubated with GTP and the fluorescence was monitored over time (Figure 4.3). The fluorescence intensity shows a significant reduction after 20 minutes of incubation with GTP, this also revealed another peak shifted towards longer

wavelength of around 420 nm which again disappeared after 40 minutes of incubation. As the incubation time increased to 120 minutes, the fluorescence intensity increased and was again quenched and stabilized at time 240 and 300 minutes (Figure 4.3).

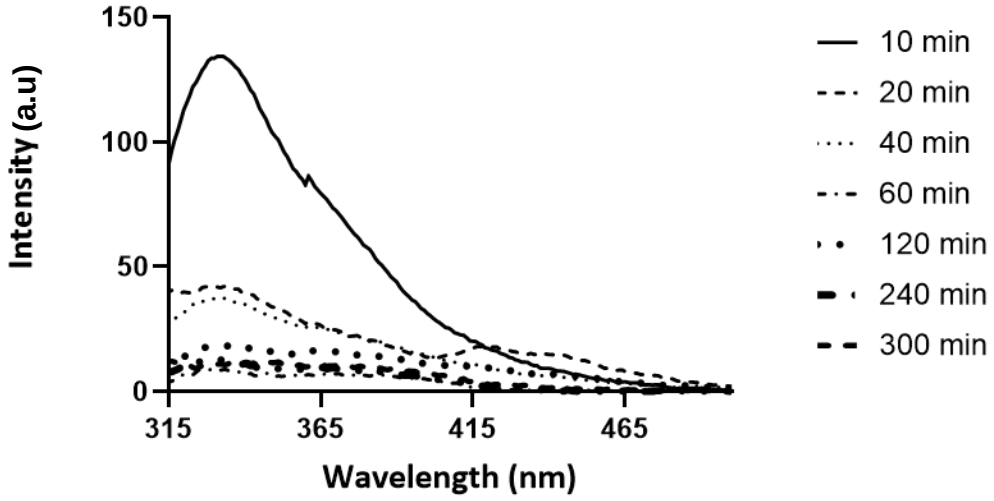


Figure 4.3. Conformation of K0513 in response to GTP presence over time.
K0513 conformation was assessed using tryptophan fluorescence spectrophotometry in the presence of GTP. The fluorescence was monitored over time for 300 minutes. Initial excitation was at 295 nm and the spectra were monitored between 280 t0 500 nm, 1000 nm/min scan speed, 3 accumulations and 10 nm emission bandwidth.

4.3.3. Determination of nucleotides effect on K0513 conformation using limited proteolysis

The effect of nucleotide on the structural conformation of K0513 was assessed using trypsin based limited proteolysis in the presence and absence of nucleotides. The resulting peptide fragments were analyzed by SDS-PAGE (Figure 4.4). Trypsin digestion after 10 minutes of incubation generated a fragment at 35 kDa and below 17 kDa which is similarly observed in the absence of nucleotides and in the presence of ATP and ADP (Figure 4.4A). The observed bands are absent in the purified K0513 (untreated) which suggest that they are a result of proteolytic digest. The degree of fragmentation follows the same pattern in the absence and presence of ATP/ADP after 20 and 40 minutes of incubation. After 80 minutes of incubation, the nucleotide free K0513 shows a greater degree of digestion compared to the ATP/GDP (Figure 4.4A and B). Similar trypsin digestion trend was observed in the presence of GTP/GDP (Figure 4.4C. fragmentation bands were observed from 10 minutes of incubation in the

presence and absence of GTP and GDP. The presence of both GTP and GDP did not show any significant effect on the conformational change of K0513 Figure 4.4A, D).

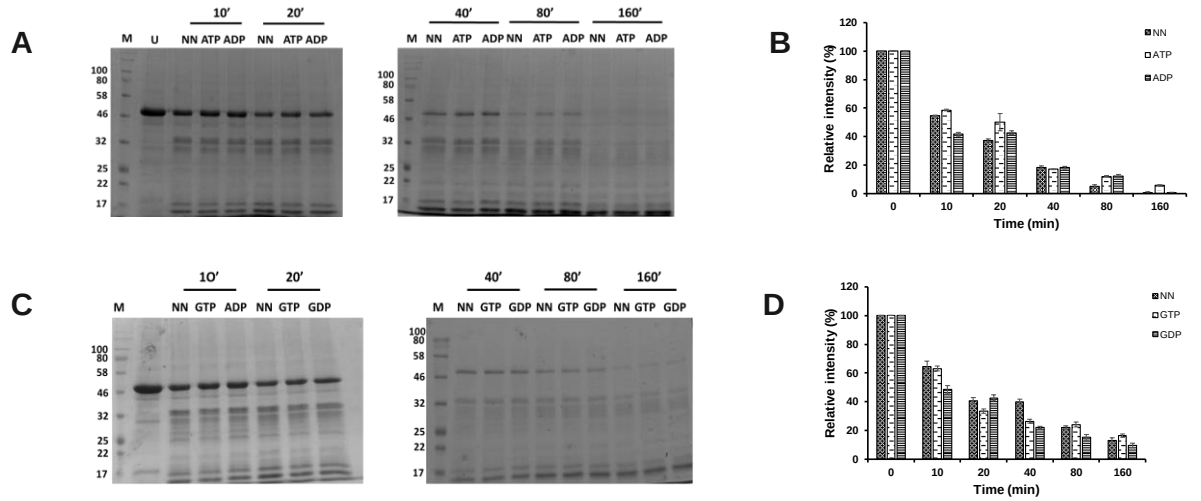


Figure 4.4. Proteolytic digest of K0513 in the absence and presence of nucleotides.
Trypsin digest of K0513 in the presence of 5 mM ATP/ADP (**A**) and in the presence of 1mM GTP/GDP (**C**). Samples were collected at each time point and analyzed using SDS-PAGE. Samples were loaded as follows; M - prestained molecular marker, U – undigested K0513, NN - trypsin digest in the absence of nucleotides, ATP/ADP/GTP/GDP - trypsin digest in the presence of the respective nucleotide. The band intensity of the digest in the presence of ATP/ADP and GTP/GDP was quantified using densitometry on Image Lab™ represented in **B** and **D**, respectively.

4.3.4. Determination of possible K0513 interactors using pulldown assay

Possible interactors of K0513 were investigated using pull down assay from fibrosarcoma cell line, the resulting interactome was analyzed using SDS-PAGE and visualized using GE Healthcare PlusOne™ Silver Stain Kit (Figure 4.5). The soluble fraction shows that presence of K0513 as 46 kDa band in its native form to allow binding with possible interactors (Figure 4.5A, Sf). The fibrosarcoma cells shows presence of many protein species, with distinct bands that are more pronounced at approximately 16 kDa, 30 kDa and 45 kDa respectively (Figure 4.5A, input). Few protein bands were detected in the first was and absent in the third wash, this indicate the unbound proteins (Figure 4.5A, W1 and W3). The eluents depict the presence of possible interactors, this is shown by the presence of two unique bands labelled 1 and 2 coming out at approximately 100 kDa and 44 kDa respectively (Figure 4.5A, E2).

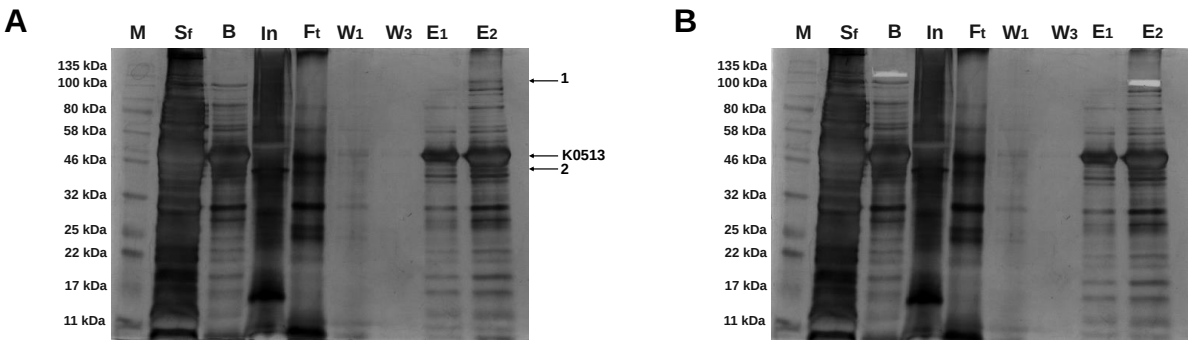


Figure 4.5: Investigation of K0513 interactors using pulldown assay.
Pull down assay was performed form fibrosarcoma cells (Input) using the recombinant His-tagged K0513 (bait) and analyzed using SDS-PAGE. Samples were loaded as follows; M - prestained molecular marker, Sf – Soluble fraction, B – Bait (immobilized K0513), In – Input (Fibrosarcoma cells), Ft – Flow through, W1 and W3 – first and third wash after cross-linking, E1 and E2 – first and second elution with 500 mM imidazole.

