Mariana Filipa de Almeida Viegas

Licenciada em Biologia

Towards determining the structure of human KIFC1, a potential target for drug development in cancer therapy

Dissertação para obtenção de Grau de Mestre em Genética Molecular e Biomedicina

Orientador: Prof. Dr. Frank Kozielski, Professor de "Pharmaceutical and Biological Chemistry", School of Pharmacy, University College London

Júri:

Presidente: Prof. Doutora Paula Maria Theriaga Mendes Bernardes Gonçalves;

Doutor Pedro Miguel Ribeiro Viana Baptista;

Arguente: Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes.

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Setembro 2017

Mariana Filipa de Almeida Viegas

Towards determining the structure of human KIFC1, a potential target for drug development in cancer therapy

Dissertação apresentada à Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa para obtenção de Grau de Mestre em Genética Molecular e Biomedicina.

Área de Concentração: Cancer Therapy

Orientador: Prof. Dr. Frank Kozielski

Lisboa

2017

Towards determining the structure of human KIFC1, a potential target for drug development in cancer therapy

Acknowledgements

My deepest appreciation to Prof. Frank Kozielski and Dr. Sandeep Talapatra for all their guidance, support and precious advice during the course of this project and writing of my thesis.

Furthermore, I would like to thank my former colleges, Federica Riccio and Chatrin Chatrin, for their help during my initial period in the laboratory.

Additionally, I would also like to thank Pedro Rodrigues, Rafaela Nogueira, Sofia Dias, Diogo Lourenço, Patrícia Alves, Madalena Santos, Martim Rodrigues and Manuel Sampaio for all their emotional and moral support throughout the whole process.

Abstract

Normal cell division is initiated upon centrosome duplication and the two centrosomes move towards the cell periphery to form the poles of a bipolar mitotic spindle, a function that is essential for accurate chromosome segregation. Aberrant centrosomes duplication leading to more than two centrosomes is referred to as centrosome amplification. Centrosome amplification causes genetic instability and is a hallmark in cancer cells. Cancer cells survive the multiple mitotic spindle formation by clustering Microtubules Organizing Centres (MTOCs) to form two supernumerary centrosomes and a single pseudo-bipolar spindle during metaphase. KIFC1, a kinesin essential for this process, is particularly interesting as a potential therapeutic target. This is due to its non-essential role in cell division in normal cells but crucial centrosome clustering function in cancer cells with supernumerary centrosomes.

The overall goal of this project was to determine the structure of the catalytic domain, also known as motor domain, of KIFC1 to aid future structure-based drug-design.

In this project, I have established the expression and purification protocol for the KIFC1 motor domain. The protocol enables to obtain large amounts of protein at highest purity. Although I obtained crystals initially, their reproduction was difficult and therefore high-resolution crystals of the KIFC1 motor domain proved difficult to obtain. As such, the search for crystallisation conditions that allow the formation of high-resolution and reproducible crystals of the KIFC1 motor domain continues. Once this bottleneck is surpassed, it will be possible to use the structure for future structure-based drug-design.

Keywords: KIFC1, Kinesin, Cancer, Structure, X-ray crystallography, Drug-design

Resumo

A divisão celular normal é iniciada quando o centrossoma sofre duplicação e os dois centrossomas se deslocam para a periferia celular para formar os polos do fuso mitótico bipolar, uma função que é essencial para a segregação cromossomal precisa. A duplicação centrossomal aberrante que resulta em mais de dois centrossomas é referida como amplificação centrossomal. A amplificação centrossomal causa instabilidade genética e é uma característica das células cancerígenas. As células cancerígenas sobrevivem ao fuso mitótico multipolar pelo agrupamento de Centros Organizadores de Microtúbulos (MTOCs), formando dois centrossomas supernumerários e um único fuso pseudo-bipolar durante a metáfase. KIFC1, uma cinesina essencial para este processo, é particularmente interessante como potencial alvo terapêutico. Isto deve-se ao seu papel na divisão celular de células normais não ser essencial, mas, no entanto, a sua função de agrupamento de centrossomas ser crucial em células cancerígenas com centrossomas supernumerários.

O objetivo geral deste projeto foi determinar a estrutura do domínio catalítico, também conhecido como domínio motor, de KIFC1 para futuro desenho de fármacos com base em estrutura.

Neste projeto, estabeleci um protocolo para expressão e purificação do domínio motor de KIFC1. O protocolo permite obter grandes quantidades de proteína com altos níveis de pureza. Apesar de ter obtido cristais inicialmente, a sua reprodução foi difícil e, portanto, cristais de alta-resolução do domínio motor de KIFC1 provaram ser difíceis de obter. Como tal, a procura por condições de cristalização que permitam a formação de cristais do domínio motor de KIFC1 de alta resolução e reprodutíveis continua. Uma vez esta barreira ultrapassada, será possível usar a estrutura para futuro desenho de fármacos com base em estrutura.

Palavras-chave: KIFC1, Cinesina, Cancro, Estrutura, Cristalografia de Raio-X, Desenho de Fármacos

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1-HB	One-Head Bound
2-HB	Two-Heads Bound
CV	Column Volume
C-terminal	Carboxyl-terminal
dsDNA	Double-Stranded DNA
EMT	Epithelial-Mesenchymal Transition
ER	**E**ndoplasmatic Reticulum
IFT	Intraflagellar Transport
K_{av}	Partition coefficient
KIFC1-MD	KIFC1 motor domain
MCS	Multiple Cloning-Site
MT(s)	Microtubule(s)
Ni^{2+}-NTA	Ni(II)-Nitrilotriacetic Acid
NL	Neck-Linker
N-terminal	Amino-terminal
NSCLC	Non-Small-Cell Lung Cancer
pDNA	plasmid DNA
rpm	rotations per minute
RT	Room Temperature
TGN	Trans-Golgi Network

List of Acronyms

βMe	β-Mercaptoethanol
ACES	N-(2-Acetamino)-2-Aminoethanesulfonic Acid
ADP	Adenosine Diphosphate
ATP	Adenosine Triphosphate
CENP-E	Centromere-associated Protein E
DNA	Deoxy-Ribonucleic Acid
DTT	Dithiotheitol
E. coli	*Escherichia coli*
EDTA	Ethylene-Diamine-Tetra-Acetic Acid
HEPES	2-[4-(2-Hydroxyethyl)piperazin-1yl] Ethanesulfonic Acid
His	Poly-histidine
IPTG	Isopyl-beta-D-thyogalactopyranoside
K-EGTA	K-Egtazic Acid
Kid	Kinesin-like DNA Binding Protein
KIFS	Kinesin Superfamily Proteins
KHC	Kinesin-1 Heavy Chain
KSP	Kinesin Spindle Protein
LB	Luria-Bertani
LDH	Lactate Dehydrogenase
MCAK	Mitotic Centromere-Associated Kinesin
MES	2-(N-Morpholino) Ethanesulfonic Acid
MKLP-1	Mitotic Kinesin-Like Protein 1
MME	Monomethyl Ether
MMT	Methylcyclopentadienyl Manganese Tricarbonyl
MOPS	(3-N-Morpholino) Propanesulfonic Acid
mRNA	Messenger RNA
MTOC(s)	Microtubule organizing centre(s)
NADH	Nicotin Amide Adenine Dinucleotide (Reduced)
NCBI	National Center for Biotechnology Information
Ncd	Non-claret disjunctional
NLS	Nuclear Localization Signal
OD	Optical Density
P-loop	Phosphate-Binding Loop
PAGE	Polyacrylamide Gel Electrophoresis
PDB	Protein Data Bank
PEG(s)	Polyethylene Glycol(s)
PEP	Phosphoenolpyruvate
PH	Pleckstrin Homology

PK	Pyruvate Kinase
PMSF	Phenyl-Methane Sulfuryl Fluoride
RNA	Ribonucleic Acid
SDS	Sodium Dodecyl Sulphate
SUMO	Small-Ubiquitin-Related Modifier
TAE	TRIS-Acetate-EDTA
TB	Terrific Broth
TLRR	Testis Leucine-rich Repeat Protein
TRIS	Tris(hydroxy-methyl)aminoethane
tRNA	Transfer RNA
UV	Ultraviolet
VSVG	Vesicular Stomatitis Virus G Protein

1. Introduction

Normal cellular function and morphology are dependent of intracellular transport (Miki et al., 2005). Most proteins are actively transported by molecular motors along cytoskeletal filaments to their site of function in the cell (Hirokawa and Tanaka, 2015; Hirokawa et al., 2009).

Three large superfamilies of molecular motors are involved in intracellular transport and work by associating with and moving along cytoskeleton filaments — dyneins, myosins and kinesins (Hirokawa et al., 2009; Kolomeisky, 2013). Dyneins move along microtubule (MT) tracks towards their minus-end and drive motility of cilia and flagella, intracellular transport and are involved in cellular division (Kolomeisky, 2013; Roberts et al., 2013; Schliwa and Woehlke, 2003). They are processive molecular motors meaning that they are able to take hundreds of discrete steps in the cytoskeleton before dissociating from it (Kolomeisky, 2013). Myosins move along actin filaments and are responsible for muscle contraction and short-range transport beneath the plasma membrane (Gross, 2004; Hirokawa et al., 2009). They are non-processive molecular motors since they typically make only one or few steps before detaching from their tracks, working in large groups in the cells to fulfil their functions. On the other hand, Kinesins transport cargo along MTs in a processive manner, being essential for proper cellular function and morphology, and have important roles during cellular division (Kolomeisky, 2013; Miki et al., 2005). They use the chemical energy of adenosine triphosphate (ATP) hydrolysis to drive internal conformational changes that generate motile force to fulfil a diverse set of functions in the cell (Kolomeisky, 2013).

1.1. Kinesin superfamily

In 1985 a motile protein was extruded from the cytoplasm of the giant axon of the squid. The purified protein is now known as kinesin-1 (Kolomeisky, 2013). For years, kinesin-1 was the only motor protein known that moved towards the MT plus-end. However, in the 1990s, many kinesin-related proteins were found by sequence homology relating to kinesin-1 heavy chain, *Drosophila melanogaster*'s KHC. Later, screening for proteins containing the highly conserved motor domain of the kinesin family, revealed a large number of kinesin-related proteins in other organisms, including humans (Kolomeisky, 2013).

Kinesins (KIFs) form a superfamily of more than 650 members identified to date in all eukaryotes (Kozielski, 2015). Currently, there are 15 phylogenetic groups termed kinesin-1 to kinesin-14B (Hirokawa and Tanaka, 2015; Hirokawa et al., 2009) and several ungrouped kinesins. Till date there are as many as ~45 kinesin-like proteins in humans (Kolomeisky, 2013).

KIFs can be grouped into three groups, as per the position of the motor domain in the molecule: N-type kinesins have the motor domain in the amino-terminal (N-terminal) region, M-type kinesins have the motor domain in the middle and C-type kinesins have the motor domain in the carboxyl-terminal (C-terminal) region (Figure 1.1) (Hirokawa and Takemura, 2004; Hirokawa and Tanaka, 2015; Hirokawa et al., 2009). Most KIFs are N-kinesins. This group consists of 11 kinesin sub-families, the major members being the KIF1, KIF3, KIF4, KIF5, KIF13, and KIF17 families. M-type kinesins are composed of only the KIF2 family. C-type kinesins include the KIFC1 and KIFC2/C3 families (Hirokawa and Takemura, 2004).

In general, N-type kinesins and C-type kinesins drive MT plus end- and minus end-directed motilities, respectively, and M-type kinesins depolymerize MTs (Hirokawa et al., 2009; Sablin, 2000).

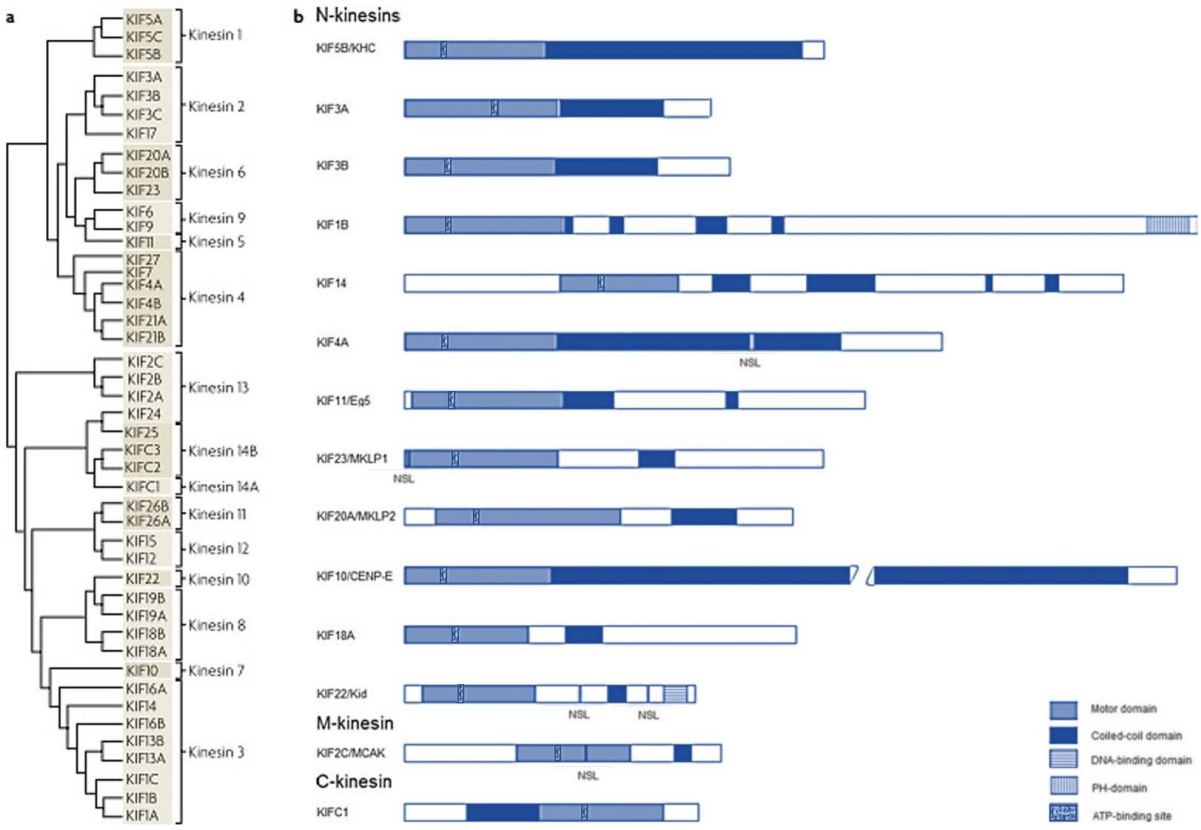

Figure 1.1: The structure and phylogeny of major KIFs. a) A phylogenetic tree of major KIFs, which are classified into 15 families; b) The functional and structural domains of major KIFs. In general, kinesins comprise a motor domain and a coiled-coil domain. Some also contain specific domains or regions, such as the pleckstrin homology (PH) domain of KIF1B, the DNA binding domain of kinesin-like DNA binding protein (Kid), and nuclear localization signals (NLSs) of KIF4A, Mitotic Kinesin-Like Protein 1 (MKLP-1), Kid, and Mitotic Centromere-Associated Kinesin (MCAK). CENP-E refers to Centrosome-associated Protein E. Adapted from Hirokawa and Tanaka, 2015; Yu and Feng, 2010.

1.1.1. Structural characteristics of kinesins

The 'conventional' kinesin (known as KIF5/kinesin-1) is a heterotetrameric protein with two heavy chains (110-120 kilo Daltons (kDa)) and two light chains (60-70 kDa). KIFs can have one of five possible conformations: monomers (KIF1), homodimers, heterodimers, heterotrimers (KIF3 has two distinct heavy chains, KIF3A and KIF3B, and one light chain, KAP3) and heterotetrameric (KIF5) (Seog et al., 2004).

All KIFs contain a motor domain, also referred to as the head, where the biochemical reactions are catalysed, usually attached to a stalk and tail (Endow et al., 2010; Kolomeisky, 2013).

The kinesin motor domain is a functional unit that includes the catalytic core and neck (Sablin, 2000). The catalytic core is a hallmark of KIFs and consists of a ~360 amino acids compact globular domain containing a catalytic pocket for the hydrolysis of ATP and the binding site for MTs. It consists of an eight-stranded β-sheet with three α-helices on each side (Endow et al., 2010; Miki et al., 2005; Sablin, 2000). The nucleotide-binding cleft is located between the β-strand 3 of the central β-sheet and

helix α2 and contains a highly conserved GxxxxGKT/S motif that forms the phosphate-binding loop (P-loop), which binds tightly to the β-phosphate of the nucleotide. Two other motifs, conserved from G-proteins, switch I (α3-L9-α3a) and II (L11, α4-L12-α5), change in conformation during the ATP hydrolysis cycle (Endow et al., 2010; Marx et al., 2009).

Since the motor domain is conserved in all KIFs, motor directionality is determined by the region adjacent to the motor core, the neck (Endow et al., 2010; Sablin, 2000). In contrast to the core, the neck is not similar in all KIFs. In C-type kinesins, the neck is helical and N-terminal to the core. In N-type kinesins, the neck is C-terminal to the core and includes a small β-sheet, known as the neck linker (NL), and an additional coiled coil helix called the neck coiled coil. In M-type kinesins, the neck is helical and N-terminal to the core (Miki et al., 2005; Sablin, 2000).

The α-helical stalk mediates oligomerization. It also coordinates the head, allowing processivity, and, along with the tail, may be involved in regulation of the motor. For example, the kinesin-1 stalk contains a flexible region – the hinge – that lets the tail fold back onto the head, inhibiting MT-binding (Endow et al., 2010; Sablin, 2000).

One of the functions of the tail is to bind to cargo, which is essential for transport, but has been less studied than the head and stalk. The light chains can exist in different isoforms that arise by alternative splicing which permits binding to different vesicles and organelles. Furthermore, adaptor proteins that bind kinesins to specific cargo have also been identified. For example, the adaptor protein Milton mediates kinesin-1 binding to mitochondria, process that is inhibited by Ca^{2+}. Other molecules facilitate kinesin-1 binding to endosomes, vesicles and organelles (Endow et al., 2010).

1.1.2. Motility of kinesins

Kinesin-1 is highly processive motor taking more than hundred steps by a ‘hand-over-hand’ mechanism each time it binds to a MT. Each step is ~8 nanometres and requires the hydrolysis of one ATP molecule (Endow et al., 2010; Schief and Howard, 2001).

On the other hand, Kinesin-14 Non-claret disjunctional (Ncd) (homolog to human HSET/KIFC1) is non-processive meaning it binds to the MT, hydrolyses a single ATP and then detaches (Endow et al., 2010; Schief and Howard, 2001). Multiple Ncd molecules probably act in arrays *in vivo* to functionally crosslink and slide MTs (Endow et al., 2010).

MCAK and other kinesin-13 motors, instead of moving directionally on MTs, bind them and diffuse to the plus- or minus-ends where they promote depolymerisation. The divergent motility mechanisms of KIFs reflects their diverse cellular functions (Endow et al., 2010).

The generic mechanism for the motility of kinesins is based on kinesin-1 family. The ‘apo’ (empty) state, with the motor strongly bound to the MT and no bound nucleotide, is the ‘ground state’ (the most stable state) of the motor. The binding of ATP (in the form of Mg^{2+}ATP) and its conversion to adenosine diphosphate (ADP) destabilizes this state, weakens the attachment to the MT and enables the motor to move to a new binding-site (Cross and McAinsh, 2014). An internal conformational switch prevents motility unless cargo is bound to the tail domain by causing the kinesin to fold up so that the tail binds to the motor domain inhibiting ATP hydrolysis (Schief and Howard, 2001). “Hand-over-hand” stepping involves an alternation between a two–heads-bound (2-HB) state, with both heads attached to

the MT, and a one–head-bound (1-HB) state, where a single head (tethered head) remains free. The processivity is allowed by gating mechanisms that maintain the catalytic cycles of the two heads out of phase (Milic et al., 2014). The NL is a key structural element for this coordination. In the 1-HB state, a conformational change of the NL is induced by nucleotide binding, immobilizing it against the MT-bound catalytic domain (Endow et al., 2010; Milic et al., 2014). This “NL docking” promotes unidirectional motility by biasing the position of the tethered head towards the next MT binding site (Milic et al., 2014). Although NL docking was present in some crystal structures, its role in generating force is controversial because it occurs with a small free energy change. Alternatively, the ‘cover strand’, a structural element at the N-terminus of the motor, was proposed act together with the NL when docking onto the catalytic domain to drive kinesin-1 steps along MTs – hypothesis that has been supported by mutant analysis showing that the cover strand is essential for kinesin-1 motility. Furthermore, mutating a head interacting conserved neck residue causes Ncd to move in either direction of MTs, proving the neck determines directionality (Endow et al., 2010). The completion of a step requires the tethered head to bind the MT, ATP hydrolysis, and trailing head detachment, thereby returning the motor to the ATP-waiting state (Milic et al., 2014). The Mg^{2+} ion of the active site is important for ATPase activity and for establishing this ADP-trapped state, in which Mg^{2+}ADP is stably bound to the active site. Purified KIFs (and most kinesins crystal structures) tend to retain Mg^{2+}ADP in their active sites since, in absence of MTs, the ATPase cycle pauses at this state (Cross and McAinsh, 2014).

The opposite directions of motion of conventional kinesin and Ncd dimers may be explained by the major differences between their structures. The kinesin dimer is asymmetric, with its two heads well separated and rotated by about 120°, while the Ncd dimer structure is symmetric, with the catalytic cores positioned much closer to each other. The tethered heads of kinesin and Ncd with bound heads positioned similarly on MTs, point towards the MT plus-end and minus-end, respectively, indicating that the different symmetries lead them to position the unbound heads in the direction of their movement. This specific positioning of the tethered heads is determined by their distinct neck regions (Sablin, 2000). The kinesin neck consists of a NL followed by a neck coiled coil which are connected by short loops while the Ncd neck forms a parallel coiled coil completely helical (Jana et al., 2012; Sablin, 2000). The different architectures of the necks respond and transmit the nucleotide-binding information differently to the partner heads, enabling movement in opposite directions along the MT. In opposition to the NL-docking that occurs for kinesin, in Ncd, nucleotide binding induces small structural changes that are amplified to the neck and stalk, which rotate, resulting in a power stroke towards the minus-end of MTs (Kull and Endow, 2013; Szczęsna and Kasprzak, 2012).

M-kinesins do not possess traditional motor activity as they function as ATP-dependent MT-destabilizing enzymes. MT depolymerisation may be induced by conformational chances caused by the energy of M-kinesins specific binding and not from ATP hydrolysis, as this is needed for dissociating M-kinesins from the complex with tubulin dimers (Sablin, 2000).

1.1.3. Functions of kinesins

KIFs have diverse functions. Kinesin-1 has a role in transporting vesicles to the presynaptic region of axons that mediate synaptic transmission. It also binds to mitochondria, lysosomes, receptor

and adaptor proteins and signalling-pathway-interacting proteins. As such, it can transport diverse cargos and spatially regulate signal transduction (Endow et al., 2010; Hirokawa, 2011).

Synapse generation and maintenance of synaptic transmission depend highly on molecular motors (Hirokawa, 2011). Kinesin-1 has been implicated in Alzheimer, Huntington and Parkinson disease (Endow et al., 2010). It transports synaptic vesicle precursors and membrane organelles that contain presynaptic plasma membrane proteins, such as syntaxin 1 and SNAP25 (Hirokawa et al., 2009). The kinesin 3 family motors KIF1A and KIF1Bβ also transport synaptic vesicle precursors that contain synaptophysin, synaptotagmin and the small GTPase RAB3A (Endow et al., 2010; Hirokawa, 2011). Another kinesin, KIF17, transports vesicles containing NMDA-type glutamate receptors to dendrites and was proved to have a role in learning and memory (Hirokawa, 2011).

KIFs are also involved in transport inside the cell body of neurons and non-neuronal cells. Specifically, they transport lysosomes, endosomes and cargo from the Golgi to the Endoplasmic Reticulum (ER) and from the Trans-Golgi Network (TGN) to the plasma membrane. For instance, the kinesin-3 family motor KIF13A binds to β1-adaptin, a subunit of the AP-1 complex that is engaged in vesicular transport from the TGN to the plasma membrane. KIF5 also transports a post-Golgi traffic marker, vesicular stomatitis virus G protein (VSVG), towards the plasma membrane (Hirokawa et al., 2009).

Additionally, KIFs regulate transport inside cilia and flagella. For example, kinesin-2 heterotrimeric motors are involved in intraflagellar transport (IFT) needed for formation and maintenance of flagella and cilia, which are essential to many human cells for generation of motility, fluid flow and mechanochemical sensation (Endow et al., 2010; Hirokawa et al., 2009). Furthermore, the heterotrimeric KIF3 complex (KIF3A/KIF3B/KAP3) functions both in IFT of ciliary components and in the exclusion of oncoproteins from the perinuclear region. It prevents the accumulation of the N-cadherin/β-catenin complex in the cytoplasm and nucleus that would enhance canonical Wnt signalling facilitating cell proliferation and reducing cell–cell adhesion by N-cadherin, transforming neuroepithelial cells into tumours (Hirokawa, 2011).

KIFs also have essential functions in both mitosis and meiosis by assembling spindles, separating centrosomes and attaching chromosomes to spindles (Endow et al., 2010; Miki et al., 2005). They produce tension on kinetochore fibres, disassemble kinetochores and depolymerize MTs driving chromosomes to move to the poles during anaphase. Female sterility, cell cycle arrest or cell death are common consequences of mutation of KIFs (Endow et al., 2010).

