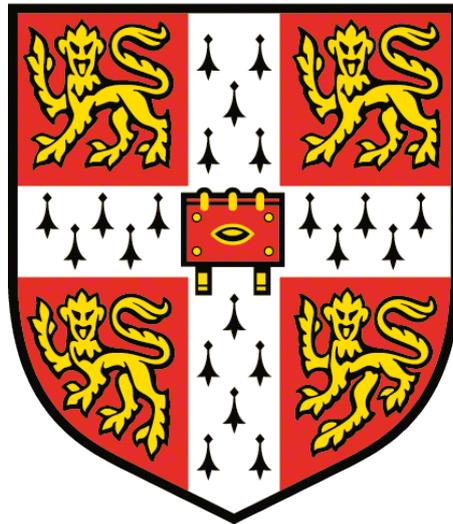# The structure of the cytoplasmic dynein tail

Aristides George Diamant

MRC Laboratory of Molecular Biology
University of Cambridge
Christ's College
United Kingdom
May 2015

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration of originality and statement of length

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the chapter introductions and specified in the text. This work is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma, or other qualification at the University of Cambridge or any other University or similar institution. This dissertation does not exceed the word limit specified by the Biology Degree Committee.

<div align="right">

Aristides Diamant
May 2015

</div>

Acknowledgements

First of all, I thank my supervisor, Andrew. It has been a real joy learning from and working with you. I am extremely grateful to have had the opportunity to take part in such an exciting project in a lab full of such wonderful people. Good labs don't run themselves, so thank you for all your hard work, careful leadership and steady support throughout my PhD.

To the entire Carter lab (past and present) thank you for maintaining the fun, supportive, and stimulating environment I have so much enjoyed the last few years. In particular, I thank Helgo, Max and Emma for their extreme patience, kindness and inexhaustible willingness to help when I was first settling into the lab. I also owe a great debt to Andrew, Max, Carina, Helgo and Ruta for all of the useful comments on my thesis.

I thank Meindert, for being my second supervisor and for all the help in the lab over my PhD. I also thank Minmin for tirelessly performing all the behind-the-scenes work necessary for using the synchrotrons, and for keeping me company while collecting data in the wee hours of the morning. There are far too many people at the LMB for me to thank individually, but I thank you all for the pleasant, collaborative atmosphere and passion for science.

Outside of the lab, I thank Matt for sharing the journey. We've been through it all together now, from the first day of orientation right through to thesis submission. It was comforting to know you were always nearby, willing to translate Britishisms and keep my work-life balance in check. Alberto, grazie per essere un "vero compagno" e un cuoco accettabile, who kept me well-nourished throughout my PhD. Not just physically, with the never-ending dinners, but also mentally and spiritually. You taught me what it means to live my life more intensely.

Beyond Cambridge, I thank John for steadfast friendship. I sleep better at night knowing that no matter what I do or where end up, I can always count on you to be there for me. Michael, thank you for being my only friend crazy enough to reassure me through his own actions that doubling up on graduate degrees is a really good idea. I'm looking forward to finally swapping roles. Fabienne, thank you for the music, and for giving me a fresh perspective when I needed it most. Even from afar your constant encouragement and motivation gave me the strength to write this thesis. Had I never met you, it might have taken me three times as long to write.

Finally, I thank my family. Nia, without the footsteps of such an exceptional older sister for me to try and follow, I might have remained more interested in video games and TV than the intricacies of intracellular motility. Dad, for giving me from a young age a sense of wonder and awe about the entire universe, ultimately pointing me toward a career in science. Mom, for never allowing me to underestimate myself and for sacrificing so much to provide me with every possible opportunity to grow and develop my curiosity.

To Mom, Dad and Nia,

Ever since I can remember, I have been both blessed and proud to call you my family. You have each been outstanding role models, without whom I would be lost. Your radiant examples of uncompromising commitment to excellence, uncontainable enthusiasm towards learning and unwavering perseverance in the face of adversity have shined forth and lit the way as I navigate my own path through life. Without the three of you as my guiding light and fountain of support, encouragement and inspiration, I would never have dreamt of embarking on a PhD, let alone finishing one!

So, with lots of love, I dedicate this work to you.

Abstract

Cytoplasmic dynein is a molecular motor that moves cargoes along microtubules. Dynein, together with its large co-factor dynactin, is responsible for the vast majority of traffic towards the centre of the cell. The largest subunit of the dynein complex is called the dynein heavy chain (DHC). The DHC includes a C-terminal motor domain, which converts ATP hydrolysis into mechanical force, an N-terminal tail domain, and a flexible linker domain to join the two together. An intermediate chain (DIC) and light intermediate chain (DLIC) bind directly to the DHC tail, while light chains (DLCs) bind to the DIC. This tail complex is important for both cargo binding as well as homodimerisation of the DHC, which is necessary for processive movement. Previous studies suggest that the DLCs play an important role in homodimerisation, but it remains unclear how else the DHCs are held together.

Using *S. cerevisiae* as a model system, I co-expressed all four dynein subunits and purified functional dynein motors. In this background, I found that truncating the DHC to include only the first 1004 residues (out of the total 4092) eliminates the motor domain as well as the flexible linker domain, while preserving binding to the DIC, DLIC and DLC. However, truncating just another 50 residues off of the C-terminus led to a loss of all accessory subunits. I developed a protocol for expressing and purifying large quantities of the 1004 residue construct, thus I provide the first description of a recombinant dynein tail domain. Using negative stain electron microscopy (EM), I also present the first 3D structural information for the tail region of the cytoplasmic dynein motor.

I then describe a construct including only the first 557 residues of the DHC, which dimerises despite not being able to bind any of the other subunits. I present a crystal structure of this smaller DHC fragment, which shows that the N-terminal 180 residues of the DHC constitute an intricate dimerisation domain made up of a β-sheet sandwiched between α-helices. Not only is this the first crystal structure of any part of the DHC N-terminus, but it reveals a previously undocumented dimerisation domain within the DHC itself. Furthermore, information garnered from this crystal structure allowed for interpretation of a recent cryo-EM structure of a triple complex containing the dynein tail, dynactin and the cargo adaptor BICD2 (TDB) that was solved by my colleagues in the Carter group. Only by docking the DHC N-terminus crystal structure within the TDB EM density did it become clear that the N-terminus of the DHC is responsible for the majority of the contacts the dynein tail makes with both dynactin and BICD2.

Therefore the work that I present here sheds new light on the unexpected importance of the DHC N-terminus and allows two important conclusions to be made. First, the N-terminal 180 residues of the DHC constitute a dimerisation domain of its own. Second, the next ~ 400 residues of the DHC form a domain that plays a key role in the complex interface between dynein, dynactin and BICD2.

Contents

List of Tables

List of Figures

Abbreviations

5-FOA	5-Fluoroorotic acid
aa	amino acids
AAA+	ATPases associated with diverse cellular activities
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BICD	Bicaudal D
Blue-Gal	5-Bromo-3-indolyl β-D-galactopyranoside
BME	β-mercaptoethanol
bp	base pairs
C	Celsius
CAPS	N-cyclohexyl-3-aminopropanesulfonic acid
DDB	dynein/dynactin/BICD2
ddH$_2$O	double distilled water
DHC	dynein heavy chain
DIC	dynein intermediate chain
DLC	dynein light chain
DLIC	dynein light intermediate chain
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside-triphosphates
DRC	dynein regulatory complex
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
ER	endoplasmic reticulum
FKBP	FK-506 binding protein
FRB	FKBP-rapamycin-binding domain
GA	Gibson assembly mix
GDP	guanosine diphosphate
GFP	green fluorescent protein
GST	glutathione S-transferase
GTP	guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hr	hours
I3C	5-Amino-2,4,6-triiodoisophtalic acid
IAB	isothermal assembly buffer
IDA	inner dynein arm
IFT	intraflagellar transport
IPTG	isopropyl β-D-thiogalactopyranoside

kb	kilobases
kDa	kilodalton
LDS	lithium dodecyl sulfate
LN_2	liquid nitrogen
MDa	megadalton
MES	2-(N-morpholino)ethanesulfonic acid
min	minutes
MOPS	3-(N-morpholino)propanesulfonic acid
MPW	Milli-Q water
mRNA	messenger RNA
MS	mass spectrometry
MT	microtubule
MW	molecular weight
NAD	nicotinamide adenine dinucleotide
Ni-NTA	nickel nitrilotriacetic acid
ODA	outer dynein arm
ODA-DC	outer dynein arm docking complex
OID	outer-inner-dynein linker
OOD	outer-outer dynein linker
ORF	open reading frame
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMSF	phenylmethylsulfonyl fluoride
RNA	ribonucleic acid
RPM	rotations per minute
RT	room temperature
s	seconds
SDS	sodium dodecyl sulfate
SDS-PAGE	denaturing polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SeMet	selenomethionine
SOB	super optimal broth
SOC	super optimal broth with catabolite repression
ssDNA	salmon sperm DNA
$[Ta_6Br_{12}]^{2+} \cdot 2Br^-$	hexatantalum tetradecabromide
TDB	dynein tail/dynactin/BICD2
TEV	tobacco etch virus
TFZ	translation function z-score
TGN	*trans*-Golgi network
TYE	tryptone yeast extract medium
UV	ultraviolet
YEPD	yeast extract peptone dextrose

Chapter 1

General Introduction

1.1 An overview of the cytoskeleton

All living cells contain a dynamic and complex intracellular matrix known as the cytoskeleton. The cytoskeleton plays an essential role in processes as diverse as chromosomal segregation, cytokinesis, endocytosis, signal transduction, intracellular transport and locomotion of the entire cell (Wittmann *et al.* 2001, Noguchi & Mabuchi 2001, Qualmann *et al.* 2000, Liu *et al.* 2013, J. Howard 1996 and T. H. Howard & Meyer 1984). The foundation of the cytoskeleton is a network of long protein filaments, but these filaments interact with a variety of other proteins, such as cross-linkers, capping proteins, regulators and motor proteins in order to achieve their myriad functions. The focus of this thesis is one of these motors, called dynein. Therefore, this brief overview of the cytoskeleton provides a general introduction to the context in which dynein is found.

1.1.1 Cytoskeletal filaments

The general eukaryotic cytoskeleton contains two fundamental types of cytoskeletal filaments: microfilaments and microtubules. There is a third type of cytoskeletal filaments present in many animal species, called intermediate filaments. These diverse filaments are essential for many cells' function and survival. The roles of intermediate filaments include providing mechanical support and integrating other cytoskeletal components (Granger & Lazarides 1979). However, as they are often specific to cell type (Lazarides 1980), they are beyond the scope of this general introduction and will not be discussed further.