4.4. Discussion

The fluorescence data revealed the tertiary structural organization of K0513. Denaturing the protein with urea and guanidine-HCl resulted in a significant red shift of the intensity maxima in a concentration dependent manner (Figure 4.1). This reveals that the fluorescing W residues are located in a nonpolar hydrophobic interior (Keshav *et al*., 2019). This is because, W has two nearby isoenergetic excitation states, ^1L_a and ^1L_b (Albani, 2014) which differs with sensitivity to solvent. In hydrophobic environments, ^1L_b dominates the W emission and display a blue shift of W fluorescence due to an increase in non-polarity (Ghisaidoobe and Chung, 2014). In contrast, denatured proteins with solvent exposed W residues displays a red shift (Eftink, 2006) indicating that the protein is now in an unfolded conformation. This further suggest that, in its native state, K0513 exist as a globular protein which has a maximum fluorescence intensity at 332 nm. The W fluorescence was quenched as the denaturant's concentration increased which suggest that the W residues on K0513 were close to the nonpolar residues.

Preliminary *in silico* studies shows that K0513 possibly interacts with Rab3a, and that it has a possible role in the GTPase cycle of Rab3a as a guanine exchange factor. GTPases transit between the active (GTP-bound) state and the inactive (GDP-bound) state. The GEF facilitate exchange of nucleotides (GTP/GDP) without binding to the nucleotides but to switch I and switch II of the GTPase (Kiontke *et al*., 2017). K0513 depicted no conformational change when exposed to nucleotides (Figure 4.2). This observation indicate that K0513 does not directly interact with nucleotides, and that its tertiary structure is not modulated by nucleotides which support the possible function as a NEF. To elucidate whether K0513 hydrolyses GTP, the fluorescence was monitored over time in the presence of a fixed GTP concentration (Figure 4.3). K0513 depicted neither red nor blue shift on its native 332 nm peak. This suggest that K0513 does not hydrolyse GTP, and further support that K0513 conformation is not altered by nucleotides. Taken together, this finding suggest that K0513 facilitates the exchange of GTP/GDP in Rab3a as a GEF and subsequently activates it.

Limited proteolysis was used to further validate the tryptophan fluorescence results, this was through trypsin digest of the recombinant K0513 and monitoring the effect of

nucleotides on its conformation (Figure 4.4). The presence of fragments on the SDS-Page indicates that K0513 is susceptible to proteolytic digest, however, the observed fragmentation patterns in the presence of nucleotides were the same. This suggests that the nucleotides did not affect the tertiary structure of the protein which is in support with the fluorescence data.

The pull-down assay was used to identify protein interactors of K0513. Since K0513 is ubiquitously expressed, fibrosarcoma cell line was used for prey proteins (Figure 4.5). The unique bands identified in the elution represent possible interactors and were sent for LCMS. Their identity may help in validating the predicted interaction network.

This finding provides preliminary results on the role of K0513 as a rab GEF, the results paves a way for further studies on this interesting molecule. Future studies will include GTPase assay to confirm the suggested role of K0513 as a GEF for Rab GTPases, specifically Rab3a, and known GEFs will be used as positive controls.

Chapter 5

Concluding remarks

In silico analysis revealed that KO513 is a globular protein which contains a conserved SBF2 domain, a transmembrane region, and suggest that it functions as a nucleotide exchange factor for small GTPase, Rab3a. Findings from this study suggest that K0513 interacts with Rab3a and induce its conformational change, facilitate the exchange of GTP/GDP, and subsequently lead to the activation of Rab3a. This suggest a role K0513 of K0513 in signalling pathways, by facilitating Rab3a in executing its role during synaptic vesicle transportation and cancer development. This is the first study to report on the recombinant expression, purification and characterization of human K0513. The biophysical characterization provided in this study is important in further elucidating the role of K0513 in signalling pathways associated with GBM as well as pancreatic cancer. Future studies may include cell-based approach such as gene knockout using GBM cell lines, identification of possible inhibitors as well as X-ray crystallography.

References

Adamson, C., Kanu, O.O., Mehta, A.I., Di, C., Lin, N., Mattox, A.K. and Bigner, D.D., 2009. Glioblastoma multiforme: a review of where we have been and where we are going. *Expert opinion on investigational drugs*, 18(8), pp.1061-1083.

Akbar, S.M., Sreeramulu, K. and Sharma, H.C., 2016. Tryptophan fluorescence quenching as a binding assay to monitor protein conformation changes in the membrane of intact mitochondria. *Journal of bioenergetics and biomembranes*, 48(3), pp.241-247.

Akiyama, T., Jin, S., Yoshida, M., Hoshino, T., Opassiri, R. and Cairns, J.R.K., 2009. Expression of an endo-(1, 3; 1, 4)-β-glucanase in response to wounding, methyl jasmonate, abscisic acid and ethephon in rice seedlings. *Journal of plant physiology*, 166(16), pp.1814-1825.

Albani, J.R., 2014. Origin of tryptophan fluorescence lifetimes part 1. Fluorescence lifetimes origin of tryptophan free in solution. *Journal of fluorescence*, 24(1), pp.93-104.

Albers, R.W.W., 2012. Cell membrane structures and functions. *In Basic Neurochemistry* (pp. 26-39). Academic Press.

Albrecht, T.R., Shevtsov, S.P., Wu, Y., Mascibroda, L.G., Peart, N.J., Huang, K.L., Sawyer, I.A., Tong, L., Dundr, M. and Wagner, E.J., 2018. Integrator subunit 4 is a 'Symplekin-like'scaffold that associates with INTS9/11 to form the Integrator cleavage module. *Nucleic acids research*, 46(8), pp.4241-4255.

Aldape, K., Zadeh, G., Mansouri, S., Reifenberger, G. and von Deimling, A., 2015. Glioblastoma: pathology, molecular mechanisms and markers. *Acta neuropathologica*, 129(6), pp.829-848.

Angov, E., Hillier, C.J., Kincaid, R.L. and Lyon, J.A., 2008. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PloS one*, 3(5).

Ardito, F., Giuliani, M., Perrone, D., Troiano, G. and Lo Muzio, L., 2017. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *International journal of molecular medicine*, 40(2), pp.271-280.

Aspenström, P., Fransson, Å. and Saras, J., 2004. Rho GTPases have diverse effects on the organization of the actin filament system. *Biochemical Journal*, 377(2), pp.327-337.

Assanah, M., Lochhead, R., Ogden, A., Bruce, J., Goldman, J. and Canoll, P., 2006. Glial progenitors in adult white matter are driven to form malignant gliomas by platelet-derived growth factor-expressing retroviruses. *Journal of Neuroscience,* 26(25), pp.6781-6790.

Barr, F. and Lambright, D.G., 2010. Rab gefs and gaps. *Current opinion in cell biology*, 22(4), pp.461-470.

Bell, E.H., Hadziahmetovic, M. and Chakravarti, A., 2011. Evolvement of molecular biomarkers in targeted therapy of malignant gliomas. *Brain Tumors—Current and Emerging Therapeutic Strategies*, pp.117-142.

Berger, P., Berger, I., Schaffitzel, C., Tersar, K., Volkmer, B. and Suter, U., 2006. Multi-level regulation of myotubularin-related protein-2 phosphatase activity by myotubularin-related protein-13/set-binding factor-2. *Human molecular genetics*, 15(4), pp.569-579.

Berger, P., Bonneick, S., Willi, S., Wymann, M. and Suter, U., 2003. Loss of phosphatase activity in myotubularin-related protein 2 is associated with Charcot–Marie–Tooth disease type 4B1. *Human molecular genetics*, 11(13), pp.1569-1579.

Bornhorst, J.A. and Falke, J.J., 2000. [16] Purification of proteins using polyhistidine affinity tags. *In Methods in enzymology* (Vol. 326, pp. 245-254). Academic Press.

Braschi, B., Denny, P., Gray, K., Jones, T., Seal, R., Tweedie, S., Yates, B. and Bruford, E., 2019. Genenames. org: the HGNC and VGNC resources in 2019. Nucleic acids research, 47(D1), pp.D786-D792.

Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H. and Beroukhim, R., 2013. The somatic genomic landscape of glioblastoma. *Cell*, 155(2), pp.462-477.

Briand, L., Marcion, G., Kriznik, A., Heydel, J.M., Artur, Y., Garrido, C., Seigneuric, R. and Neiers, F., 2016. A self-inducible heterologous protein expression system in Escherichia coli. *Scientific reports*, 6, p.33037.

Brondyk, W.H., 2009. Selecting an appropriate method for expressing a recombinant protein. *In Methods in enzymology* (Vol. 463, pp. 131-147). Academic Press.

Brown, G.R., Hem, V., Katz, K.S., Ovetsky, M., Wallin, C., Ermolaeva, O., Tolstoy, I., Tatusova, T., Pruitt, K.D., Maglott, D.R. and Murphy, T.D., 2015. Gene: a gene-centered information resource at NCBI. *Nucleic acids research*, 43(D1), pp.D36-D42.

Cancer Genome Atlas Research Network, 2013. Correction: Corrigendum: Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 494(7438), pp.506-506.

Carney, D.S., Davies, B.A. and Horazdovsky, B.F., 2006. Structural clues to the mechanism of Rab5 nucleotide exchange. *Trends in Cell Biology*, 1(16), pp.27-35.

Carson, D.A. and Lois, A., 1995. Cancer progression and p53. *The Lancet*, 346(8981), pp.1009-1011.

Chen, J., McKay, R.M. and Parada, L.F., 2012. Malignant glioma: lessons from genomics, mouse models, and stem cells. *Cell*, 149(1), pp.36-47.

Chen, K., He, Y., Liu, Y. and Yang, X., 2019. Gene signature associated with neuro-endocrine activity predicting prognosis of pancreatic carcinoma. *Molecular Genetics and Genomic Medicine*, 7(7), p.e00729.