Finally, KIFs are also involved in MTs dynamics. For example, kinesin-13 motors depolymerize MT and regulate their dynamics, which is essential for cellular function. Specifically, kinesin-13 MCAK contributes to chromosome dynamics, pole-directed MT motion and regulation of spindle positioning and length during mitosis. Furthermore, together with kinesin-13 KIF2A, MCAK is involved in spindle bipolarity (Endow et al., 2010). Kinesin-14 Kar3 has a function in joining nuclei during karyogamy possibly by depolymerizing MTs at their minus-ends. Other kinesins, such as kinesin-14 Ncd, affect spindle MT dynamics by stabilizing rather than destabilizing them (Endow et al., 2010).

1.2. Human KIFC1

Kinesin-14 family members are C-terminal motor proteins, which move towards the minus-end of MTs in an ATP-dependent, non-processive manner (Braun et al., 2013; Xiao and Yang, 2016). In humans, this family comprises only three known elements: KIFC1 (HSET), KIFC2, and KIFC3. Focus has been on KIFC1, which has a function in mitotic spindle organization. Homologues of human KIFC1 include Ncd in *Drosophila sp.*, XCTK2 in *Xenopus sp.* and Kar3 in *Saccharomyces sp.* (Xiao and Yang, 2016).

1.2.1. Structural organisation of human KIFC1

Structurally, kinesin-14 motors, including KIFC1, have a homodimeric conformation and each monomer has a C-terminal conserved kinesin-like motor domain, a central coiled-coil stalk, and an N-terminal tail (Figure 1.2) (Cai et al., 2009; Wu et al., 2013; Yang et al., 2014). The motor domain processes an ATP-dependent MT-binding site and the tail possesses a second MT-binding site that binds MT in an ATP-independent fashion, which enables kinesin-14 family members to cross-link and slide MTs (Cai et al., 2009).

As referred in section 1.1.2., the KIFC1 homologue Ncd dimer structure is symmetric, with the catalytic cores positioned much closer to each other when compared to kinesin-1. Furthermore, also unlike kinesin-1, the Ncd neck is entirely helical forming a parallel coiled coil (Jana et al., 2012; Sablin, 2000).

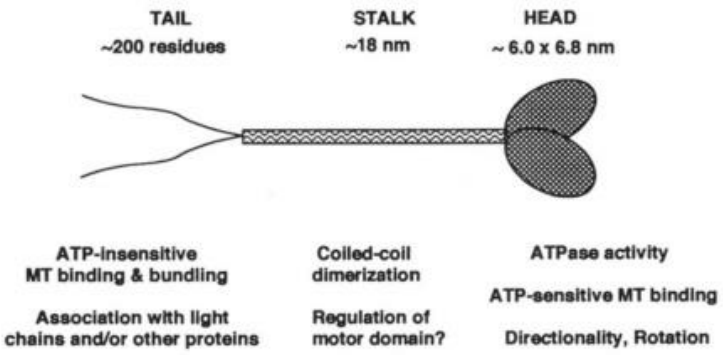

Figure 1.2: Diagrammatic representation of Ncd. Ncd has three distinct domains: an N-terminal tail is followed by a central stalk, which is joined to the globular C-terminal motor domain. The proposed functions of each of the domains are listed. The structure of the tails is based on the predicted amino acid sequence of Ncd. Adapted from Chandra et al., 1993.

1.2.2. Functions of KIFC1

The human kinesin-14 KIFC1, *in vivo*, is predominantly present in MT lattices where it cross-links MTs and, during mitosis, regulates spindle length by sliding MTs relative to each other (Braun et al., 2013).

KIFC1 also has a function in the transport of vesicles and organelles. For example, movement and fission of early endocytic vesicles require co-function of KIF5B and KIFC1. Furthermore, KIFC1 was

found to be co-localized with KIFC2 and dynein during endocytic vesicle transport in human liver cells (Xiao and Yang, 2016).

One of the most important mammalian processes that require KIFC1 is spermatogenesis. It has vital roles in acrosome biogenesis and nucleus deformation. In the first case, KIFC1 helps in Golgi apparatus transport through its vesicle association tail domain. In the second case, with the aid of testis leucine-rich repeat protein (TLRR), KIFC1 coordinates the positioning of the regulatory factors on the nuclear membrane that mediate the deformation of the nucleus and even facilitate acrosomal biogenesis (Xiao and Yang, 2016).

KIFC1 also has a central role in oocyte division and embryonic development (Xiao and Yang, 2016). Specifically, in oocytes, cells devoid of centrosomes, KIFC1 is indispensable for assembling a fusiform bipolar spindle (Pannu et al., 2015). Moreover, KIFC1, along with KIF17 is periodically expressed in the early stage human placenta, rising suspicion of a role in gestation (Xiao and Yang, 2016).

Finally, it has been reported that KIFC1 actively transports and binds double-stranded deoxy-ribonucleic acid (DNA) (dsDNA) (Farina et al., 2013; Pannu et al., 2015; Xiao and Yang, 2016). Gene therapy relies on transportation of exogenous DNA into the nucleus. Yet, whether this occurred by diffusion or via motor transport remained unknown until a study demonstrated that the motility of DNA in HeLa cells *in vitro* dramatically decreased in the absence of KIFC1 (Farina et al., 2013; Xiao and Yang, 2016).

1.2.3. KIFC1 as a novel target for cancer chemotherapy

The centrosome is a cytoplasmic organelle formed by two centrioles surrounded by an amorphous mass of protein – the pericentriolar material. Normal cell division is initiated when centrosomes undergo duplication and the two centrosomes form the poles of a bipolar mitotic spindle, a function that is essential for accurate chromosome segregation (Li et al., 2015; Wu et al., 2013). Tight control of centrosome duplication limits it to one per cell cycle and warrants that normal cells enter mitosis with two centrosomes or MTOCs. Aberrant centrosome duplication leading to more than two centrosomes is referred to as centrosome amplification (Li et al., 2015; Xiao and Yang, 2016). It can arise from numerous cell division errors: centrosome overduplication, *de novo* synthesis of centrosomes, cell fusion or cytokinesis failure (Kwon et al., 2008). A variety of tumour suppressor genes and oncogenes are linked to centrosome amplification and genetic instability levels rise along with centrosome amplification, paving the way for tumorigenesis. As such, centrosome amplification is a hallmark in cancer cells. However, the existence of multiple centrosomes can be lethal (Xiao and Yang, 2016). Multipolar metaphase arrangements can result in cell unviability due to mitotic catastrophe, multipolar cell divisions, or whole chromosome loss or gains caused by merotelic kinetochore attachments (an error in which a single kinetochore is attached to MTs emanating from both spindle poles (Gregan et al., 2011)) (Kwon et al., 2008; Wu et al., 2013). Cancer cells are able to survive this by clustering multiple centrosomes together during prometaphase to form two supernumerary centrosomes leading to a single pseudo-bipolar spindle during metaphase (Kozielski, 2015; Pawar et al., 2014; Xiao and Yang, 2016). Via this mechanism, KIFC1, which is activated by genetic instability

signals such as DNA-damage, is essential for the viability of cancer cells bearing extra centrosomes (Pawar et al., 2014; Xiao and Yang, 2016). This results in lower missegregation levels allowing survival of cancer cells with a certain degree of genetic instability, augmenting the likelihood of mutation and enhancing tumour malignancy (Xiao and Yang, 2016).

A model was proposed to explain KIFC1's role in this process. Centrosome amplification occurs during interphase causing a transient multipolar spindle during prometaphase. This creates chromosome instability. Then, a merotelic kinetochore attachment to MTs forms and begins driving centrosome clustering. Shortly afterwards, syntelic attachments with kinetochores (in which both sister kinetochores interact with MTs that emanate from the same spindle pole (Gregan et al., 2011)) further promote formation of the pseudo-bipolar spindles and centrosome clustering. During this process, KIFC1, which binds to plus ends of the MTs, crosslinks and then slides along the antiparallel MTs while it moves towards the spindle pole, creating MT locking forces that cause centrosomes to cluster together. Increased expression of some cyclins causes an anaphase delay giving KIFC1 the time required to transform the transient multipolar spindle into a bipolar spindle, whilst shortening the whole cell cycle (Xiao and Yang, 2016). Furthermore, KIFC1 protects cancer cells surviving signals, further enhancing malignancy – it is capable of binding to survivin, inhibiting its poly-ubiquitination and subsequent degradation, therefore, protecting the cells from apoptosis. At the point of mitotic exit, KIFC1 is degraded (Mittal et al., 2016; Xiao and Yang, 2016).

KIFC1 is abundantly expressed in cancer cells of the ovary, breast, bladder, lung, kidney and other cancers (Mittal et al., 2016; Pannu et al., 2015; Pawar et al., 2014; Xiao and Yang, 2016). Particularly, in non-small-cell lung cancer (NSCLC), KIFC1 is an indicator of brain metastasis by real-time quantitative reverse transcriptase PCR screening analysis (Pannu et al., 2015; Pawar et al., 2014; Xiao and Yang, 2016). Likewise, KIFC1 is also reported as a candidate for a prognostic and metastases onset biomarker in ovarian cancers. Details on the role of KIFC1 in metastasis are still unknown but suspicion is that its function in the survival of multi-centrosome cells enhances cancer cell polarity, hence powering the epithelial-mesenchymal-transition (EMT) and cell motility (Mittal et al., 2016; Pawar et al., 2014; Xiao and Yang, 2016).

Moreover, KIFC1 is among the factors relating to drug resistance in breast and prostate cancer. Docetaxel, along with paclitaxel, bind to β-tubulin to prevent the depolymerisation of MTs disrupting their proper dynamics (Kozielski, 2015; Xiao and Yang, 2016). KIFC1, along with three other KIFs, KIFC3, KIF1A, and KIF5A, were found to be overexpressed in docetaxel resistant breast cancer cell lines, indicating that the binding of KIFs to MTs opposes the stabilizing effect of the drug. This suggests that a combination therapy of kinesin inhibitors and taxol could potentially overcome resistance (Kozielski, 2015; Pannu et al., 2015; Xiao and Yang, 2016). Consistently, inhibiting both KIFC1 and MCAK has been seen to increase prostate cancer cells' sensitivity to taxane (Xiao and Yang, 2016).

The mitotic spindle is a validated target in chemotherapy. Anti-mitotic agents that target tubulin, such as taxanes and vinca-alkaloids, are clinically successful agents against certain types of cancer. Nevertheless, there are limitations to these drugs, like innate or acquired resistance and dose-limiting toxicities. Consequently, another strategy that targets kinesins on the MTs has emerged and has achieved some progress so far (Kozielski, 2015; Xiao and Yang, 2016). Inhibitors targeting Eg5 (kinesin-

5 family) and CENP-E (kinesin-7 family) have proceeded into clinical trials and positive reports have been published (Xiao and Yang, 2016). Now, the focus in this area falls upon KIFC1 due to the fact that it is indispensable for survival of cancer cells regardless of normal or supernumerary centrosome number (Kozielski, 2015; Li et al., 2015; Xiao and Yang, 2016). However, the other roles that KIFC1 plays during vesicular and organelle trafficking, spermatogenesis, oocyte development and dsDNA transportation, raise concerns about using KIFC1 inhibitors medically. These concerns could be disregarded since KIFC1 is actually dispensable in ordinary somatic cells which have only one pair of non-supernumerary centrosomes, as KIFC1 and NuMA are thought to be redundant in MT minus-end organization (Xiao and Yang, 2016). It is also reported that the viability of non-multiple centrosome MCF-7 cell lines is not significantly influenced by depletion of KIFC1 (Kwon et al., 2008; Xiao and Yang, 2016; Yang et al., 2014). As such, KIFC1 is a particularly interesting therapeutic target because normal cells division does not require it and KIFs are amenable to inhibition by small molecules (Kwon et al., 2008).

To this point, three small-molecule KIFC1 inhibitors were found (Xiao and Yang, 2016). Two directly inhibit KIFC1, AZ82 and CW069, causing centrosome de-clustering in cancer cells with amplified centrosomes (Pawar et al., 2014; Xiao and Yang, 2016). AZ82 binds to the KIFC1-MT complex, inhibiting the binding of ATP and release of ADP. When cancer cells with extra centrosomes are treated with AZ82, fatal multipole spindles appear (Wu et al., 2013; Xiao and Yang, 2016; Yang et al., 2014). In addition, a study showed AZ82 yielded high and sustained exposure in mouse after intraperitoneal injection, being, therefore, suitable for further *in vivo* studies (Yang et al., 2014). However, an overdose of AZ82 can lower the selective efficiency of the drug. CW069 is a highly selective small-molecule KIFC1 inhibitor with an affinity to the loop-5 cleft of the motor domain, which disrupts KIFC1's motility. CW069 proved lethal to breast cancer cell lines while the spindle shape of normal dermal fibroblast cells was not significantly altered. In addition, the specificity of CW069 is sufficient to avoid any mitotic phenotype that would occur upon the inhibition of kinesin spindle protein (KSP) even though it has up to 80% similarity with KIFC1. This makes its action more predictable making it a desirable clinical drug candidate (Xiao and Yang, 2016). Another inhibitor, PJ34, may also be a promising option. KIFC1's messenger ribonucleic acid (RNA) (mRNA) level is shown to be significantly reduced in various PJ34 treated breast cancer cell lines, suggesting PJ34 transcriptionally suppresses expression of KIFC1 (Li et al., 2015; Xiao and Yang, 2016).

1.3. Aims of the project

KIFC1 is considered as a potential and novel target for drug development in cancer chemotherapy. The overall goal of this project was to determine the structure of the KIFC1 motor domain for future structure-based drug-design.

The first aim was to clone, express and purify the human KIFC1 motor domain (KIFC1-MD) in large amounts and high purity.

The second goal of the project was to set up crystallisation trials for KIFC1-MD and if crystals were obtained, to optimize the crystallisation conditions to attain well diffracting crystals. This will lead to high-resolution structure of KIFC1 to be used in future structure-based drug design.

2. Materials and Methods

2.1. Materials

2.1.1. Consumables

The pDHT-2 vector, ppSUMO-2 vector and pDHT-KIFC1-MD plasmid were provided by GenScript® (for details on vectors and predicted KIFC1-MD amino acid sequence see Attachments sections 7.1 and 7.2, respectively). *Escherichia coli* (*E. coli*) DH5α, BL21 (DE3), Rosetta™ BL21 (DE3) and Rosetta™ BL21 (DE3) pLysS cells were bought from Novagen. *E. coli* BL21 Codon Plus (DE3) RIPL Cells were obtained from Agilent Technologies (for details on *E. coli* strains see section 7.3). The following kits were obtained from Qiagen®: Hi-Speed Plasmid Midi Kit, Gel Extraction Kit, MiniPrep Kit and Ni(II)-Nitrilotriacetic Acid (Ni^{2+}-NTA) columns. CutSmart Buffer (10x concentrated), Restriction Enzymes (*XhoI* at 20.000 U/mL and *NcoI* at 10.000 U/mL), 1kb DNA ladder (500 µL/mL), Purple Gel Loading Dye (6x concentrated) and Pre-Stained Protein ladder (11-245 kDa) were bought from New England Biolabs® Inc. SyBrSafe Dye and Protein Loading Dye (4x concentrated) were obtained from Invitrogen™. DNA Rapid Ligation Kit was obtained from Roche Diagnostics, Bugbuster Reagent was bought from Milipore Corp., Instant Blue™ was obtained from Expedeon Ltd. and 1x Bradford Reagent was bought from BIO-RAD®. SnakeSkin® Dialysis Tubing 10 kDa was obtained from Thermo Scientific. His-trap™ FF Crude 5 mL, 96 clear wells Microplates, Masterblock 96 well 2 mL and 96 wells microplate µClear® were bought from Greiner Bio-One. Amicon® Ultra-15 Centrifugal Filter Ultracel® 3 kDa was obtained from Milipore Corp. ASNA Screen, RdRP Screen and PEG ION Screen were provided by Dr. Sandeep Talapatra. MORPHEUS® Screen, JCSG® Screen, PACT® Screen and PGA® Screen were bought from Molecular Dimensions. MRC 3 drop plates (96 wells), ClearVue™ Sheets, MORPHEUS® 1 M Buffer System 2 pH 7.5 (1 M Sodium 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) and 1 M 3-(N-morpholino)propanesulfonic acid (MOPS)), MORPHEUS® Ethylene Glycols Mix (0.3 M di-ethylene glycol, 0.3 M tri-ethylene glycol, 0.3 M tetra-ethylene glycol, and 0.3 M penta-ethylene glycol), 24 Well Linbro Plates and Microbridges were also obtained from Molecular Dimensions. 22 mm circular coverslips were bought from Jena Bioscience©. PD10 3 mL Desalting column was bought from GE Healthcare Life Sciences. Ulp1 (5 mg/mL), Mg^{2+}ATP (100 mM), Phosphoenolpyruvate (PEP) (200 mM), reduced Nicotinamide adenine dinucleotide (NADH) (70 mM), Pyruvate kinase (PK) (10 mg/mL) and Lactate dehydrogenase (LDH) (5 mg/mL) were all provided by Sigma-Aldrich®. Additional materials can be found in Attachments section 7.4.

2.1.2. Equipment

U:GENIUS3 Gel imaging system (Syngene Ltd.), Soniprep 150 (Richmon Scientific Ltd.), Avanti® J-E Centrifuge with a 25.5 JA rotor (Beckman Coulter Inc.), ÄKTA™ start machine (GE Healthcare Life Sciences), HiLoad™ 16/600 Superdex 200 prep grade column (GE Healthcare Life Sciences), XK 16/100 column (GE Healthcare Life Sciences), Spectrophotometer Ultrospec 2000

(Sinteck Instruments), Mosquito® LCP Robot (Labtech Ltd), Absorbance Microplate Reader (TECAN). Other equipment included: Agarose gel Tank and Power Supply, Sodium Dodecyl Sulfate (SDS)-Polyacrylamide Gel Electrophoresis (PAGE) Gel Electrode Chamber, Tank and Power Source, Sonicator (Model CL-18) and Variable Flow Mini-peristaltic Pump, microscope, protein crystals collection loops, loop holding packs and Synchrotron.

2.1.3. Software packages

The following software packages were utilized:

Hampton Make Tray Website (https://hamptonresearch.com/make_tray.aspx)

ImageJ, which can be downloaded from https://imagej.nih.gov/ij/download.html

KaleidaGraph, which can be downloaded from http://www.synergy.com/wordpress_650164087/kaleidagraph/free/

BLAST® tool in National Center for Biotechnology Information (NCBI) Website (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome)

2.1.4. List of buffers

The following buffers were used:

Tris(Hydroxy-Methyl)Amino-Methane (TRIS) – Acetate – Ethylene-Diamine-Tetra-Acetic acid (EDTA) (TAE) Buffer 1x: 40 mM TRIS (pH 7.6), 20 mM Acetic acid, 1 mM EDTA

TRIS-Glycine-SDS Buffer 1x: in 1 L there is: 3 g TRIS, 14.4 g Glycine and 1 g SDS

Small-scale purification buffers:

Lysis Buffer/Equilibrium Buffer: 50 mM TRIS, pH 8.0, 500 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole

Wash Buffer: 50 mM TRIS, pH 8.0, 500 mM NaH_2PO_4, 300 mM NaCl, 25 mM Imidazole

Elution Buffer: 50 mM TRIS, pH 8.0, 500 mM NaH_2PO_4, 300 mM NaCl, 300 mM Imidazole

Large-scale purification buffers:

Lysis Buffer: 50 mM TRIS pH 7.5, 300 mM NaCl, 10 mM Imidazole, 10 % Glycerol

Wash Buffer: 50 mM TRIS pH 7.5, 300 mM NaCl, 1 mM $MgCl_2$, 30 mM Imidazole

Elution Buffer: 20 mM HEPES pH 7.5, 500 mM NaCl, 0.01% Tween-20, 250 mM Imidazole

Dialysis Buffer: 50 mM TRIS pH 7.6, 150 mM NaCl, 1 mM $MgCl_2$, 5% Glycerol

Gel Filtration Buffer: 20 mM HEPES pH 7.6, 500 mM NaCl, 0.01% Tween-20

Desalting Buffer: 20 mM HEPES, pH 7.6, 250 mM NaCl, 0.01% Tween-20

A25 Buffer for ATPase assay: 25 mM N-(2-Acetamido)-2-Amino-Ethane-Sulfonic acid (ACES)-KOH pH 6.9, 2 mM $MgAc_2$, 2 mM K-Egtazic acid (K-EGTA), 0.1 mM EDTA and 1 mM β-Mercapto-Ethanol (βMe)

2.2. Methods

2.2.1. Transformation for obtaining plasmid DNA (pDNA) and protein expression

Respective *E. coli* competent cells were thawed on ice for 10 min. 1 µL of each pDNA was added separately to 100 µL of the respective competent cells (Table 2.1) and left for 30 min on ice (for details on vectors and *E. coli* strains used see Attachments sections 7.1 and 7.3). The cells were heat-shocked at 42ºC for 45 s and immediately transferred to ice to recover for 2 min. 200 µL of pre-warmed (37ºC) Luria-Bertani (LB) medium were added to each of the tubes. The cells were incubated for 1 h at 37ºC and 700 rotations per min (rpm). The cells were pipetted onto Kanamycin supplemented LB agar plates, spread using a sterile spreader and incubated overnight at 37ºC to obtain colonies.

Table 2.1: Summary of DNA transformed into each *E. coli* strain and respective purpose.

***E. coli* Competent cells**	**DNA transformed into the cells**	**Objective**
E. coli DH5α cells	ppSUMO-2 and pDHT-2 vectors	DNA replication
	Ligated DNA from section 2.2.4	DNA replication
	Purified DNA from section 2.2.5	DNA replication
E. coli BL21 (DE3) cells	ppSUMO-2-KIFC1-MD and pDHT-2-KIFC1-MD constructs (obtained in section 2.2.5)	Protein expression and purification
E. coli Rosetta™ BL21 (DE3) cells	ppSUMO-2-KIFC1-MD and pDHT-2-KIFC1-MD constructs (obtained in section 2.2.5)	Protein expression and purification
E. coli Rosetta™ BL21 (DE3) pLysS cells	ppSUMO-2-KIFC1-MD and pDHT-2-KIFC1-MD constructs (obtained in section 2.2.5)	Protein expression and purification
E. coli BL21 Codon Plus (DE3) RIPL cells	ppSUMO-2-KIFC1-MD and pDHT-2-KIFC1-MD constructs (obtained in section 2.2.5)	Protein expression and purification

2.2.2. Midi-prep protocol for pDNA purification

A single colony from the agar plate was transferred into 200 mL of LB medium containing 200 µL of Kanamycin and incubated overnight at 37ºC at 200 rpm. The cells were harvested by centrifugation at 4000 rpm for 10 min at 4ºC. The supernatant was discarded and the pellet was used for further pDNA purification. The rest of the protocol was performed with the Hi-Speed Plasmid Purification Midi Kit. The bacterial pellets were re-suspended in 6 mL of buffer P1, each. 6 mL of buffer P2 were added and the solution mixed by inversion. The tubes were incubated at Room Temperature (RT) for 5 min. 6 mL of chilled buffer P3 were added to the lysate and the solution was mixed by inversion. The lysate was poured into the barrel of the QIAfilter Cartridge and incubated at RT for 10 min. A Hi-Speed Midi Tip was equilibrated by applying 4 mL of buffer QBT and allowing the column to empty by gravity flow. The cap from the Cartridge outlet nozzle was removed, the plunger inserted and the cell lysate filtered into the previously equilibrated Hi-Speed Tip and allowed to enter the resin by gravity flow. The Hi-Speed Midi Tip was washed by applying 20 mL of buffer QC. A new sterile falcon 50 mL tube was prepared

with 3.5 mL of isopropanol. The pDNA was eluted with 5 mL of buffer QF into this falcon, mixed and incubated at RT for 5 min. The plunger from a 20 mL syringe was removed and the QIAprecipitator Midi Module attached onto the outlet nozzle. The eluate/isopropanol mixture was transferred into the 20 mL syringe and the plunger inserted to filter the eluate/isopropanol mixture using constant pressure. The QIAprecipitator was removed from the syringe and the plunger pulled out. The QIAprecipitator was re-attached and 2 mL of 70% ethanol were added to the syringe. The pDNA was washed by inserting the plunger and pressing the ethanol through using constant pressure. The QIAprecipitator was removed from the 20 mL syringe and the plunger pulled out. The QIAprecipitator was re-attached, the plunger inserted, and the membrane dried by pressing air through the QIAprecipitator quickly and forcefully, four times. The outlet nozzle of the QIAprecipitator was dried with absorbent paper to prevent ethanol carryover. The plunger was removed from a new 5 mL syringe and the QIAprecipitator attached onto the outlet nozzle. 1 mL of buffer TE was added to the 5 mL syringe, the plunger was inserted and the pDNA eluted into a microtube, using constant pressure. The eluate was transferred back into the 5 mL syringe and eluted for a second time into the same 1.5 mL microtube. The pDNA was stored at -20ºC.

2.2.3. Restriction and purification of the restricted DNA

45 µL of each pDNA (vectors ppSUMO-2 and pDHT-2 and pDHT-KIFC1-MD plasmid), 5.5 µL of CutSmart buffer and 2 µL of each of the *NcoI* and *XhoI* restriction enzymes were mixed in microtubes and incubated for 2 h at 37°C. 5 µL of 6x sample loading dye were added to each of the restriction reactions. Each sample/sample loading buffer mixture, as well as 15 µL of 1 kb DNA ladder, were pipetted into separate wells in a 1% agarose gel stained with SyBrSafe Dye and covered with TAE buffer 1x. The gel was run at 200 V and 125 mA. The gel was observed under Ultraviolet (UV) light using the U:GENIUS3 Gel imaging System to visualise DNA bands. Then, the desired bands were cut from the gel to extract the restricted DNA. The rest of the protocol was performed with the Gel Extraction Kit. 550 µL of QG buffer were added to each cut from the gel placed in microtubes and were incubated at 65ºC until the gel dissolved. The solutions were transferred to the purification columns and centrifuged at 10000 rpm for 2 min. 700 µL of PE buffer were added and the columns were centrifuged at 10000 rpm for 1 min. The columns were dried by spinning-down at 10000 rpm for 2 min. To elute the DNA, 50 µL of EB buffer were added and incubated for 1 min after which the columns were centrifuged at 10000 rpm for 1 min. The results from the agarose gel were analysed with the ImageJ software.