Microfilaments

Microfilaments (also known as actin filaments) are the narrowest of the cytoskeletal filaments, measuring only $\sim$ 7nm in diameter (Kishino & Yanagida 1988). They are long polymers of a single protein, which is known as F-actin (filamentous actin) when it is found within a microfilament or G-actin (globular actin) when it is in its free monomeric form (Korn 1982). Each microfilament contains two helical strands wrapped around each other and is formed when G-actin monomers polymerise in an ATP-dependent manner (Fig. 1.1A). The actin monomers within both strands are oriented in the same way, which gives the microfilament an overall polarity. The two ends, which were named for their appearance in electron micrographs, are
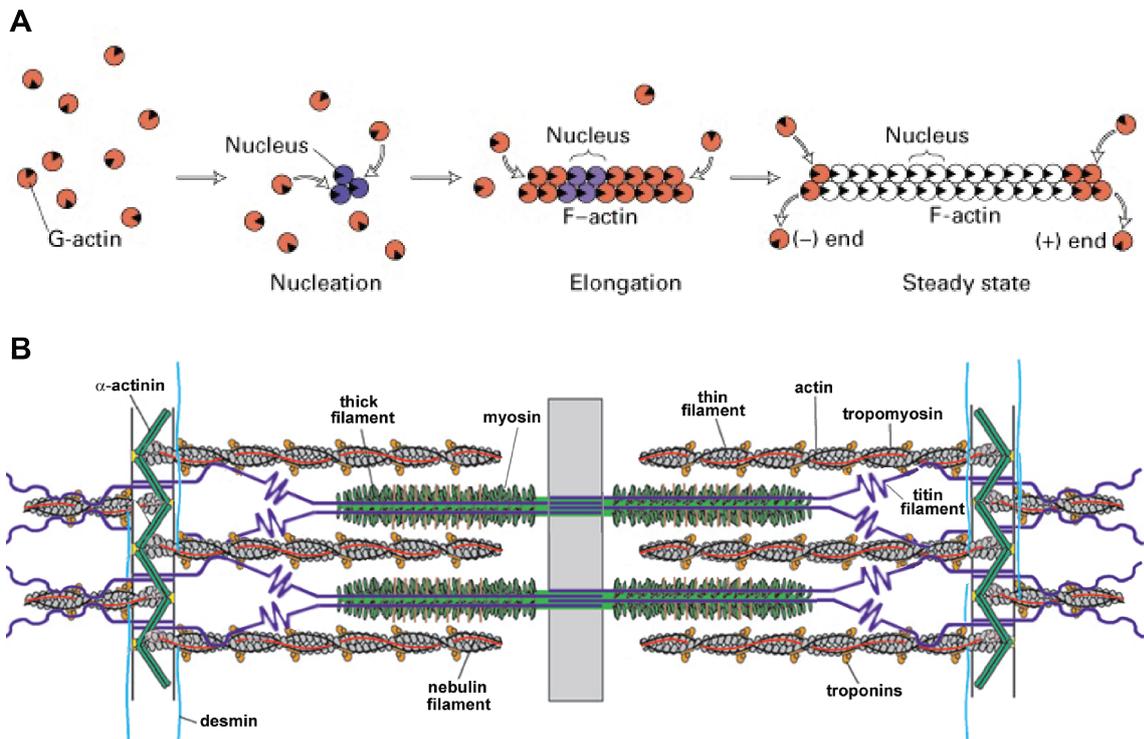
Figure 1.1: **Formation of microfilaments and their role in muscle contraction. A)** Polymerisation of microfilaments begins when ATP–G-actin monomers (pink) form stable trimers (purple). The filament then extends in both directions, but it extends more rapidly from the barbed (+) end. The ends of the filament reach a steady state where ATP-G-monomers are added and released at the same rate. After incorporation into the filament, subunits hydrolyse ATP and become the more stable ADP–F-actin (white). Figure reproduced from Lodish *et al.* 2004 with permission of the author. **B)** A model of skeletal muscle sarcomeres. The sarcomere is made up of thin filaments (microfilaments), thick filaments, and titin. When the myosin heads of the thick filament move along the anchored microfilaments, the entire sarcomere contracts. It is the coordinated contraction of several thousand sarcomeres in series that results in the contraction of skeletal muscle fibres. Figure reproduced from Ottenheijm *et al.* 2008 with permission of the author.

the barbed (or plus) end and the pointed (or minus) end. Upon nucleation of a microfilament, G-actin associates with the barbed end at a much higher rate than at the pointed end (Korn 1982). Migrating cells exploit the resulting directional bias of microfilament growth by orienting the barbed ends toward the cell's leading edge. The rapid polymerisation of microfilaments exerts sufficient force on the cell membrane to create protrusions such as lamellipodia and filopodia, which allow the cell to creep forward along a substrate (Westphal *et al.* 1997).

Microfilaments are essential for many fundamental cellular processes, including cytokinesis, chemotaxis and signal transduction (Noguchi & Mabuchi 2001, T. H. Howard & Meyer 1984 and Liu *et al.* 2013). One well-known role of microfilaments is within sarcomeres, the basic contractile unit of skeletal muscle. In this context, microfilaments are "thin filaments" that the thick filaments slide in relation to (Fig. 1.1B). As the myosin heads of the thick filaments undergo coordinated, ATP-dependent power strokes, the interlaced thin and thick filaments slide with re-

lation to one another, increasing the overlap between them and therefore decreasing the overall length of the sarcomere (Huxley & Hanson 1954). Each skeletal muscle fibre contains thousands of sarcomeres arranged in series. When these microscopic units contract in unison, the muscle fibre shortens, resulting in the macroscopic movements observed at the organismal level.

Microtubules

Microtubules (MTs) are the thickest of the cytoskeletal filaments, measuring ~ 24nm in diameter (Olmsted & Borisy 1973), and their structure is very different from that of microfilaments (Stephens & Edds 1976). The basic subunit of MTs is a heterodimer of two related, yet distinct, globular proteins called α-tubulin and β-tubulin. The GTP-bound form of these heterodimers polymerise head-to-tail and form linear protofilaments (Bergen *et al.* 1980). The α- and β-tubulins alternate down the length of each protofilament. Therefore, the protofilament, like a microfilament, has a plus end and a minus end (Bergen *et al.* 1980). *In vivo*, a ring-shaped template ensures that exactly 13 protofilaments associate with one another laterally to form a hollow tube (Moritz *et al.* 2000). All the protofilaments in a MT are oriented in the same direction, meaning the MT as a whole is polar as well. Once the tube has been formed, GTP-α/β-tubulin dimers may continue to add to the plus end, extending the length of the MT further (Akhmanova & Steinmetz 2008)(Fig. 1.2). GTP-α/β-tubulin dimers that have already been incorporated into a MT slowly hydrolyse their GTP into GDP. GTP-α/β-tubulin favours a straight protofilament conformation and maintains a protective cap at the plus end of growing MTs. If all of the GTP-α/β-tubulin at the plus end hydrolyses the GTP to GDP, a curved protofilament is favoured. This introduces strain within the MT, causing it to spontaneously fray into individual protofilaments, which then depolymerise back into GDP-α/β-tubulin dimers (Akhmanova & Steinmetz 2008)(Fig. 1.2).

MTs play a central role in various cellular processes, including intracellular transport and chromosomal segregation during nuclear division (J. Howard 1996). The MT network in most cells is formed by the radiation of MTs outward from the microtubule organising centre (MTOC). The minus ends are anchored at the MTOC, whereas the plus ends are oriented outward (Bergen *et al.* 1980). This highly polarised organisation supports a system whereby one set of unidirectional motors carries cargoes out to the cell periphery and a different set of unidirectional motors carries cargoes back toward the cell centre (Vale 1987). This polarity is particularly crucial for efficient transport in neuronal axons, which can span distances several magnitudes of order longer than the diameter of the cell body (Goldstein & Z. Yang 2000). MTs also play a key role in nuclear division, as they form the mitotic spindle. This MT-based apparatus aligns and segregates chromosomes into two daughter nuclei, ensuring an equal division of genetic material is distributed to each one (Wittmann *et al.* 2001). Finally, MTs comprise the core structure of axonemes, the highly specialised structures that power both cilia and flagella (Stephens & Edds 1976). A more detailed discussion of the axoneme follows in Section 1.2.2.

1.1.2   Cytoskeletal motors

In order for the cytoskeleton to carry out the majority of its functions, it relies not just on the filaments described above, but also on the molecular motors that move
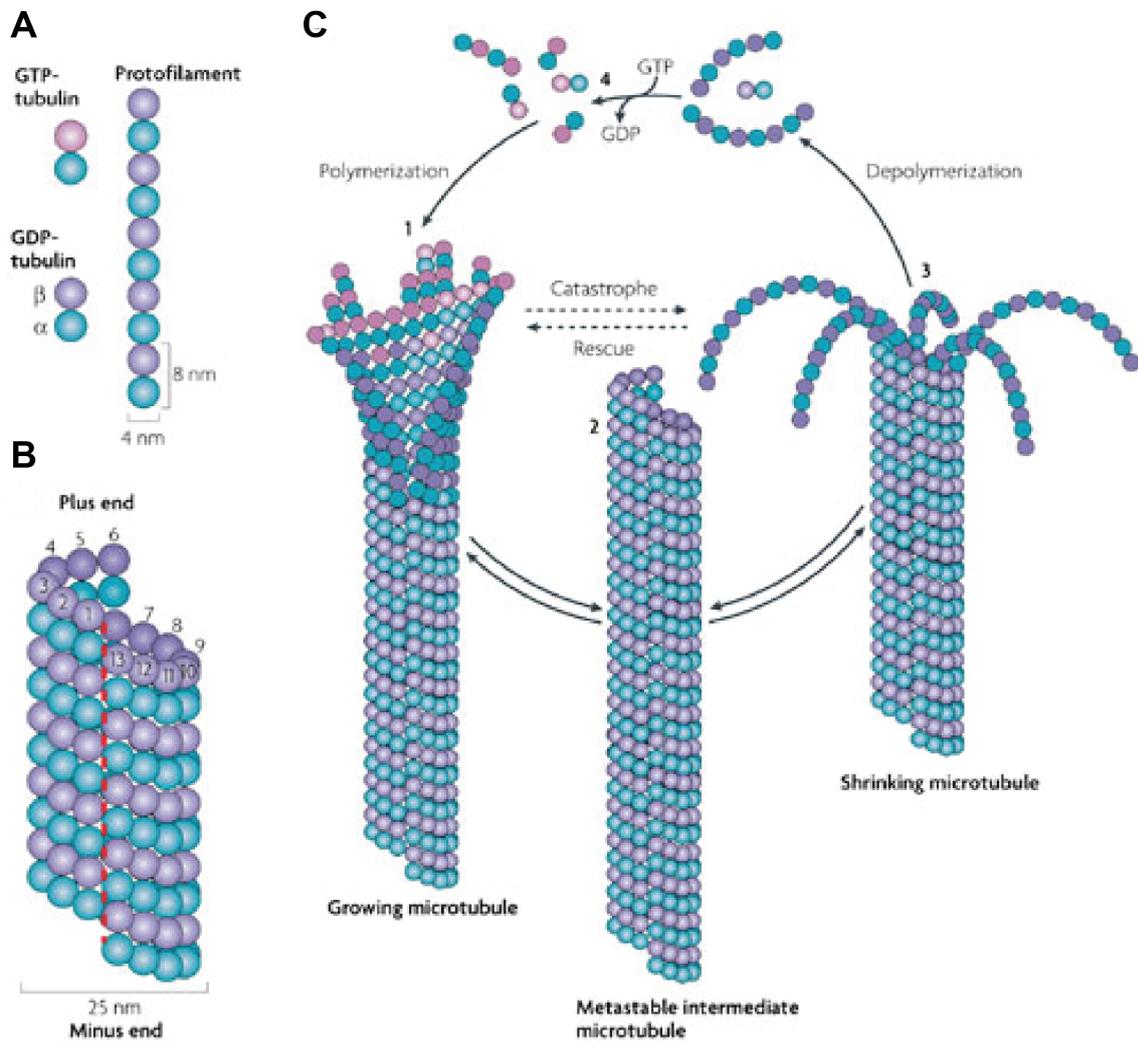
Figure 1.2: **Microtubule polymerisation. A)** Polymerisation of microtubules (MTs) begins when GTP-α/β-tubulin dimers join end-to-end and form polar protofilaments. **B)** A hollow MT is formed when protofilaments interact with each other laterally in a helical pattern. The closing of the MT results in a lattice seam (red dashed line). **C)** An overview of MT turnover. Once a MT is formed as described in **A** and **B**, GTP-α/β-tubulin dimers may continue to add to the plus end, extending the MT length. Over time, GTP-α/β-tubulin that has been incorporated into the MT hydrolyses its GTP and becomes GDP-α/β-tubulin, which favoured a curved protofilament rather than a straight one. If all theGTP-α/β-tubulin at the growing end hydrolyses, this strain overwhelms the lateral interactions between protofilaments and the MT separates into individual protofilaments, which are ultimately depolymerised back into GDP-α/β-tubulin dimers. Figure reproduced from Akhmanova & Steinmetz 2008 with permission of the author.