Chen, Y. and Barkley, M.D., 1998. Toward understanding tryptophan fluorescence in proteins. *Biochemistry*, 37(28), pp.9976-9982.

Chen, Y., Ng, F. and Tang, B.L., 2016. Rab23 activities and human cancer—emerging connections and mechanisms. *Tumor Biology*, 37(10), pp.12959-12967.

Cherfils, J. and Zeghouf, M., 2013. Regulation of small gtpases by gefs, gaps, and gdis. *Physiological reviews*, 93(1), pp.269-309.

Chung, R., Whaley, J., Kley, N., Anderson, K., Louis, D., Menon, A., Hettlich, C., Freiman, R., Hedley-Whyte, E.T., Martuza, R. and Jenkins, R., 1991. TP53 gene mutations and 17p deletions in human astrocytomas. *Genes, Chromosomes and Cancer*, 3(5), pp.323-331.

Colman, H. and Aldape, K. (2011). Molecular Pathogenesis. *Primary Central Nervous System Tumors*, pp.27-44.

Colman, H. and Aldape, K., 2011. Molecular Pathogenesis. *In Primary Central Nervous System Tumors* (pp. 27-44). Humana Press.

Corbier, C. and Sellier, C., 2017. C9ORF72 is a GDP/GTP exchange factor for Rab8 and Rab39 and regulates autophagy. *Small GTPases*, 8(3), pp.181-186.

Da Silva, A.R., Malafaia, G. and Menezes, I.P.P., 2017. biotools: an R function to predict spatial gene diversity via an individual-based approach. *Genetic and molecular research*, 16, p.gmr16029655.

da Silveira Tomé, C., Foucher, A.E., Jault, J.M. and Housset, D., 2018. High concentrations of GTP induce conformational changes in the essential bacterial GTP ase EngA and enhance its binding to the ribosome. *The FEBS journal*, 285(1), pp.160-177.

Dumon-Seignovert, L., Cariot, G. and Vuillard, L., 2004. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21 (DE3), C41 (DE3), and C43 (DE3). *Protein expression and purification*, 37(1), pp.203-206.

Eftink, M.R., 1991. Fluorescence techniques for studying protein structure. *Methods of biochemical analysis*, pp.127-205.

Eftink, M.R., 2006. The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophysical journal*, 66(2), pp.482-501.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A. and Sonnhammer, E.L.L., 2019. The Pfam protein families database in 2019. *Nucleic acids research*, 47(D1), pp.D427-D432.

Ferrer-Miralles, N., Saccardo, P., Corchero, J.L., Xu, Z. and García-Fruitós, E., 2015. General introduction: recombinant protein production and purification of insoluble proteins. *In Insoluble Proteins* (pp. 1-24). Humana Press, New York, NY.

Fischer, I. and Aldape, K., 2010. Molecular tools: biology, prognosis, and therapeutic triage. *Neuroimaging Clinics*, 20(3), pp.273-282.

Fisher, J.L. and Schwartzbaum, J.A., 2007. Wrensch 4. M, Wiemels JL. Epidemiology of brain tumors. *Neuroimaging Clinics*, 25, pp.867-90.

Friedman, A., Bigner, D. and Van Meir, E.G., 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. The Cancer Genome Atlas Research Network. *Nature*, 455, pp.1061-1068.

Fukuda, M., 2008. Regulation of secretory vesicle traffic by Rab small GTPases. *Cellular and Molecular Life Science*, 65(18), pp.2801-2813.

Furnari, F.B., Fenton, T., Bachoo, R.M., Mukasa, A., Stommel, J.M., Stegh, A., Hahn, W.C., Ligon, K.L., Louis, D.N., Brennan, C. and Chin, L., 2007. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes and development*, 21(21), pp.2683-2710.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D. and Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy server. *In The proteomics protocols handbook* (pp. 571-607). Humana press.

Gerondopoulos, A., Langemeyer, L., Liang, J.R., Linford, A. and Barr, F.A., 2012. BLOC-3 mutated in Hermansky-Pudlak syndrome is a Rab32/38 guanine nucleotide exchange factor. *Current Biology*, 22(22), pp.2135-2139.

Ghisaidoobe, A.B. and Chung, S.J., 2014. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *International journal of molecular sciences*, 15(12), pp.22518-22538.

Goody, R.S., Rak, A. and Alexandrov, K., 2005. The structural and mechanistic basis for recycling of Rab proteins between membrane compartments. *Cellular and Molecular Life Sciences CMLS*, 62(15), pp.1657-1670.

Gopal, G.J. and Kumar, A., 2013. Strategies for the production of recombinant protein in Escherichia coli. *The protein journal*, 32(6), pp.419-425.

Guo, X., Li, X., Jiang, Y., Yi, L., Wu, Q., Chang, H., Diao, X., Sun, Y., Pan, X. and Zhou, N., 2014. A spectroscopic study on the interaction between p-nitrophenol and bovine serum albumin. *Journal of luminescence*, 149, pp.353-360.

Gustafsson, C., Govindarajan, S. and Minshull, J., 2004. Codon bias and heterologous protein expression. *Trends in biotechnology*, 22(7), pp.346-353.

Han, M.Z., Huang, B., Chen, A.J., Zhang, X., Xu, R., Wang, J. and Li, X.G., 2017. High expression of RAB43 predicts poor prognosis and is associated with epithelial-mesenchymal transition in gliomas. *Oncology reports*, 37(2), pp.903-912.

Hartmann, C., Meyer, J., Balss, J., Capper, D., Mueller, W., Christians, A., Felsberg, J., Wolter, M., Mawrin, C., Wick, W. and Weller, M., 2009. Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. *Acta neuropathologica*, 118(4), pp.469-474.

Hasan, N., Corbin, D. and Hu, C., 2010. Fusogenic pairings of vesicle-associated membrane proteins (VAMPs) and plasma membrane t-SNAREs–VAMP5 as the exception. *PloS one*, 5(12).

Hodgson, L., 2014. Forword. Regulation of RhoGTPases in motility: *A fine balancing act*.

Hortsch, R. and Weuster-Botz, D., 2011. Growth and recombinant protein expression with Escherichia coli in different batch cultivation media. *Applied microbiology and biotechnology*, 90(1), pp.69-76.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M. and Jensen, L.J., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic acids research*, 44(D1), pp.D286-D293.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J. and von Mering, C., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research*, 47(D1), pp.D309-D314.

Huse, J.T. and Aldape, K.D., 2014. The evolving role of molecular markers in the diagnosis and management of diffuse glioma.

Iacob, G. and Dinca, E.B., 2009. Current data and strategy in glioblastoma multiforme. *Journal of medicine and life*, 2(4), p.386.

Ishida, M., Oguchi, M.E. and Fukuda, M., 2016. Multiple types of guanine nucleotide exchange factors (GEFs) for Rab small GTPases. *Cell structure and function*, p.16008.

Itzen, A., Pylypenko, O., Goody, R.S., Alexandrov, K. and Rak, A., 2006. Nucleotide exchange via local protein unfolding—structure of Rab8 in complex with MSS4. *The EMBO Journal*, 25(7), pp.1445-1455.

Iwakuma, T. and Lozano, G., 2003. MDM2, an introduction. *Molecular Cancer Research*, 1(14), pp.993-1000.

Jean, S., Cox, S., Schmidt, E.J., Robinson, F.L. and Kiger, A., 2012. Sbf/MTMR13 coordinates PI (3) P and Rab21 regulation in endocytic control of cellular remodeling. *Molecular biology of the cell*, 23(14), pp.2723-2740.

Jones, S., 2000. Newman C, Liu F, and Segev N. The TRAPP complex is a nucleotide exchanger for Ypt1 and Ypt31/32. *Molecular Biology Cell*, 11, pp.4403-4411.

Kanu, O.O., Hughes, B., Di, C., Lin, N., Fu, J., Bigner, D.D., Yan, H. and Adamson, C., 2009. Glioblastoma multiforme oncogenomics and signaling pathways. *Clinical medicine*. Oncology, 3, pp.CMO-S1008.

Khan, F., Legler, P.M., Mease, R.M., Duncan, E.H., Bergmann-Leitner, E.S. and Angov, E., 2012. Histidine affinity tags affect MSP142 structural stability and immunodominance in mice. *Biotechnology journal*, 7(1), pp.133-147.

Kim, J.K., Lee, S.Y., Park, C.W., Park, S.H., Yin, J., Kim, J., Park, J.B., Lee, J.Y., Kim, H. and Kim, S.C., 2014. Rab3a promotes brain tumor initiation and progression. *Molecular biology reports*, 41(9), pp.5903-5911.

Kiontke, S., Langemeyer, L., Kuhlee, A., Schuback, S., Raunser, S., Ungermann, C. and Kümmel, D., 2017. Architecture and mechanism of the late endosomal Rab7-like Ypt7 guanine nucleotide exchange factor complex Mon1–Ccz1. *Nature communications*, 8(1), pp.1-10.

Koch, D., Rai, A., Ali, I., Bleimling, N., Friese, T., Brockmeyer, A., Janning, P., Goud, B., Itzen, A., Müller, M.P. and Goody, R.S., 2016. A pull-down procedure for the identification of unknown GEFs for small GTPases. *Small GTPases*, 7(2), pp.93-106.