2.2.4. Ligation and transformation of restricted DNA

The protocol was performed with the DNA Rapid Ligation Kit. Four different ratios of μL_{vector}:μL_{insert} were prepared: 1:1, 1:10, 1:20 and 10:1. 2 µL of buffer 2, 10 µL of buffer 1 and 1 µL of T4 DNA Ligase were added to all and mixed thoroughly. Distilled water was added to make up 25 µL total volume in each solution and the solutions were mixed and incubated for 2 h at RT. Thereafter the ligated pDNA was transformed into *E. coli* DH5α cells using the protocol described in the section 2.2.1., the only difference being the amount of pDNA added to each 100 µL of cells was 10 µL of the ligations product instead of 1 µL described before.

2.2.5. Purification of ligated pDNA using the Mini-Prep Kit

Only the ratios 1:10 and 1:20 for both vectors resulted in colony growth. As such, 4 colonies of each of these were used for this protocol. The colonies were grown in suspension in 5 mL of LB medium supplemented with 5 µL of Kanamycin at 37ºC at 200 rpm overnight. The cultures which grew after overnight incubation were then spun-down for 10 min at 4000 rpm at 4°C and the supernatant discarded. The Qiagen® MiniPrep Kit was used to purify the pDNA. The bacterial pellets were re-suspended in 250 µL of cold buffer P1 and transferred into microtubes. 250 µL of buffer P2 was then added and mixed by inversion. Thereafter, 350 µL of buffer N3 were added and the solution was mixed by inversion. The tubes were centrifuged for 10 min at 10000 rpm. The supernatants were applied to QIAprep spin columns that were centrifuged for 60 s at 10000 rpm. The QIAprep spin columns were washed by adding 500 µL of buffer PB and centrifuging for 60 s and then adding 750 µL of buffer PE and centrifuging for 60 s. The empty QIAprep spin columns were centrifuged at 10000 rpm for an additional 2 min to remove residual wash buffer. To elute the pDNA, 50 µL of buffer EB were added to each QIAprep spin column placed in clean microtubes which were incubated for 1 min and then centrifuged for 1 min.

2.2.6. Diagnostic restriction, transformation for protein expression and purification and Midi-prep protocol for pDNA purification

To 15 µL of each purified pDNA, 2 µL of CutSmart buffer, 0.3 µL of each of the *NcoI* and *XhoI* restriction enzymes and 7.4 µL of distilled water were added. The mixture was incubated for 1 h at 37°C. The samples were then run on a 1% agarose gel as described in section 2.2.3. The gel image was taken as described in section 2.2.3 to find the positive colonies where the ligation had worked. It was also determined qualitatively which sample had yielded the larger amount of pDNA. The mini-prep samples with maximum amount of positive pDNA were selected for transformation: ppSUMO-2-KIFC1-MD 1:20 colony 3 and pDHT-2-KIFC1-MD 1:10 colony 2 (section 3.2). Thereafter, the purified pDNA was transformed into *E. coli* competent cells using the protocol described in the section 2.2.1. The Midi-Prep Protocol described in section 2.2.2. was used to purify the pDNA from the *E. coli* DH5α cells.

2.2.7. Small-scale protein expression and purification

The colonies from transformation into different expression cells were transferred into 40 mL of Terrific Broth (TB) medium supplemented with 40 µL of the appropriate antibiotics (Table 2.2.) and incubated overnight at 37ºC at 200 rpm. Next morning, 40 µL of 0.5 M Isopyl-beta-D-thyogalactopyranoside (IPTG) were added to the cultures and incubated at 20ºC at 200 rpm, overnight. Two flasks broke during the incubation period: *E. coli* BL21 Codon Plus (DE3) RIPL cells with pDHT-2-KIFC1-MD and *E. coli* Rosetta™ BL21 pLysS cells with ppSUMO-2-KIFC1-MD. The remaining cultures were centrifuged at 4000 rpm for 10 min at 4ºC the next day to obtain the bacterial pellets. 250 µL of Bugbuster reagent, 250 µL of small-scale Equilibrium buffer/Lysis buffer (see section 2.1.4.) and 15 µL DNaseI (2 mg/mL) were added to the cell pellets, mixed thoroughly and transferred into microtubes. Sonication was performed for 5 rounds in Sonicator Model CL-18, each consisting of 15 s ON/OFF cycle. The sonicated cells were then centrifuged at 13000 rpm at 4ºC. At the same time, the Qiagen®

Ni^{2+}-NTA columns were equilibrated with 500 µL of small-scale Equilibrium buffer/Lysis buffer and centrifuged at 4500 rpm for 1 min. The samples were then transferred to the Ni^{2+}-NTA columns, incubated on ice for 5 min and then centrifuged at 3000 rpm at 4ºC for 2 min. The Ni^{2+}-NTA columns were washed three times with 500 µL of small-scale Wash buffer (see section 2.1.4.) by centrifuging at 3000 rpm at 4ºC for 2 min. The protein samples were eluted into new microtubes by adding 150 µL of small-scale Elution buffer (see section 2.1.4.), incubating on ice for 2 min and then centrifuging at 3000 rpm for 2 min at 4ºC. The protein elutes were transferred back to the same Ni^{2+}-NTA columns and incubated again on ice for 2 min, then centrifuged again. This step was repeated four times in total. 5 µL of 4x Loading Dye were added to 25 µL of the protein samples obtained above and the mixture, along with 7 µL of Pre-Stained Protein Ladder (11-245 kDa), were loaded on a 12% SDS-PAGE gel covered in TRIS-Glycine-SDS buffer 1x. The gel was run at 180 V and 125 mA for 45 min to 1 h. The gel was stained with Instant Blue™. To de-stain the gel, it was washed with distilled water a couple of times and then placed on the shaker overnight. The gel was analysed with the ImageJ software. The cells with maximum protein expression as observed on the gel (qualitatively) (*E. coli* BL21 (DE3) and *E. coli* BL21 Codon Plus (DE3) RIPL with ppSUMO-2-KIFC1-MD, section 3.3.1) were subjected to cleaving of the Poly-Histidine (His) and Small-Ubiquitin-Related Modifier (SUMO) tags. For this, to 40 µL of each of the protein solutions, 8 µL of Ulp1 protease (5 mg/mL) were added. The samples were incubated overnight at 4ºC. 5 µL of 4x Loading Dye were added to the uncleaved and cleaved protein samples and to 8 µL of Ulp1 sample and, along with 7 µL of Pre-Stained Protein ladder (11-245 kDa), were loaded on a 10% SDS-PAGE gel. The gel was run, stained, distained and analysed as described before in this section. The cells with maximum protein expression as observed qualitatively on the gels were selected for large scale expression (sections 3.3.1 and 3.3.2).

Table 2.2: Antibiotic supplements provided for each *E. coli* strain. For details on the *E. coli* strains used see Attachments section 7.3.

***E. coli* cells**	**Antibiotic Supplement**
E. coli BL21 (DE3) cells	Kanamycin (50 mg/mL)
E. coli Rosetta™ BL21 and Rosetta™ BL21 pLysS cells	Kanamycin (50 mg/mL) and Chloramphenicol (34 mg/mL)
E. coli BL21 Codon Plus (DE3) RIPL cells	Kanamycin (50 mg/mL), Streptomycin (50 mg/mL) and Chloramphenicol (34 mg/ mL)

2.2.8. Large-scale protein expression and purification

2.2.8.1. Large-scale KIFC1-MD expression

A single colony from the transformation plates of *E. coli* BL21 (DE3) cells was incubated in 40 mL of TB medium supplemented with 40 µL of Kanamycin at 37ºC at 200 rpm overnight. 2 mL of the grown culture were transferred to each of six flasks with 1 L TB medium supplemented with 1 mL of Kanamycin and incubated overnight at 37ºC at 200 rpm. 1 mL of 0.5 M IPTG was added to induce the protein expression next morning and the cultures were incubated at 20ºC at 200 rpm for 24 h. The cultures were centrifuged for 20 min at 4000 rpm at 4ºC and the supernatant discarded. The final pellets

were collected by re-suspending in 100 mL of large-scale Lysis buffer (see section 2.1.4.), snap-frozen in liquid nitrogen and stored at -80ºC for subsequent purification.

2.2.8.2. First purification step by nickel-affinity column

The frozen pellet was thawed in warm water and then placed on ice. 1 mM of Phenyl-Methane Sulfonyl Fluoride (PMSF) was added to the mixture of pellet and Lysis buffer, which was then sonicated with the Soniprep 150, for 10 rounds, with cycles of 30 s ON and 45 s OFF. The cells were centrifuged for 1 h 15 min at 4ºC at 20000 rpm in an Avanti® J-E Centrifuge with a 25.5 JA rotor. The supernatant from the centrifugation was loaded into the His-trap™ FF Crude 5 mL column pre-equilibrated with 50 mL of Lysis buffer (see section 2.1.4.) and the flow-through was collected. The column was thereafter washed with large-scale Wash buffer (see section 2.1.4.) for around 50 Column Volume (CV) and the flow-through collected into a bottle. The protein was eluted with large-scale Elution buffer (see section 2.1.4.) in 2 mL fractions. The protein elution was qualitatively checked with Bradford reagent by mixing 10 µL of eluted sample from each well with 100 µL Bradford. To check for the protein from this affinity purification, samples were run on a 12% SDS-PAGE gel as described in the section 2.2.7. Additionally, the percentage of purification of the target protein was determined with recourse of ImageJ software.

To perform the next step of cleavage of the fusion tag, the amount of protein obtained was measured using Braford reagent and the following formula:

$\frac{mg}{mL} = \frac{OD_{595}}{\mu L\ of\ protein\ used} x16$, where OD_{595} corresponds to the optical density measured at 595λ.

Ulp1 protease was used in the ratio of 1 mg for each 50 mg of protein. 3 mM Dithiothreitol (DTT) were also added to the cleavage solution. At the same time of protein cleavage, the protein was dialysed overnight against 1 L of Dialysis buffer (see section 2.1.4.) using a 10 kDa SnakeSkin® Dialysis membrane, at 4ºC, overnight, to remove excess imidazole for subsequent purification step.

2.2.8.3. Second purification step by nickel-affinity column

The dialysed protein was loaded into the pre-equilibrated His-trap™ FF Crude 5 mL column followed by 20 CV wash and subsequent elution as mentioned above in section 2.2.8.2. All of the different fractions were collected separately and the protein concentration determined using Bradford reagent. To verify for protein cleavage success and protein purity level, samples were run on a 12% SDS-PAGE gel. The gel was run, stained, distained and analysed as described in section 2.2.8.2.

2.2.8.4. Final purification step by gel filtration

The HiLoad™ 16/600 Superdex 200 prep grade column was equilibrated with the Gel Filtration buffer (see section 2.1.4.) at 1 mL/min flow rate, for 140 mL (1.2 CV), at 4ºC. Since it is only possible to load 2 mL into the injection loop in the ÄKTA start machine, the protein was concentrated to 2 mL using a 3 kDa Amicon® Ultra-15 Centrifugal Filter. The protein sample was injected into the loop using a 2.5 mL syringe. The protein was loaded into the column. The elution from the gel filtration column was collected into tubes in a fraction collector attached to the AKTA machine at 4ºC (a scheme representing the gel filtration process is presented in Figure 2.1). The peak fractions were run on a 12% SDS-PAGE

gel to check for the quality of the protein. The protein fractions with >90% purity were concentrated to ~16 mg/mL and supplemented with 2 mM ADP and 20 mM $MgCl_2$. The protein samples were then snap frozen in liquid nitrogen in 50 µL aliquots for future crystallisation trials.

The large-scale expression/purification protocol was repeated as requirement for new protein emerged. The following differences occurred: 1) due to equipment failure, in batches Nº 2-3, gel filtration was made in a column with lower resolution for separation (XK 16/100); 2) in batches Nº 2-3, the protein was not supplemented with ADP or $MgCl_2$; 3) in batches Nº 2-5 the elution fractions volumes changed to 5 mL. The results of KIFC1-MD batches Nº 2-5 can be seen in Attachments sections 7.5-7.8 and the calibration curves for each gel filtration column used can be found in Attachments sections 7.9-7.10.

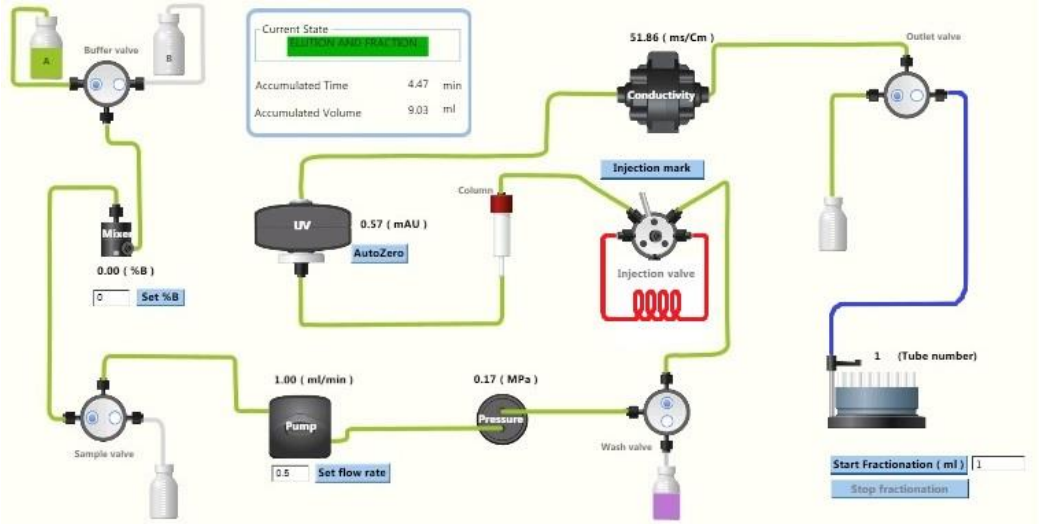

Figure 2.1: Scheme representing the process of gel filtration. The equilibration process is represented in green. The path for injection of the sample corresponds to the green and red marked tubing. During elution, the sample follows the green marked route, except for the outlet valve that switches to the collection tubes, as marked in blue. As the sample elutes from the column, a U.V. profile is traced by the ÄKTA start machine by measuring the absorbance at 280 nm.

2.2.9. Crystallisation trials using a nanodrop robot

Nanodrop crystallisation trails (100 nL:100 nL) were set up using the Mosquito® LC Robot for KIFC1-MD at 8 mg/mL and 16 mg/mL for 7 different screens (MORPHEUS®, JCSG®, PACT®, PGA®, PEG ION, ASNA and RdRP Screens) at two different temperatures (18ºC and 4ºC). The presence/absence of crystals was verified by observing the plates under the microscope frequently for about 2.5 months after the setup.

A second crystallisation screen was prepared similarly, after the failure to reproduce the crystals in optimization plates Nº 2-11 (section 2.2.12), with the following differences: 1) instead of drops with different protein concentrations, 100 nL:100nL and 200 nL:100 nL protein:well solution ratios were created, with protein at 16 mg/mL; 2) only one set of screens was prepared, to be incubated at 4ºC.

2.2.10. Determination of KIFC1-MD crystal type and quality

The crystals were collected onto a loop by adding 1 µL of the cryoprotectant to the drop containing the crystals. Once collected, the crystals were either immediately measured in-house or

immediately placed into a loop holding pack in liquid nitrogen and sent to measure at the Synchrotron. Diffraction data for individual crystals was collected at beamlines at Diamond Light Source. Table 2.3 shows the cryoprotectant used for each crystal and where it was measured. Each diffraction pattern obtained allowed to determine if the crystals were protein or salt crystals and how well they diffracted. Data was processed using the CCP4 suite (Winn et al., 2011). An attempt to solve the structure of KIFC1-MD was made for the crystal with 3.2Å resolution (see table 3.1) by molecular replacement (PHASER MR in CCP4 suite) (Protein Data Bank (PDB) code 2REP as a search model). Electron density and difference density maps, all σA-weighted, were inspected, and the models were improved using Coot from CCP4 suite. The calculation of R_{free} used 5% data.

Table 2.3: Summary of crystals measured, cryoprotectant used and place of measurement.

Screen/Optimization Plate	Crystal Measured	Cryopreservant	Measured in
First Crystallisation Screen	MORPHEUS® Screen H6	15% erythritol	In-house
	MORPHEUS® Screen B6 and E6	20% glycerol	Synchrotron
	RdRP Screen A1, A2 and A3	15% glycerol	
Optimization plate Nº 1 (section 2.2.12)	D6 2:1 drop	20% PEG 400	
	B1 1:1 drop	20% ethylene glycol	
	D2 2:1 drop, D5 2:1 drop, D6 2:1 drop	15% glycerol	
Second Crystallisation Screen	RdRP Screen F10 2:1 drop, G12 2:1 drop, H10 2:1 drop, H11 2:1 drop	20% glycerol	
	ASNA Screen E12 2:1		
	MORPHEUS® Screen C10 2:1 drop, C12 2:1 drop, D10 1:1 drop		
	PACT® Screen F6 1:1 drop		
	PEG ION Screen C11 1:1 drop and 2:1 drop, D12 2:1 drop, G7 2:1 drop		
	PGA® Screen B2 2:1 drop		

2.2.11. Desalting of KIFC1-MD

The PD10 gravity flow desalting column was initially equilibrated using Desalting buffer (see section 2.1.4.). 1 mL of the protein (KIFC1-MD batch Nº 5) was loaded into the column. The protein was then eluted with 3 mL of the Desalting buffer. As the protein was diluted in the process it was concentrated back to ~1 mL and the protein concentration was remeasured to be ~11 mg/mL.

2.2.12. Optimization of the crystallisation conditions from nanodrops to microdrops

Using the Hampton Make Tray Website (https://hamptonresearch.com/make_tray.aspx), the conditions were set for 24 well Linbro plates. For each plate, sitting (1 µL:1µL) and/or hanging drops (1 µL:1µL; 1.5 µL:1µL; 2 µL:1µL) were created as shown in table 2.4. Sitting drops were prepared using microbridges. Hanging drops were created on coverslips that were placed on top of the wells facing down. The plate was incubated at respective temperature indicated in table 2.4. The presence/absence of crystals was verified by observing the drops under a microscope frequently for about 1.5 months.

Table 2.4: Summary of optimization plates setup. Protein batch used, type of drops created, temperature of incubation, aim of the setup and number of table where the respective conditions can be found are indicated.

Optimization Plate Number (Nº)	Protein Batch	Type of drops	Temperature	Aim	Notes	Table with condition set
1	1	Sitting drops; Hanging drops 1:1, 2:1	4ºC	Optimization of the MORPHEUS® Screen (First Screen) conditions where crystals were found		2.5
2	1	Hanging drops 1:1, 1.5:1, 2:1	10ºC	Optimization of the D6 well condition of plate Nº 1	Amino acids Mix without alanine	2.6
3	2			Smaller variations of the D6 condition due to lack of crystals in plate Nº 2		2.7
4	2					
5	2	Hanging drops 1:1, 2:1	4ºC	Repeat plates Nº 1 and 3 to exclude experimental error as cause of lack of crystals		2.5
6	2					2.7
7	2	Sitting drops; Hanging drops 1:1, 2:1		Optimization of the RdRP Screen (first screen) where larger crystals were found (A1-A3)	----------	2.8
8	3	Hanging drops 1:1, 2:1		Assess the role of ADP and $MgCl_2$ in crystallisation success	Amino acids Mix without alanine	2.9
9	3					2.10
10	4			Repeat plates Nº 1 and 2 to assess the effect of higher purity in crystallisation success	Amino acids mix with alanine	2.5
11	4					2.6
12	4 and 5			Optimization of the RdRP Screen (second screen) conditions where crystals were found; Assess batch variation effect in crystallisation success	----------	2.11
13	4 and 5			Optimization of the PEG ION Screen (second screen) G7 condition where crystals were found	----------	2.12
14	5 desalted			Repeat plate Nº 1 to assess if lowering salt concentration of protein solution would allow to decrease solubility and reach supersaturated state required for crystal formation	Amino acids mix with alanine	2.5

Table 2.5: Crystallisation conditions optimization plan for optimization plate Nº 1. The first two columns correspond to the optimization of the MORPHEUS® Screen B6 Condition. The second two columns correspond to the optimization of the MORPHEUS® Screen E6 Condition. The last two columns correspond to the optimization of the MORPHEUS® Screen H6 Condition.

0.1 M Buffer System 2, pH 7.5 45% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 45% Precipitant Mix 2 0.1 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 45% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 45% Precipitant Mix 2 0 M Halogens Mix 0.17 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 45% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.1 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 45% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.2 M Amino acids Mix
0.1 M Buffer System 2, pH 7.5 50% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 50% Precipitant Mix 2 0.1 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 50% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 50% Precipitant Mix 2 0 M Halogens Mix 0.17 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 50% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.1 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 50% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.2 M Amino acids Mix
0.1 M Buffer System 2, pH 7.5 55% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 55% Precipitant Mix 2 0.1 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 55% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 55% Precipitant Mix 2 0 M Halogens Mix 0.17 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 55% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.1 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 55% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.2 M Amino acids Mix
0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.1 M Halogens Mix 0 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.17 M Ethylene Glycols Mix 0 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.1 M Amino acids Mix	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0 M Ethylene Glycols Mix 0.2 M Amino acids Mix

Table 2.6: Crystallisation conditions optimization plan for optimization plate Nº 2. It corresponds to variations of the condition present in well D6 of the optimization plate Nº 1.

0.13 M Amino acids Mix 54% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.13 M Amino acids Mix 56% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.13 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.13 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.13 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.13 M Amino acids Mix 64% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5
0.15 M Amino acids Mix 54% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 56% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 64% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5
0.17 M Amino acids Mix 54% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.17 M Amino acids Mix 56% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.17 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.17 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.17 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.17 M Amino acids Mix 64% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5
0.19 M Amino acids Mix 54% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.19 M Amino acids Mix 56% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.19 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.19 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.19 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.19 M Amino acids Mix 64% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5

Table 2.7: Crystallisation conditions optimization plan for optimization plates Nº 3/4. It corresponds to smaller variations of the condition present in well D6 of the optimization Plate Nº 1.

0.1 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.1 M Amino acids Mix 59% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.1 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.1 M Amino acids Mix 61% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.1 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.1 M Amino acids Mix 63% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5
0.15 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 59% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 61% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.15 M Amino acids Mix 63% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5
0.2 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.2 M Amino acids Mix 59% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.2 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.2 M Amino acids Mix 61% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.2 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.2 M Amino acids Mix 63% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5
0.215 M Amino acids Mix 58% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.215 M Amino acids Mix 59% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.215 M Amino acids Mix 60% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.215 M Amino acids Mix 61% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.215 M Amino acids Mix 62% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5	0.215 M Amino acids Mix 63% Precipitant Mix 2 0.1 M Buffer System 2, pH 7.5

Table 2.8: Crystallisation conditions optimization plan for optimization plate Nº 7. It corresponds to variations of the conditions A1, A2 and A3 of the RdRP Screen. MES refers to 2-(*N*-Morpholino)Ethane-Sulfonic acid.

0.05 M MES, pH 5.0 15% Glycerol 3.4 M NaCl	0.05 M MES, pH 5.0 15% Glycerol 3.5 M NaCl	0.05 M MES, pH 5.0 15% Glycerol 3.6 M NaCl	0.05 M MES, pH 5.0 15% Glycerol 3.655 M NaCl	0.05 M MES, pH 5.0 15% Glycerol 3.7 M NaCl	0.05 M MES, pH 5.0 15% Glycerol 3.8 M NaCl
0.05 M MES, pH 5.5 15% Glycerol 3.4 M NaCl	0.05 M MES, pH 5.5 15% Glycerol 3.5 M NaCl	0.05 M MES, pH 5.5 15% Glycerol 3.6 M NaCl	0.05 M MES, pH 5.5 15% Glycerol 3.655 M NaCl	0.05 M MES, pH 5.5 15% Glycerol 3.7 M NaCl	0.05 M MES, pH 5.5 15% Glycerol 3.8 M NaCl
0.05 M MES, pH 6.0 15% Glycerol 3.4 M NaCl	0.05 M MES, pH 6.0 15% Glycerol 3.5 M NaCl	0.05 M MES, pH 6.0 15% Glycerol 3.6 M NaCl	0.05 M MES, pH 6.0 15% Glycerol 3.655 M NaCl	0.05 M MES, pH 6.0 15% Glycerol 3.7 M NaCl	0.05 M MES, pH 6.0 15% Glycerol 3.8 M NaCl
0.05 M MES, pH 6.5 15% Glycerol 3.4 M NaCl	0.05 M MES, pH 6.5 15% Glycerol 3.5 M NaCl	0.05 M MES, pH 6.5 15% Glycerol 3.6 M NaCl	0.05 M MES, pH 6.5 15% Glycerol 3.655 M NaCl	0.05 M MES, pH 6.5 15% Glycerol 3.7 M NaCl	0.05 M MES, pH 6.5 15% Glycerol 3.8 M NaCl

Table 2.9: Crystallisation conditions optimization plan for optimization plate Nº 8. It corresponds to a constant condition D6 from optimization plate Nº 1 and variations of the ADP and $MgCl_2$ concentrations.