4

processively along them. Cytoskeletal motors can be divided into three families: myosins, kinesins and dyneins. Each motor includes at least one heavy chain, which contains a motor domain that converts the chemical energy of ATP into mechanical work as well as binding sites for various accessory chains. Despite these similarities in overall structure and function, each family has unique characteristics that relate to its members' specific functions. The focus of this section is to give a broad structural and functional overview of each of the three families, while making use of relevant examples to compare and contrast them.

Myosins

Myosins are cytoskeletal motors that move along microfilaments. Although myosins were first studied in the context of muscle contraction, there are many other types of myosin with a host of functions outside of skeletal muscle cells. There are even myosins present in unicellular organisms. For example, *S. cerevisiae* has five myosins, which play roles in processes ranging from endocytosis to the generation of asymmetry between mother and daughter cells (S. S. Brown 1997).

There are at least 17 known classes of myosins, each being structurally and/or functionally distinct from the others (Hodge *et al.* 2000). Myosin II includes the motors that assemble into the thick filaments of sarcomeres (Fig. 1.1B) and are responsible for muscle contraction (Reggiani & Bottinelli 2008). Myosin VI includes motors that move toward the pointed end of microfilaments, whereas all other characterised myosins move toward the barbed end (Wells *et al.* 1999). Myosin V will be described in greater detail as it appears to be present in nearly all non-plant eukaryotes and is well characterised as a processive motor involved in long-range cellular transport (Reck-Peterson *et al.* 2000).

Structurally, myosin V is a homodimer of two heavy chains, which bind each other via an extended coiled-coil called the stalk (Fig. 1.3). At the N-terminus of the myosin heavy chain is a globular motor domain, which is the site of ATP hydrolysis and actin binding. Immediately C-terminal to the motor domain is a neck domain, which consists of six IQ motifs. Five of the IQ motifs bind calmodulin, and the last binds an essential light chain. The stalk connects the neck domain to the C-terminal tail domain. The tail domain mediates cargo binding via adaptor proteins (Vale 2003).

Compared to conventional myosin II, which is characterised by high-speed, ensemble based motility, myosin V's high duty ratio and 36nm step size make it highly processive (Enrique *et al.* 1999). Thus, it is uniquely suited for membrane trafficking and long-range cargo transport along microfilaments (Mehta *et al.* 1999). Myosin V has also been shown to walk using a "hand-over-hand" mechanism (Yildiz *et al.* 2003). This means that one motor is always attached to the microfilament and that the two motors alternate between leading and trailing, as opposed to an "inchworm" mechanism whereby one motor always leads.

Kinesins

Kinesins move along MTs rather than microfilaments. The kinesin family is large, with at least 14 distinct families (Lawrence *et al.* 2004), and 45 known kinesin genes in the human genome (Miki *et al.* 2001). Of these, conventional kinesin, or kinesin I, is the best characterised. Kinesin I is processive (Vale *et al.* 1985a) and moves toward

Motor **Neck** **Stalk** **Tail**
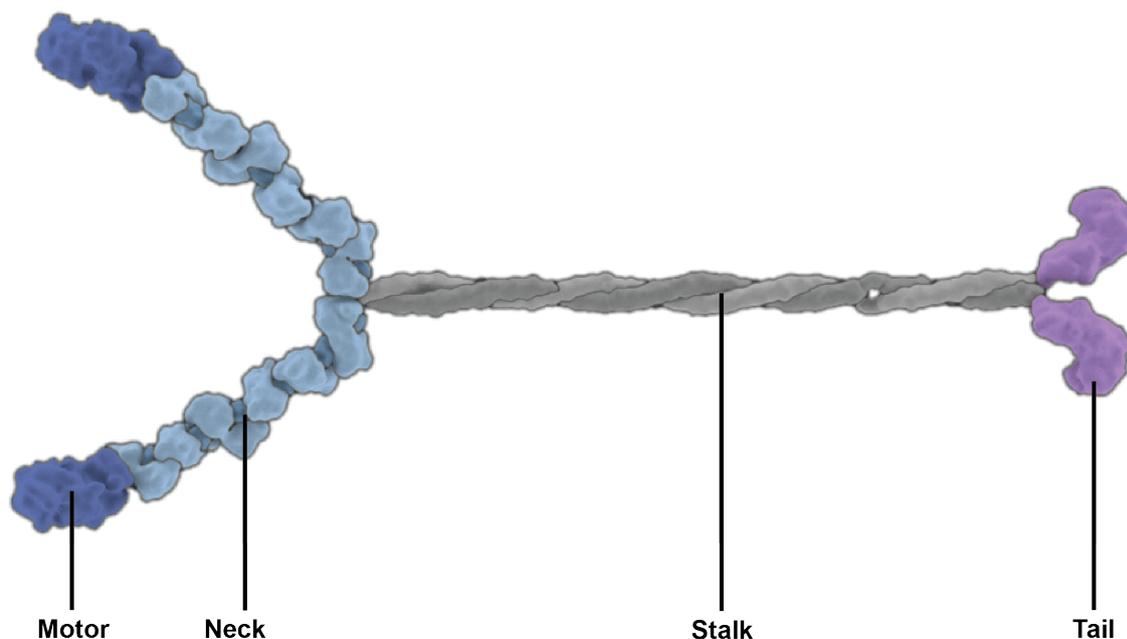
Figure 1.3: **A model of myosin V.** Myosin V consists primarily of two myosin heavy chains, which dimerise via a long coiled-coil stalk. At the heavy chain N-terminus there is a motor domain, which is connected to the stalk by a neck domain. The neck domain contains six IQ motifs, which provide binding sites for five calmodulin molecules and one essential light chain. The globular tail domain at the heavy chain C-terminus binds cargo adaptors, which link the motor to its various cargoes. Figure adapted with permission of the author from a model illustrated by Janet Iwasa.

Neck linker **Hinge** **KLC**

Motor **Stalk 1** **Stalk 2** **Tail**
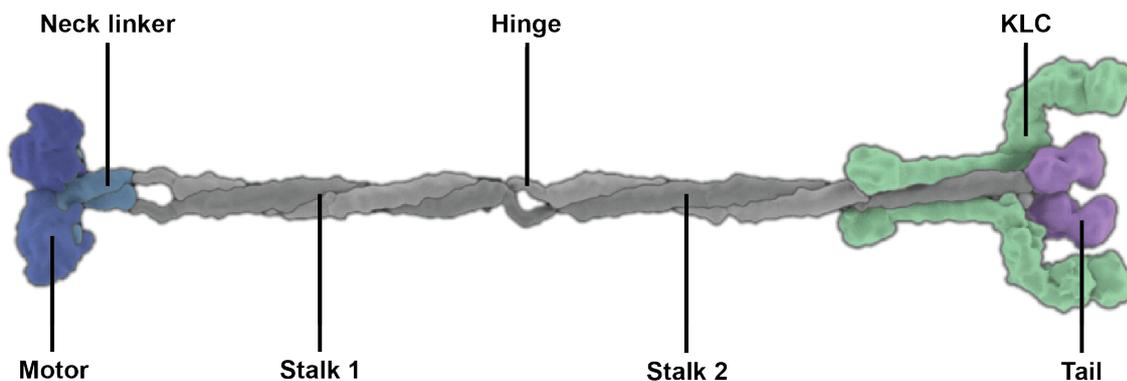
Figure 1.4: **A model of conventional kinesin (kinesin I).** Kinesin I consists primarily of two heavy chains, which dimerise via a long coiled-coil stalk. At the heavy N-terminus there is a motor domain, which is connected to the stalk by a neck linker. At the heavy chain C-terminus is the tail domain, which binds to the light chains (KLCs). The heavy chain's globular tail domain and KLCs bind cargo adaptors, which link the motor to its various cargoes. The stalk is interrupted in the middle by a hinge, which allows kinesin to fold in half and adopt an autoinhibited conformation where the tail is bound to the motor. Figure adapted with permission of the author from a model illustrated by Janet Iwasa.

the plus end of MTs (Vale *et al.* 1985b). Other kinesin variants include some that are non-processive (Rogers *et al.* 2001), some that are monomeric (Okada *et al.* 1995), and some that move toward MT minus ends (McDonald *et al.* 1990). There are even kinesins that are not motors, but rather MT polymerases and depolymerases (Hunter *et al.* 2003, Varga *et al.* 2009 and Grava & Philippsen 2010).

Structurally, kinesin I is a homodimer of two heavy chains, bound together via an extended coiled-coil called the stalk (Fig. 1.4). The N-terminal motor domain, which is connected to the stalk by a neck linker, is the site of ATP hydrolysis and MT binding. The C-terminal tail of the heavy chain binds to kinesin light chains (KLCs), which are important for cargo binding (Pernigo *et al.* 2013). The coiled-coil of the stalk is interrupted halfway through by a hinge region, which allows kinesin I to fold in half. When folded, the motor domain and tail domain can interact to adopt an autoinhibited state (Kaan *et al.* 2011), the advantage of which is to prevent idle motors from wasting precious ATP (Vale 2003).

Kinesin I functions via a "hand-over-hand" mechanism, like myosin V (Yildiz *et al.* 2004). However, the step-size of kinesin I is only 8nm (N. J. Carter & Cross 2005), in contrast to the 36nm steps that myosin V takes. This results from the kinesin I neck linker being much shorter than the myosin V neck, and corresponds to the spacing between available binding sites on MTs compared to microfilaments. As a processive cytoskeletal motor, kinesin I is also involved with long-range transport of cargoes. For example, kinesin I transports organelles such as mitochondria and lysosomes, as well as anterograde-directed vesicles in axons (Hirokawa 1998). In light of all the structural, mechanistic and functional similarities between kinesins and myosins, it has been proposed that the kinesins and myosins share a common ancestor (Kull *et al.* 1998 and Vale & Milligan 2000).

Dyneins

All known dyneins are motors that drive movement toward the minus end of MTs. In many ways, dyneins are the odd ones out of the three cytoskeletal motor families. At over 500kDa, dynein heavy chains (DHCs) are much larger than those of kinesin ($\sim 70 - 170$kDa) and myosin ($\sim 120 - 250$kda). In addition to their massive size, in DHCs, the N-terminus forms the tail domain and the C-terminus forms the motor domain, while in the heavy chains of most myosins and kinesins, the orientation is reversed. This disparity in both size and domain organisation reflects the difference in evolutionary origin between dynein and the other two motor families. This difference is further highlighted by the fact that the highly conserved nucleotide binding pockets of kinesin and myosin are homologous to the active site of G proteins (Kull *et al.* 1998), whereas dynein's hydrolytic activity relies on AAA+ domains (Neuwald *et al.* 1999).