Korshunov, A., Ryzhova, M., Hovestadt, V., Bender, S., Sturm, D., Capper, D., Meyer, J., Schrimpf, D., Kool, M., Northcott, P.A. and Zheludkova, O., 2015. Integrated analysis of pediatric glioblastoma reveals a subset of biologically favorable tumors with associated molecular prognostic markers. *Acta neuropathologica*, 129(5), pp.669-678.

Kotlyar, M., Pastrello, C., Malik, Z. and Jurisica, I., 2019. IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species. *Nucleic acids research*, 47(D1), pp.D581-D589.

Kotlyar, M., Pastrello, C., Malik, Z. and Jurisica, I., 2019. IID 2018 update: context-specific physical protein–protein interactions in human, model organisms and domesticated species. *Nucleic acids research*, 47(D1), pp.D581-D589.

Kuehn, B.M., 2010. Genomics illuminates a deadly brain cancer. *Jama*, 303(10), pp.925-927.

Laporte, J., Blondeau, F., Buj-Bello, A., Tentler, D., Kretz, C., Dahl, N. and Mandel, J.L., 2003. Characterization of the myotubularin dual specificity phosphatase gene family from yeast to human. *Human molecular genetics*, 7(11), pp.1703-1712.

Lauriat, T.L., Dracheva, S., Kremerskothen, J., Duning, K., Haroutunian, V., Buxbaum, J.D., Hyde, T.M., Kleinman, J.E. and McInnes, L.A., 2006. Characterization of KIAA0513, a novel signaling molecule that interacts with modulators of neuroplasticity, apoptosis, and the cytoskeleton. *Brain research*, 1121(1), pp.1-11.

Letunic, I. and Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research*, 44(W1), pp.W242-W245.

Levivier, E., Goud, B., Souchet, M., Calmels, T.P., Mornon, J.P. and Callebaut, I., 2001. uDENN, DENN, and dDENN: indissociable domains in Rab and MAP kinases signaling pathways. *Biochemical and biophysical research communications*, 287(3), pp.688-695.

Li, Y., Kelly, W.G., LOGSDON JR, J.M., Schurko, A.M., Harfe, B.D., Hill-Harfe, K.L. and Kahn, R.A., 2003. Functional genomic analysis of the ADP-ribosylation factor family of GTPases: phylogeny among diverse eukaryotes and function in C. elegans. *The FASEB journal*, 18(15), pp.1834-1850.

Liu, X.Y., Gerges, N., Korshunov, A., Sabha, N., Khuong-Quang, D.A., Fontebasso, A.M., Fleming, A., Hadjadj, D., Schwartzentruber, J., Majewski, J. and Dong, Z., 2012. Frequent ATRX mutations and loss of expression in adult diffuse astrocytic tumors carrying IDH1/IDH2 and TP53 mutations. *Acta neuropathologica*, 124(5), pp.615-625.

Louis, D.N., 1994. The p53 gene and protein in human brain tumors. *Journal of neuropathology and experimental neurology*, 53(1), pp.11-21.

Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W. and Kleihues, P., 2007. The 2007 WHO classification of tumours of the central nervous system. *Acta neuropathologica*, 114(2), pp.97-109.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R., Potter, S.C., Finn, R.D. and Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research*, 47(W1), pp.W636-W641.

Marat, A.L., Dokainish, H. and McPherson, P.S., 2011. DENN domain proteins: regulators of Rab GTPases. *Journal of Biological Chemistry*, 286(16), pp.13791-13800.

Mathys, H., Basquin, J., Ozgur, S., Czarnocki-Cieciura, M., Bonneau, F., Aartse, A., Dziembowski, A., Nowotny, M., Conti, E. and Filipowicz, W., 2014. Structural and biochemical insights to the role of the CCR4-NOT complex and DDX6 ATPase in microRNA repression. *Molecular cell*, 54(5), pp.751-765.

Matsumura, F., Ono, S., Yamakita, Y., Totsukawa, G. and Yamashiro, S., 1998. Specific localization of serine 19 phosphorylated myosin II during cell locomotion and mitosis of cultured cells. *The Journal of cell biology*, 140(1), pp.119-129.

Murata, T., Delprato, A., Ingmundson, A., Toomre, D.K., Lambright, D.G. and Roy, C.R., 2006. The Legionella pneumophila effector protein DrrA is a Rab1 guanine nucleotide-exchange factor. *Nature cell biology*, 8(9), pp.971-977.

Nagase, T., Koga, H. and Ohara, O., 2006. Kazusa mammalian cDNA resources: towards functional characterization of KIAA gene products. *Briefings in Functional Genomics*, 5(1), pp.4-7.

Nilsson, J., Ståhl, S., Lundeberg, J., Uhlén, M. and Nygren, P.Å., 1997. Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein expression and purification*, 11(1), pp.1-16.

Ohgaki, H. and Kleihues, P., 2005. Epidemiology and etiology of gliomas. *Acta neuropathologica*, 109(1), pp.93-108.

Ohgaki, H. and Kleihues, P., 2009. Genetic alterations and signaling pathways in the evolution of gliomas. *Cancer science*, 100(12), pp.2235-2241.

Ohgaki, H. and Kleihues, P., 2013. The definition of primary and secondary glioblastoma. *Clinical cancer research*, 19(4), pp.764-772.

Ohgaki, H., 2009. Epidemiology of brain tumors. *In Cancer Epidemiology* (pp. 323-342). Humana Press.

Olsen, J.V., Ong, S.E. and Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular and Cellular Proteomics*, 3(6), pp.608-614.

Ozawa, T., Brennan, C.W., Wang, L., Squatrito, M., Sasayama, T., Nakada, M., Huse, J.T., Pedraza, A., Utsuki, S., Yasui, Y. and Tandon, A., 2010. PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas. *Genes and development*, 24(19), pp.2205-2218.

Parsons, D.W., Jones, S., Zhang, X., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.M., Gallia, G.L. and Olivi, A., 2008. Hartigan J. Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. *Science*, 321(5897), pp.1807-12.

Pawlowska, E., Szczepanska, J., Szatkowska, M. and Blasiak, J., 2018. An interplay between senescence, apoptosis and autophagy in glioblastoma multiforme—role in pathogenesis and therapeutic perspective. *International journal of molecular sciences*, 19(3), p.889.

Petrotchenko, E.V. and Borchers, C.H., 2014. Modern mass spectrometry-based structural proteomics. *In Advances in protein chemistry and structural biology* (Vol. 95, pp. 193-213). Academic Press.

Pfeffer, S. and Aivazian, D., 2004. Targeting Rab GTPases to distinct membrane compartments. *Nature reviews Molecular cell biology*, 5(11), pp.886-896.

Phillips, H.S., Kharbanda, S., Chen, R., Forrest, W.F., Soriano, R.H., Wu, T.D., Misra, A., Nigro, J.M., Colman, H., Soroceanu, L. and Williams, P.M., 2006. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer cell*, 9(3), pp.157-173.

Rawlings, A.E., 2016. Membrane proteins: always an insoluble problem?. *Biochemical Society Transactions*, 44(3), pp.790-795.

Riemenschneider, M.J., Jeuken, J.W., Wesseling, P. and Reifenberger, G., 2010. Molecular diagnostics of gliomas: state of the art. *Acta neuropathologica*, 120(5), pp.567-584.

Robinson, F.L. and Dixon, J.E., 2005. The phosphoinositide-3-phosphatase MTMR2 associates with MTMR13, a membrane-associated pseudophosphatase also mutated in type 4B Charcot-Marie-Tooth disease. *Journal of Biological Chemistry*, 280(36), pp.31699-31707.

Rock, K., McArdle, O., Forde, P., Dunne, M., Fitzpatrick, D., O'Neill, B. and Faul, C., 2012. A clinical review of treatment outcomes in glioblastoma multiforme—the validation in a non-trial population of the results of a randomised Phase III clinical trial: has a more radical approach improved survival?. *The British journal of radiology*, 85(1017), pp.e729-e733.

Rosano, G.L. and Ceccarelli, E.A., 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in microbiology*, 5, p.172.

Sahdev, S., Khattar, S.K. and Saini, K.S., 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and cellular biochemistry*, 307(1-2), pp.249-264.

Sakaguchi, A., Sato, M., Sato, K., Gengyo-Ando, K., Yorimitsu, T., Nakai, J., Hara, T. and Sato, K., 2015. REI-1 is a guanine nucleotide exchange factor regulating RAB-11 localization and function in C. elegans embryos. *Developmental cell*, 35(2), pp.211-221.

Sathornsumetee, S., Rich, J.N. and Reardon, D.A., 2007. Diagnosis and treatment of high-grade astrocytoma. *Neurologic clinics*, 25(4), pp.1111-1139.

Schlegel, J., Merdes, A., Stumm, G., Albert, F.K., Forsting, M., Hynes, N. and Kiessling, M., 1994. Amplification of the epidermal-growth-factor-receptor gene correlates with different growth behaviour in human glioblastoma. *International journal of cancer*, 56(1), pp.72-77.

Schlenker, O., Hendricks, A., Sinning, I. and Wild, K., 2006. The structure of the mammalian signal recognition particle (SRP) receptor as prototype for the interaction of small GTPases with Longin domains. *Journal of Biological Chemistry*, 281(13), pp.8898-8906.