0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.002 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.004 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.008 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.016 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.032 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.064 M $MgCl_2$ 0.002 M ADP
0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.002 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.004 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.008 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.016 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.032 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.064 M $MgCl_2$ 0.004 M ADP
0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.002 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.004 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.008 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.016 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.032 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.064 M $MgCl_2$ 0.008 M ADP
0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.002 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.004 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.008 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.016 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.032 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.15 M Amino acids Mix 0.064 M $MgCl_2$ 0.016 M ADP

Table 2.10: Crystallisation conditions optimization plan for optimization plate Nº 9. The first three columns correspond to a constant condition B6 from MORPHEUS® Screen and variations of the ADP and $MgCl_2$ concentrations; the last three columns correspond to a constant condition E6 from MORPHEUS® Screen and variations of the ADP and $MgCl_2$ concentrations.

0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.002 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.002 M ADP
0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.004 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.004 M ADP
0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.008 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.008 M ADP
0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0.09 M Halogens Mix 0 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.002 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.008 M $MgCl_2$ 0.016 M ADP	0.1 M Buffer System 2, pH 7.5 60% Precipitant Mix 2 0 M Halogens Mix 0.12 M Ethylene Glycols Mix 0.032 M $MgCl_2$ 0.016 M ADP

Table 2.11: Crystallisation conditions optimization plan for optimization plate Nº 12. The first three columns correspond to variations of the condition H10 from RdRP Screen and the last three columns correspond to a variation of the condition H11 from RdRP Screen.

0.2 M TRIS, pH 7.5 5% PEG 20000	0.2 M TRIS, pH 7.6 5% PEG 20000	0.2 M TRIS, pH 7.9 5% PEG 20000	0.2 M TRIS, pH 8.0 5% PEG 20000	0.2 M TRIS, pH 8.1 5% PEG 20000	0.2 M TRIS, pH 8.5 5% PEG 20000
0.2 M TRIS, pH 7.5 6% PEG 20000	0.2 M TRIS, pH 7.6 6% PEG 20000	0.2 M TRIS, pH 7.9 6% PEG 20000	0.2 M TRIS, pH 8.0 6% PEG 20000	0.2 M TRIS, pH 8.1 6% PEG 20000	0.2 M TRIS, pH 8.5 6% PEG 20000
0.2 M TRIS, pH 7.5 7% PEG 20000	0.2 M TRIS, pH 7.6 7% PEG 20000	0.2 M TRIS, pH 7.9 7% PEG 20000	0.2 M TRIS, pH 8.0 7% PEG 20000	0.2 M TRIS, pH 8.1 7% PEG 20000	0.2 M TRIS, pH 8.5 7% PEG 20000
0.2 M TRIS, pH 7.5 8% PEG 20000	0.2 M TRIS, pH 7.6 8% PEG 20000	0.2 M TRIS, pH 7.9 8% PEG 20000	0.2 M TRIS, pH 8.0 8% PEG 20000	0.2 M TRIS, pH 8.1 8% PEG 20000	0.2 M TRIS, pH 8.5 8% PEG 20000

Table 2.12: Crystallisation conditions optimization plan for optimization plate Nº 13. It corresponds to variations of the condition G7 from the PEG ION Screen. MME refers to Monomethyl Ether.

0.1 M Imidazole, pH 7.0 19% PEG 5000 MME 1% PEG 400	0.1 M Imidazole, pH 7.0 21.5% PEG 5000 MME 1% PEG 400	0.1 M Imidazole, pH 7.0 24% PEG 5000 MME 1% PEG 400	0.1 M Imidazole, pH 7.0 26.5% PEG 5000 MME 1% PEG 400	0.1 M Imidazole, pH 7.0 29% PEG 5000 MME 1% PEG 400	0.1 M Imidazole, pH 7.0 31.5% PEG 5000 MME 1% PEG 400
0.1 M Imidazole, pH 7.0 19% PEG 5000 MME 2% PEG 400	0.1 M Imidazole, pH 7.0 21.5% PEG 5000 MME 2% PEG 400	0.1 M Imidazole, pH 7.0 24% PEG 5000 MME 2% PEG 400	0.1 M Imidazole, pH 7.0 26.5% PEG 5000 MME 2% PEG 400	0.1 M Imidazole, pH 7.0 29% PEG 5000 MME 2% PEG 400	0.1 M Imidazole, pH 7.0 31.5% PEG 5000 MME 2% PEG 400
0.1 M Imidazole, pH 7.0 19% PEG 5000 MME 4% PEG 400	0.1 M Imidazole, pH 7.0 21.5% PEG 5000 MME 4% PEG 400	0.1 M Imidazole, pH 7.0 24% PEG 5000 MME 4% PEG 400	0.1 M Imidazole, pH 7.0 26.5% PEG 5000 MME 4% PEG 400	0.1 M Imidazole, pH 7.0 29% PEG 5000 MME 4% PEG 400	0.1 M Imidazole, pH 7.0 31.5% PEG 5000 MME 4% PEG 400
0.1 M Imidazole, pH 7.0 19% PEG 5000 MME 8% PEG 400	0.1 M Imidazole, pH 7.0 21.5% PEG 5000 MME 8% PEG 400	0.1 M Imidazole, pH 7.0 24% PEG 5000 MME 8% PEG 400	0.1 M Imidazole, pH 7.0 26.5% PEG 5000 MME 8% PEG 400	0.1 M Imidazole, pH 7.0 29% PEG 5000 MME 8% PEG 400	0.1 M Imidazole, pH 7.0 31.5% PEG 5000 MME 8% PEG 400

2.2.23. ATPase Assay

To 50 mL of the A25 buffer (see section 2.1.4.), 500 µL of PEP (200 mM), 180 µL of NADH (70 mM), 25 µL of PK (10 mg/mL) and 40.5 µL of LDH (5 mg/mL) were added. A concentration series of Mg^{2+}ATP was used was indicated in table 2.13. For the ATPase assay, 98 µL of the individual ATP concentration containing ATPase buffer were transferred into corresponding wells in the microplate µClear®, along with 2 µL of the KIFC1-MD protein solution (from batch Nº 5 desalted; resulting in 5.58 µM of protein concentration in the assay), in triplicate. The microplate µClear® was placed in the Absorbance Microplate Reader that had been previously cooled to 25ºC. A scheme representing the reactions occurring in the microplate can be seen in Figure 2.3. The absorbance was read at 340 nm for 30 min.

The OD/s was calculated based on the following formula:

$$OD/\,\mathrm{s}(s^{-1}) = \frac{-\Delta A/min\ x2x10^{6}}{\varepsilon_{NADH} x60x[protein]_{\mu M}},$$

where ΔA/min is the absorbance decrease at 340 nm per min, ε_{NADH} corresponds to the molar extinction coefficient of NADH (6220 ΔA/mol/cm) (Horecker and Kornberg, 1948) and $[protein]_{\mu M}$ is the molar concentration of KIFC1-MD in the assay (5.58 µM).

Average and Standard Deviation of the OD/min of the triplicates were calculated and the results were plotted and analysed using the program KaleidaGraph.

Table 2.13: ATP concentration series (mM) used in the ATPase assay.

ATP concentration series (mM)
10
5
2.5
1.25
0.625
0.3125
0.15625
0.078125
0.039063
0.019531
0.009766
0

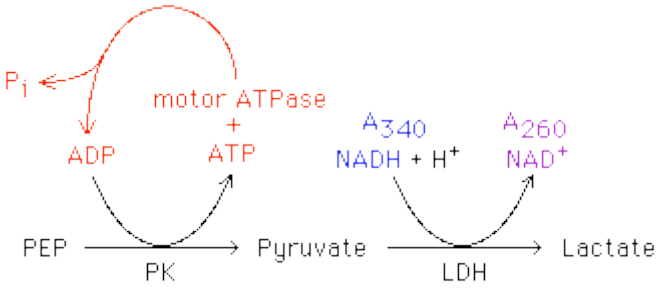

Figure 2.2: Scheme representing the reactions involved in the ATPase assay. Adapted from https://labs.cellbio.duke.edu/kinesin/Methods/ATPase_assay.html.

2.2.24. Amino acids sequence alignment

Using the BLAST® tool in NCBI, the amino acids sequence was compared between the predicted sequence for the KIFC1-MD purified during this project and the KIFC1-MD whose structure is available in the PDB. The sequence alignment can be found in Attachments section 7.11.

2.2.25. Fragment-based screening

1 mL of the KIFC1-MD batch Nº 4 was sent to Monash University for fragment-based screening. This is still in progress.

3. Results

3.1. Restriction of the vectors and the insert containing plasmid for sub-cloning of KIFC1-MD

The restricted pDNA of the vectors and the insert used for cloning KIFC1-MD for this project is shown in Figure 3.1. The sizes of the bands resulting from the restriction of pDHT-KIFC1-MD (Figure 3.1a) correspond exactly to the expected sizes for the vector pDHT (~5.3 kb) and the insert coding KIFC1-MD (~1.1 kb). One band resulted from both ppSUMO-2 and pDHT-2 restrictions (Figure 3.1), with ~5.4 kb and ~5.3 kb, respectively. Although a restriction with two enzymes should result in two pDNA fragments, it is not unlikely that only one band would be visible from these restrictions since the second fragment is expected to have very low molecular size (~10 base pairs (bp) for both vectors) and rapidly go through the pores of the gel. Nevertheless, the bands in the gel correspond approximately to the expected restricted sizes of the vectors (~5.3 kb for pDHT-2 and ~5.6 kb for ppSUMO-2).

In Figure 3.1.b the marker is difficult to see, being impossible to distinguish the bands corresponding to the 1.5 kb, 1 kb and 0.5 kb sizes. This could be due to errors in the pipetting of the SyBr Safe Dye or in the mixing of the same in the gel. Another possibility is that the gel was exposed to the UV light for too long and the DNA started degrading, since UV light is known for inducing DNA damage, including DNA single-strand breaks (Roos and Kaina, 2006). For this reason, the calibration curve used to calculate the molecular size of the band had less points to consider comparing to the one used for in Figure 3.1.a, which could have induced slight deviations to the resulting equation and determined size of the ppSUMO-2 restricted band.

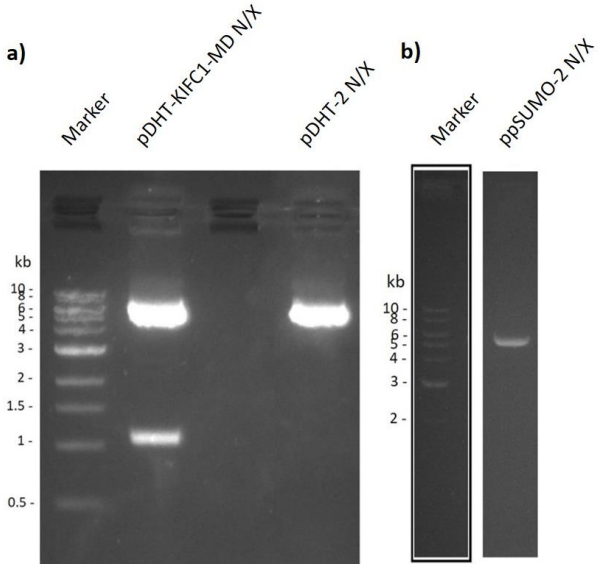

Figure 3.1: Agarose gels containing the restricted DNA from a) pDHT-KIFC1-MD plasmid and pDHT-2 vector and b) ppSUMO-2 vector. N/X refers to the restriction enzymes used: *NcoI* and *XhoI*. The marker used was a 1 kb ladder and the molecular sized of the marker DNAs are indicated on the left.

3.2. Diagnostic restriction digest to check for positive sub-cloning of KIFC1-MD

The restricted pDNA from the colonies transformed with the ligation product of ppSUMO-2 and KIFC1-MD insert at the ratio of 1:20 gave origin to two visible bands on the agarose gel with the respective sizes of ~5.7 kb and ~1.1 kb (Figure 3.2a), which correspond approximately to the expected sizes of the ppSUMO-2 cleaved vector (~5.6 kb) and the KIFC1-MD insert (~1.1 kb). On the other hand, from restriction of the pDNA from the mini-prep of the colonies transformed with the ligation product of ppSUMO-2 and KIFC1-MD at the ratio of 1:10, no visible bands appeared on the gel (Figure 3.2a). This indicates that the mini-prep did not yield any pDNA because colonies did not contain the ligation's product. This would mean the colonies were actually kanamycin resistant contamination colonies and not *E. coli* DH5α cells transformed with the desired construct. This could explain why the colonies labelled number one for both ratios did not grow on liquid media – potentially the colonies picked were also contamination colonies with a minor resistance to kanamycin that was enough to survive in the solid medium but not on the liquid medium.

All the diagnostic restriction solutions from colonies transformed with the ligation product of pDHT-2 and KIFC1-MD insert gave origin to two bands on the agarose gel, regardless of the ratio vector:insert (Figure 3.2b), with the following sizes: ~5.6 kb and ~1.1 kb, respectively. These sizes correspond approximately to the expected sizes for the cleaved pDHT-2 vector (~5.3 kb) and the KIFC1-MD insert (~1.1 kb).

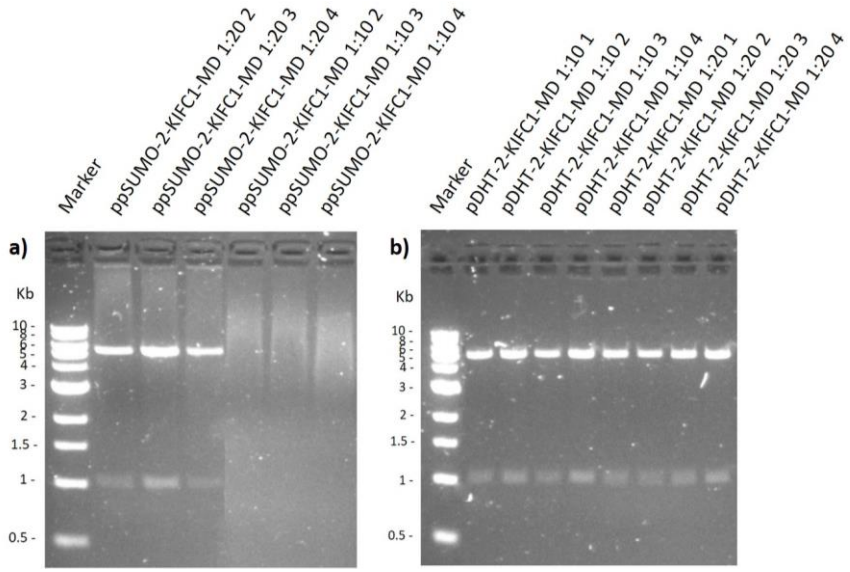

Figure 3.2: Agarose gels containing the pDNA from the diagnostic restrictions of a) the ppSUMO-2 constructs and b) the pDHT-2 constructs. The marker used for both cases was a 1 kb ladder and the molecular sizes DNA markers are indicated on the left. 1:10 and 1:20 refers to the ratios of vector:insert used and 1,2,3,4 refers to the number attributed to each colonies picked from the transformation plate.

ppSUMO-2-KIFC1-MD 1:20 colony number 3 (lane 3, Figure 3.2a) and pDHT-2-KIFC1-MD 1:10 colony number 2 (lane 3, Figure 3.2b) samples qualitatively appear to have greater amount of DNA. Because of this, these were selected to be used for small-scale KIFC1-MD expression (section 2.2.7.). The differences in DNA yield verified for both vectors' different colonies are probably due to different

amount of total cell growth during the incubation step, due to different number of cells picked from the colonies. Small differences in the incubation time during the mini-prep's elution step could also be the cause.

3.3. Small-scale purification to determine the best conditions of KIFC1-MD expression

3.3.1. Small-scale expression of pDHT-2-KIFC1-MD and ppSUMO-2-KIFC1-MD in different *E. coli* strains

The ~53 kDa bands in *E. coli* BL21 (DE3) and *E.* coli BL21 Codon Plus (DE3) RIPL cells transformed with ppSUMO-2-KIFC1-MD (Figure 3.3., lanes 3 and 4, respectively) correspond exactly to the expected size of KIFC1-MD with the His and SUMO tags. The ~55 kDa protein band in lane 5 of Figure 3.3. (*E. coli* Rosetta™ BL21 (DE3) cells transformed with ppSUMO-2-KIFC1-MD) is also presumed to be KIFC1-MD but with lower expression levels, since 55 kDa is approximate to the expected molecular size and the dye front is not entirely straight (probably due to the fact that the gel and buffer are self-made instead of commercials) which could have affected the calculation of the molecular sizes. The remaining bands correspond to contamination proteins naturally expressed in the *E. coli* cells (Figure 3.3). Based on this, ppSUMO-2-KIFC1-MD construct in *E. coli* BL21 (DE3) was selected for subsequent large-scale purification.

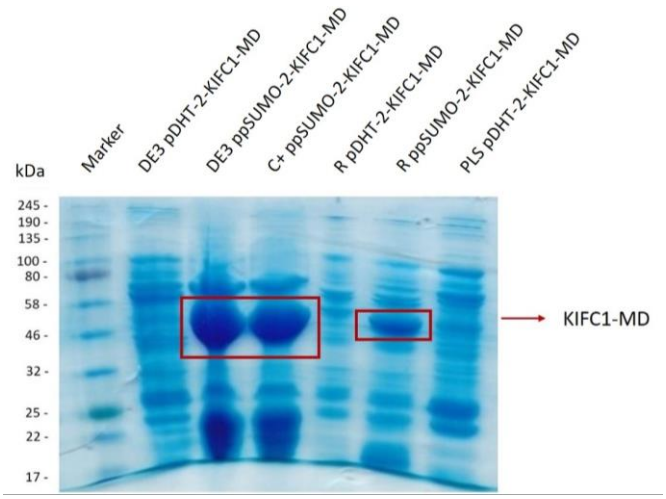

Figure 3.3: SDS-PAGE gel with small-scale expression/purification samples. The marker proteins molecular sizes are indicated on the left. DE3 refers to *E. coli* BL21 (DE3) cells, C+ refers to *E. coli* BL21 Codon Plus (DE3) RIPL cells, R refers to *E. coli* Rosetta™ BL21 (DE3) cells and PLS refers *E. coli* Rosetta™ BL21 (DE3) pLysS cells. The ~53 kDa and ~55 kDa bands correspondent to KIFC1-MD are indicated by the red boxes and arrow.

3.3.2. Assessment of the His and SUMO tags cleavage efficiency

E. coli BL21 (DE3) and *E. coli* BL21 Codon Plus (DE3) RIPL cells purification samples reveal bands with ~56 kDa and ~53 kDa, respectively (lanes 1 and 3 of Figure 3.4, respectively) that approximately correspond to the expected size of KIFC1-MD with the respective tags (~53 kDa). Although the first sample appears to have a higher molecular size than expected, it is still possible to

infer it's KIFC1-MD for three reasons: 1) The molecular weight of the bands is difficult to estimate due to the size of the bands. Nevertheless, the high size of the bands is a sign of good expression; 2) In the previous gel the same sample was verified to have a band at exactly ~53 kDa; 3) The gel, once again does not present a straight dye front, being the dye front from the marker very different from the dye front in the samples (again, this is probably due to the fact that both the gel and buffer are self-made instead of commercials), which can lead to errors in the calculated molecular sizes.

The Ulp1 treated samples show bands with ~36 kDa and ~34 kDa for *E. coli* BL21 (DE3) and *E. coli* BL21 Codon Plus (DE3) RIPL cells, respectively, which corresponds approximately to the expected size of the cleaved KIFC1-MD (~39 kDa). The bands with ~24 kDa in lane 2 and ~23 kDa in the lanes 4 and 5 of Figure 3.4, correspond approximately to the expected molecular size of to Ulp1 protease (~26 kDa).

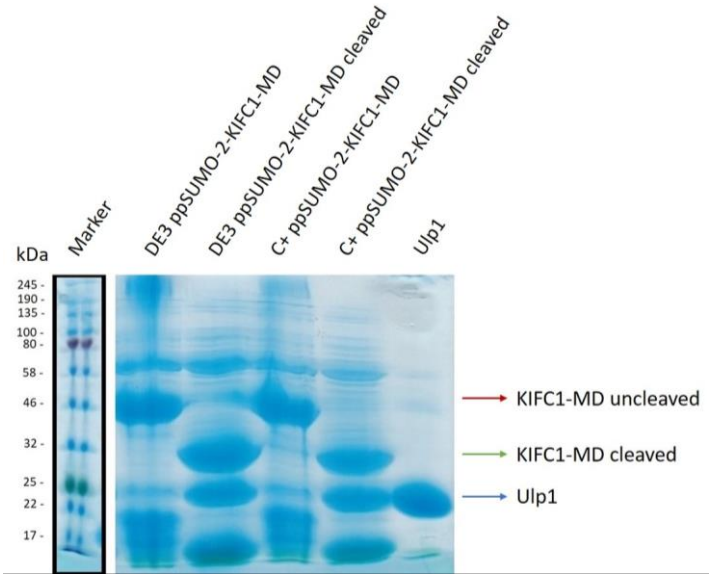

Figure 3.4: Assessment of the His and SUMO tags cleavage efficiency. The molecular sizes of the marker proteins are indicated on the left. DE3 refers to BL21 DE3 *E. coli* cells and C+ refers to the Codon Plus *E. coli* cells. The ~56 and 53 kDa bands that correspond to KIFC1-MD with the respective His and SUMO tags are indicated by the red arrow. The ~36 and ~34 kDa bands that correspond to KIFC1-MD without the respective tags are indicated by the green arrow. The ~24 and ~23 kDa bands that correspond to Ulp1 are indicated by the blue arrow.

3.4. Large-scale purification to obtain large amounts of KIFC1-MD at high purity

3.4.1. First purification step by nickel-affinity column

The bands at ~54 kDa correspond approximately to the predicted molecular size of KIFC1-MD with SUMO and His tags (~53 kDa) (Figure 3.5, lanes E1-E6). By comparing the elution fractions (lanes E1-E6 of Figure 3.5) with the lysate and wash flow-through samples (lanes L and W of Figure 3.5), it is possible to deduce that the purification was successful since most of the bands visible in lanes L and W are not present in the elution fractions.

The calculated purity of the protein at this stage was estimated to be ~50%. The total protein concentration calculated was 6.6 mg/mL. The total amount of protein calculated was 492 mg (total = 75

mL volume x 6.6 mg/mL). Since the quantity of protease necessary was 1 mg per each 50 mg of protein, the total amount used was 9.8 mg or ~2 mL of Ulp1 (5 mg/mL).

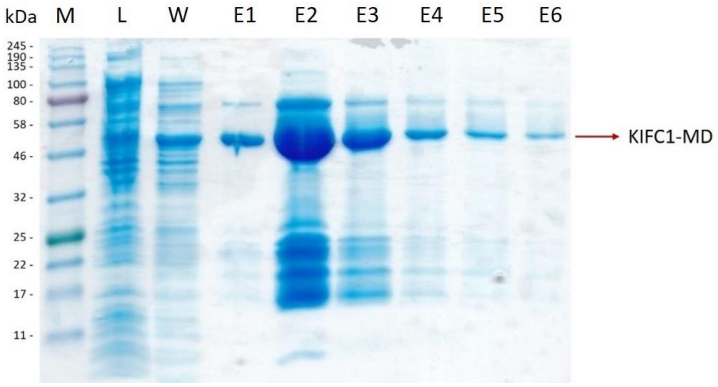

Figure 3.5: SDS-PAGE gel of KIFC1-MD large-scale first purification step by nickel-affinity column. This gel shows the marker (M), lysate flow-through (L), wash (W) and elution fractions (E1-E6) from the first purification step. The molecular sizes of the marker proteins are indicated on the left. The ~54 kDa bands corresponding to KIFC1-MD are indicated by the red arrow. The remaining bands are contamination bands corresponding to proteins naturally expressed in *E. coli*.

3.4.2. Second purification step by nickel-affinity column

After cleavage with Ulp1, a ~39 kDa band is visible in all lanes containing cleaved flow-through and wash samples and corresponds exactly to the expected molecular size of cleaved KIFC1-MD (i.e. without the SUMO and His tags) (Figure 3.6, lanes C1-W3). In the elution fractions (lanes E1-E3 of Figure 3.6) exist contamination bands that are not present in the remaining lanes (or that are present but in much lesser extent), indicating that most contaminations were eliminated.

Furthermore, the calculated purity of the protein in the cleaved and wash flow-throughs was ~86%, which is higher purity than verified in the previous purification step, confirming the success of the purification process. However, as protein purity is very important for the crystallisation success (McPherson and Gavira, 2014), typically protein samples that are less than 90% pure go through an additional purification step (Khurshid et al., 2014; Sauder et al., 2008). Therefore, a third purification step was undertaken.

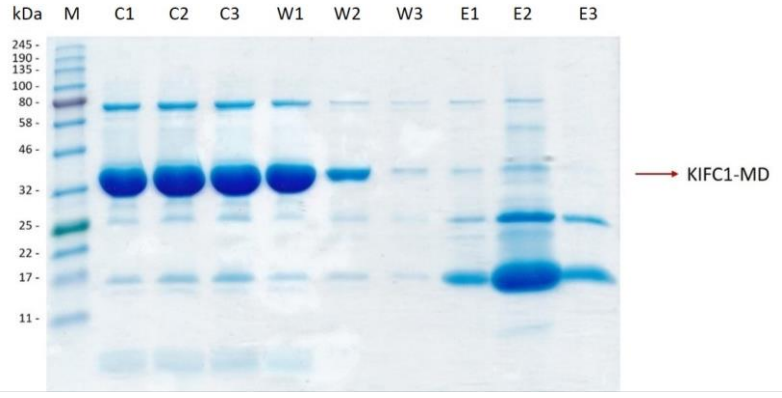

Figure 3.6: SDS-PAGE gel of KIFC1-MD after the second purification step by nickel-affinity column. This gel shows the marker (M), cleaved flow-through (C1-C3), wash (W1-W3) and elution (E1-E3) fractions from the second step of the purification procedure. The molecular sizes of the marker proteins are indicated on the left. The ~39 kDa band corresponding to KIFC1-MD is indicated by the red arrow. The remaining bands are minor contaminations bands corresponding to proteins naturally expressed in *E. coli*.