The overall structure of dyneins (Fig. 1.5) is also quite different from myosins and kinesins. For example, in myosins and kinesins, the motor domain itself binds directly to the filament. On the other hand, dynein's microtubule binding domain (MTBD) is located at the end of a stalk, which extends ~ 15nm away from the motor domain. Furthermore, the dimer interface of both myosin and kinesin are simple coiled-coils. Although dynein's mechanisms for oligomerisation are not well understood, there is evidence that the homodimerisation of cytoplasmic dynein-1 involves accessory subunits as well as the DHCs (see Section 1.3.2 for a more detailed discussion). Dyneins also tend to have many stably bound accessory subunits. The
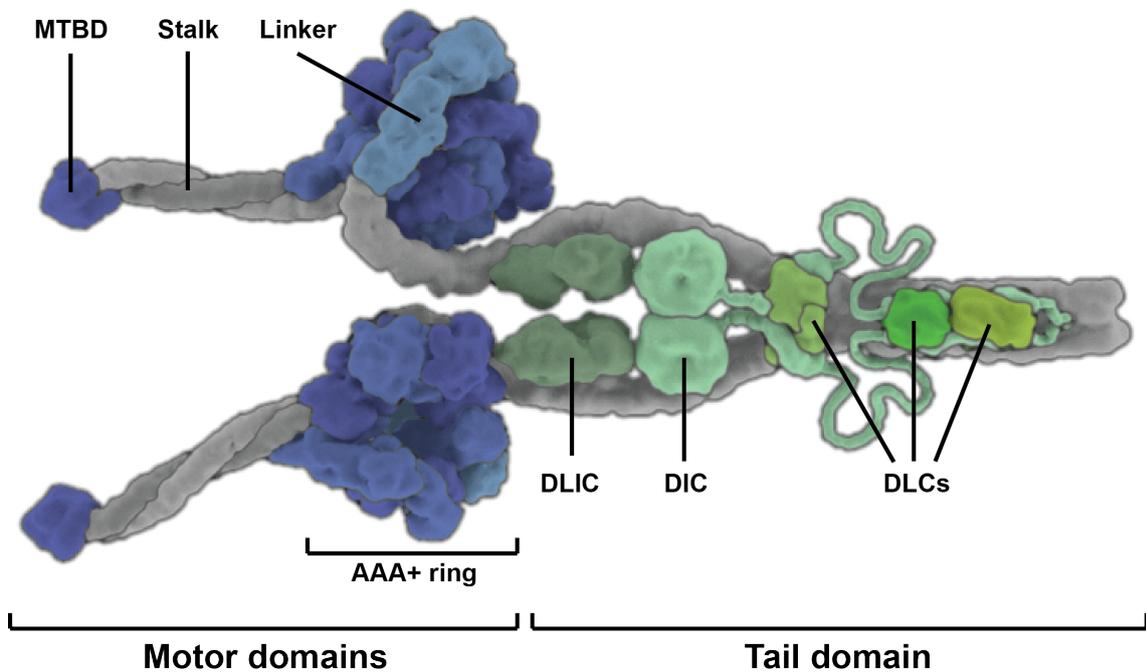
Figure 1.5: **A model of cytoplasmic dynein-1.** Cytoplasmic dynein-1 consists primarily of two dynein heavy chains (DHCs). The C-terminal two thirds of the DHC (blue and grey) form the motor domain, while the N-terminal one-third and various accessory chains (shades of green) form the tail domain. The ATPase activity of the motor domain occurs within a hexameric ring of AAA+ domains, and results in mechanical motion that is communicated through the linker domain. The motor domain binds microtubules through a globular microtubule binding domain (MTBD), which is located at the end of stalk, a 15nm coiled-coil that projects out of AAA+ domain 4. Far less is known about the organisation of the tail domain than the motor domain, which is the site of oligomerisation, cargo binding and regulation of the dynein motor. In mammals, the tail domain consists of the two DHC N-termini as well as two copies each of the dynein intermediate chain (DIC), dynein light intermediate chain (DLIC), and three dynein light chains (DLCs), called LC8, Tctex-1 and Roadblock/LC7. The DLCs bind to the DICs, rather than the DHCs directly. The DLICs and DICs both contact the DHCs directly. Figure adapted with permission of the author from a model illustrated by Janet Iwasa.

specific subunits vary depending on species and dynein type, but the tail domain typically includes intermediate chains (DICs), light intermediate chains (DLICs) and light chains (DLCs) in addition to the DHCs (Fig. 1.5).

Dynein functions are diverse, but generally related to either long-range transport of various cargoes or axonemal bending, the latter being a mechanism for translating dynein's movements on the molecular scale into locomotion of the entire cell. The following section provides a more detailed description of dynein's functions while discussing the various members of the dynein family.

1.2 Dynein family classification

Some text and the figures in this section have been adapted from the following article. Figures that have already been published are indicated by citing Diamant & A. P. Carter 2013 in the figure legends.

Diamant, A. G. & Carter, A. P. Dynein Family Classification in *Encyclopedia of Biophysics*. (ed. Roberts, G. C. K.) 552-558. (Springer Heidelberg, 2013).

1.2.1 Overview of dynein families

The dynein family is divided into axonemal and cytoplasmic dyneins. Axonemal dyneins reside within cilia and flagella and are responsible for the rhythmic and coordinated bending of these organelles. Cytoplasmic dynein-1 is found across the cellular MT network and is responsible for retrograde transport of various cargoes toward the minus ends of MTs, while cytoplasmic dynein-2 predominantly transports cargoes along MTs within the axoneme. The axonemal dyneins consist of one known outer dynein arm (ODA) and at least seven distinct inner dynein arms (IDAs), with the exact number varying between species. The ODA can contain either two or three heavy chains, which are further defined as the α, β and γ DHCs. Each IDA contains a single DHC except for one known as inner arm I1 or f (hereafter referred to as I1/f), which is a heterodimer. Both cytoplasmic dyneins are homodimers of their respective DHCs (Fig. 1.6).

Structurally, the dyneins differ not only in the number and sequences of their DHCs, but also in the composition and organisation of the smaller subunits. These differences are described below, but first there are some key characteristics that all dyneins share. Every dynein complex contains at least one DHC. The C-terminal two-thirds of the DHC is known as the motor domain. It consists of six AAA+ domains that form a ring. The stalk, a 15nm antiparallel coiled-coil, projects out of the fourth AAA+ domain and bears the globular microtubule-binding domain (MTBD) at its tip (Fig. 1.7). A shorter coiled-coil, known as the buttress or the strut, extends from the fifth AAA+ domain to the middle of the stalk. N-terminal to the hexameric ring is an element known as the linker, which plays a pivotal role in translating the ATPase action of the motor domain into mechanical motion (Burgess *et al.* 2003). The N-terminal one-third of the DHC is known as the tail domain and serves as a major site of protein-protein interactions. It contains the site of DHC oligomerisation (homo and heterodimerisation, as well as trimerisation) and binds accessory proteins, such as dynein intermediate chains (DICs) and dynein light chains (DLCs). The resulting tail subcomplex is responsible for attachment of
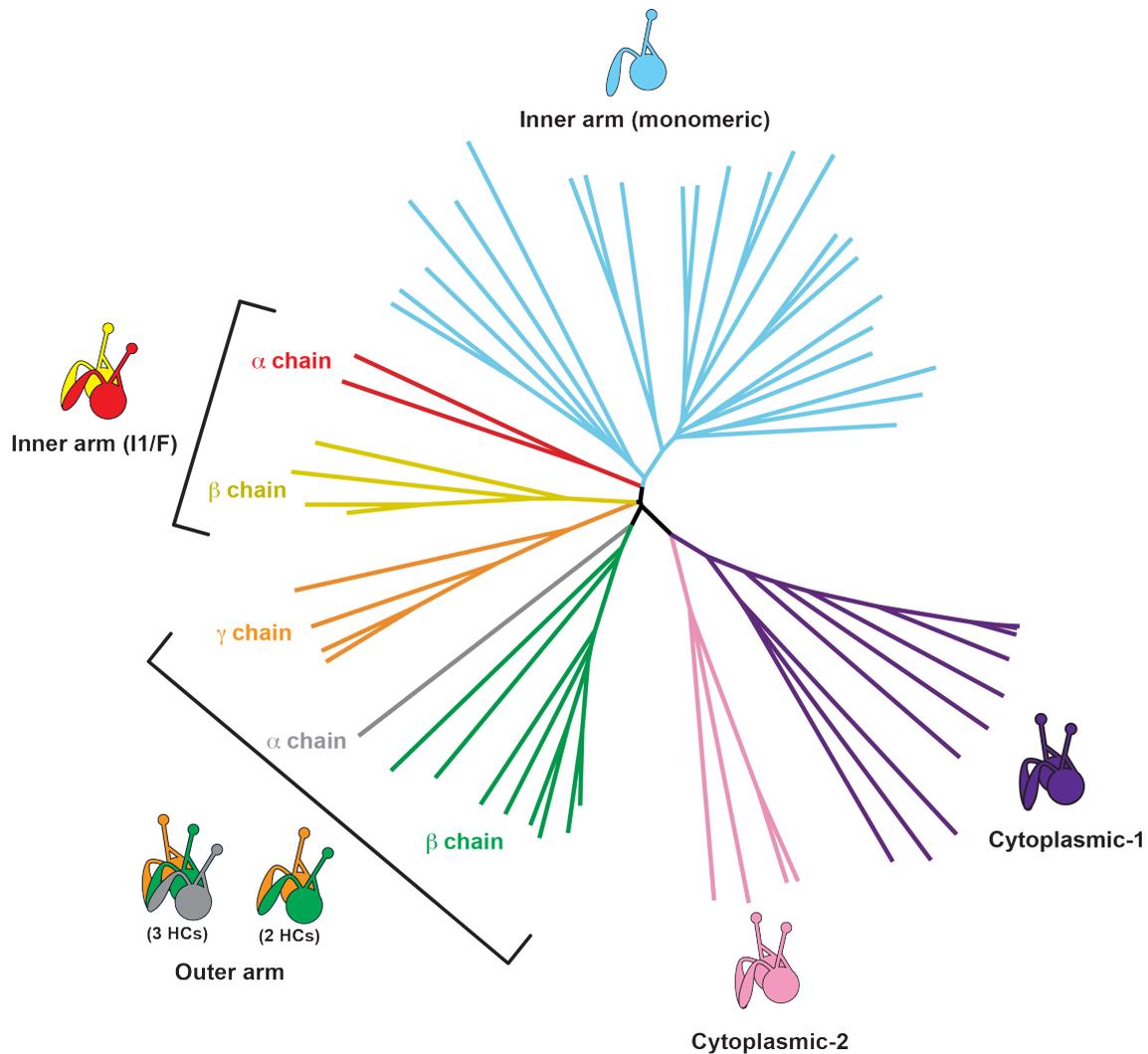
Figure 1.6: **Classification of dynein.** A phylogenetic tree of ∼ 60 distinct dynein heavy chains. Sequences cover all known types of dynein heavy chains from several species, including *A. gossypii, C. elegans, C. reinhardtii, D. rerio, D. melanogaster, G. gallus, H. sapiens, K. lactis, N. crassa, S. cerevisiae, T. thermophila.* Surrounding the phylogenetic tree are schematic representations of the heavy chains that make up each type of dynein motor. Figure adapted from Diamant & A. P. Carter 2013 with permission of the author.
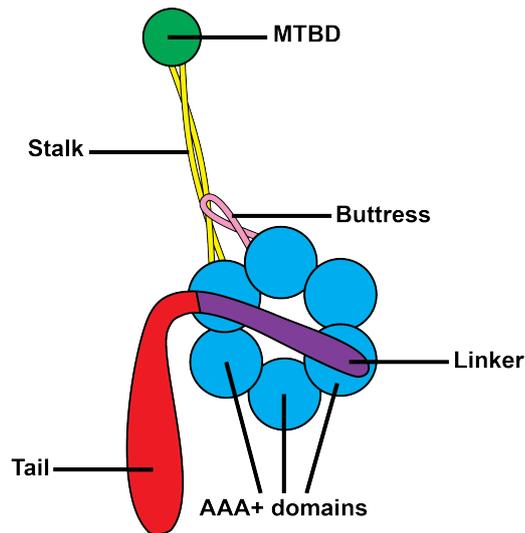
Figure 1.7: **Overview of dynein heavy chain structure.** Starting at its N-terminus, the heavy chain includes a cargo-binding tail domain (red), a mechanical linker (purple), and a motor domain (blue), which consists of a hexameric ring of AAA+ domains. A 15 nm coiled-coil known as the stalk (yellow) projects out of the ring from the fourth AAA+ domain and bears the globular microtubule-binding domain (green) at its tip. A shorter coiled-coil known as the buttress or strut (pink) similarly extends out from the fifth AAA+ domain and interacts with the middle of the stalk, possibly playing a role in the mechanochemical cycle. Figure adapted from Diamant & A. P. Carter 2013 with permission of the author.

cytoplasmic dynein to cargoes and anchors the motor to doublet MTs in axonemal dyneins.