Schwartzbaum, J.A., Fisher, J.L., Aldape, K.D. and Wrensch, M., 2006. Epidemiology and molecular pathology of glioma. *Nature clinical practice Neurology*, 2(9), pp.494-503.

Senderek, J., Bergmann, C., Weber, S., Ketelsen, U.P., Schorle, H., Rudnik-Schöneborn, S., Büttner, R., Buchheim, E. and Zerres, K., 2003. Mutation of the SBF2 gene, encoding a novel member of the myotubularin family, in Charcot–Marie–Tooth neuropathy type 4B2/11p15. *Human molecular genetics*, 12(3), pp.349-356.

Shapiro, G.I., 2006. Cyclin-dependent kinase pathways as targets for cancer treatment. *Journal of clinical oncology*, 24(11), pp.1770-1783.

Sherr, C.J. and Weber, J.D., 2000. The ARF/p53 pathway. *Current opinion in genetics and development*, 10(1), pp.94-99.

Shiloach, J. and Fass, R., 2005. Growing E. coli to high cell density—a historical perspective on method development. *Biotechnology advances*, 23(5), pp.345-357.

Simpson, L. and Parsons, R., 2001. PTEN: life as a tumor suppressor. *Experimental cell research*, 264(1), pp.29-41.

Song, J.M., An, Y.J., Kang, M.H., Lee, Y.H. and Cha, S.S., 2012. Cultivation at 6–10 C is an effective strategy to overcome the insolubility of recombinant proteins in Escherichia coli. *Protein expression and purification*, 82(2), pp.297-301.

Spiess, C., Beil, A. and Ehrmann, M., 1999. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell*, 97(3), pp.339-347.

Standfuß, C., Parczyk, J., Ruhnau, J. and Klein, A., 2019. Genome reorganization in different cancer types: detection of cancer specific breakpoint regions. *Molecular cytogenetics*, 12(1), p.25.

Stano, N.M. and Patel, S.S., 2004. T7 lysozyme represses T7 RNA polymerase transcription by destabilizing the open complex during initiation. *Journal of Biological Chemistry*, 279(16), pp.16136-16143.

Stenmark, H. and Olkkonen, V.M., 2001. The rab gtpase family. *Genome biology*, 2(5), pp. reviews3007-1.

Stenmark, H., 2009. Rab GTPases as coordinators of vesicle traffic. *Nature reviews Molecular cell biology*, 10(8), pp.513-525.

Studier, F.W., 2005. Protein production by auto-induction in high-density shaking cultures. *Protein expression and purification*, 41(1), pp.207-234.

Swaminathan, R., Krishnamoorthy, G. and Periasamy, N., 1994. Similarity of fluorescence lifetime distributions for single tryptophan proteins in the random coil state. *Biophysical journal*, 67(5), pp.2013-2023.

Tanaka, M., Okamura, Y., Arakaki, A., Tanaka, T., Takeyama, H. and Matsunaga, T., 2006. Origin of magnetosome membrane: proteomic analysis of magnetosome membrane and comparison with cytoplasmic membrane. *Proteomics*, 6(19), pp.5234-5247.

Terpe, K., 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*, 60(5), pp.523-533.

Thakkar, J.P., Dolecek, T.A., Horbinski, C., Ostrom, Q.T., Lightner, D.D., Barnholtz-Sloan, J.S. and Villano, J.L., 2014. Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiology and Prevention Biomarkers*, 23(10), pp.1985-1996.

The UniProt Consortium, 2017. UniProt: the universal protein knowledgebase. *Nucleic acids research*., 45, D158–D169.

Thomas, C., Fricke, I., Scrima, A., Berken, A. and Wittinghofer, A., 2007. Structural evidence for a common intermediate in small G protein-GEF reactions. *Molecular cell*, 25(1), pp.141-149.

Van de Weert, M. and Stella, L., 2011. Fluorescence quenching and ligand binding: A critical discussion of a popular methodology. *Journal of Molecular Structure*, 998(1-3), pp.144-150.

Van Weering, J.R., Toonen, R.F. and Verhage, M., 2007. The role of Rab3a in secretory vesicle docking requires association/dissociation of guanidine phosphates and Munc18-1. *PLoS One*, 2(7).

Velásquez, C., Mansouri, S., Mora, C., Nassiri, F., Suppiah, S., Martino, J., Zadeh, G. and Fernández-Luna, J.L., 2019. Molecular and Clinical Insights into the Invasive Capacity of Glioblastoma Cells. *Journal of oncology*, 2019.

Venkataraman, A., Nevrivy, D.J., Filtz, T.M. and Leid, M., 2012. Grp1-associated scaffold protein (GRASP) is a regulator of the ADP ribosylation factor 6 (Arf6)-dependent membrane trafficking pathway. *Cell biology international*, 36(12), pp.1115-1128.

Vera, A., González-Montalbán, N., Arís, A. and Villaverde, A., 2006. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnology and bioengineering*, 96(6), pp.1101-1106.

Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 17: 98–110. doi: 10.1016/j.ccr.2009.12.02

Vivian, J.T. and Callis, P.R., 2001. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical journal*, 80(5), pp.2093-2109.

Wandinger-Ness, A. and Zerial, M., 2014. Rab proteins and the compartmentalization of the endosomal system. *Cold Spring Harbor perspectives in biology*, 6(11), p.a022616.

Wang, H. and Jiang, C., 2013. RAB38 confers a poor prognosis, associated with malignant progression and subtype preference in glioma. *Oncology reports*, 30(5), pp.2350-2356.

Weller, M., Felsberg, J., Hartmann, C., Berger, H., Steinbach, J.P., Schramm, J., Westphal, M., Schackert, G., Simon, M., Tonn, J.C. and Heese, O., 2009. Molecular predictors of progression-free and overall survival in patients with newly diagnosed glioblastoma: a prospective translational study of the German Glioma Network. *Journal of Clinical Oncology*, 27(34), pp.5743-5750.

Wen, P.Y. and Kesari, S., 2008. Malignant gliomas in adults. *New England Journal of Medicine*, 359(5), pp.492-507.

Wennerberg, K., Rossman, K.L. and Der, C.J., 2005. The Ras superfamily at a glance. *Journal of cell science*, 118(5), pp.843-846.

Wiech, M., Olszewski, M.B., Tracz-Gaszewska, Z., Wawrzynow, B., Zylicz, M. and Zylicz, A., 2012. Molecular mechanism of mutant p53 stabilization: the role of HSP70 and MDM2. *PLoS One*, 7(12), p.e51426.

Wong, A.J., Bigner, S.H., Bigner, D.D., Kinzler, K.W., Hamilton, S.R. and Vogelstein, B., 1987. Increased expression of the epidermal growth factor receptor gene in malignant gliomas is invariably associated with gene amplification. *Proceedings of the National Academy of Sciences*, 84(19), pp.6899-6903.

Wu, X., Hu, A., Zhang, M. and Chen, Z., 2013. Effects of Rab27a on proliferation, invasion, and anti-apoptosis in human glioma cell. *Tumor Biology*, 34(4), pp.2195-2203.

Xu, C., Gu, J., Ma, X., Dong, T. and Meng, X., 2014. Investigation on the interaction of pyrene with bovine serum albumin using spectroscopic methods. Spectrochimica Acta Part A: *Molecular and Biomolecular Spectroscopy*, 125, pp.391-395.

Yoshimura, S.I., Gerondopoulos, A., Linford, A., Rigden, D.J. and Barr, F.A., 2010. Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. *Journal of Cell Biology*, 191(2), pp.367-381.

Zarghooni, M., Bartels, U., Lee, E., Buczkowicz, P., Morrison, A., Huang, A., Bouffet, E. and Hawkins, C., 2010. Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor α and poly (ADP-ribose) polymerase as potential therapeutic targets. *Journal of clinical oncology*, 28(8), pp.1337-1344.

Zeinoddini, M., Khajeh, K., Hosseinkhani, S., Saeedinia, A.R. and Robatjazi, S.M., 2013. Stabilisation of recombinant aequorin by polyols: activity, thermostability and limited proteolysis. *Applied biochemistry and biotechnology*, 170(2), pp.273-280.

Zhang, F.L. and Casey, P.J., 1996. Protein prenylation: molecular mechanisms and functional consequences. *Annual review of biochemistry*, 65(1), pp.241-269.

Zhen, Y. and Stenmark, H., 2015. Cellular functions of Rab GTPases at a glance. *Journal of Cell Science*, 128(17), pp.3171-3176.

Zhou, X., Wang, W., Zhang, S., Wang, X., Tang, Z., Gu, J., Li, J. and Huang, J., 2017. CACNA1B (Cav2. 2) overexpression and its association with clinicopathologic characteristics and unfavorable prognosis in non-small cell lung cancer. *Disease markers*, 2017.

Zininga, T., Achilonu, I., Hoppe, H., Prinsloo, E., Dirr, H.W. and Shonhai, A., 2015. Overexpression, purification and characterisation of the Plasmodium falciparum Hsp70-z (PfHsp70-z) protein. *PLoS One*, 10(6).

Zininga, T., Achilonu, I., Hoppe, H., Prinsloo, E., Dirr, H.W. and Shonhai, A., 2016. Plasmodium falciparum Hsp70-z, an Hsp110 homologue, exhibits independent chaperone activity and interacts with Hsp70-1 in a nucleotide-dependent fashion. *Cell Stress and Chaperones*, 21(3), pp.499-513.