3.4.3. Gel filtration chromatography

A single peak at ~82 mL is present in Figure 3.7 and corresponds approximately to the expected volume of elution of a globular ~39 kDa protein (section 7.9, Figure 7.10) and, as such, represents the purified KIFC1-MD protein. The single peak is indication of high purity. A small peak at ~43 mL is also present (Figure 3.7.). This peak corresponds to the void volume of the column (section 7.9). Any contents of the sample that, due to very high molecular size, do not enter the column's pores, flow-through at this volume. Usually, this represents agglomerates of proteins. The small peak at this volume indicates that the protein in the sample was starting to aggregate, but with the purification, these aggregates were separated from the soluble KIFC1-MD.

Upon running the samples of KIFC1-MD on a SDS-PAGE gel, a band at ~38 kDa is visible in lanes 2-10 of Figure 3.8. that corresponds to the large peak obtained at ~82 mL and approximately to the predicted molecular size of a monomeric KIFC1-MD (~39 kDa). A contamination ~70 kDa band is also visible in lanes 2-4 of Figure 3.8. For this reason, only the protein elution fractions corresponding to lanes 5-10 in Figure 3.8 were used for crystallisation trials. The calculated purity was ~93%, which is high purity, consistent with the single peak observed in Figure 3.7. Moreover, it is above the minimum desired 90% (Khurshid et al., 2014; Sauder et al., 2008). As the purity increased in this final step, the purification process was successful.

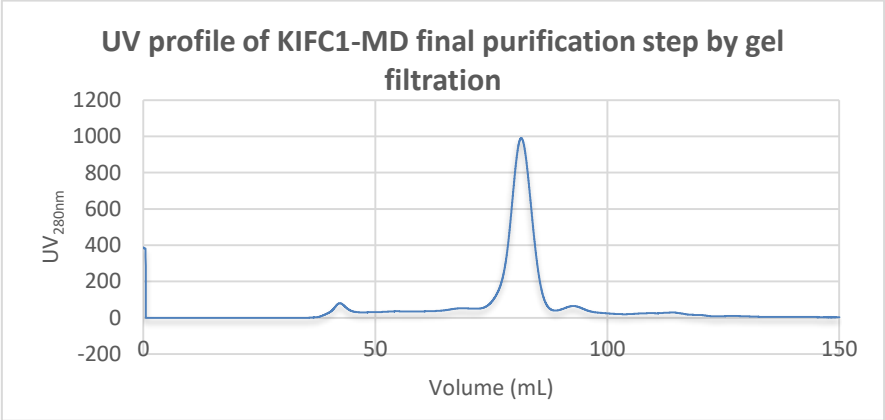

Figure 3.7: UV profile of KIFC1-MD final purification step by gel filtration. x axis – Volume (mL); y axix – UV_{280nm}**.** A peak at ~82 mL is visible corresponding to purified KIFC1-MD.

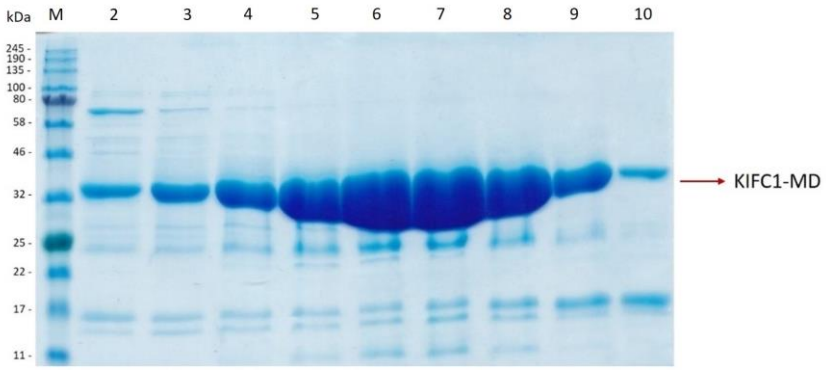

Figure 3.8: SDS-PAGE gel of purified KIFC1-MD after gel filtration. This gel shows the marker proteins (M) and the gel filtration elution fractions (lanes 2-10). The molecular sizes of the marker proteins are indicated on the left. The ~38 kDa band corresponding to KIFC1-MD is indicated by the red arrow. The remaining bands are minor contaminations corresponding to proteins naturally expressed in the *E. coli*.

3.5. First crystallisation screen

Six days after setting up various crystallisation screens, a total of four crystals was found in the MORPHEUS® Screen at 4ºC, all in drops with ~16 mg/mL of protein: one in the B6 condition (Figure 3.9a), one in E6 condition (Figure 3.9b) and two in the H6 condition (Figure 3.9c). The shapes of the crystals in the different conditions are slightly different. Two and a half months after the screen was setup, several crystals were found in the RdRP Screen at 4ºC, all in drops with ~16 mg/mL of protein: 1) one crystal, similar to the ones in the condition H6 of MOPHEUS® Screen, was found in the condition A1 (Figure 3.10a); 2) Over 50 crystals were found in A2 and A3 conditions (Figure 3.10b-c); 3) total eight very small crystals were found in A6 and A7 conditions (Figure 3.10d-e). The crystals found in A2, A3, A6 and A7 had different shapes from the ones previously seen but were similarly shaped to one another.

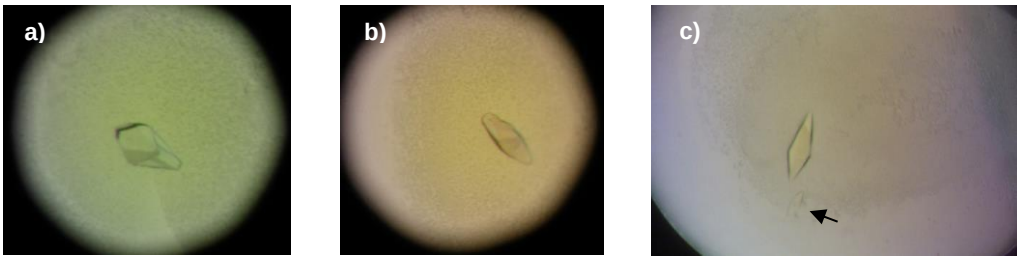

Figure 3.9: Crystals found in MORPHEUS® Screen at 4ºC, six days after the screen setup. a) Crystal in condition B6; b) Crystal in condition E6; c) Two crystals in condition H6; the smaller one is indicated by the black arrow. All crystals were found in the wells with ~16 mg/mL of protein. The pictures are zoomed in from the view on the microscope on 50X amplification.

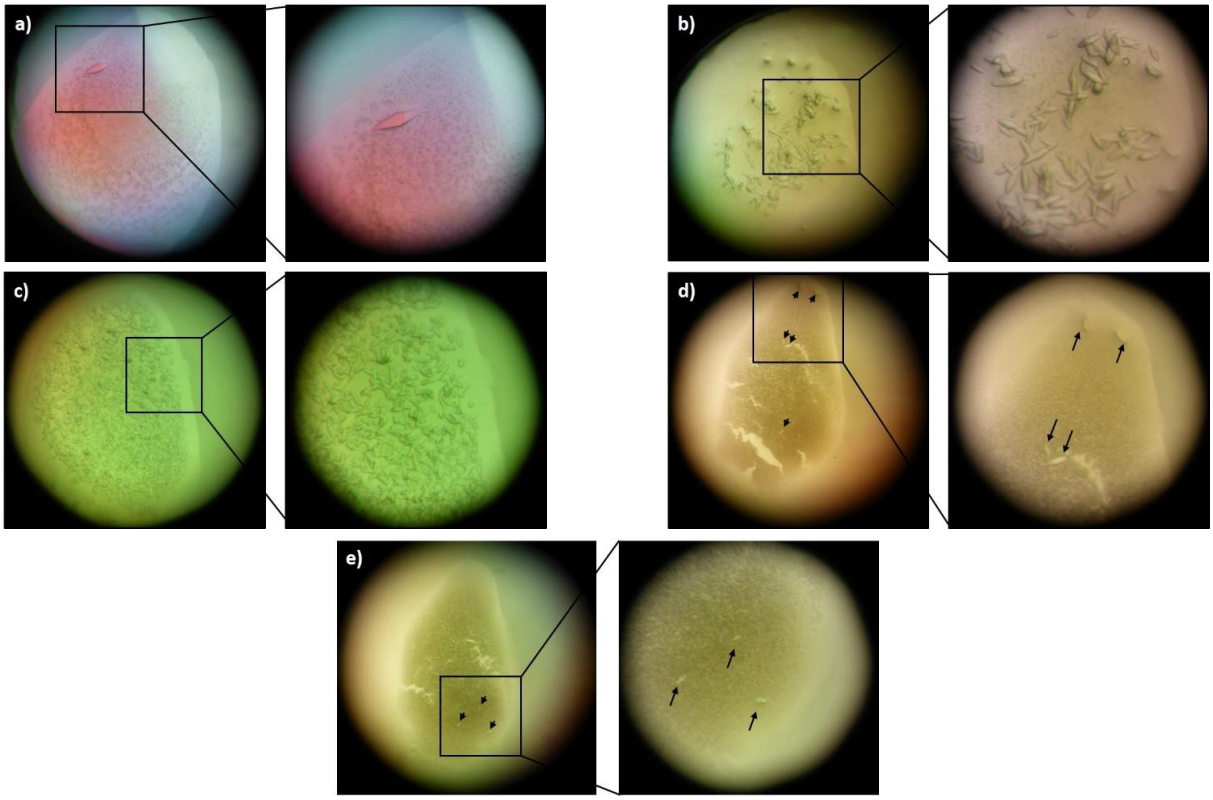

Figure 3.10: Crystals found in the RdRP Screen at 4ºC, 2.5 months after the setup of the crystallisation screen. All pictures are 50X amplified. All crystals were present in the drops with ~16 mg/mL of protein. A) A single crystal present in condition A1; b) Over 50 crystals present in condition A2; c) Over 50 crystals present in condition A3; d) Five very small crystals present in condition A6; e) three very small crystals present in condition A7.

The range of the diffraction pattern from a crystal is associated with the level of its internal order. The more structurally uniform the molecules are in the crystal, the vaster the pattern becomes, extending to higher resolution. This degree of order relates closely to how detailed the determination of atomic positions can be. Although salt crystals often diffract to their theoretical limit of resolution, protein crystals diffract with a more limited range (McPherson, 2004; McPherson and Gavira, 2014). Upon measuring, all crystals found in the first crystallisation screen, revealed low-resolution diffraction patterns, which indicates they were protein crystals. For example of a diffraction pattern obtained, see Figure 3.11.

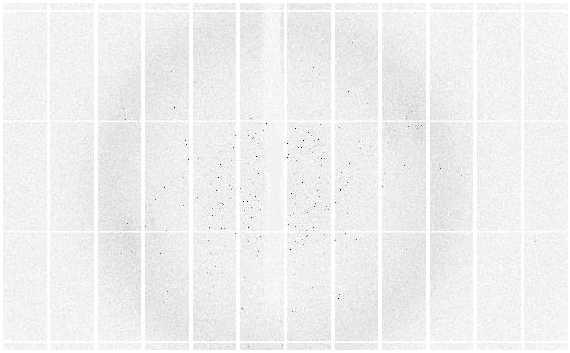

Figure 3.11: Diffraction pattern obtained from the crystal of H6 condition of MORPHEUS® Screen. Several dots can be seen at low resolution, and the maximum resolution obtained was 3.4Å.

3.6. Optimization plate Nº 1

Four days after the optimization plate Nº 1 was setup, crystals were found in well D6, in hanging drop ratio 2:1. Total of three small crystals could be observed (Figure 3.12). Even smaller crystals seemed to be trying to grow around these, all over the drop. No crystals were found in any other drop, at this point of time. Upon measuring, the crystals in D6 condition revealed a diffraction pattern with low resolution, indicating they were protein crystals. (For an example of diffraction pattern, see Figure 3.11).

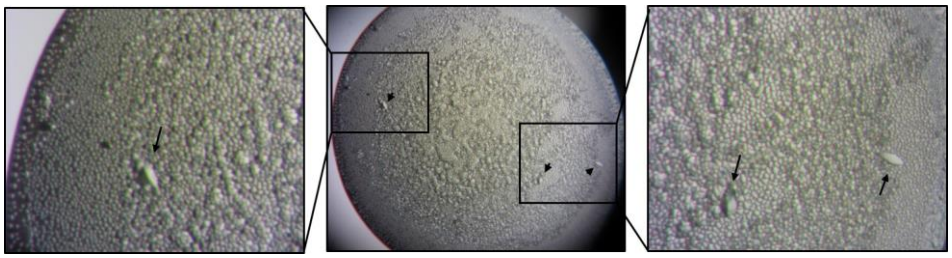

Figure 3.12: Crystals found in well D6 of optimization plate Nº 1, in 2:1 ratio hanging drop, 4 days after the plate setup. The crystals are indicated with black arrows. 50X amplification.

Over 50 crystals were found in each of the 1:1 drops of the first two columns of the optimization plate Nº 1, four weeks after the plate setup (Figure 3.13a-h). These crystals were very small and rectangular-shaped, unlike the ones previously found. Upon measuring some of these crystals, their diffraction patterns revealed very high resolution, indicating they were salt crystals. On the other hand, small crystals shaped similarly to the ones found before were observed in the 2:1 ratio drops of the wells D2 (Figure 3.13i), D5 (Figure 3.13j) and D6 (Figure 3.13k). Upon measuring, these crystals revealed diffraction patterns at low resolution, indicating they were protein crystals (see Figure 3.11).

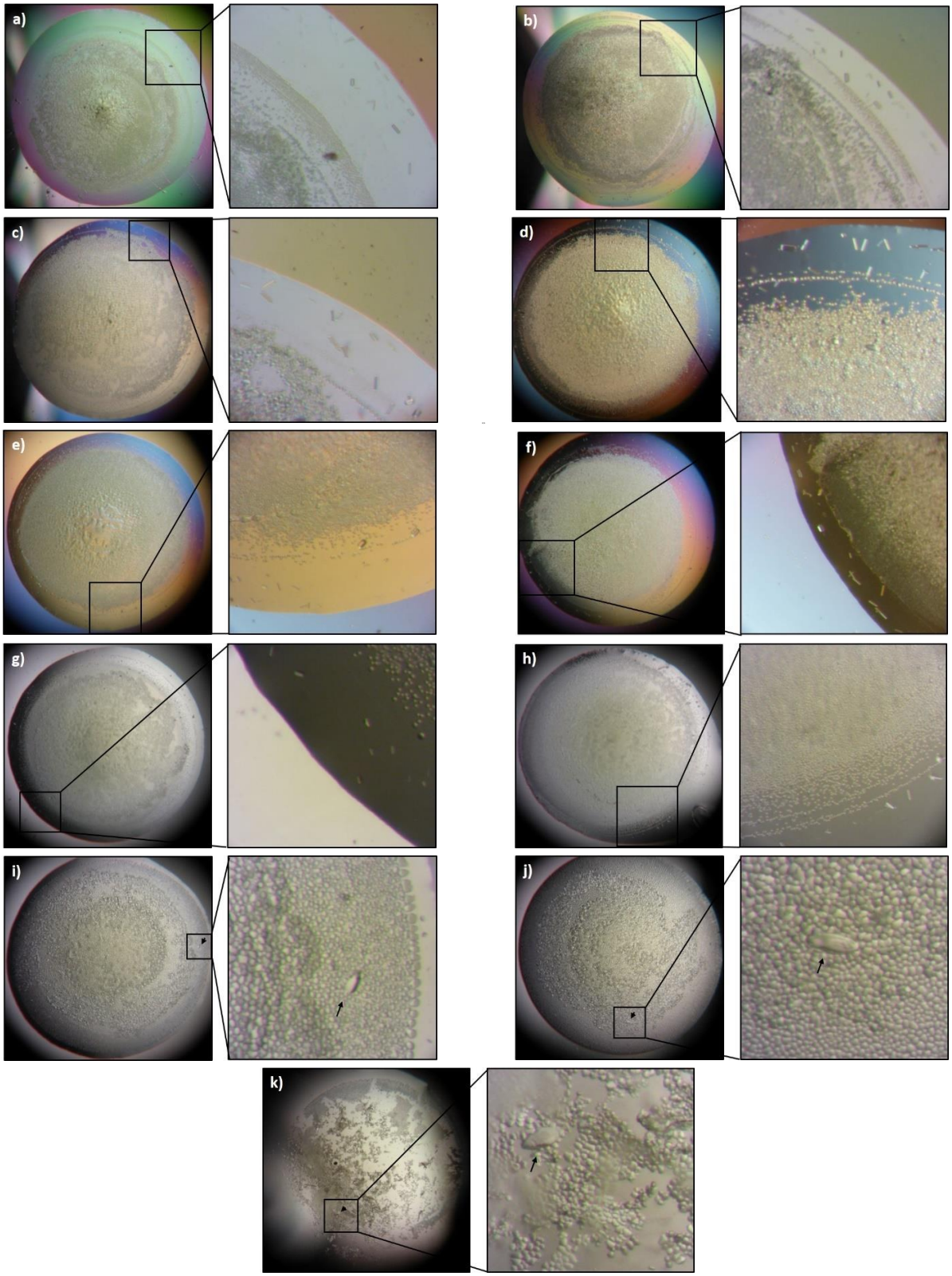

Figure 3.13: Crystals found in the optimization plate Nº 1, four weeks after the plate setup. In i), j) and k) the black arrows indicate crystals. a) Well A1, 1:1 protein:well solution ratio; b) Well A2, 1:1 protein:well solution ratio; c) Well B1, 1:1 protein:well solution ratio; d) Well B2, 1:1 protein:well solution ratio; e) Well C1, 1:1 protein:well solution ratio; f) Well C2, 1:1 protein:well solution ratio; g) Well D1, 1:1 protein:well solution ratio; h) Well D2, 1:1 protein:well solution ratio; i) Well D2, 2:1 protein:well solution ratio; j) Well D5, 2:1 protein:well solution ratio; k) Well D6, 2:1 protein:well solution ratio. 50X amplification.

3.7. Optimization plates Nº 2-11

No crystals were found in these plates, even 1.5 months after they were setup.

3.8. Second crystallisation screen

Four days after the setup of the screen, a total of 16 crystals was found in the RdRP Screen, all in drops with a protein:well solution ratio of 2:1: one in the F10 condition (Figure 3.14a), two in the G12 condition (Figure 3.14b), one in H10 condition (Figure 3.14c) and eleven in the H11 condition (Figure 3.14d). The shapes of the crystals in the distinct conditions are all different. Upon measuring, all crystals revealed diffraction patterns to very high resolution, indicating they were salt crystals.

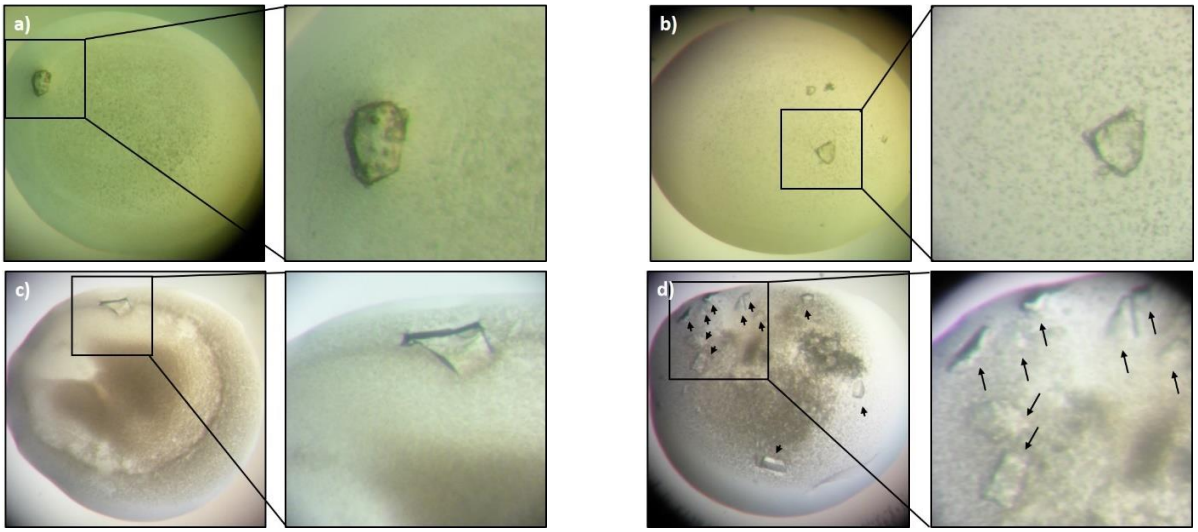

Figure 3.14: Crystals found in the RdRP Screen of the second crystallisation screen, 4 days after the screen setup. All crystals were found in the drops with 2:1 protein:well solution ratio, a) one crystal in F10 condition; b) two crystals found in G12 condition; c) one crystal found in H10 condition; d) eleven crystals found in H11 condition, indicated by black arrows. 50X amplification.

Fifteen days after the screen was setup, several crystals were found: 1) one crystal was found in the condition E12 of ASNA Screen (Figure 3.15a); 2) three crystals were found in C10 condition of MORPHEUS® Screen (Figure 3.15b) 3) Over 50 crystals were found in C12 condition of MORPHEUS® Screen (Figure 3.15c); 4) one crystal was found in D10 condition of MORPHEUS® Screen (Figure 3.15d); 5) one crystal found in F6 condition of PACT® Screen (Figure 3.15e); 6) one crystal found in each of the drops of the C11 condition of the PEG ION Screen (Figure 3.15f-g); 7) one crystal found in condition D12 of the PEG ION Screen (Figure 3.15h); 8) Over 50 crystals found in G7 of the PEG ION Screen (Figure 3.15i); 9) one crystal found in B2 condition of the PGA® Screen (Figure 3.15j). All crystals had different shapes from one another. Particularly, one crystal in condition C12 of MORPHEUS® Screen had similar shape to the ones found in H6 condition of MORPHEUS® Screen of the initial crystallisation screen. As many of these crystals as possible were measured (except or the ones in G7 condition of PEG ION Screen, since they were too small to collect), revealing diffraction patterns with very high resolution, indicating they were salt crystals.

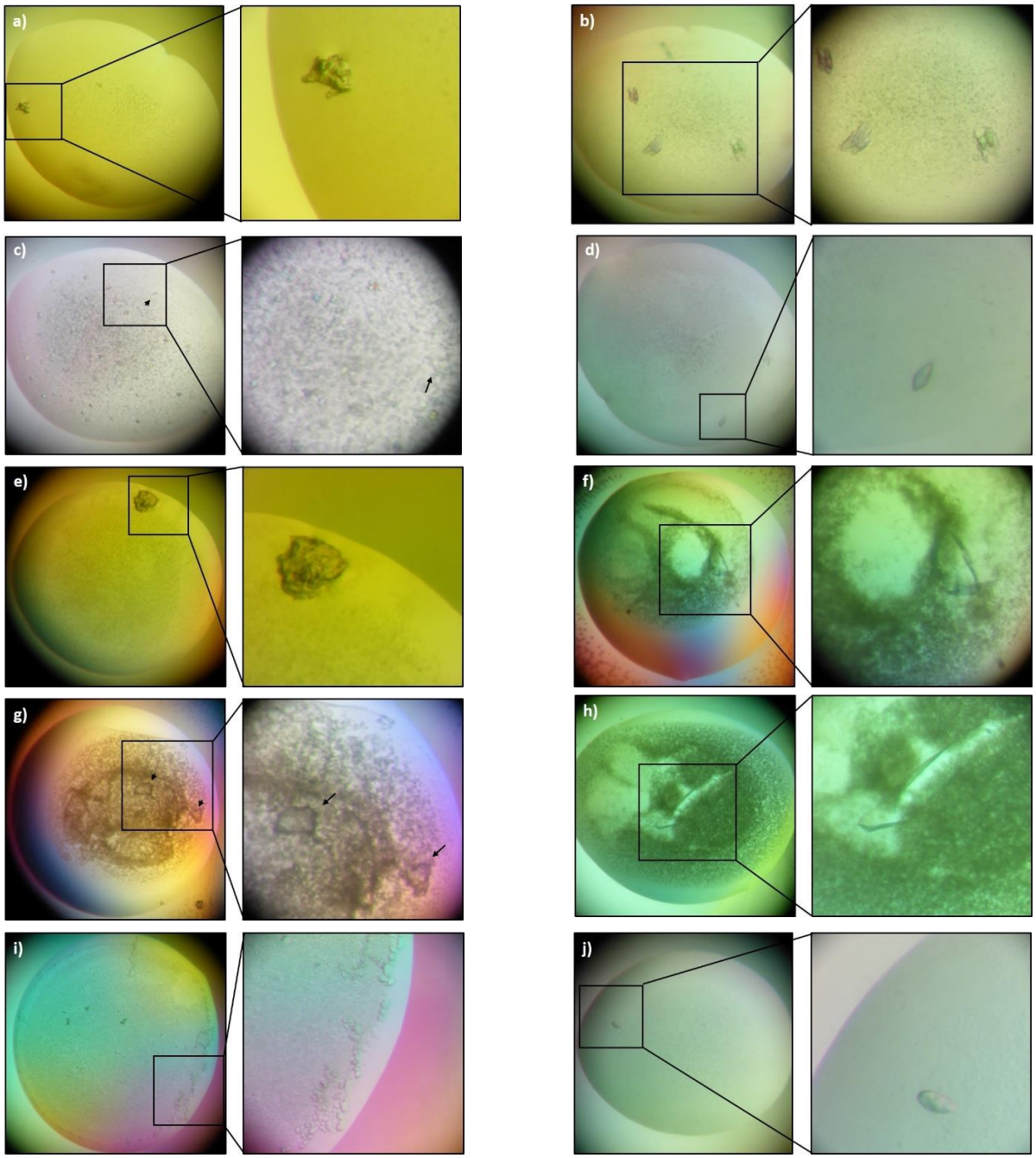

Figure 3.15: Crystals found in the second crystallisation screen, 15 days after the screen setup. a) one crystal in E12 condition of ASNA Screen (drop 2:1 ratio); b) three crystals found in C10 condition of MORPHEUS® Screen (drop 2:1 ratio); c) Over 50 crystals found in C12 condition of MORPHEUS® Screen (drop 2:1 ratio), the black arrow indicates one particular crystal that similarly shaped to the one found in H6 condition of the initial screen; d) one crystal found in D10 condition of MORPHEUS® Screen (drop 1:1 ratio); e) one crystal found in F6 condition of PACT® Screen (drop 1:1 ratio); f) one crystal found in C11 condition of PEG ION Screen (drop 1:1 ratio); g) three crystals found in C11 condition of PEG ION Screen (drop 2:1 ratio); h) one crystal found in D12 condition of PEG ION Screen (drop 2:1 ratio); i) Over 50 crystals found in G7 condition of PEG ION Screen (drop 2:1 ratio); j) one crystal found in B2 condition of PGA® Screen (drop 2:1 ratio); 50X amplification.