1.2.2 Axonemal dyneins

The axoneme is a highly conserved structure found in cilia and eukaryotic flagella. In motile cilia and flagella, each axoneme is made up of a bundle of nine peripheral doublet MTs and two central singlet MTs that run longitudinally throughout the organelle. Each doublet MT has several arms projecting toward neighbouring doublet MTs as well as radial spokes that span the gap between the peripheral and central MTs (Fig. 1.8). In 1963, Ian Gibbons set out to identify the protein responsible for axonemal bending. He dissolved the membrane of *Tetrahymena* cilia and treated the exposed axonemes with various solutions. He found that after treatment with EDTA, the axonemes lost their ATPase activity, while the supernatant gained it. Using an electron microscope, Gibbons noticed that the arms that extend between adjacent peripheral MTs were missing from these inactive axonemes and named the large protein purified from the supernatant "dynein" after the dyne, a unit of force (S. M. King 2011).

The unicellular green alga *Chlamydomonas* has served as an important model organism for the study of axonemal dyneins. Several dynein deficient mutants that exhibit reduced flagellar motility or are completely paralysed have been described. By comparing their physiology, subcellular anatomy and genotypes, it was demonstrated that there are several types of inner dynein arms (IDAs), which are structurally and functionally distinct from the single outer dynein arm (ODA). The IDAs
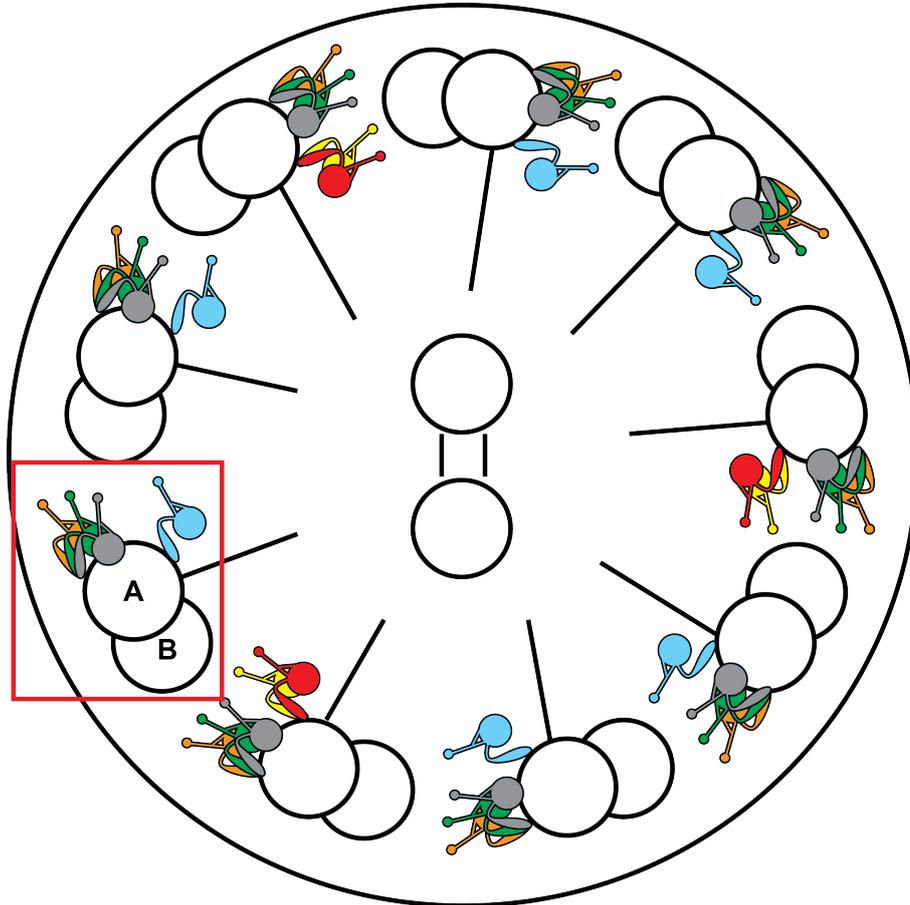
Figure 1.8: **Arrangement of various dyneins within the axoneme.** Cross-sectional view of an axoneme showing the 9 + 2 arrangement of microtubules and the location of various dyneins. The outer dynein arms are heterotrimers of three distinct heavy chains (two in some species). The inner dynein arms are monomers aside from the heterodimeric I1/f dynein. Colours correspond to Fig. 1.6. The red box outlines the doublet microtubule that is viewed in Fig. 1.9. Figure reproduced from Diamant & A. P. Carter 2013 with permission of the author.
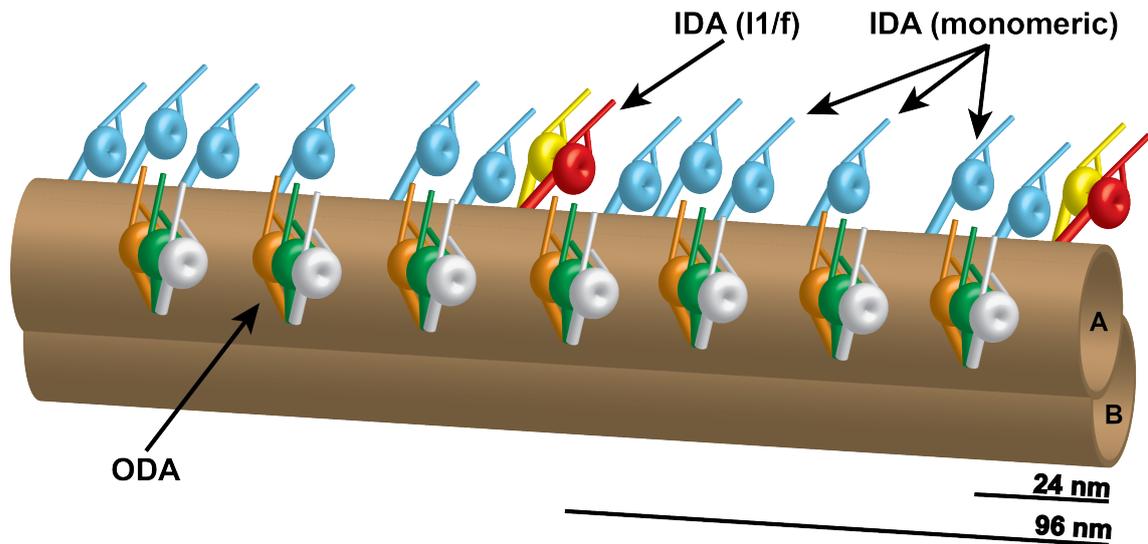
Figure 1.9: **Longitudinal view of the spacing and orientation of dyneins along a doublet microtubule within the axoneme.** Longitudinal view of a single doublet microtubule (enclosed in a red box in Fig. 1.8) demonstrating the spacing of outer dynein arms (ODAs) and inner dynein arms (IDAs) along the axoneme. There is one outer dynein arm every 24 nm, while a complex pattern of several inner dynein arms repeats every 96 nm. Colours correspond to Fig. 1.6. Figure reproduced from Diamant & A. P. Carter 2013 with permission of the author.

can be further resolved into seven distinct dyneins by using high-pressure liquid chromatography to fractionate high salt extracts of *Chlamydomonas* axonemes. The seven resulting peaks have been named a–g (Kagami & Kamiya 1992). Full genome sequencing predicted at least three more axonemal DHCs, which have only recently been purified due to a lower level of expression. These minor IDAs were shown to localise specifically to either the proximal or distal region of the axoneme, demonstrating yet another level of complexity in the coordination between the various IDAs (Yagi *et al.* 2009).

The outer dynein arm

The ODA can be found on the side of the doublet MT farthest from the central pair of MTs (Fig. 1.8). Each doublet MT is made of one complete MT, known as the A-tubule, and one incomplete MT, known as the B-tubule. The ODA is situated such that its tail subcomplex associates with the A-tubule of one doublet MT, and its stalk stretches across to the B-tubule of the neighbouring doublet MT. The link between the ODA and A-tubule remains intact throughout the mechanochemical cycle, while the MTBD at the tip of the stalk dissociates and reassociates with the B-tubule each step. Looking down the length of the axoneme, there is one ODA every 24 nm (Nicastro *et al.* 2006 and Bui *et al.* 2010)(Fig. 1.9).

Depending on the species, each ODA can contain either two or three DHCs. Electron tomography of *Chlamydomonas* axonemes shows the planes of the AAA+ rings of three DHCs (α, β and γ) stack parallel to the A-tubule they are attached to (Nicastro *et al.* 2005 and Takazaki *et al.* 2010). Phylogenetically, the α and β chains are very similar and probably result from gene duplication, while the γ chain

is more distantly related. Animal ODAs contain only two DHCs, which are most similar to the β and γ chains of *Chlamydomonas* (Fig. 1.6).

Each ODA also contains two DICs and eight DLCs, which associate with the tail portion of the DHC to form the tail subcomplex. Several of the DLCs are homologues of well-known proteins and these relationships have been used to suggest potential mechanisms for regulation of the ODA by these DLCs. For example, LC3 and LC5 are closely related to thioredoxin, and allow control of activity in response to changes in redox potential (Kamiya 2002).

There are associated complexes that are not considered part of the ODA, but are found adjacent to it and are thought to play a role in its regulation. One subcomplex of three proteins, known as the ODA docking complex (ODA-DC) is made primarily of coiled-coils that are roughly 24 nm in length (Casey *et al.* 2003). The ODA-DCs lay end-to-end along the A-tubule, which suggests that its primary function is to measure out the appropriate distance between adjacent ODAs (Bui *et al.* 2009). There are also two linkers, which are thought to contribute to the tight inter-dynein regulation necessary for axonemal beating. The first, known as the outer-outer-dynein linker (OOD) connects adjacent (down the length of the axoneme) ODA HCs, and is thought to help propagate ODA activity necessary for the production of the observed waveform. The second is the outer-inner-dynein linker (OID), which connects two out of every 4 ODAs to the IDAs and is thought to help coordinate activity between the ODAs and IDAs (Nicastro *et al.* 2006). In summary, although there is only one type of ODA, it is a highly complex molecular machine made up of several individual polypeptides and forming direct interactions with at least a dozen others.

The inner dynein arms

In contrast to the ODA, each IDA is much simpler in structure, but the interplay of the different types of IDAs results in complexity. The IDAs are found on the side of the doublet MT closest to the central pair. They are oriented in the same way as the ODAs, with their tail attached to the A-tubule of one doublet MT and the stalk attached to the B-tubule of its neighbour. Rather than repeating every 24 nm like the ODA, a complex pattern consisting of several IDAs repeats every 96 nm (Fig. 1.9).

IDA I1/f is phylogenetically, structurally and functionally distinct from the rest of the IDAs. It is a heterodimer of two DHCs, also referred to as α and β (not to be confused with the α and β DHCs of the ODA), which form distinct clusters in Fig 1.6. I1/f contains three ICs (140, 138, and 97 kDa). The 140-kDa IC anchors IDA I1/f to the doublet MT, while the 138-kDa either promotes or inhibits the activity of this dynein depending on its state of phosphorylation. The monomeric IDAs do appear to form some phylogenetic clusters, but these groups do not correlate well with groups that are based on structural, biochemical, and physiological properties (Hom *et al.* 2011). Structurally, each monomeric IDA has a single DHC, the tail of which interacts with an actin monomer. The actin subunit in turn interacts with a centrin DLC in IDAs b, e, and g and a p28 DLC in IDAs a, c, and d (Kamiya 2002). Therefore actin may provide a scaffold for DLCs to bind to the DHC tail in these dyneins.