Appendices

Appendix A: Supplementary data

Table A1. Human Rab proteins and their identified nucleotide exchange factors

Rabs	GEFs	References	Rabs	GEFs	References
Rab1a	TRAPP I DrrA (Pneumophila)	Yoshimura *et al*., 2010; Ishida *et al*., 2016 Murata *et al*., 2006	**Rab17**	Vps9	**Yoshimura *et al*., 2010; Ishida *et al*., 2016**
Rab1b	TRAPP I DrrA (Pneumophila)	Yoshimura *et al*., 2010; Ishida *et al*., 2016 Murata *et al*., 2006	**Rab18**		
Rab2a			**Rab20**		
Rab2b			**Rab21**	Rabex-5 (Vps9)	**Yoshimura *et al*., 2010; Ishida *et al*., 2016**
Rab3a	MADD (DENN)	Yoshimura *et al*., 2010, Ishida *et al*., 2016	**Rab22a**	Rabex-5 (Vps9)	**Yoshimura *et al*., 2010; Ishida *et al*., 2016**
Rab3b	MADD (DENN)	Yoshimura *et al*., 2010 Ishida *et al*., 2016	**Rab23**		
Rab3c	MADD (DENN)	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab24**		
Rab3d	MADD (DENN)	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab25**		
Rab4a			**Rab26**		
Rab4b			**Rab27a**	MADD (DENN)	**Yoshimura *et al*., 2010; Ishida *et al*., 2016**
Rab5a	Rabex-5 (Vps9)	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab27b**		
Rab5b			**Rab28**	SBF1 (DENN)	**Yoshimura *et al*., 2010; Ishida *et al*., 2016**
Rab5c			**Rab29**		
Rab6a	Ric1-Rgp1	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab30**		
Rab6b	Ric1-Rgp1	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab31**		
Rab6c			**Rab32**	BLOC-3	**Gerondopoulos *et al*., 2012**
Rab7a	Mon1/Ccz1	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab33a**		
Rab7b	Mon1/Ccz1	Yoshimura *et al*., 2010, Ishida *et al*., 2016	**Rab33b**		
Rab8a	Rabin-8 GRAB Mss4 C9Orf72	 Yoshimura *et al*., 2010; Ishida *et al*., 2016 Itzen *et al*., 2006 Corbier and Sellier, 2016	**Rab34**		
Rab8b	C9Orf72	Corbier and Sellier, 2016	**Rab35**	DENND1A-C	**Yoshimura *et al*., 2010; Ishida *et al*., 2016**

Rab9a	DENND2	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab36**		
Rab9b	DENND2	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab37**		
Rab10	DENND4	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab38**	BLOC-3	**Gerondopoulos *et al*., 2012**
Rab11a	SH3BP5 (REI-1)	Sakaguchi *et al*., 2015	**Rab39a**	DENND5A-B C9Orf72	**Yoshimura *et al*., 2010; Ishida *et al*., 2016 Corbier and Sellier, 2016**
Rab11b	SH3BP5 (REI-1)	Sakaguchi *et al*., 2015	**Rab39b**	DENND5A-B C9Orf72	**Yoshimura *et al*., 2010; Ishida *et al*., 2016 Corbier and Sellier, 2016**
Rab12	DennD3	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab40a**		
Rab13	DENND1C	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab40b**		
Rab14	DENND6	Yoshimura *et al*., 2010; Ishida *et al*., 2016	**Rab40c**		
Rab15			**Rab41**		
			Rab43		

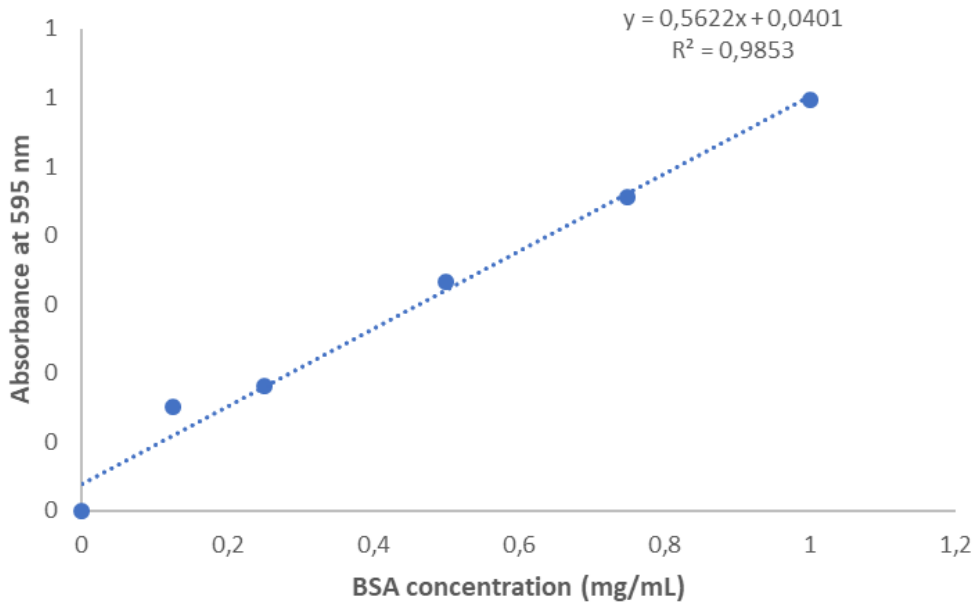


Figure A1. BSA standard curve generated from 10 mg/mL BSA stock.
Absorbance of different BSA dilution read at 595 nm was plotted against respective concentration. Linear regression and the R2 value were generated using Excel.

Table A2. Physicochemical characteristics of K0513 homologues

	DrK0513	XtK0513	GgK0513	MmK0513	RnK0513	MamK0513	PtK0513	ClK0513	BtK0513
Molecular wight	47975.42	48059.25	45383.72	46318.39	46275.34	46583.54	46648.59	46837.19	46869.00
Theoretical pI	5.1888	4.8843	5.8166	4.6636	4.6560	4.7406	4.6994	5.0558	5.2156
Residues	424	419	400	407	407	411	411	410	413
Net charge	-11.0	-15.5	-5.0	-18.5	-18.0	-16.5	-17.5	-11.0	-9.0

Table A3: Predicted interaction partners of K0513

UniProt ID	Name	Description	Localization	PMIDs
Q9NTI2	ATP8A2	Phospholipid-transporting ATPase IB, serves as a Catalytic component of the P4-ATPase flippase complex	Golgi, endosomes	25402006
P60880	SNAP25	Synaptosomal-associated protein 25, involved in the molecular regulation of neurotransmitter release	Cell membrane	25402006
Q2M2I8	AAK1	AP2-associated protein kinase 1, Regulation of clathrin-mediated endocytosis	Cell membrane	25402006
Q9NWB1	RBFOX1	RNA binding protein fox-1 homolog 1, involved in the regulation of alternative splicing events	Nucleus, cytoplasm	25402006
P14867	GABRA1	Gamma-aminobutyric acid receptor subunit alpha-1, Ligand-gated chloride channel which forms part of heteropentameric receptor for GABA	postsynaptic cell membrane	25402006
P09471	GNAO1	Guanine nucleotide-binding protein G(o) subunit alpha, modulator of various transmembrane signalling systems	Cell membrane	25402006
P61764	STXBP1	Syntaxin-binding protein 1, regulates synaptic vesicle docking and fusion through its interaction with the GTP-binding proteins	Cytosol	25402006
P37840	SNCA	Alpha-synuclein, Neuronal protein that plays several roles in synaptic activity	Nucleus, cytosol	25402006
Q59EK9	RUNDC3A	RUN domain-containing protein 3A, in neural cells, it serves as an effector of RAP2A	Cytosol	25402006
Q70YC5	ZNF365	Protein ZNF365, Involved in the regulation of neurogenesis	Centrosome	25402006
Q9BWQ8	FAIM2	Protein lifeguard 2, functions as an antiapoptotic protein from Fas-induced apoptosis	Cell membrane	25402006
Q9UI12	ATP6V1H	V-type proton ATPase subunit H, a subunit of the peripheral V1 complex of vacuolar ATPase	Cytosol	25402006
Q9UM19	HPCAL4	Hippocalcin-like protein 4, Implicated in the calcium-dependent regulation of rhodopsin phosphorylation	Cytosol	25402006
P78357	CNTNAP1	Contactin-associated protein 1, facilitates the radial and longitudinal organization of myelinated axons	Cell membrane	23023127
Q9NZU7	CABP1	Calcium-binding protein 1, regulates the activity of inositol 1,4,5-triphosphate receptors	Cell membrane	25402006
Q9UPR5	SLC8A2	Sodium/calcium exchanger 2, facilitates the exchange of Ca^{2+} and Na^{+} ions across the membrane	Cell membrane	25402006
Q9UI15	TAGLN3	Transgelin-3, central nervous system development	Nucleus	25402006
Q9UBL0	ARPP21	cAMP-regulated phosphoprotein 21, may act on of calmodulin-dependent enzymes as a competitive inhibitor	Cytosol	25402006
P17600	SYN1	Synapsin-1, Neuronal phosphoprotein responsible for coating synaptic vesicles as well as binding to the cytoskeleton	Golgi apparatus	25402006
O60359	CACNG3	Voltage-dependent calcium channel gamma-3 subunit, Facilitates the trafficking to the AMPA-selective glutamate receptors (AMPARs)	Cell membrane	25402006
P08247	SYP	Synaptophysin, membrane organizing protein which facilitates organization of other membrane components or target vesicles to the plasma membrane	Cell membrane	25402006
Q3SXP7	SHISAL1	Protein shisa-like-1,	Cell membrane	25402006
P62760	VSNL1	Visinin-like protein 1, Regulates the inhibition of rhodopsin phosphorylation in a calcium-dependent manner	Cytosol	25402006