3.9. Optimization plates Nº 12-14

No crystals were found in these plates, even 1.5 months after they were setup.

A summary of the crystals obtained during the course of this project can be seen in Table 3.1. The data from the crystal with 3.2Å resolution was processed to try to solve the structure of KIFC1-MD without success, since the calculated R_{free} was not inferior to 30%.

Table 3.1: Summary of crystals obtained. The plate and conditions where the crystals were found and the protein batch used are shown. The results from the diffraction patterns are also summarized, indicating which crystals were protein- and salt-based and the maximum resolution obtained for the protein crystals.

Plate	Condition were crystal was found	Protein batch used	Salt/protein	Maximum resolution of protein crystals
First Crystallisation screen (nanodrops)	MORPHEUS® Screen B6		Protein	3.2Å
	MORPHEUS® Screen E6			4.5Å
	MORPHEUS® Screen H6			3.4Å
	RdRP Screen A1			>10Å
	RdRP Screen A2			>10Å
	RdRP Screen A3			>10Å
	RdRP Screen A6 and A7		Not measured	Not measured
Optimization Plate Nº 1 (microdrops)	B1 1:1 drop	1	Salt	----------
	A1 1:1 drop, A2 1:1 drop, B2 1:1 drop, C1 1:1 drop, C2 1:1 drop, D1 1:1 drop and D2 1:1 drop		Not measured	Not measured
	D2 2:1 drop		Protein	15-20Å
	D5 2:1 drop			15-20Å
	D6 2:1 drop			15-20Å
Second Crystallisation Screen (nanodrops)	RdRP Screen F10 2:1 drop, G12 2:1 drop, H10 2:1 drop, H11 2:1 drop; ASNA Screen E12 2:1; MORPHEUS® Screen C10 2:1 drop, C12 2:1 drop, D10 1:1 drop; PACT® Screen F6 1:1 drop; PEG ION Screen C11 1:1 drop and 2:1 drop, D12 2:1 drop; PGA® Screen B2 2:1 drop	4	Salt	----------
	PEG ION Screen G7 2:1 drop		Not measured	Not measured

3.6. ATPase Assay

The V_{max} and K_m values obtained were 0.009±0.0004 and 0.28±0.05, respectively.

The V_{max}, or maximal rate, is attained when the enzyme is saturated with substrate. The V_{max}, reveals the turnover number of the kinesin (k_{cat}), which is the number of substrate molecules converted into product by the kinesin per second when it is fully saturated. This is because k_{cat} equals the division of V_{max} by the concentration of protein in Molars (Berg et al., 2002). As such, the V_{max} value obtained reveals a turnover of ~1.6 ATP molecules per second (since the concentration of protein in the assay

was 5.58 µM). The turnover number of most enzymes with their physiological substrates ranges from 1 to 10^4 per second (Berg et al., 2002). As such, the k_{cat} value is within the expected parameters.

The Michaelis constant, K_M, represents the substrate concentration at which the reaction rate is half its maximal value (V_{max}). Thus, it reveals the substrate concentration required for significant catalysis to occur. The K_M values of enzymes vary extremely and depend on the substrate and on environmental conditions such as pH and temperature, but for most enzymes, it ranges from 10^{-1} to 10^{-7} M (Berg et al., 2002). However, the K_M obtained for KIFC1-MD was higher than this (0.28). Such a high K_M value implies a high concentration of substrate is needed for KIFC1-MD to have a significantly active ATPase activity. This is perhaps not surprising considering that the assay was performed in the absence of tubulin and, in the absence of MTs, kinesins tend to be less active since the ATPase cycle pauses at a trapped-ADP state (Cross and McAinsh, 2014). As such, a higher ATP (substrate) concentration must be necessary to escape this ADP-trapped state. Furthermore, It has been observed that unpolymerized tubulin heterodimers stimulate the ATPase activity of several KIFs (Cochran, 2015).

This assay proved that the protein was active, excluding the possibility that difficulty in reproducing the crystals was due to lack of activity of the purified KIFC1-MD in the different purification batches. Furthermore, determining the Michaelis-Menten steady state kinetic parameters (K_M and k_{cat}) is extremely important for achieving a robust and sensitive assay suitable for use in compound screening and drug discovery efforts. The Michaelis-Menten constants are critical in the interpretation of IC_{50}s determined for inhibitors used in the enzyme assay (Acker and Auld, 2014). As such, with this ATPase assay, the basal parameters for future screening of KIFC1-MD inhibiting compounds were established.

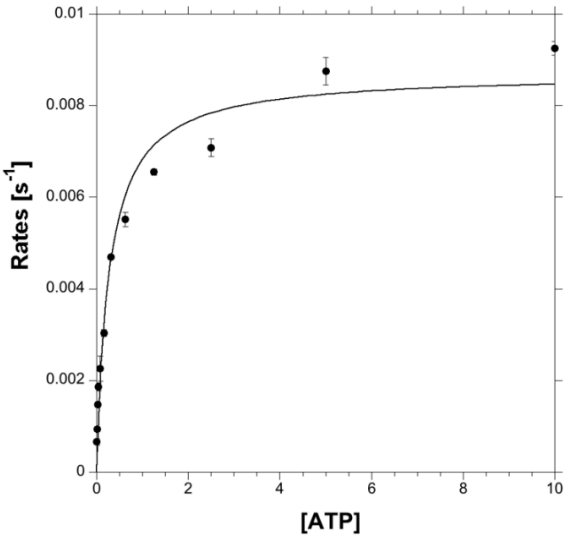

Figure 3.16: Characterization of the basal ATPase activity of KIFC1-MD. The average of the obtained OD/s [s^{-1}] is represented in the y axis as a function of the substrate (ATP) concentration, which is represented in the x axis. Error bars representing the standard deviation of the triplicates measured are presented.

4. Discussion

4.1. Sub-cloning, expression and purification of human KIFC1-MD

Low expression levels in *E. coli* may be caused by codon bias. Heterologous genes with abundancy of codons that are rare in *E. coli*, may not be efficiently expressed and may lead to translation errors like mistranslational amino acid substitutions, frameshifting events or premature translational termination. Usually, this can be dealt with by using strains with additional copies of rare transfer RNA (tRNA) genes (Derewenda, 2004; Khow and Suntrarachun, 2012; Rosano and Ceccarelli, 2014). However, in the present study, the KIFC1-MD (cloned in ppSUMO-2 vector) expression was largest in *E. coli* BL21 (DE3) cells, which do not have any extra tRNA expressing genes. As such, we can infer this is not the cause of lower expression in the *E. coli* Rosetta™ BL21 (DE3) cells transformed with ppSUMO-2-KIC1-MD or for absence of expression in all the cells transformed with the pDHT-2-KIFC1-MD.

The lower expression levels in *E. coli* Rosetta™ BL21 (DE3) cells transformed with ppSUMO-2-KIFC1-MD might be due to the fact that these cells have one extra plasmid, comparing to BL21 (DE3) cells, that allows them to express extra tRNAs and antibiotic resistance proteins (reason why they are resistant to chloramphenicol). Although these extra tRNAs can be useful to improve the protein's expression rate, this is only true if its' sequence contains rare codons (Derewenda, 2004; Khow and Suntrarachun, 2012). When this is not the case, the construct, in this case ppSUMO-2-KIFC1-MD, is in competition with this plasmid for the use of the transcription and translation machineries, causing lower expression. However, even though these also contain one extra plasmid, the BL21 Codon Plus (DE3) RIPL cells had almost as much expression as the BL21 (DE3) cells. This might be due to the inactivation of the gene encoding endonuclease I (*endA*), responsible for degrading pDNA, in these cells. This allows for the ppSUMO-2-KIFC1-MD construct to be present in the cells for a longer time and, therefore, express a higher amount of protein even when the competition for the transcription and translation machineries is present.

Another possible reason for low expression in *E. coli* are subtle issues involving mRNA secondary structure that inhibit its translation (Derewenda, 2004). It is possible that the sequences coding for the different tags used in the pDHT-2 vector, comparing to the ppSUMO-2 vector, introduced same kind of secondary structure in the mRNA. This structure might have prevented it from being translated and resulted in the absence of expression in all cells lines transformed with the pDHT-2-KIFC1-MD construct. Sequencing of these constructs is now in progress.

BL21 (DE3) cells and construct ppSUMO-2-KIFC1-MD proved to be a good choice for large-scale expression/purification of KIFC1-MD as the scaling-up from small-scale purification was successfully achieved, granting large amounts of protein at high purity (>90%).

4.2. Crystallisation

The KIFC1-MD crystals found six days after the setup of the first crystallisation screen, were found in very similar conditions: all had 0.1 M MORPHEUS® Buffer System 2 pH 7.5 and 50%

MORPHEUS® Precipitant Mix 2, being the only difference in the last solution present (0.09 M Halogens Mix in well B6, 0.12 M Ethylene Glycols Mix in well E6 and 0.1 M Aminoacids Mix in well H6). This could indicate that the first two factors (precipitant type and concentration and buffer's pH) are the most important, amongst these three factors, for the protein to form crystals at a fast rate. In fact, apart from the protein it-self, precipitant type and concentration and pH, have been considered the most important variables in a crystallisation trial, along with temperature (McPherson, 2004; McPherson and Gavira, 2014). On the other hand, the crystals that were found later on, were found to crystallize under very different conditions, which could explain the difference in time for the crystals to form. Although different from the conditions on the MORPHEUS® Screen, some of these conditions were similar amongst themselves. All three A1-A3 conditions of the RdRP Screen had 15% Glycerol and 3.655 M NaCl, being the only difference the buffer used: A1 – 0.05 M MES pH 5.0; A2 – 0.05 M Methylcyclopentadienyl manganese tricarbonyl (MMT) pH 5.5 in the A2 condition; and A3 – 0.05 M $(CH_3)2AsO_2Na$ pH 6.5. The A6 and A7 conditions are somewhat different from these, but similar to each other: both contain 24.5% Glycerol and 1.29 M $(NH4)_2SO_4$, the only difference being the pH of buffer used: both have 0.05 M TRIS but in A6 the pH is 7.5 and in A7 the pH is 8.0. These results appear to indicate that KIFC1-MD can form crystals under very different pH conditions when in the presence of glycerol and high salt concentrations. Furthermore, the difference in pH appears to allow the formation of higher number of crystals at a higher pH up to a certain point (pH 6.5), when in the presence of 15% glycerol and high salt concentration. Moreover, the crystals were all found in the drops that had high protein concentration (~16 mg/mL). 10 mg/mL is often referred as being the minimum concentration required for crystallisation of proteins. However, there are examples of successful crystallisation with concentrations as low as 2 mg/mL. The concentration required depends, therefore, on the individual protein (Bergfors, 2009). For KIFC1-MD, it appears that higher protein concentrations are required to reach the supersaturation state that is necessary for nucleation and growth to occur. Considering that in the exact same condition, crystals only formed at the highest protein concentration employed, it is possible to conclude that the protein concentration is the most important factor of all the factors involved. This is expected since it has long been recognized that, among all the factors that affect a crystallisation experiment, the biggest importance is by far attributed to the macromolecule (McPherson, 2004; McPherson and Gavira, 2014).

Several protein crystal morphologies were observed, even though some crystals were similarly shaped to one another. This is not an unexpected outcome since polymorphism is common for protein crystals. Probably this is due to their high conformational dynamicity and to the weakness of the lattice contacts between molecules in the crystal. In fact, differently shaped crystals are sometimes found coexisting in the same sample (McPherson, 2004; McPherson and Gavira, 2014). As one of the crystals in C12 condition of the MORPHEUS® Screen of the second crystallisation screen had a similar shape to the KIFC1-MD crystals previously found (Figure 3.15c), it could be expected for it to turn out as being a protein crystal. However, crystals of macromolecule and salt origins can be similar in morphology. On the other hand, there are significant differences between them. Salt crystals have firm lattice forces, are relatively highly ordered, physically hard and stiff, easily manipulated, can usually be exposed to air, have strong optical properties and diffract X-rays intensely. Contrarily, macromolecular crystals are

usually smaller, very soft and easily crushed, disintegrate when dehydrated, exhibit weak optical properties and diffract X-rays poorly (McPherson, 2004; McPherson and Gavira, 2014).

The low-resolution level of the diffraction patterns of protein crystals is because, on average, they consist of ~50% solvent, and the protein occupies the remaining volume. The high amount of fluid canals and solvent-filled craters are the primary reason for their limited resolution. Furthermore, because of the relatively large distance between adjacent molecules and the consequent weak lattice forces, the molecules may not exist in exactly similar orientations and positions inside the crystal. Also, due to their complex structures and their potential for dynamic conformational changes, in the same crystal, slight variations in the polypeptide chains or in the positions of lateral groups of the protein molecules may occur as well (McPherson, 2004; McPherson and Gavira, 2014). Although the low resolution allows to deduce the crystals are protein-based, the resolution is too low for determining the protein's structure. Resolution is measured in Angstroms (Å) and defines the amount of detail, or the minimum distance between structural features that can be distinguished in the electron-density maps. The lower the distance, the higher the resolution, which is desirable because it provides independent reflections that allow to determine the structure (Wlodawer et al., 2015). A typical diffraction resolution is 2.0–2.5Å (Durbin and Feher, 1996; Wlodawer et al., 2015). Since, the resolution for all the crystals from the first crystallisation screen was >3Å, optimization was required.

Regarding the second crystallisation screen, a failure in obtaining any protein crystals in a crystallisation screen is also not unexpected. The success rates of crystallisation screening revealed by the statistics assembled by various Structural Genomics Centres range from 10-30% for small prokaryotic proteins to only a few percent for a wide variety of eukaryotic proteins, including human proteins (Derewenda, 2011).

As the crystals in well D6 of optimization plate Nº 1 appeared only after four days, this condition seemed to allow the reproduction of the crystals found in the first screen at a fast rate. As drop size increases (from 200 nL to 3 µL), different conditions appeared to benefit the formation of KIFC1-MD crystals: 1) The crystals were found in a hanging drop (which contrasts with the initial screening were all drops were sitting drops). 2) The final protein concentration was increased by 1/6 (in the initial screen the ratio protein:well solution was 1:1 and in the optimization drop where crystals were found the ratio was 2:1); 3) The final precipitant concentration was 5% lower (in the initial screening the well solution has 50% MORPHEUS® Precipitant Mix 2, which means in the 1:1 drop the concentration was 25%; in the Linbro plate well D6 the precipitant concentration was 60%, which means that in the 2:1 drop the concentration was 20%). This is not unexpected since the best method for screening and obtaining an initial crystallisation condition is not always the best means for optimization (McPherson, 2004; McPherson and Gavira, 2014). Sitting and hanging-drop trials differ in drop support and geometry of the setup (Bergfors, 2009). In this case, at large-scale, the support and geometry associated to hanging drops appear to be beneficial comparing to sitting drops and, as such, further optimization trials only included hanging drops (Table 2.4). Since even a small difference in the geometry, such as the angle of inclination of a support, can disturb the vapor-phase equilibration rate, it is expected to find differences between hanging and sitting drop experiments. A slower equilibration rate can usually be seen in sitting drops comparing to hanging drops in equivalent trials (Bergfors, 2009), so it appears that in this case a

faster equilibration process is beneficial for the formation of KIFC1-MD crystals. On one hand, even higher protein concentration than before appears to be required for the formation of crystals in larger drops, since there were no crystals found in the drops with 1:1 ratio. On the other, in Figure 3.12, it seems that a lot of very small crystals were trying to grow around the three indicated ones, which could mean that too many crystals were trying to form, reducing the space for the growth of each one, inhibiting the formation of more crystals and of larger crystals. This could indicate that the protein concentration was too high in these drops. As such, maybe the ideal concentration would be somewhere in between the two ratios experimented in this trial. Therefore, in the next optimization trials, a 1.5:1 protein:well solution ratio was experimented along with the 1:1 and 2:1 ratios (Table 2.4). Furthermore, since the crystals only formed at the highest precipitant concentration tested in the trial, higher precipitant concentration seemed to be preferred for the crystallisation of KIFC1-MD. Additionally, the lack of alanine in the Amino acids Mix does not appear to influence greatly the crystallisation of KIFC1-MD, since in its absence crystals still formed. This is consistent with the conclusion made previously that the precipitant type and concentration and buffer's pH are much more important factors. Nonetheless, it is possible its absence might have made a difference in the quality of the crystals. Furthermore, because the crystals in D2 (condition that contains Halogens) and D5 (condition that contains amino acids at a lower concentration comparing to the D6 condition) needed a longer time to form, it is possible to deduce that the amino acids are playing a role in the rate of nucleation and growth of crystals. Moreover, because in the D6 well an extra crystal was formed after the collection of the three first ones (and addition of cryoprotectant PEG 400), it is possible that, in this case, the addition of PEG 400 compensated the dilution of the protein and helped it reach the supersaturated state, even at a lower concentration. PEGs are well known additives in crystallisation and PEGs with molecular weights from 400 to 20000 have successfully provided protein crystals in the past (McPherson, 2004; McPherson and Gavira, 2014). Nevertheless, the quality of all these KIFC1-MD crystals was still poor, and further optimization was required.

Apart from the first optimization plate, no other crystallisation trial successfully reproduced the crystals found in the MORPHEUS® Screen. Furthermore, the crystals found in the RdRP Screen of the first crystallisation screen did not reveal to be more easily reproducible. There are many factors that affect the crystallisation of proteins, which are summarised in Table 4.1. These may affect the probability of it happening at all, the nucleation and crystal-growth likelihood and rates and/or the final sizes and quality of the crystals. In general, it is not possible to predict which of the many variables is important for a particular protein, and the suspected influence of each needs to be assessed empirically (McPherson, 2004; McPherson and Gavira, 2014). As such, several hypotheses can be thought to explain the difficulty in reproducing the KIFC1-MD crystals:

1) pH, salt and the concentrations of precipitants are very important (McPherson and Gavira, 2014). Therefore, one possibility could be a different pH in optimization plate Nº 2 – if by any chance overtime the pH of the stock MORPHEUS® Buffer System 2 was altered, this could explain the lack of crystals in this trial. However, the remaining trials made with new buffer exclude this premise.

2) Another possibility is that the range of concentrations for the MORPHEUS® Precipitant Mix 2 experimented in optimization plate Nº 2 was too wide and the differences in concentrations between

each column was too high. This too was rejected as next optimization trials, which had 1% difference (instead of 2%) in the precipitant concentration, also failed to reproduce the crystals.

Table 4.1: Physical, chemical and biochemical factors affecting crystallisation. Adapted from McPherson and Gavira, 2014

Physical	Chemical	Biochemical
1. Temperature	1. pH	1. Purity of the macromolecule
2. Methodology	2. Precipitant type	2. Ligands, inhibitors, effectors
3. Mother liquor volume	3. Final precipitant concentration	3. Aggregation state
4. Geometry of chamber or capillary	4. Ionic strength	4. Post-translational modifications
5. Gravity	5. Cation type and concentration	5. Source of macromolecule
6. Pressure	6. Anion type and concentration	6. Proteolysis / hydrolysis
7. Time	7. Degree of supersaturation	7. Chemical or genetic modifications
8. Vibration/mechanical perturbation	8. Reductive / oxidative environment	8. Inherent macromolecule symmetry
9. Electrostatic / magnetic fields	9. Macromolecule concentration	9. Degree of denaturation
10. Dielectric properties of the medium	10. Metal ions	10. Isoelectric point
11. Viscosity of the medium	11. Initial precipitant concentration	11. Unsaturated regions
12. Rate of equilibration	12. Cross-linkers / polyions	12. His tags, purification tags
	13. Detergents	13. α-Helical content
	14. Non-macromolecular impurities	14. Conformational states
		15. Thermal stability
		16. Allowable pH range
		17. History of the sample

3) Small molecules (additives) can affect dramatically the success of protein crystallisation (McPherson and Gavira, 2014). Even though it did not appear to make a difference in the first optimization plate, there is still a possibility that the cause for the lack of crystals might be due to the absence of alanine in the Amino acids Mix. However, the presence of alanine in the last repetition plates, that did not yield any crystals either, seems to indicate that, even if it was one of the factors impeding crystallisation, it was not the main factor and not enough to lead to crystal formation. On the other hand, the concentration of the same might have just been too low in optimization plate Nº 2 (considering that the highest concentration used in this trial was 0.19 M, which, even if not by a lot, is still lower than the 0.2 M that was present in the condition were crystals were found in optimization plate Nº 1). But, once again, the next optimization trials with higher Amino acids mix concentration indicates this was not the case.

4) On the count of additives, particular importance is given to physiologically or biochemically relevant small molecules like coenzymes, substrate analogues, inhibitors, metal ions, etc. They bind to the active sites of enzymes, or at specific sites elsewhere in the molecule, and often promote stability and homogeneous conformations or induce conformational changes that facilitate molecule-molecule interactions (Luft et al., 2014; McPherson and Gavira, 2014). The lack of the additives ADP and $MgCl_2$ in the protein batch Nº 2 could very probably be the cause for lack of crystal formation. As referred in section 1.1.2., the Mg^{2+} ion of the active site is important for establishing an ADP-trapped state, in which Mg^{2+}ADP is stably bound to the active site. Since, in absence of MTs, the ATPase cycle pauses at this state, purified KIFs (and most kinesins crystal structures) tend to retain Mg^{2+}ADP in their active sites (Cross and McAinsh, 2014). For this reason, it is quite possible that, if the protein does not bind to Mg^{2+}ADP inside the *E. coli* cells, the $MgCl_2$ and ADP added to the protein after the purification are essential for the protein to acquire a stable state. Stability of the molecule is an important factor affecting crystallisation (Luft et al., 2014; McPherson and Gavira, 2014). Nevertheless, the next crystallisation

trials in which ADP or $MgCl_2$ were added to the well solutions so as to assert their role in the crystallisation process, also did not lead to KIFC1-MD crystallisation. The same is true for following optimization trials with protein batches that contained ADP and $MgCL_2$. As such, it is possible to conclude that these molecules absence was not the main factor hindering the crystals' growth, although it is still likely that these molecules play a role in the process.

5) Another possible cause was experimental error in the pipetting of the solutions when preparing the individual well conditions. To exclude this possibility the optimization plates Nº 1 and Nº 3 were repeated (Table 2.4). As these repetition plates also did not produce any crystals, this hypothesis was rejected.

6) Furthermore, increasing the incubation temperature to 10ºC (optimization plate Nº 4) did not appear to make any difference.

7) Many researchers state that if crystallisation is irreproducible, the protein sample is simply lacking sufficient purity (Luft et al., 2014; Weber, 1991), since probability of success in crystallisation trials is highly augmented by increased homogeneity of the sample (Khurshid et al., 2014; McPherson and Gavira, 2014). There can be various causes of impurity, such as natural isoforms, fragments resulting from proteolytic reactions, misfolding, random oligomerization or simply co-elution of unrelated proteins (Khurshid et al., 2014). Although the purity of the protein solution used in the first optimization trials was high (~93%), it is still possible that the few contaminants that are present induced proteolysis of the protein overtime, resulting in lack of crystals in optimization plate Nº 2. If this was the case, a new batch of the protein could've potentially been enough to solve the issue. However, repetition of optimization plates with new protein batches did not result in crystals. This could be because some of those batches had lower purity (~83-85%, i.e. 8-10% less pure than the first batch; sections 7.5 and 7.6), but since the last repetition plates were setup with protein batches of very high purity (~90%) (sections 7.7. and 7.8.), this also appeared not to be the main factor for the crystals irreproducibility.