Each 96 nm repeat of IDAs also contains a crescent-shaped complex called the dynein regulatory complex (DRC). This complex is situated directly on the doublet

MT and interacts with the radial spokes, the OID mentioned above and nexin, an elastic protein that ties together the bundle of nine peripheral doublet MTs. Because of its central location and interaction with both the radial spokes and the OID, the DRC plays a key role in coordinating the activity of the axoneme (Barber *et al.* 2012). The OIDs themselves promote communication between the ODAs and IDAs. Since there are two OIDs for every 96 nm unit in the IDA pattern, the two OIDs must interact with different parts of the inner arm. One OID forms a connection from an ODA to the DRC, while the other links an ODA to IDA I1/f (Nicastro *et al.* 2006). The discovery of these OID connections underscored the critical roles that the DRC and IDA I1/f have in regulating axonemal activity.

Functional characteristics of axonemal dyneins

Functional studies of various *Chlamydomonas* mutants revealed that not all axonemal dyneins are necessary for motility and helped to define a minimum combination of dyneins that must be present for axonemal beating to take place. The first, somewhat unexpected finding was that mutants lacking ODAs altogether still swim, but at less than half the speed of the wild type. Mutants lacking certain combinations of IDAs were also able to swim at a reduced velocity, but other combinations were paralysed. Specifically, if a mutant lacked two of the following three items, it was paralysed completely: (1) The entire ODA, (2) IDA I1/f, (3) IDAs a, c, and d. Analysis of these same mutants also helped characterise the types of motion each type of dynein is responsible for. For example, ODAs produce a high beat frequency, while IDAs promote the large amplitude of each beat.

In vitro motility assays performed with isolated dyneins provided further insights into the velocity and force produced by each specific dynein. For example, IDA I1/f, which appears to be very important to the function of the axoneme *in vivo*, moves MTs 95–99% slower than the other isolated dyneins (E. F. Smith & Sale 1991). Additionally, IDAs a, c, d, e, and g rotate MTs in addition to translocating them (Kagami & Kamiya 1992), which suggests that one of the key roles of these dyneins is to produce torque. Furthermore, IDA c has been reported to be a processive motor (Sakakibara *et al.* 1999), which was surprising both because it is a monomer, and because there is no immediately obvious benefit for such a motor in the axoneme.

The current model for reconciling all of these structural and functional observations describes two independent systems working in parallel within the axoneme. The relationship between the two remains unclear, but it appears that the IDA system is responsible for initiating axonemal beating, while the ODA amplifies the power. This would account for the fact that mutants completely lacking ODAs can still swim, while those lacking all IDAs cannot. It is also supported by the layout of the regulatory network, which seems to suggest that the signal originates in the central pair MTs and radial spokes, and moves outward through the IDAs, before arriving at the ODAs through the DRC and OIDs (Movassagh *et al.* 2010). Another perplexing observation is the large discrepancy in velocity between individual dyneins. One possibility is that some may only be involved in beat initiation while others are required for propagation, similar to how one shifts gears on a bicycle when accelerating from a complete stop. It has also been shown that different dyneins are active on each side of the axoneme (Hayashi & Shingyoji 2008). This combination of temporal and spatial coordination of the various axonemal dyneins is likely to play a critical role in producing the beating pattern of the motile cilium or flagellum.

1.2.3 Cytoplasmic dynein-1

Whereas all other dyneins are found in the axoneme, a single dynein, cytoplasmic dynein-1 (hereafter, dynein-1) carries out all dynein functions in the cytoplasm. In contrast to the axonemal dyneins, which collectively promote cellular locomotion (Milisav 1998), dynein-1's many functions involve transport within the cell. Like certain myosins and kinesins, dynein-1 transports vesicles, organelles (*e.g.*, mitochondria and nuclei), RNAs, and proteins to specific locations in the cell (Caviston *et al.* 2007, Varadi *et al.* 2004, Xiang *et al.* 1994, Dienstbier *et al.* 2009 and Z. Y. Yang *et al.* 2003). It is also responsible for the collection of misfolded and aggregated proteins (Ravikumar *et al.* 2005). Furthermore, due to its retrograde direction of motility, dynein-1 has been co-opted for a parasitic function. Through an example of co-evolution, many different viruses are capable of hijacking dynein-1 and taking advantage of its retrograde motion to increase their chances of infecting the cell nucleus (Döhner *et al.* 2002). These functions lead to dynein-1 playing a key role in processes as diverse as fast axonal transport, golgi dynamics, protein recycling, viral infection and several distinct stages of mitosis including spindle orientation, nuclear migration, spindle assembly, and chromosomal segregation during anaphase (Karki & Holzbaur 1999).

Dynein-1 forms a homodimer made up of DHCs, DICs, DLCs, and light intermediate chains (DLICs) (Fig. 1.5). Each motor contains a dimer of DHCs, which binds directly to a pair of one of two DLIC homologues (either DLIC1 or DLIC2). The DHCs also bind a dimer of DICs, which provide a scaffold for dimers of one to three different DLCs depending on the species. The role of the DLCs has been controversial, and is still not entirely clear. As the purpose of this section is to briefly introduce all the various dyneins, a more in-depth discussion of this controversy, and many other aspects of dynein-1 is provided in Section 1.3.

1.2.4 Cytoplasmic dynein-2

Even though cytoplasmic dynein-2 (hereafter, dynein-2) is most closely related to dynein-1 in phylogeny, general function, and structure, it is primarily found within or near cilia and flagella where it carries out its primary function of axoneme maintenance. Fig. 1.6 shows that the two cytoplasmic dyneins cluster together and are more closely related than either is to the axonemal dyneins. Functionally, dynein-2 is also similar to dynein-1 in that it participates in long-range transport. Dynein-2's main function is to bring intraflagellar transport (IFT) particles back to the cell body where they can be recycled (Rosenbaum & Witman 2002). The mechanism is thought to be similar to that used by dynein-1 to carry cargoes across the cell, except in this case the cargo is a form of proteinaceous raft (IFT particle) and the track is the axoneme (Mikami *et al.* 2002). When dynein-2 is defective, flagella are extremely short and IFT particles accumulate in their tips (Pazour *et al.* 1999). Dynein-2 also plays a role in signal transduction by relaying signals from the tip of a nonmotile cilium back toward the cell body. For example, mouse embryos lacking dynein-2 are not responsive to the Hedgehog signalling pathway and are therefore unable to specify ventral neural cell types in the neural tube (Scholey & Anderson 2006).

Structurally, dynein-2 is similar to dynein-1 in that it also contains a dimer of DHCs and two DLICs; however, it is only known to bind one type of DLIC called

DLIC3 (Tan *et al.* 2011). A recent study demonstrated that two DICs called WDR34 and WDR60 bind to dynein-2's DHC (Asante *et al.* 2014). Interestingly, WDR60 is larger than both WDR34 and the DIC that binds dynein-1, which suggests that dynein-2 may be asymmetrical. Furthermore, the same study also confirmed that the DLCs Tctex-1, Tctex-3, Roadblock-1, Roadblock-2, LC8-1 and LC8-2 all interact with dynein-2, in addition to dynein-1. Finally, they identified a new DLC called Tctex-2, which is unique to dynein-2 (Asante *et al.* 2014). Much less is known about the regulation of dynein-2 than that of dynein-1, but some likely adaptors that are necessary for proper function include IFT complex A (Jonassen *et al.* 2012), IFT complex B (Follit *et al.* 2009), and a subunit of kinesin-2 (which is required for anterograde IFT) called FLA10 (Kozminski *et al.* 1995).

1.3 The state of cytoplasmic dynein-1 research

Because of the central role that cytoplasmic dynein-1 (hereafter "dynein", unless otherwise specified) plays in so many essential cellular processes, it has been the subject of intense study ever since its discovery in 1987 (Paschal *et al.* 1987). However, dynein proved a much more elusive subject of scientific investigation than kinesin I or myosin V and progress in understanding dynein lagged far behind the other cytoskeletal motors for years. This was largely due to dynein's large size and the fact that it is a complex consisting of two copies of each of four to six protein subunits. Both of these characteristics made every step of expressing recombinant dynein a formidable challenge. Cloning, sequencing, expressing and purifying large protein complexes only became much more feasible in the last ten to fifteen years thanks to advances in molecular biology. Before these advances, most dynein studies were carried out either *in vivo*, or using protein purified from endogenous sources. Much was learnt about dynein despite these obstacles, but the eventual ability to design and create recombinant constructs accelerated dynein research.

1.3.1 Dynein motor domain

Early work on the motility of dynein was done using samples purified from endogenous sources. The first cytoplasmic dynein was purified from cow brains. Its *in vitro* motility was confirmed by observing that it could translocate MTs after being adsorbed onto glass slides (Paschal *et al.* 1987). Another early study discovered a large protein that was responsible for transporting vesicles in squid axons. This molecule's dynein-like structure was confirmed by negative stain EM (Gilbert & Sloboda 1989). A more detailed analysis of dynein's motility was made possible using latex beads coated with dynein purified from chick embryo brains. This study was precise enough to see that dynein sometimes moved across MT protofilaments and often took backward steps when doing so (Wang *et al.* 1995). Furthermore, native cow dynein was used to show that the addition of a co-factor called dynactin improves motor processivity (S. J. King & Schroer 2000).

A critical step toward understanding dynein's mechanochemical cycle came after a series of studies where groups introduced specific mutations into the various AAA+ sites of recombinant motor domains. This allowed them to measure the relative importance of each AAA+ site for ATPase activity, MT binding and MT gliding ability (Silvanovich *et al.* 2003, Kon *et al.* 2004, Reck-Peterson & Vale 2004 and
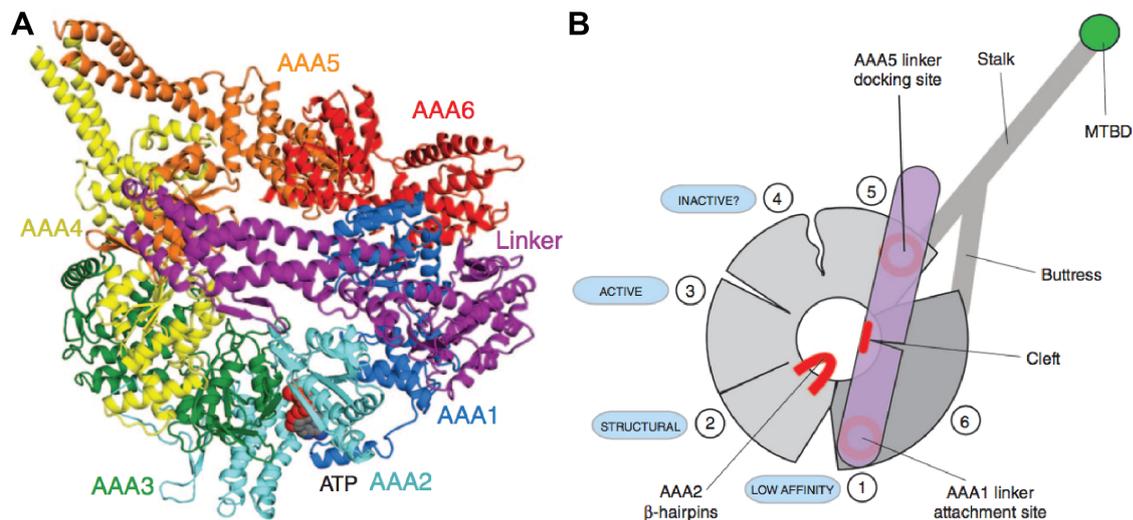
Figure 1.10: **The 3.3Å crystal structure of the *S. cerevisiae* dynein motor domain. A)** The 3.3Å crystal structure of the *S. cerevisiae* dynein motor domain (PDB accession code 4AKI). The construct began at residue 1364 of the *S. cerevisiae* DHC and eliminated the majority of the stalk and the MTBD. In this crystal, the motor domain is in an open conformation and the only AAA+ domain with ATP bound is AAA2. The linker is in the post-powerstroke position. **B)** A schematic summary of all the information gained from the crystal structure in **A**. The linker was discovered to have a cleft, which allows it to bend between the AAA5 docking site where it binds in the post-powerstroke state and the AAA2 β-hairpins, where it was predicted to bind during the pre-powerstroke state. A summary of what was learnt about ATP binding in each AAA+ site was included as well. Figures reproduced from Schmidt *et al.* 2012 with permission of the author.