Q99784	OLFM1	Noelin, serves as a regulatory protein of the axonal growth in both adult and embryonic central nervous system	Endoplasmic reticulum	25402006
Q96HW7	INTS4	Integrator complex subunit 4, forms part of the Integrator (INT) complex, which is involved in the small nuclear RNAs U1 and U2 transcription	Nucleus	17010949;21836163
O75493	CA11	Carbonic anhydrase-related protein 11,	Extracellular	25402006
P85037	FOXK1	Forkhead box protein K1, Regulator of transcription involved in various processes including muscle cell differentiation, autophagy, glucose metabolism, and aerobic glycolysis	Nucleus	25609649
O95741	CPNE6	Copine-6, Phospholipid-binding protein implicated in the calcium-mediated intracellular processes	Cell membrane	25402006
Q9UPA5	BSN	Protein bassoon, is a scaffold protein which forms part of the presynaptic cytomatrix at the active zone (CAZ)	Cytoskeleton, Cytoplasm	25402006
Q92581	SLC9A6	Sodium/hydrogen exchanger 6, serves in the exchange Na^+ and K^+ protons across the early and recycling endosome membranes	Endosome	25402006
P21579	SYT1	Synaptotagmin-1, Calcium sensor that participates in triggering neurotransmitter release at the synapse	Cell membrane	25402006
Q9UPV7	PHF24	PHD finger protein 24	Cytosol	25402006
Q9P2U7	SLC17A7	Vesicular glutamate transporter 1, facilitates glutamate uptake into synaptic vesicles at the presynaptic nerve terminals of the excitatory neural cells	Synaptic vesicle membrane	25402006
Q7L0J3	SV2A	Synaptic vesicle glycoprotein 2A, serves to maintain a balanced and regulated secretion in endocrine and cells	Synaptic vesicle membrane	25402006
Q9Y4E6	WDR7	WD repeat-containing protein 7		25402006
Q05193	DNM1	Dynamin-1, Involved in the production of microtubule bundles, it is also known to bind and hydrolyse GTP	Cytoskeleton, Cytosol	25402006
Q16799	RTN1	Reticulon-1, May be implicated in membrane trafficking in neuroendocrine cells	Golgi apparatus	25402006
Q9UPP2	IQSEC3	IQ motif and SEC7 domain-containing protein 3, Functions as a guanine nucleotide exchange factor for ARF1	Cytosol	25402006
Q01167	FOXK2	Forkhead box protein K2, A transcription regulator playing a role in processes such as aerobic glycolysis, autophagy and glucose metabolism	Nucleus	25609649
Q15818	NPTX1	Neuronal pentraxin-1, Facilitates the uptake of synaptic material during synapse remodelling	Cytoplasmic vesicle	25402006
Q92561	PHYHIP	Phytanoyl-CoA hydroxylase-interacting protein, May have a role in the development of the central nervous system	Cytosol	25402006
O60307	MAST3	Microtubule-associated serine/threonine-protein kinase 3	Cytosol	25402006
Q8TDI0	CHD5	Chromodomain-helicase-DNA-binding protein 5, Chromatin-remodelling protein which binds to DNA through histones and facilitates gene transcription	Nucleus	25402006
Q13367	AP3B2	AP-3 complex subunit beta-2, Subunit of clathrin- and non-clathrin-associated adaptor protein complex 3	Golgi apparatus	25402006
Q13387	MAPK8IP2	C-Jun-amino-terminal kinase-interacting protein 2, the JNK-interacting protein (JIP) group of scaffold proteins that mediates the JNK signalling	Cytosol	25402006
Q9BR01	SULT4A1	Sulfotransferase 4A1, member of the atypical sulfotransferase family known to bind 3'-phospho-5'-adenylyl sulfate (PAPS) at a very low affinity	Cytosol	25402006

O60641	SNAP91	Clathrin coat assembly protein AP180, Adaptins are components of the adapter complexes which link clathrin to receptors in coated vesicles	Cell membrane	25402006
Q9UPP5	KIAA1107	AP2-interacting clathrin-endocytosis protein, involved in endocytosis mediated by clathrin at the synapse	Presynapse	25402006
Q7L1I2	SV2B	Synaptic vesicle glycoprotein 2B, serves in the control of regulated secretion in endocrine neural cells	Synaptic vesicle membrane	25402006
Q9NYX4	CALY	Neuron-specific vesicular protein calcyon, interacts with clathrin light chain A and induce the clathrin-mediated endocytosis and clathrin self-assembly	Cell membrane, cytosol	25402006
P46459	NSF	Vesicle-fusing ATPase, involved in vesicle-mediated transport	Cytosol, neurofilament	25402006
P07196	NEFL	Neurofilament light polypeptide, consists of three intermediate filament proteins: L, M, and H that are known to have role in maintenance of neuronal caliber	Cytoskeleton, cytosol	25402006
P49418	AMPH	Amphiphysin, involved in the regulatory mechanisms of exocytosis in synapses and endocrine cells	Cytoskeleton	25402006
Q16143	SNCB	Beta-synuclein, Non-amyloid component of senile plaques found in Alzheimer disease	Cytosol	25402006
Q08209	PPP3CA	Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform, Calcium-dependent, calmodulin-stimulated protein phosphatase	Cell membrane	25402006
Q8IX03	WWC1	Protein KIBRA, plays a pivotal role in tumour suppression through restriction of proliferation and apoptosis promotion	Nucleus, cell membrane	17010949;21836163
P63027	VAMP2	Vesicle-associated membrane protein 2, Involved in the fusion transport of vesicles to their target membrane	Cell membrane	25402006
Q9P2S2	NRXN2	Neurexin-2, Neuronal cell surface protein implicated in cell adhesion and recognition	Cell membrane	23023127
Q96BY2	MOAP1	Modulator of apoptosis 1, Plays a role in receptor-dependent apoptosis	Cytosol	25402006
Q99819	ARHGDIG	Rho GDP-dissociation inhibitor 3, known to Inhibit the exchange of GDP/GTP in RhoB	Cytoplasm	25402006
P05771	PRKCB	Protein kinase C beta type, Calcium-activated, phospholipid- and diacylglycerol (DAG)-dependent serine/threonine-protein kinase	Cell membrane, cytosol	25402006
Q9H2X9	SLC12A5	Solute carrier family 12 member 5, required for neuronal Cl- homeostasis, and facilitates electroneutral potassium-chloride cotransport in mature neurons	Cell membrane	25402006
P05067	APP	Amyloid-beta precursor protein, Serves as a cell surface receptor	Cell membrane, nucleus	21832049
O14576	DYNC1I1	Cytoplasmic dynein 1 intermediate chain 1, Functions as a non-catalytic accessory components of the cytoplasmic dynein 1 complex	Cytosol	25402006
O15079	SNPH	Syntaphilin, absorbing free syntaxin-1 and iInhibits SNARE complex formation	Cell membrane	25402006
O15197	EPHB6	Ephrin type-B receptor 6, Kinase-defective receptor for members of the ephrin-B family	Cell membrane	25402006
O43505	B4GAT1	Beta-1,4-glucuronyltransferase 1, Beta-1,4-glucuronyltransferase implicated in O-mannosylation of alpha-dystroglycan (DAG1)	Golgi apparatus membrane	25402006

Appendix B: Methodology

B1. DNA extraction and restriction digestion analysis

A single colony from successful transformants was inoculated into 2xYT broth supplemented with 100 µg/mL ampicillin and incubated with shaking (250 rpm) at 37 °C. The pQE30-K0513$_W$ extraction followed the GeneJET Plasmid Miniprep Kit (Thermo scientific, USA) protocol. Briefly, 4 mL of the culture was harvested into sterilized eppendorf tube by centrifuging sterilized for 60 seconds at 14000 xg. The cells were resuspended by the Resuspension Solution containing RNAse A by vigorously mixing and pipetting up and down. The cells were lysed by adding Lysis Buffer from the kit and thoroughly mixed by inverting the tube 6 times until the solution becomes viscous and slightly clear. The reaction was neutralized by adding Neutralization Buffer and thoroughly mixing followed by centrifugation at 14000 xg for 5 minutes. The supernatant was carefully transferred into a GeneJET spin column, avoiding disturbance of white precipitate. The supernatant was collected by placing the column in the collection tube and centrifuged for 1 minute. The flow through was decanted, and in the same collection tube, the column was washed with the Wash Solution. After the second wash with Wash Solution, the flow through was decanted and the tube was spun for 1 minute to remove the residual Wash Solution. The plasmid DNA was eluted with pre-warmed GeneJET Elution Buffer into a sterile Eppendorf tube. The integrity of the DNA was confirmed using restriction digest following preparation described in table below and analyzed using agarose gel electrophoresis.