8) The simple fact that the optimization experiments were made with different batches of protein might be the cause of the lack of crystals. Different batches of protein commonly show different crystallisation behaviours: a second batch, even if prepared from the same construct, may crystallize in conditions where the first one does not or vice versa (Bergfors, 2009; Luft et al., 2014). Proteins gain all kinds of hitchhikers (such as cofactors, detergents, lipids, or membrane components), often through their biding sites, during purification that vary from batch to batch (Bergfors, 2009). Moreover, batch variation is probably the most common cause for failure to successfully optimize crystals after the initial screening hits (Bergfors, 2009; Luft et al., 2014). Consistently, even for the KIFC1-MD batches with similar purity levels to the first batch (batches Nº 4 and 5), it is possible to see differences in the SDS-PAGE gels from the gel filtration purifications: on the first batch, most contamination bands are of lower molecular size than KIFC1-MD, while in the last the reverse is true (see sections 3.4.3., 7.7. and 7.8). The different proteins present in the solution, even if they are present at low concentrations, might be influencing the crystallisation success, particularly if these proteins are interacting with the target KIFC1-MD in any way. Batch variation is, therefore the most likely cause for the difficulty in reproducing the KIFC1-MD crystals both in the optimization trials and in the second crystallisation screen (where the protein did not form any crystals, not even in the same conditions as it had in the first screen).

9) The attempt to help the protein become less soluble and potentially reach the supersaturated state by lowering the salt concentration also did not result in any crystals and, as such, was not what the protein required to be able to crystallize. The protein itself, instead of the precipitating agent, is the most important variable in the crystallisation process (Dale et al., 2003; Luft et al., 2014).

10) Finally, one last hypothesis was that, it was possible that some problem had occurred during the expressions/purifications of the different batches that had affected the proper folding of the protein, hindering the crystallisation success. This premise was tested with the ATPase assay – if the protein was not properly folded it would likely not be active (Rosano and Ceccarelli, 2014). As it proved that the protein was active, this possibility was also excluded.

4.3. Final remarks

More than 60 structures of kinesins motor domain are known from 9 classes of the kinesin superfamily (Marx et al., 2009). Till date, only one unpublished structure for KIFC1-MD is available in the PDB (Figure 4.1) (Zhu et al.). The fact that only one structure is available might reflect the difficulty in obtaining crystals and high-resolution structure of KIFC1-MD. Consistent with this idea, a group reported that an inhibitor-bound KIFC1 X-ray crystal structure proved problematic to obtain (Yang et al., 2014). This contrasts with the observation that many proteins that have not crystallized in their native state could be readily crystallized as complexes (with cofactors, inhibitors, or even antibody fragments). This is due to the conformational changes induced that were favourable to crystallisation by exposing new possible molecule-molecule contacts or by stabilizing the protein (Dale et al., 2003).

As demonstrated during this project, crystallisation constitutes a bottleneck for structure-based drug-design. Statistics from the numerous structure–genomics programs show divergent success rates from the cloned protein to the structure determination. For example, during the Human Proteome Structural Genomics pilot project, of the 120 proteins expressed, 19 (15.8%) yielded crystals suitable for structure determination. The program at the Berkeley Structural Genomics Centre obtained similar results, with 19.4% of the soluble expressed proteins providing diffraction-quality crystals. Varying degrees of success were found in other programs, from 31.2% for identifying crystallisation conditions of the expressed proteins to 3.1% of the proteins providing diffraction data. Most programs obtained diffraction-quality crystals for ~10% of the expressed proteins. Awareness is rising that the success of a crystallisation trial is not directly proportional to the number of conditions tested (Dale et al., 2003). The protein itself is the most important variable in the screening process. Even in cases where the protein is both soluble and suitable for crystallisation many times it proves difficult or impossible to crystallize. For this reason, since 1972 it has been proposed that if crystals of a target protein cannot be grown, homologous proteins from another species can be alternatives. Generally, these proteins have small differences in only a few amino acid residues, often on the surface of the protein; nevertheless, the effect on crystallisation may be drastic. For example, some proteins crystallize rather easily and under various conditions (e.g., porcine pancreatic elastase, *Staphylococcus aureus* DHNA, *S. aureus* DHFR), while their homologues proved to be difficult to crystallize (e.g., human neutrophil elastase, *E. coli* DHNA, *Streptococcus pneumoniae* DHFR). Those proteins that do not crystallize readily may have surface properties that do not favour molecule-molecule contacts (Dale et al., 2003). Consistent with

this idea, at least 9 Ncd (*Drosophila melanogaster* homolog of human KIFC1) structures are available in the PDB, while only one unpublished structure for the motor domain of human KIFC1 is available.

The KIFC1-MD structure available in the PDB (Figure 4.1) was obtained with a protein expressed in insect cells (Zhu et al.). *E. coli* has several advantages as an expression system comparing to insect cells: it is easily manipulated, inexpensive, has fast growth and protein production rates and has a variety of vectors for expression of numerous fusion proteins (Cain et al., 2014; Derewenda, 2004; Khow and Suntrarachun, 2012). Nevertheless, *E. coli* has limited eukaryotic post-translational machinery function (Khow and Suntrarachun, 2012), which can translate in differences on the activity of the final protein, including biomolecule interactions (Cain et al., 2014). As such, the fact that the only structure available was obtained with a protein produced in an eukaryotic expression system might imply that the protein undergoes post-translational modifications when in an eukaryotic organism that it does not in an bacterial system and that might favour the crystallisation process.

It remains important to find a condition to obtain reproducible high quality crystals of KIFC1-MD. One possibility on going forward can be to produce a modified protein construct with single or multiple point mutations, specially of surface residues, hoping the new protein will crystallize (Chayen and Saridakis, 2008; Derewenda, 2011; McPherson and Gavira, 2014). Furthermore, the sequence alignment, using the BLAST® tool of the NCBI website, between the crystallized KIFC1-MD in the PDB (PDB ID: 2REP) and the predicted amino acid sequence for the KIFC1-MD purified during this project revealed a similarity of 99% (Attachments section 7.11). The difference was found on amino acid 65 of the predicted sequence, where it had threonine instead of a proline, like in the sequence from the KIFC1-MD structure in the PDB. As such, this amino acid may be playing a key role in the crystallisation process and the point mutation of its coding codon might increase the propensity for crystal growth. However, there are no sure shot ways to be successful in obtaining crystals. Moreover, this strategy demands lot of time, effort and expense, with no guarantee that crystals will be obtained (Chayen and Saridakis, 2008). On the other hand, various studies demonstrated that only a limited number of mutations is necessary to improve the yield and/or quality of crystals. However, although there are many examples of mutagenesis as an instrument for improving crystallisation, it has not been used broadly, probably because there are no established protocols regarding the type of mutation that is more worth pursuing and defining a rational strategy is no trivial task (Dale et al., 2003).

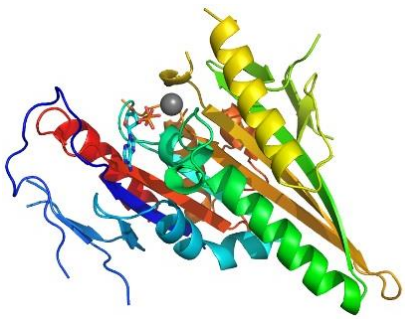

Figure 4.1: Structure of the motor domain of KIFC1 in complex with Mg^{2+} and ADP (PDB ID: 2REP). The motor domain of KIFC1 adapts a similar folding topology to other kinesin motor domains: the three-layer (αβα) sandwich architecture. Three α-helices are at each side of an eight β-stranded sheet. The Mg^{2+} is represented by the grey sphere. Surface coloured with a colour spectrum from N-terminus (blue) to C-terminus (red). Adapted from Zhu et al.

5. Conclusion

KIFC1 is an interesting cancer therapeutic target because of its non-essential role in normal cell division in contrast to its crucial role for multiplication of cancer cells with supernumerary centrosomes.

In this project, I sub-cloned the DNA coding for the human KIFC1 motor domain into two expression vectors and developed a protocol for small-scale followed by large-scale expression and purification to high purity.

As high quality crystals proved difficult to obtain, it was not possible to determine the structure of the target protein. As such, it is still necessary to optimize crystallisation conditions to obtain reproducible high quality KIFC1-MD crystals. Once this bottleneck is surpassed, it will be possible to move forward with structure-based drug-design to find inhibitors for KIFC1. Various avenues can be tried to obtain better crystals and determine the structure of KIFC1-MD.

Once the native protein's structure in known, new and previously identified inhibitors can be used in co-crystallisation attempts of KIFC1-MD bound to inhibitors. These structures will pinpoint the binding-site and key interactions of the protein with the inhibitors, which can be used to develop strategies for making rational modifications for future drug-design.

6. References

Acker, M.G., and Auld, D.S. (2014). Considerations for the design and reporting of enzyme assays in high-throughput screening applications. Perspect. Sci. *1*, 56–73.

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes. In Biochemistry, (New York: W H Freeman).

Bergfors, T. (2009). Protein Crystallization (CA, USA: International University Line).

Braun, M., Lansky, Z., Bajer, S., Fink, G., Kasprzak, A.A., and Diez, S. (2013). The human kinesin-14 HSET tracks the tips of growing microtubules in vitro. Cytoskeleton *70*, 515–521.

Cai, S., Weaver, L.N., Ems-McClung, S.C., and Walczak, C.E. (2009). Kinesin-14 Family Proteins HSET/XCTK2 Control Spindle Length by Cross-Linking and Sliding Microtubules. Mol. Biol. Cell *20*, 1348–1359.

Cain, J.A., Solis, N., and Cordwell, S.J. (2014). Beyond gene expression: The impact of protein post-translational modifications in bacteria. J. Proteomics *97*, 265–286.

Chandra, R., Salmon, E.D., Erickson, H.P., Lockhart, A., and Endow, S.A. (1993). Structural and functional domains of the Drosophila ncd microtubule motor protein. J. Biol. Chem. *268*, 9005–9013.

Chayen, N.E., and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. Nat. Methods *5*, 147–153.

Cochran, J. (2015). Kinesin Motor Enzymology: Chemistry, Structure, and Physics of Nanoscale Molecular Machines. Biophys. Rev. *7*.

Congreve, M., Murray, C.W., and Blundell, T.L. (2005). Keynote review: Structural biology and drug discovery. Drug Discov. Today 895–907.

Cross, R. a., and McAinsh, A. (2014). Prime movers: the mechanochemistry of mitotic kinesins. Nat. Rev. Mol. Cell Biol. *15*, 257–271.

Dale, G.E., Oefner, C., and Dõarcy, A. (2003). The protein as a variable in protein crystallization. J. Struct. Biol. *142*, 88–97.

Derewenda, Z.S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. Methods *34*, 354–363.

Derewenda, Z.S. (2011). It's all in the crystals... Acta Crystallogr. *67*, 243–248.

Durbin, S.D., and Feher, G. (1996). Protein Crystallization. Annu. Rev. Phys. Chem *47*, 171–204.

Endow, S.A., Kull, F.J., and Liu, H. (2010). Kinesins at a glance. J. Cell Sci. *123*, 3420–3424.

Farina, F., Pierobon, P., Delevoye, C., Monnet, J., Dingli, F., Loew, D., Quanz, M., Dutreix, M., and Cappello, G. (2013). Kinesin KIFC1 actively transports bare double-stranded DNA. Nucleic Acids Res. *41*, 4926–4937.

Gregan, J., Polakova, S., Zhang, L., Tolić-Nørrelykke, I.M., and Cimini, D. (2011). Merotelic kinetochore attachment: Causes and effects. Trends Cell Biol. *21*, 374–381.

Gross, S.P. (2004). Hither and yon: a review of bi-directional microtubule-based transport. Phys. Biol. *1*, 1–11.

Hirokawa, N. (2011). From electron microscopy to molecular cell biology, molecular genetics and structural biology: Intracellular transport and kinesin superfamily proteins, KIFs: Genes, structure, dynamics and functions. J. Electron Microsc. (Tokyo). *60*, 63–92.

Hirokawa, N., and Takemura, R. (2004). Kinesin superfamily proteins and their various functions and dynamics. Exp. Cell Res. *301*, 50–59.

Hirokawa, N., and Tanaka, Y. (2015). Kinesin superfamily proteins (KIFs): Various functions and their relevance for important phenomena in life and diseases. Exp. Cell Res. *334*, 16–25.

Hirokawa, N., Noda, Y., Tanaka, Y., and Niwa, S. (2009). Kinesin superfamily motor proteins and intracellular transport. Nat. Rev. Mol. Cell Biol. *10*, 682–696.

Horecker, B.L., and Kornberg, A. (1948). The extinction coeficients o the reduced band of pyridine nucleotides. J.

Biol. Chem. *175*, 385–390.

Jana, B., Hyeon, C., and Onuchic, J.N. (2012). The Origin of Minus-end Directionality and Mechanochemistry of Ncd Motors. PLoS Comput. Biol. *8*.

Khow, O., and Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pac. J. Trop. Biomed. *2*, 159–162.

Khurshid, S., Saridakis, E., Govada, L., and Chayen, N.E. (2014). Porous nucleating agents for protein crystallization. Nat. Protoc. *9*, 1621–1633.

Kolomeisky, A.B. (2013). Motor proteins and Molecular Motors: How to Operate Machines at Nanoscale. J Phys Condens Matter *25*, 1–23.

Kozielski, F. (2015). Kinesins and cancer. Kinesins and Cancer *12*, 1–271.

Kull, F.J., and Endow, S.A. (2013). Force generation by kinesin and myosin cytoskeletal motor proteins. J. Cell Sci. *126*, 9–19.

Kwon, M., Godinho, S.A., Chandhok, N.S., Ganem, N.J., Azioune, A., Thery, M., and Pellman, D. (2008). Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. Genes Dev. *22*, 2189–2203.

Li, Y., Lu, W., Chen, D., Boohaker, R.J., Zhai, L., Padmalayam, I., Wennerberg, K., Xu, B., and Zhang, W. (2015). KIFC1 is a novel potential therapeutic target for breast cancer. Cancer Biol. Ther. *16*, 1316–1322.

Luft, J.R., Newman, J., and Snell, E.H. (2014). Crystallization screening: The influence of history on current practice. Acta Crystallogr. Sect. FStructural Biol. Commun. *70*, 835–853.

Marx, A., Hoenger, A., and Mandelkow, E. (2009). Structures of kinesin motor proteins. Cell Motil. Cytoskeleton *66*, 958–966.

McPherson, A. (2004). Introduction to protein crystallization. Methods *34*, 254–265.

McPherson, A., and Gavira, J. (2014). Introduction to protein crystallization. Acta Crystallogr. *70*, 2–20.

Miki, H., Okada, Y., and Hirokawa, N. (2005). Analysis of the kinesin superfamily: Insights into structure and function. Trends Cell Biol. *15*, 467–476.

Milic, B., Andreasson, J.O.L., Hancock, W.O., and Block, S.M. (2014). Kinesin processivity is gated by phosphate release. Proc. Natl. Acad. Sci. *111*, 14136–14140.

Mittal, K., Choi, D.H., Klimov, S., Pawar, S., Kaur, R., Mitra, A.K., Gupta, M. V, Sams, R., Cantuaria, G., Rida, P.C.G., et al. (2016). A centrosome clustering protein, KIFC1, predicts aggressive disease course in serous ovarian adenocarcinomas. J. Ovarian Res. *9*, 1–11.

Novagen pET-28a-c(+) Vectors. https://plasmid.med.harvard.edu/PlasmidRepository/file/map/pET28.pdf. Date Accessed: 25-08-2017.

Pannu, V., Rida, P.C.G., Ogden, A., Turaga, R.C., Donthamsetty, S., Bowen, N.J., Rudd, K., Gupta, M. V, Reid, M.D., Cantuaria, G., et al. (2015). HSET overexpression fuels tumor progression via centrosome clustering-independent mechanisms in breast cancer patients. Oncotarget *6*, 6076–6091.

Pawar, S., Donthamsetty, S., Pannu, V., Rida, P., Ogden, A., Bowen, N., Osan, R., Cantuaria, G., and Aneja, R. (2014). KIFCI, a novel putative prognostic biomarker for ovarian adenocarcinomas: delineating protein interaction networks and signaling circuitries. J. Ovarian Res. *7*, 1–9.

Roberts, A.J., Kon, T., Knight, P.J., Sutoh, K., and Burgess, S.A. (2013). Functions and mechanics of dynein motor proteins. Nat. Rev. Mol. Cell Biol. *14*, 713–726.

Roos, W.P., and Kaina, B. (2006). DNA damage-induced cell death by apoptosis. Trends Mol. Med. *12*, 440–450.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. Front. Microbiol. *5*, 1–17.

Sablin, E.P. (2000). Kinesins and microtubules: their structures and motor mechanisms. Curr. Opin. Cell Biol. *12*, 35–41.

Sauder, M.J., Rutter, M.E., Bain, K., Rooney, I., Gheyi, T., Atwell, S., Thompson, D. a, Emtage, S., and Burley, S.K.

(2008). High throughput protein production and crystallization at NYSGXRC. In Methods in Molecular Biology: Structural Proteomics: High-Throughput Methods, B. Kobe, M. Guss, and T. Huber, eds. (Human Press, Totowa, NJ), pp. 561–575.

Schief, W.R., and Howard, J. (2001). Conformational changes during kinesin motility. Curr. Opin. Cell Biol. *13*, 19–28.

Schliwa, M., and Woehlke, G. (2003). Molecular motors. Nature *422*, 759–765.

Seog, D.H., Lee, D.H., and Lee, S.K. (2004). Molecular Motor Proteins of the Kinesin Superfamily Proteins (KIFs): Structure, Cargo and Disease. J. Korean Med. Sci. *19*, 1–7.

Szczęsna, E., and Kasprzak, A.A. (2012). The C-terminus of kinesin-14 Ncd is a crucial component of the force generating mechanism. FEBS Lett. *586*, 854–858.

Weber, P.C. (1991). Physical Principles of Protein Crystallization. Adv. Protein Chem. *41*, 1–36.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*, 235–242.

Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M., and Physics, B. (2015). Protein crystallography for non-crystallographers, or how to get the (but not more) from published macromolecular structures. Febs J. *275*, 1–21.

Wu, J., Mikule, K., Wang, W., Su, N., Petteruti, P., Gharahdaghi, F., Code, E., Zhu, X., Jacques, K., Lai, Z., et al. (2013). Discovery and mechanistic study of a small molecule inhibitor for motor protein KIFC1. ACS Chem. Biol. *8*, 2201–2208.

Xiao, Y.-X., and Yang, W.-X. (2016). KIFC1: a promising chemotherapy target for cancer treatment? Oncotarget *7*, 48656–48670.

Yang, B., Lamb, M.L., Zhang, T., Hennessy, E.J., Grewal, G., Sha, L., Zambrowski, M., Block, M.H., Dowling, J.E., Su, N., et al. (2014). Discovery of potent KIFC1 inhibitors using a method of integrated high-throughput synthesis and screening. J. Med. Chem. *57*, 9958–9970.

Yu, Y., and Feng, Y.M. (2010). The role of kinesin family proteins in tumorigenesis and progression. Cancer *116*, 5150–5160.

Zhu, H., Tempel, W., He, H., Shen, Y., Wang, J., Brothers, G., Landry, R., Arrowsmith, C., Edwards, A., and Park, H. Motor domain of human kinesin family member C1 PBD Structure: 2REP.

7. Attachments

7.1. Vectors' details

pDHT-2 and ppSUMO-2 vectors are modified versions from the pET28a vector (Figure 7.1), which has an IPTG inducible promoter (T7 promoter followed by *lacO* operator sequence), a *lacI* repressor coding sequence, a Kanamycin resistance gene (*KanR*) and a multiple cloning-site (MCS) that includes sequences for cleaving with *NcoI* (5'C|CATG3') and *XhoI* (5'C|TCGAG3').

pDHT-2 has a sequence for a double His tag (each with six histidines, with a linker between the two His tags), meaning the target protein will have this double His-tag at its N-terminus.

ppSUMO-2 has a sequence for a 8 histidine His-tag and a sequence for a SUMO tag and respective the Gly-Gly motif that is the Ulp1 cleaving site, also positioning the tags at the N-terminus of the protein.

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites	
(*Bam*H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

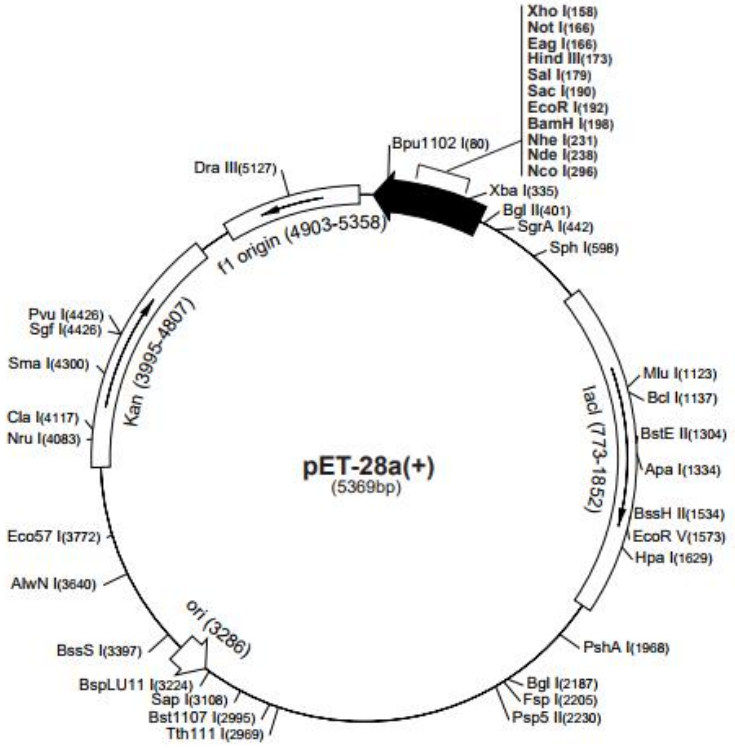

Figure 7.1: Scheme representing vector pET28a. It includes a replication origin (pBR32), a Kanamycin resistance gene, a *lacI* gene, a T7 promoter associated to a *lacO* operator, a multiple cloning site and a T7 terminator sequence. Adapted from Novagen.

7.2. Amino acids sequences of the purified KIFC1-MD before and after cleavage

The predicted amino acids sequence coded in the codon optimized insert sub-cloned into the vector ppSUMO-2 is as follows:

With the His and SUMO tags:

MGSSHHHHHHHHSSGLVPRGSHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKT
TPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGTMGLKGNIRVFC
RVRPVLPGEPTPPPGLLLFPSGPGGPSDPPTRLSLSRSDERRGTLSGAPAPPTRHDFSFDRVFPPGS
GQDEVFEEIAMLVQSALDGYPVCIFAYGQTGSGKTFTMEGGPGGDPQLEGLIPRALRHLFSVAQELS
GQGWTYSFVASYVEIYNETVRDLLATGTRKGQGGECEIRRAGPGSEELTVTNARYVPVSCEKEVDAL
LHLARQNRAVARTAQNERSSRSHSVFQLQISGEHSSRGLQCGAPLSLVDLAGSERLDPGLALGPGER
ERLRETQAINSSLSTLGLVIMALSNKESHVPYRNSKLTYLLQNSLGGSAKMLMFVNISPLEENVSESLN
SLRFASKVNQC

The His and SUMO tags are highlighted in red and green, respectively; in blue is highlighted the Gly-Gly motif recognized by the Ulp1 protease.

After the tags are cleaved:

TMGLKGNIRVFCRVRPVLPGEPTPPPGLLLFPSGPGGPSDPPTRLSLSRSDERRGTLSGAPAPPTRH
DFSFDRVFPPGSGQDEVFEEIAMLVQSALDGYPVCIFAYGQTGSGKTFTMEGGPGGDPQLEGLIPRA
LRHLFSVAQELSGQGWTYSFVASYVEIYNETVRDLLATGTRKGQGGECEIRRAGPGSEELTVTNARY
VPVSCEKEVDALLHLARQNRAVARTAQNERSSRSHSVFQLQISGEHSSRGLQCGAPLSLVDLAGSER
LDPGLALGPGERERLRETQAINSSLSTLGLVIMALSNKESHVPYRNSKLTYLLQNSLGGSAKMLMFVNI
SPLEENVSESLNSLRFASKVNQC

7.3. *E. coli* cells details

T7 expression strains are lysogens of bacteriophage DE3, as indicated by the (DE3). These hosts carry a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter (a mutated version of the *lac* promoter that allows higher expression). Such strains are suitable for production of protein from target genes cloned in appropriate T7 expression vectors, using IPTG as an inducer. Furthermore, these strains are deficient in lon and ompT proteases.

Rosetta™ host strains are BL21 derivatives designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*. These strains supply tRNAs for AGG, AGA, AUA, CUA, CCC, GGA codons on a compatible chloramphenicol-resistant plasmid.

Rosetta™ BL21 (DE3) pLysS strains express T7 lysozyme, which further suppresses basal expression of T7 RNA polymerase prior to induction, thus stabilizing pET recombinants encoding target proteins that affect cell growth and viability. In Rosetta™ BL21 (DE3) pLysS, the rare tRNA genes are present on the same plasmid that carries the T7 lysozyme gene.

BL21 Codon Plus (DE3) RIPL cells contain extra copies of rare *E. coli* tRNA genes for AGA, AGG, AUA, CUA, and CCC codons in a streptomycin and chloramphenicol-resistant plasmid. These cells have the high transformation efficiency (*Hte*) allele, which helps increase transformation efficiency to > 1 x 10^7 cfu/µg for large and ligated DNA. Additionally, the gene encoding endonuclease I (*endA*), which rapidly degrades pDNA isolated by most miniprep procedures, is inactivated.