Kon *et al.* 2005). Another major step resulted from the creation of a minimal motor domain that could be chemically induced to artificially dimerise and was fluorescently labelled. This degree of control during *in vitro* assays was essential in determining the effect of dimerisation on motility. It also provided a means for a much more precise characterisation of dynein's stepping behaviour than was possible with latex beads (Reck-Peterson *et al.* 2006).

Over the same time period, there was much work being done to elucidate the dynein motor domain's structure. Genetic analysis of the dynein gene first provided evidence that is made up of four or five AAA+ domains, although the actual number was not settled until much later (Gibbons *et al.* 1991 and Ogawa 1991). Recombinant fragments of dynein from *R. norvegicus* were used to define the C-terminal two-thirds of the DHC as the motor domain. This study also first provided evidence of the MTBD and stalk structures using sequence-based predictions and rotary shadowing EM (Gee *et al.* 1997). The first low-resolution model of the dynein motor AAA+ domains was proposed using a recombinant motor domain structure from *D. discoideum*. Using negative stain EM, it was possible to see a ring-like structure, with a central cavity (Samsó *et al.* 1998). This provided clear evidence that dynein was structurally distinct from myosin and kinesin.

One study that revealed much about both the structure and function of dynein made use of the axonemal IDA c from *C. reinhardtii*. By performing negative stain EM on this intact, endogenous dynein in various nucleotide conditions, it was possi-

ble to directly observe the movement of the linker relative to the tail domain (Burgess *et al.* 2003). Finally, the ability to fine-tune and overexpress recombinant dynein motors in large quantities culminated in crystal structures of the entire dynein motor domain (A. P. Carter *et al.* 2011, Kon *et al.* 2011, Kon *et al.* 2012, Schmidt *et al.* 2012 and Schmidt *et al.* 2014). These crystal structures provided unprecedented levels of detail about dynein motility, such as the mechanism by which the linker domain bends during different stages of the mechanochemical cycle.

One theme that is clearly visible throughout the history of dynein motor domain research is that many of the biggest turning points came from the freedom introduced by new recombinant expression techniques. Many of the questions that were answered using recombinant dyneins, through site-directed mutagenesis and protein engineering, would have been much more difficult to probe using endogenous samples only.

1.3.2   Tail domain

Recombinant studies of the dynein tail domain lagged far behind those of the motor because in addition to being larger than a single motor domain, the tail domain is also a multi-subunit complex. This makes recombinant overexpression and purification, which would be necessary for many potential biochemical and structural experiments, even more challenging. When I began my PhD in October, 2011, the first crystal structures of the minimal dynein motor domain crystal had just been published (A. P. Carter *et al.* 2011 and Kon *et al.* 2011), but there were still no reports of a purified minimal tail complex. One group had successfully purified endogenous levels of recombinant full-length dynein for *in vitro* motility assays (Kardon *et al.* 2009), and another had expressed tail complexes *in vivo* to observe the effects by fluorescence microscopy (Markus *et al.* 2009). Relative to the progress being made on the motor domain, there was little known about the structure and functions of the dynein tail domain, and one of the primary obstacles was the lack of a recombinant minimal tail sample.

Nevertheless, a basic knowledge of the components of the dynein tail, with some hints to its organisation were already available. As the work presented in this thesis was done using the *S. cerevisiae* (yeast) model system, the differences between yeast and mammalian dynein will be highlighted when applicable. Both yeast and mammalian dynein consist of a homodimer of DHCs, which binds two DICs and two DLICs (Fig. 1.5). However, in yeast, each DIC binds to two copies of the same DLC, called Dyn2, while in mammalian dynein, each DIC binds to three different DLCs, called light chain 8 (LC8), light chain 7 (LC7 or roadblock) and t-complex testis specific protein 1 (Tctex-1) (Kardon & Vale 2009). Furthermore, in yeast, there is just one gene for each dynein subunit, whereas in mammalian cells, there are often multiple genes for each dynein subunit, which are expressed as various isoforms. The resulting combinatorial possibilities give rise to a large number of potential dynein complexes of differing compositions (Pfister *et al.* 2006 and Palmer *et al.* 2009).

Functionally, there is a large difference between yeast dynein and mammalian dynein. Yeast dynein has only one known function, which is to position the mitotic spindle to ensure proper nuclear segregation (Eshel *et al.* 1993). Furthermore, it is not essential for this role as there is a redundant pathway for this function. On

the contrary, dynein is absolutely essential for mammalian cells, and it has multiple roles (listed above in Section 1.2.3). Despite this apparent difference between yeast dynein and mammalian dynein as a whole, the tail domain is thought to have at least three key functions common to all species: dynein homodimerisation, dynein cargo binding and dynein regulation. Indeed, the simplicity of the yeast system could be seen as a benefit of using it as a model system as there is only one known cargo and consequently less regulation required, compared to the dozens of proposed cargoes and extremely convoluted regulation present in mammalian cells.

The remainder of this introduction addresses each of the three general functions of the dynein tail in turn. In each section, the focus is on what was known at the beginning of my PhD, in October 2011, with an emphasis on the controversies and gaps in understanding that existed at the time. This is in order to provide the context and motivation for the work that I carried out during my PhD. However, at the end of each section, any major developments that occurred over the course of my PhD will also be summarised.

Dynein homodimerisation

One of the characteristics that processive motors have in common is that they are dimeric. This is expected as a dimeric motor is able to keep one head bound to the track at all times, preventing the motor from diffusing away (Vale & Milligan 2000). While it was known that dynein forms a homodimer, the exact mechanism of dimerisation had remained elusive. The first study to map a site for homodimerisation looked for a region within the DHC that was necessary and sufficient for co-immunoprecipitating another DHC out of solution (Habura *et al.* 1999). They concluded that the region between residues 627-780 in the *D. discoideum* DHC was the site of dynein homodimerisation.

That same year, the first crystal structure of any dynein component was published: the structure of LC8, one of the DLC (Liang *et al.* 1999). This paper suggested that LC8 forms a homodimer in solution, and the crystal structure confirmed the presence of an intricate dimer interface that includes domain swapping of a β-strand. Several years later, the structure of Tctex-1, another DLC, was solved (Williams *et al.* 2005). Despite having no sequence identity to LC8, the Tctex-1 structure adopted a similar fold and also formed a homodimer. Finally, just two years later, a crystal structure was solved of the ternary complex formed by mixing LC8, Tctex-1 and a peptide fragment of the DIC that included both of their binding sites (Williams *et al.* 2007)(Fig. 1.11A,B).

All three structures suggest that DLCs form homodimers, even in the context of the DIC. In fact, the consensus model suggested that one of the primary functions of the DLCs is to drive the homodimerisation of the entire dynein complex. A directed evolution study determined what the LC8 binding motif preference is, and then searched the human proteome for novel LC8 interactors (Rapali *et al.* 2011). This list included several proteins that homodimerise, so it was suggested that LC8 actually serves as a "dimerisation engine" and it's primary function is to bind other proteins and dimerise them. The specific case of dynein homodimerisation was further addressed by Nyarko & Barbar 2011 and Stuchell-Brereton *et al.* 2011. Nyarko & Barbar 2011 used NMR and fluorescence quenching to establish that the DIC N-terminus dimerises only in the presence of the DLCs. Using yeast as a model, Stuchell-Brereton *et al.* 2011 determined that the two DLCs that bind the DIC sta-
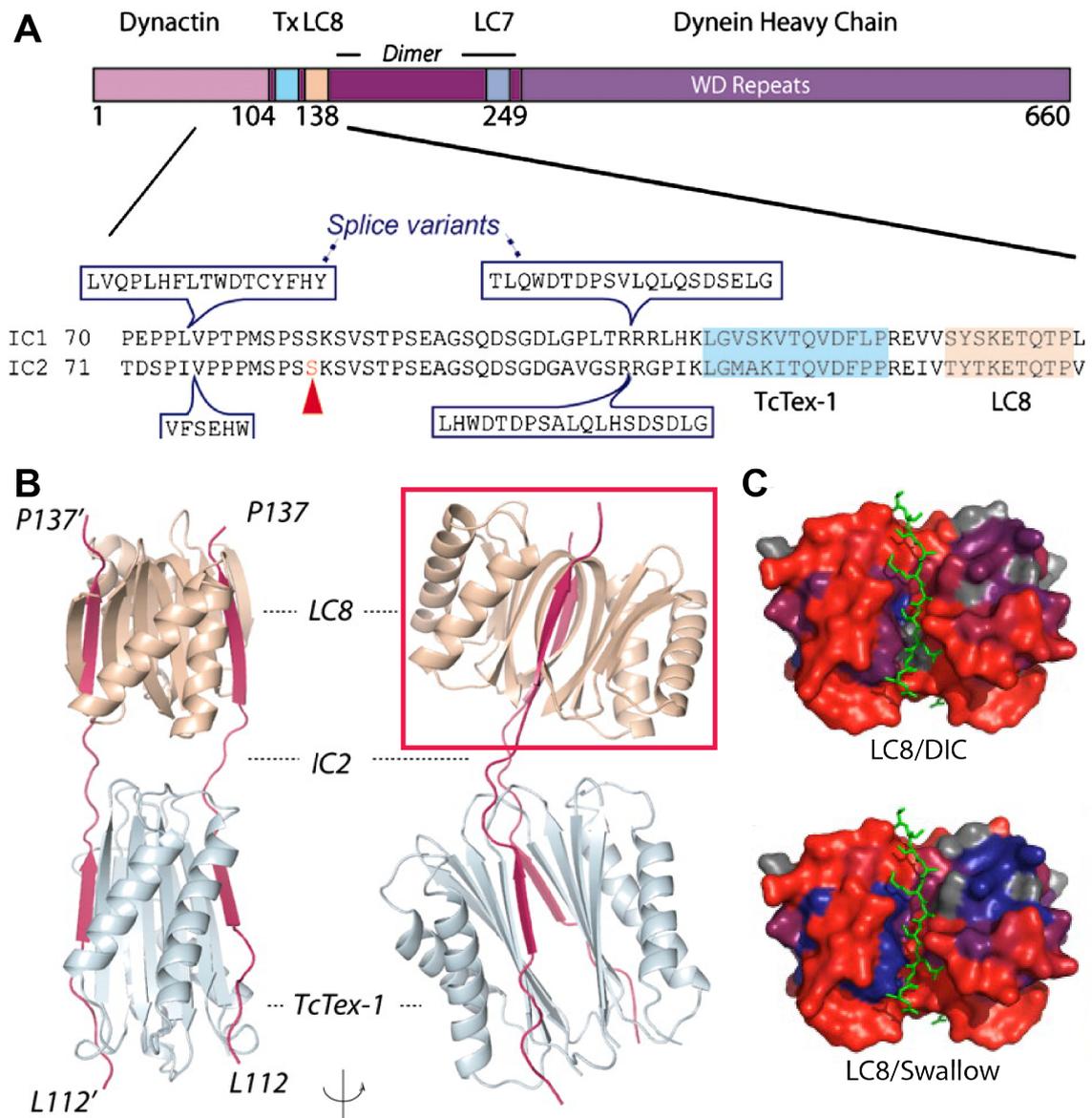
Figure 1.11: **Homodimerisation of a LC8/Tctex-1/DIC complex. A)** Above
is a schematic diagram of the DIC. It shows various binding sites including the two
DLCs bound in the crystal structure. Below are the sequences of two different DIC
variant peptides, as well as splice isoforms of each one. IC2 was the peptide used
in the crystal structure. The DLC binding sites are highlighted in cyan and wheat.
B) The crystal structure (PDB accession code 2PG1) of Tctex-1 (cyan) and LC8
(wheat) homodimers bound to two DIC peptides (residues 104–138; magenta). Each
DIC peptide binds to one side of each LC dimer producing a 2:2:2 complex. Residues
112–124 of the IC contact Tctex-1, and residues 129–137 contact LC8. Two views
are shown, rotated by 90° about the vertical axis. **C)** Two crystal structures of
LC8 (red surface representation) bound to different peptide fragments (green stick
representation). The view of both structures is similar to the view of LC8 in the red
box in **B**. Above is LC8 bound to a DIC peptide (PDB accession code 2P2P) and
below is LC8 bound to a potential cargo (Swallow) peptide (PDB accession code
2P1K). The two structures show that both DIC and Swallow occupy the same cleft
of the LC8 dimer. **A** and **B** adapted from Williams *et al.* 2007 with permission of
the journal and **C** adapted from Benison *et al.* 2007 with permission of the authors.

bilise the DIC dimer, and in doing so support the DIC's interaction with dynactin, an important co-factor of dynein.