Following DNA extraction, restriction analysis was performed using *Bam* HI and *Hind* III for pQE30-K0513$_W$, *Nco* I and *Bgl* II were used for pQE60-K0513$_H$ (New England Biolabs) restriction enzymes. Table 1.1 shows how samples were prepared. The 0.8 % agarose gel was prepared by dissolving 0.32 g agar in 1 x TAE buffer and heat with frequent agitation. Once cooled, 4 µL of ethidium bromide was added and the gel was gently casted and allowed to solidify.

Table B1. Restriction digest preparation

	Control/Uncut)	(*Bam* HI)	(*Hind* III)	(*Bam* HI + *Hind* III)
H_2O	14 µL	12 µL	12 µL	10 µL
Buffer 2.1/3.1	5 µL (3.1)	5 µL (3.1)	5 µL (2.1)	5 µL (3.1)
DNA	2 ng	2 ng	2 ng	2 ng
Enzyme	-	1 µL	1 µL	1 + 2 µL
Total	20 µL	20 µL	20 µL	20 µL
	Control/Uncut)	**(*Nco* I)**	**(*Bgl* II)**	**(*Nco* I + *Bgl* II)**
H_2O	17 µL	16 µL	16 µL	16 µL
Buffer 2.1/3.1	2 µL (3.1)	2 µL (CutSmart)	2 µL (3.1)	2 µL (CutSmart)
DNA	1 µL	1 µL	1 µL	µL
Enzyme	-	1 µL	1 µL	1 + 1 µL
Total	20 µL	20 µL	20 µL	20 µL

B2. Competent cells preparation

A single colony of *E. coli* JM109 and XL1-Blue were inoculated each into 5 mL double strength yeast-tryptone broth (2 x YT media) and allowed to grow overnight with shaking at 37 ºC. The overnight starter culture was transferred by diluting it into 45 mL of fresh 2xYT media. The culture was allowed to grow with shaking to an early log phase of absorbance 0.3-0.6 measured at OD_{600}. The cells were each harvested into a 50 mL centrifuge tube and spun at 5000 xg for 10 minutes at 4 ºC. From this point, all the steps were carried out on ice. After centrifugation, the supernatant was discarded and the cells were chemically treated by resuspending in 10 mL of ice-cold 0.1 M $MgCl_2$, this was left on ice for 30 minutes. The suspensions were each centrifuged for 10 minutes at 4000 xg at 4 ºC. The cells were pelleted as before and gently resuspended in 10 ml ice cold 0.1 M $CaCl_2$ which was subsequently followed by incubation on ice for 5 hours. The suspensions were centrifuged at 5000 xg at 4 ºC for 10 minutes. The pellets were resuspended by adding 200 µL of sterile 30 % glycerol

and 200 µL of 0.1 M $CaCl_2$ and incubated on ice for 10 min. Following incubation, the competent cells were aliquoted into pre-chilled Eppendorf tubes and stored at -80 ºC.

B3. Transformation of K0513 into the competent cells

A volume of 50 µL of *E. coli* JM109 competent cells from -80 °C were placed in Eppendorf tubes. 2 µL of pQE30-K0513$_W$/pQE60-K0513$_H$ construct was added in the test and 2 µL of water was added in the control tube. The tubes were left to stand on ice for 15 minutes. Samples were heat shocked for 2 minutes at 42 °C. The samples were chilled on ice for 2 minutes prior to the addition of 800 µL of 2xYT media into each tube. Tubes were incubated with shaking (200 rpm) at 37 °C for 40 minutes. About 100 µL of each sample was spread on 2xYT-agar plates supplemented with 100 µg/mL ampicillin, and the remaining sample was centrifuged at 5000 xg for 2 minutes. The 650 µL of the supernatant was discarded and pellets were resuspended in the remaining 100 µL supernatant which was then spread on 2xYT-agar plates supplemented with 100 µg/mL ampicillin. All plates were incubated at 37 °C overnight. pQE30-K0513$_W$/pQE60-K0513$_H$ construct was transformed following the same procedure in *E. coli* XL1-Blue cells.

B4. SDS-PAGE and Western blot analysis of the protein samples

Protein samples were treated by boiling in SDS sample buffer (0.25 % Coomasie Brilliant blue [R250]; 2 % SDS; 10 % glycerol [v/v]; 100 mM Tris; 1 % β-mercaptoethanol) in a ratio of 4:1 for 10 minutes at 100 °C. The samples were analyzed using SDS-PAGE by resolving them through 12 % acrylamide resolving gel (Table 3). This was done by loading the boiled samples in respective wells as well as the prestained protein molecular marker. The electrophoresis was done at 120 volts and 150 amperes for 1 hour 30 minutes using the Bio-Rad Mini protein 3 electrophoresis system (BioRad, U.S.A).

On completion of electrophoresis, SDS-PAGE gels were removed from glass plates and the stacking gel was cut out. Western blotting was done using wet transfer. The nitrocellulose membrane, filter papers and fibre pads were soaked in Western transfer buffer (25 mM Tris; 192 mM glycine; 20 % methanol). The gel was placed on the filter

paper on top of the fibre pad. The nitrocellulose membrane was then laid over the gel, followed by the filter paper and the other fibre pad. The transfer was run at 100 volts for an hour. The membrane blocked with 5 % BSA for 1 hour on ice. BSA was decanted and the membrane was incubated with the primary antibody on ice with shaking. The membrane was wash twice with TBST (50 mM Tris; 150 mM NaCl; 0.1 % Tween 20). Membrane was incubated with the secondary antibody for 45 minutes and subsequently washed four times with TBST. The visualization was done using Enhanced Chemiluminescent (ECL) on a ChemiDoc (Bio-Rad, USA).

Table B2. Preparation of SDS-PAGE 12 % resolving gel and 5 % stacking gel

Running Gel	**X1**	**X2**	**X3**	**X4**	**Stacking Gel**	**X1**	**X2**	**X3**	**X4**
40 % Bis (mL)	1.5	3	4.5	6	**40 % Bis (mL)**	0.25	0.5	0.75	1
1.5 M Tris pH 8.8 (mL)	1.25	2.5	3.75	5	**0.5 M Tris pH 6.8 (mL)**	0.25	0.5	0.75	1
10 % APS (mL)	0.05	0.1	0.15	0.2	**10 % APS (mL)**	0.02	0.04	0.06	0.08
10 % SDS (mL)	0.05	0.1	0.15	0.2	**10 % SDS (mL)**	0.02	0.04	0.06	0.08
TEMED (mL)	0.002	0.004	0.006	0.008	**TEMED (mL)**	0.002	0.004	0.006	0.008
H_2O (mL)	2.15	4.3	6.45	8.6	**H_2O (mL)**	1.46	2.92	4.38	5.89

B5. Quantification of proteins

Protein concentration was determined using Bradford's method (Bradford, 1976). Briefly, BSA standard was set up by diluting 10 mg/mL BSA in 0.01 M Tris-HCl, pH 8 to concentration ranging from 0 to 1 mg/mL as shown in Table 2. A volume of 1 mL of the Bradford's reagent (10 % Coomassie G250 in 95 % ethanol, 85 % (w/v) phosphoric acid) was added to 20 µL of each solution and the reaction was incubated at room temperature for 5 minutes. 250 µL of each sample was transferred int triplicates to a 96-well plate. Absorbance reading was taken at 595 nm using Spectramax M3. The

protein eluents (20 µL) were similarly treated, and the protein concentration determined by extrapolation from the standard curve.

Table B3. Protein quantification by Bradford's assay

Standard dilution	Volume 10 mg/mL BSA	Volume dialysis (µL)	Final BSA concentration (mg/mL)
Blank	0	100	0
S1	1,25	98,75	0,125
S2	2,5	97,5	0,25
S3	5	95	0,5
S4	7,5	92,5	0,75
S5	10	90	1

Table B4: Special chemical reagents

Name of reagents	Supplier
Glacial acetic acid	Merck, Germany
Nucleotides; Adenosine triphosphate (ATP), Adenosine diphosphate (ATP), Guanosine triphosphate (GTP) and Guanosine triphosphate (GTP)	Sigma-Aldrich, USA
Agarose	Merck, Germany
Ammonium persulphate	Sigma-Aldrich, USA
Ampicillin	Melford, UK
Bovine serum albumin	Melford, UK
Bromophenol blue	Merck, Germany
Calcium chloride	Merck, Germany
Chemiluminescence Western blotting kit	Amersham, USA
Nitrocellulose membrane	Thermo Scientific, USA
Coomasie Brilliant Blue R250	Merck, Germany
Ethidium bromide	Merck, Germany
Glycerol	Merck, Germany
Glycine	Merck, Germany
Glucose	Merck, Germany
Imidazole	Merck, Germany
Isopropyl-1-thio-D-galacopyranoside	Sigma-Aldrich, USA
GeneRuler 1 Kb DNA ladder	Thermo Scientific, USA
Lysozyme	Merck, Germany
Magnesium chloride	Merck, Germany
Methanol	Merck, Germany
HisPur™ NI-NTA Resin	Thermo Scientific USA
Phenylmethylsufonyl fluoride	Merck, Germany
Polyacrylamide	Merck, Germany
Restriction enzymes	Thermo Scientific, USA
Amicon® Ultra-15 10K centrifuge filter device	Merck, Germany
Snakeskin™ pleated dialysis tubing	Thermo Scientific, USA
Sodium chloride	Merck, Germany
Sodium dodecyl sulphate	Merck, Germany