7.4. Additional Materials

LB medium, LB agar medium, TB medium, Kanamycin (50 mg/mL), Streptomycin (50 mg/mL), Chloramphenicol (34 mg/mL), LB agar plates supplemented with Kanamycin, Kanamycin and Chloramphenicol or Kanamycin, Chloramphenicol and Streptomycin (self-prepared), 1% Agarose gels (self-prepared), IPTG (500 mM), DNase I (2 mg/mL), 10% and 12% SDS-PAGE gels (self-prepared), PMSF (200 mM), DTT (1 M), EDTA (250 mM), Nickel Sulfate (100 mM), Liquid Nitrogen, Erythritol (15%), Glycerol (20%, 40% and 100%), MORPHEUS 100% Precipitant Mix 2 (40% v/v Ethylene glycol and 20% w/v PEG 8000) (self-prepared), 0.9 M MORPHEUS Halogens Mix (0.3 M Sodium Fluoride, 0.3 M Sodium Bromate and 0.3 M Sodium Iodate) (self-prepared), 0.8 M Aminoacids Mix (0.2 M DL-Glutamic Acid Monohydrate, 0.2 M Glycine, 0.2 M DL-Lysine Monohydrochloride and 0.2 M DL-Serine) (self-prepared) and MORPHEUS 1 M Aminoacids Mix (0.2M DL-Glutamic Acid Monohydrate, 0.2 M Glycine, 0.2M DL-Lysine Monohydrochloride, 0.2 M DL-Serine and 0.2M L-Alanine) (self-prepared), 1 M MES (at four different pH: 5, 5.5, 6 and 6.5) (self-prepared), 5 M NaCl (self-prepared), 1 M TRIS (at 6 different pH: 7.5, 7.6, 7.9, 8.0, 8.1 and 8.5) (self-prepared), PEG 20000 (50%) (self-prepared), 1M Imidazole pH 7.0 (50%) (self-prepared), PEG 5000 MME (50%) (self-prepared), PEG 400 (100%).

7.5. Large-scale protein expression and purification (batch Nº 2) results

A single peak at ~109 mL is present in Figure 7.2. and corresponds to expected volume of elution of a globular ~39 kDa globular protein (section 7.10., Figure 7.11) and, as such, represents the purified KIFC1-MD protein. The difference in volume of elution comparing to the first purification is due to the different characteristics of the column used for the final purification. Unlike the first time the protein was purified, the peak is not single peak indicating the protein is less pure.

A small peak at ~68 mL is also present (Figure 7.2). This peak corresponds to the void volume of the column. Any contents of the sample that, due to very high molecular size, do not enter the column's pores, flow-through at this volume. Usually, this represents agglomerates of proteins. The small peak at this volume indicates that part of the protein of the sample was starting to aggregate, but with the purification these aggregates were separated from the soluble KIFC1-MD.

Upon running the samples of KIFC1-MD on a SDS-PAGE gel, a band with ~37 kDa is visible in lanes 7-10 of Figure 7.3. that corresponds to the large peak at ~109 mL and approximately to the predicted size of a monomeric KIFC1-MD (~39 kDa). A contamination ~71 kDa band is also visible in the lanes 4-7 of Figure 7.3. For this reason, only the protein elution fractions visible on lanes 8-10 of Figure 7.3. were used for crystallisation trials.

The calculated purity was ~83%, which is high purity but not as high as the first time the protein was purified (~93%) and under the minimum desired 90% (Khurshid et al., 2014; Sauder et al., 2008).

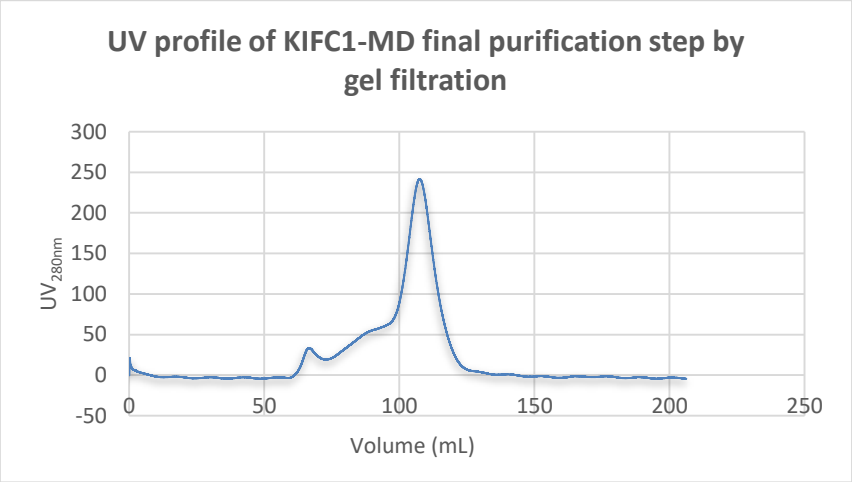

Figure 7.2: UV profile of KIFC1-MD final purification step by gel filtration (batch Nº 2). x axis – volume (mL); y axis – UV_{280nm}. A peak at ~109 mL is visible corresponding to the purified KIFC1-MD.

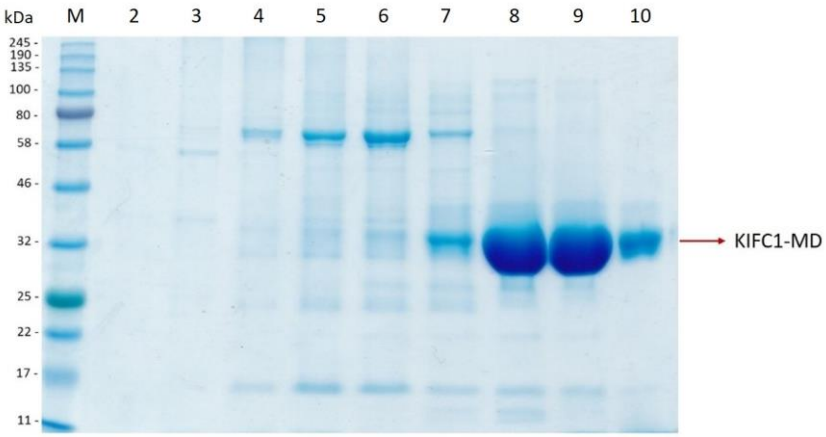

Figure 7.3: SDS-PAGE gel of purified KIFC1-MD after gel filtration (batch Nº 2). This gel shows the marker proteins (M) and the gel filtration elution fractions (lanes 2-10). The molecular sizes of the marker proteins are indicated on the left. The ~37 kDa bands corresponding to KIFC1-MD are indicated by the red arrow. The remaining bands are minor contaminations corresponding to proteins naturally expressed in the *E. coli*.

7.6. Large-scale protein expression and purification (batch Nº 3) results

A single peak at ~109 mL is present in Figure 7.4. and corresponds to the expected elution volume of a ~39 kDa globular protein (section 7.10. and Figure 7.11) and, as such, represents the purified KIFC1-MD protein. The difference in volume of elution to the first purification is due to the different characteristics of the column used for the final purification. Unlike the first time the protein was purified, the peak is not single peak indicating the protein is less pure.

A small peak at ~68 mL is also present (Figure 7.4). This peak corresponds to the void volume of the column. Any contents of the sample that, due to very high molecular size, do not enter the column's pores, flow-through at this volume. Usually, this represents agglomerates of proteins. The

small peak at this volume indicates that a small part of the sample was starting to aggregate, but with the purification these aggregates were separated from the soluble KIFC1-MD.

Upon running the sample of KIFC1-MD on a SDS-PAGE gel, a band at ~37 kDa is visible in lanes 7-10 of Figure 7.5. that corresponds to the large peak obtained at 109 mL and approximately to the predicted molecular size of a monomeric KIFC1-MD (~39 kDa).

The calculated purity was ~85%, which is high purity but not as high as the first time the protein was purified (~93%) and under the minimum desired 90% (Khurshid et al., 2014; Sauder et al., 2008).

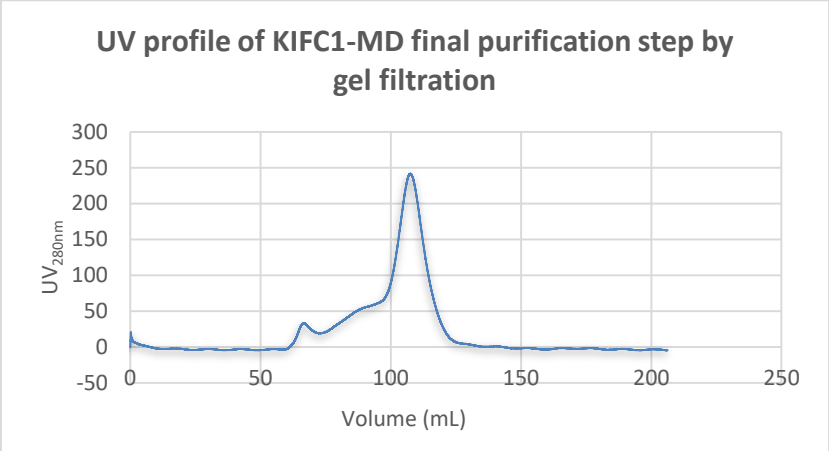

Figure 7.4: UV profile of KIFC1-MD final purification step by gel filtration (batch Nº 3). x axis – volume (mL); y axis – UV_{280nm}. A peak at ~109 mL is visible corresponding to the purified KIFC1-MD.

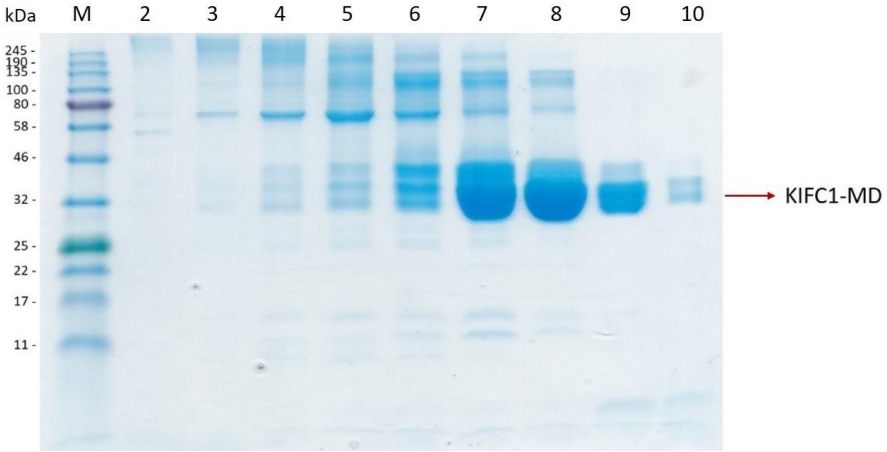

Figure 7.5: SDS-PAGE gel of purified KIFC1-MD after gel filtration (batch Nº 3). This gel shows the marker proteins (M) and the gel filtration elution fractions (lanes 2-10). The molecular sizes of the marker proteins are indicated on the left. The ~37 kDa bands corresponding to KIFC1-MD are indicated by the red arrow. The remaining bands are minor contaminations corresponding to proteins naturally expressed in the *E. coli*.

7.7. Large-scale protein expression and purification (batch Nº 4) results

A single peak at ~85 mL is present in Figure 7.6 and corresponds to the expected elution volume of a ~39 kDa globular protein (section 7.9, Figure 7.10) and, as such, represents the purified KIFC1-MD. As the column used for this purification is similar to the first one, the difference is volume of elution is no longer present. Like the first time the protein was purified, the single peak indication of high purity.

A small peak at ~96 mL is also present (Figure 7.6). This peak corresponds to the contamination proteins with higher molecular size that can be seen in the SDS-PAGE Gel.

Upon running the samples of KIFC1-MD on a SDS-PAGE gel, a band at ~37 kDa is visible in the lanes 4-7 of Figure 7.7 that corresponds to the large peak obtained at ~85 mL and approximately to the predicted molecular size of a monomeric KIFC1-MD (~39 kDa).

The calculated purity ~90%, which is very high purity, consistent with the single peak observed in Figure 7.5. Moreover, it corresponds to the minimum purity desired for crystallisation trials of 90% (Khurshid et al., 2014; Sauder et al., 2008).

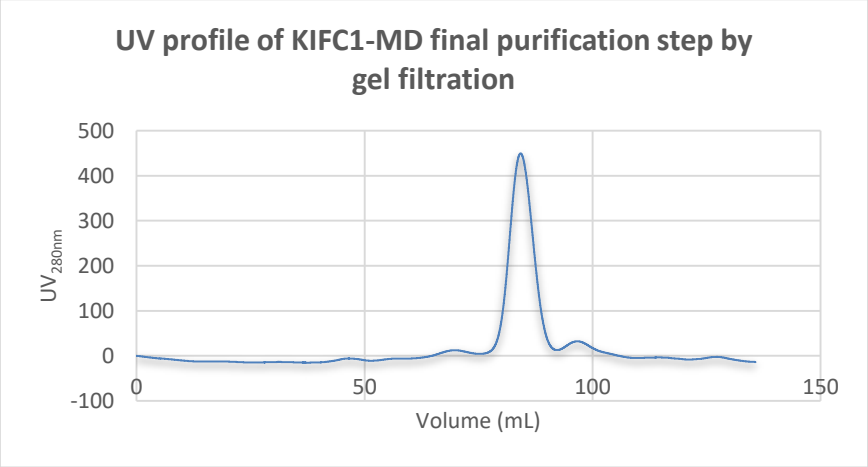

Figure 7.6: UV profile of KIFC1-MD final purification step by gel filtration (batch Nº 4). x axis – volume (mL); y axis – UV_{280nm}. A peak at ~85 mL is visible corresponding to the purified KIFC1-MD.

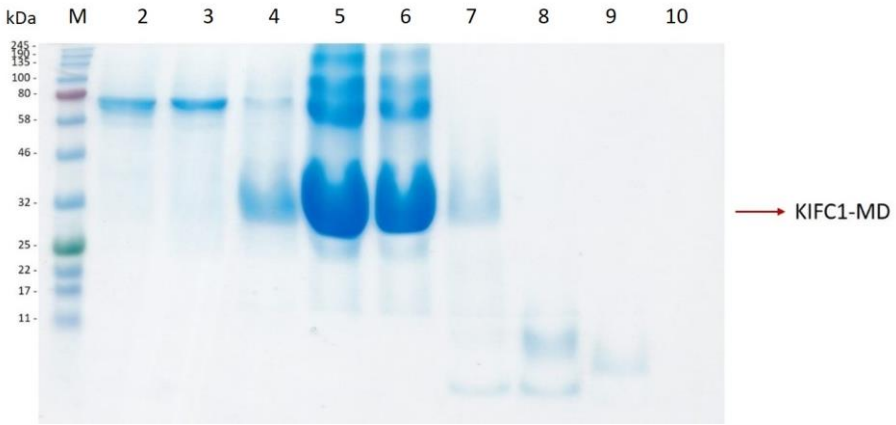

Figure 7.7: SDS-PAGE gel of purified KIFC1-MD after gel filtration (batch Nº 4). This gel shows the marker proteins (M) and the gel filtration elution fractions (lanes 2-10). The molecular sizes of the marker proteins are indicated on the left. The ~37 kDa bands corresponding to KIFC1-MD are indicated by the red arrow. The remaining bands are minor contaminations corresponding to proteins naturally expressed in the *E. coli*.

7.8. Large-scale protein expression and purification (batch Nº 5) results

A single peak at ~80 mL is present in Figure 7.8. and corresponds to the expected elution volume of a ~39 kDa globular protein (section 7.9, Figure 7.10) and, as such, represents the purified KIFC1-MD. As the column used for this purification is similar to the first one, the difference is volume of

elution is no longer present. Like the first time the protein was purified, the single peak is indication of high purity.

A small peak at ~41 mL is also present (Figure 7.8). This peak corresponds to the void volume of the column (section 7.9). Any contents of the sample that, due to very high molecular size, do not enter the column's pores, flow-through at this volume. Usually, this represents agglomerates of proteins. The small peak at this volume indicates that a small part of the sample was starting to aggregate, but with the purification, these aggregates were separated from the soluble KIFC1-MD. Another small peak at ~96 mL is also present (Figure 7.8). This peak corresponds to the contamination bands at higher molecular size that can be seen in the SDS-PAGE gel in lanes 2-3 (Figure 7.9).

Upon running the samples of KIFC1-MD on a SDS-PAGE gel, a band with ~37 kDa is visible in lanes 4-7 of Figure 7.9 that corresponds to the large peak at ~80 mL and approximately to the predicted molecular size of a monomeric KIFC1-MD (~39 kDa).

The calculated purity was ~90%, which is very high purity, consistent with the single peak observed in Figure 7.8. Moreover, it corresponds to the minimum purity desired for crystallisation trials of 90% (Khurshid et al., 2014; Sauder et al., 2008).

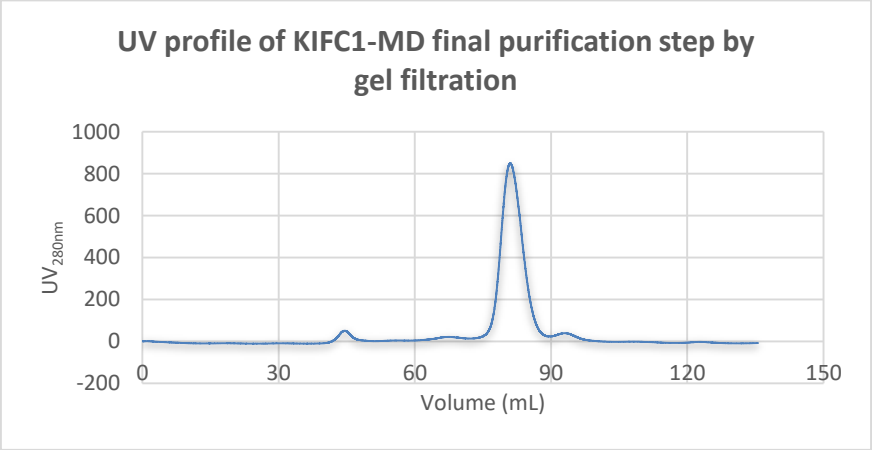

Figure 7.8: UV profile of KIFC1-MD final purification step by gel filtration (batch Nº 5). x axis – volume (mL); y axis – UV_{280nm}. A peak at ~80 mL is visible corresponding to the purified KIFC1-MD.

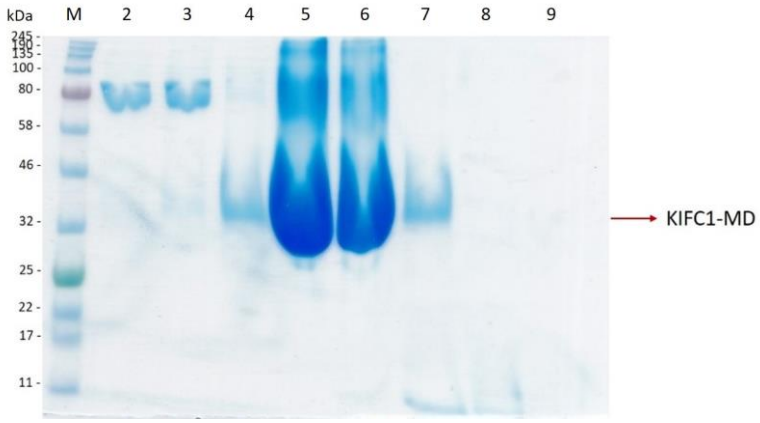

Figure 7.9: SDS-PAGE gel of purified KIFC1-MD after gel filtration (batch Nº 5). This gel shows the marker proteins (M) and the gel filtration elution fractions (lanes 2-9). The molecular sizes of the marker proteins are indicated on the left. The ~37 kDa bands corresponding to KIFC1-MD are indicated by the red arrow. The remaining bands are minor contaminations corresponding to proteins naturally expressed in the *E. coli*.

7.9. Calibration curve of HiLoadTM 16/600 Superdex 200 prep grade column

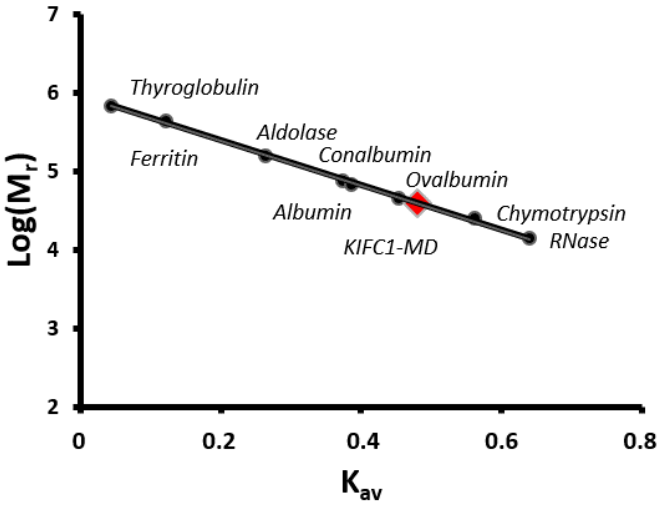

Figure 7.10: Calibration curve of HiLoad™ 16/600 Superdex 200 prep grade column. x axis – partition coefficient (K_{av}), which is calculated based on the volume of elution of each protein; y axis – Logarithm base 10 of the molecular weight of each protein.

From the following equation, it is possible to deduce the volume of elution of the target protein:

$$V_e = K_{av}(V_c - V_0) + V_0,$$

where, V_e = volume of elution; V_c = volume of the column (120 mL); V_0 = void volume = volume of elution of Dextran (43 mL).

From the equation of the calibration curve, the calculated K_{av} was 0.48. As such, V_e of KIFC1-MD is 80 mL.

7.10. Calibration curve of XK 16/100 column

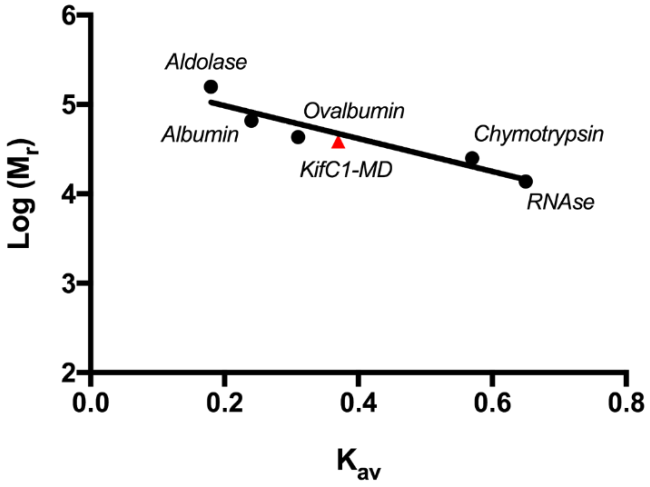

Figure 7.11: Calibration curve of XK 16/100 column. x axis - K_{av}, which is calculated based on the volume of elution of each protein; y axis – Logarithm base 10 of the molecular weight of each protein.

From the following equation, it is possible to deduce the volume of elution of the target protein:

$$V_e = K_{av}(V_c - V_0) + V_0,$$

where, V_e = volume of elution; V_c = volume of the column (180 mL); V_0 = void volume = volume of elution of Dextran (68 mL).

From the equation of the calibration curve, the calculated K_{av} was 0.37. As such, V_e of KIFC1-MD is 109 mL.

7.11. Amino acids sequence alignment

The sequence alignment between the crystallized KIFC1-MD whose structure is available in the PDB (PDB ID: 2REP) and the predicted amino acid sequence for the KIFC1-MD purified during this project, revealed a similarity of 99% (356/357 identities) and 0% gaps. The obtained alignment is presented below. In the alignment, the predicted amino acids sequence for the KIFC1-MD purified during this project is presented on the top, while the amino acid sequence visible on the KIFC1-MD structure in the PDB is presented on the bottom. The different amino acid is highlighted in red and corresponds to a threonine on position 65 of the predicted sequence of the KIFC1-MD purified during this project and to a proline on position 81 of the visible amino acids sequence on the KIFC1-MD structure in the PDB.

```
Predicted_KIFC1-MD  4    LKGNIRVFCRVRPVLPGEPTPPPGLLLFPSGPGGPSDPPTRLSLSRSDERRGTLSGAPAP  63
PDB_KIFC1-MD        20   LKGNIRVFCRVRPVLPGEPTPPPGLLLFPSGPGGPSDPPTRLSLSRSDERRGTLSGAPAP  79

Predicted_KIFC1-MD  64   PTRHDFSFDRVFPPGSGQDEVFEEIAMLVQSALDGYPVCIFAYGQTGSGKTFTMEGGPGG  123
PDB_KIFC1-MD        80   PPRHDFSFDRVFPPGSGQDEVFEEIAMLVQSALDGYPVCIFAYGQTGSGKTFTMEGGPGG  139

Predicted_KIFC1-MD  124  DPQLEGLIPRALRHLFSVAQELSGQGWTYSFVASYVEIYNETVRDLLATGTRKGQGGECE  183
PDB_KIFC1-MD        140  DPQLEGLIPRALRHLFSVAQELSGQGWTYSFVASYVEIYNETVRDLLATGTRKGQGGECE  199

Predicted_KIFC1-MD  184  IRRAGPGSEELTVTNARYVPVSCEKEVDALLHLARQNRAVARTAQNERSSRSHSVFQLQI  243
PDB_KIFC1-MD        200  IRRAGPGSEELTVTNARYVPVSCEKEVDALLHLARQNRAVARTAQNERSSRSHSVFQLQI  259

Predicted_KIFC1-MD  244  SGEHSSRGLQCGAPLSLVDLAGSERLDPGLALGPGERERLRETQAINSSLSTLGLVIMAL  303
PDB_KIFC1-MD        260  SGEHSSRGLQCGAPLSLVDLAGSERLDPGLALGPGERERLRETQAINSSLSTLGLVIMAL  319

Predicted KIFC1-MD  304  SNKESHVPYRNSKLTYLLQNSLGGSAKMLMFVNISPLEENVSESLNSLRFASKVNQC     360
PDB_KIFC1-MD        320  SNKESHVPYRNSKLTYLLQNSLGGSAKMLMFVNISPLEENVSESLNSLRFASKVNQC     376
```