Evidence that the DLCs drive dimerisation of the DICs does not exclude the possibility that the DHCs also homodimerise within the region determined by Habura *et al.* 1999. Nevertheless, this was the current state of knowledge about dynein homodimerisation as I began my PhD, and so one obvious question we hoped to explore with a minimal tail complex was how the tail dimerises.

During my PhD, at least two more publications came out, both reinforcing the model that the DLCs are promiscuous proteins that drive the dimerisation of various other proteins, of which dynein is just one example (Romes *et al.* 2012 and Rao *et al.* 2013).

Dynein cargo binding

An early model for how dynein binds cargoes was suggested by the fact that LC8 is known to bind both dynein, and many other proteins, which suggests that it could be a cargo adaptor. When two possible LC8 binding motifs, KXTQTX or XGIQVD, were identified using a pepscan technique (Rodríguez-Crespo *et al.* 2001), the number of LC8 binding partners (and therefore potential dynein cargoes) increased greatly. Some proteins that were thought to be dynein cargoes based on LC8 binding or the pepscan results include neuronal nitric oxide synthase (nNOS) (Jaffrey & Snyder 1996), inhibitor of NF-κB transcription factor (IκBα) (Crépieux *et al.* 1997), p53 binding protein 1 (Lo *et al.* 2005), oestrogen receptor (Rayala *et al.* 2005), and even the microfilament-based motor described above, myosin V (Espindola *et al.* 2000 and Hódi *et al.* 2006). Furthermore, most of these proteins were somehow inhibited upon LC8 binding, suggesting that dynein may inhibit the cargo as it is being transported.

However, in light of structural information that revealed how LC8 binds the DIC, serious doubts were cast on the role of LC8 as a dynein cargo adaptor. Another study presented two crystal structures similar to the one shown in Fig. 1.11B (Benison *et al.* 2007). The first of these new structures showed a LC8 dimer bound to a DIC peptide (Fig. 1.11C, top). The second showed a LC8 dimer bound to a peptide from Swallow, a potential dynein cargo that contains the same KXTQTX motif as the DIC (Fig. 1.11C, bottom). A comparison of these two structures revealed that Swallow and the DIC bind in the same cleft of the LC8 dimer. Therefore, LC8 cannot bind both dynein and swallow at the same time. This finding lent further credence to the idea that LC8 is only associated with dynein due to its capacity to drive dimerisation, and that it might drive dimerisation for many other proteins as well.

Many other mechanisms for cargo binding in mammalian cells were proposed. One example involves a cargo adaptor discovered in *D. melanogaster* called Bicaudal D, (BICD), which interacts with dynein (Bullock & Ish-Horowicz 2001). Its mammalian homologues, BICD1 and BICD2, were shown to colocalise with both dynein and its activator, dynactin (Hoogenraad *et al.* 2001). Finally, it was shown that BICD1 and BICD2 also co-localise with Rab6 on the *trans*-Golgi network (TGN)(Matanis *et al.* 2002). With this, a complete connection could finally be traced between dynein and one of its known cargoes.

In yeast, there is only one known cargo, called Num1. Num1 is a large cortical protein that interacts with the DIC and the MT network (Farkasovsky & Küntzel

2001). An offloading model was proposed to explain how dynein achieves its role in nuclear migration with yeast cells. Inactive dynein is targeted to the growing plus end of a MT, which it tracks until it reaches the cell cortex. At the cortex, it becomes anchored to the cortical Num1 and is activated. It then pulls the MT toward itself as if it were walking toward the minus end of the MT. Since dynein is anchored in place, the MT moves toward it, bringing the nucleus with it, which is attached via the centrosome at the minus MT minus end (Lee *et al.* 2005).

Concerning dynein's mechanisms for binding cargoes, one extremely important advance took place during my PhD. Two different groups found that dynein only binds its activator, dynactin, in the presence of a cargo adaptor, such as BICD2 (Schlager *et al.* 2014 and McKenney *et al.* 2014). Furthermore, mammalian dynein on its own *in vitro* never displayed the processivity observed *in vivo*. But upon forming a dynein-dynactin-adaptor triple complex, the dynein became super processive and displayed motility characteristics similar to *in vivo* motors. This is discussed further in the context of dynein regulation in the following section.

Dynein regulation

The dynein tail is the primary site of regulation for dynein. Regulation of dynein can be divided into two important aspects. The first applies to all dyneins, and is concerned with if and how dynein is switched from an active state to an inactive one. As described above, kinesin-1 is capable of autoinhibition via a dramatic conformational change, and so it is at least plausible that dynein may have a similar mechanism. The second aspect, which only applies to dyneins with multiple functions (and therefore not yeast dynein) is concerned with cargo specificity. In humans, there are at least 29 different cytoplasmic kinesins, each specialised for its role as an anterograde MT motor within the cytoskeleton. However, there is only one cytoplasmic dynein to perform most of the retrograde MT-associated motility within the cytoskeleton. Therefore, the question arises of how dynein is controlled to ensure that it always functions with the correct cargo, at the correct place and at the correct time.

One proposed mechanism for the complete activation/inactivation of dynein is through differential phosphorylation of specific sites within the dynein tail. An early study separated two distinct pools of dynein from cow brains and determined their phosphorylation states. The total cellular pool had a higher degree of phosphorylation on the DHC than a pool consisting of only anterograde moving dyneins (*i.e.*, inactive dyneins being transported by kinesin), suggesting that phosphorylation of the DHC may be correlated with activation (Dillman & Pfister 1994). Another study found that Casein Kinase II, which is a highly conserved serine/threonine kinase, binds to and phosphorylates dynein (Karki *et al.* 1997). Over the course of my PhD, three more studies were published, all suggesting regulation of dynein by differential phosphorylation of the DIC, although some of these proposed changes in the specific behaviour of dynein rather than a simple activation/deactivation switch (D. J. Mitchell *et al.* 2012, Pullikuth *et al.* 2013 and Blasier *et al.* 2014).

Another way that dynein may be activated and deactivated is through the presence of its most important co-factor, dynactin. Dynactin is a 1 MDa, multi-subunit complex (Fig. 1.12) (Schroer 2004). It has been shown that for *in vitro* transport of vesicles, dynein requires dynactin (S. R. Gill *et al.* 1991). Furthermore, *in vivo* experiments suggested that dynactin is required for all of dynein's cellular functions
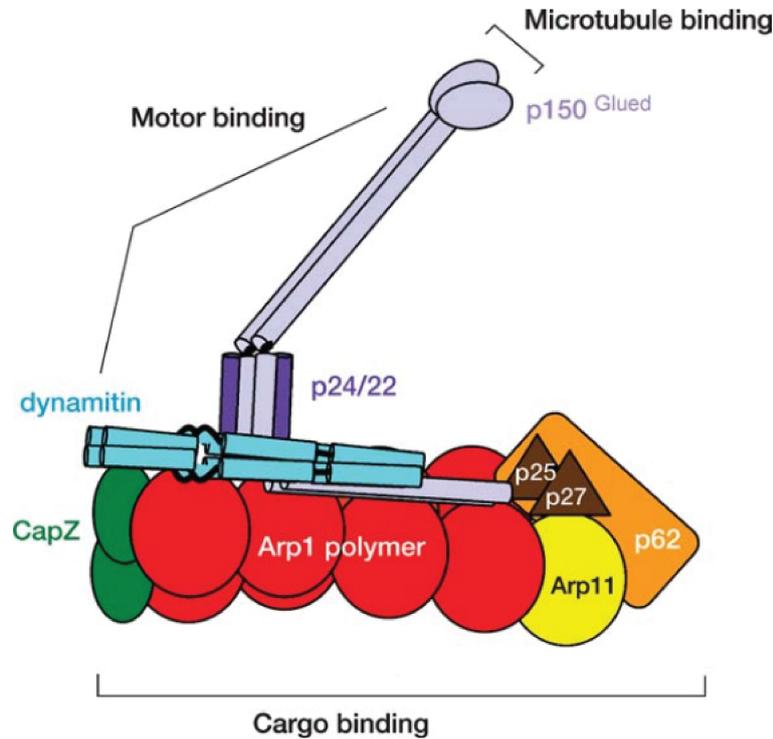
Figure 1.12: **Schematic model of the dynactin complex.** An overview of the structure of dynactin, as it was understood when I began my PhD. An Arp1 filament is capped on either end by two different capping subcomplexes. The shoulder (blue and purple) contains the largest dynactin subunit, p150$^{\text{Glued}}$, which was thought to be the primary site of dynein interactions. p150$^{\text{Glued}}$ also contains a MTBD at its tip. Figure reproduced from Schroer 2004 with permission of the author. See Fig. 4.18 for the cryoEM structure of dynactin published in 2015.

(Schroer 1996). Both of these results provide evidence for dynactin as a dynein activator (in fact this is the origin of its name: *dyn*ein *act*ivator, S. R. Gill *et al.* 1991). However, further experiments complicated the matter. For example, recombinant mammalian dyneins were generally not processive on their own when tested *in vitro* and mixing dynactin and dynein together *in vitro* failed to produce processive motors, as they did not form a stable complex. As described above, it was only upon the inclusion of a cargo adaptor, such as BICD2, that a stable complex formed, and a processive motor could be observed (Schlager *et al.* 2014 and McKenney *et al.* 2014). It is therefore possible that the S. J. King & Schroer 2000 study may have only seen processive movement because the beads had both dynein and dynactin bound directly to them and that in reality they were not interacting directly. In contrast to mammalian dynein, it has long been thought that yeast dynein does not require dynactin to be processive, although it also displays an increase in processivity when dynactin is present (Kardon *et al.* 2009).

There are also many possible mechanisms by which mammalian dynein might achieve cargo specificity. As mentioned above, in mammalian cells, there are multiple genes expressed as various isoforms for all but one dynein subunit. So one idea is that the heterogeneity required for dynein to perform so many different roles comes from mixing and matching gene products to create different variants of dynein (Pfister *et*

al. 2006). Some preliminary work has been done to explore this possibility. Dyneins that were formed with different DLICs seemed to show some different characteristics; DLIC1 was important for ER-to-Golgi transport, whereas DLIC2 was important for recycling endosome distribution and cytokinesis (Palmer *et al.* 2009). However, the different DICs seem to be more tissue-specific rather than having separate pools of distinct dyneins within a single cell (Ha *et al.* 2008).

Another way that one dynein might be adapted into functionally distinct motors would be through the binding of different cofactors. In addition to dynactin, dynein requires various adaptors in order to recruit cargoes, such as BICD and the Rod-ZW10-Zwilch (RZZ) complex (Kardon & Vale 2009). Therefore, depending on the adaptor present, dynein might only bind specific types of cargoes. Another important complex, made of LIS1 and NUDE or LIS1 and NUDEL, was known to bind dynein in at least two different places and to have multiple regulatory effects on the motor (Kardon & Vale 2009). During my PhD, a lot of exciting work has been published on LIS1. It was first shown that LIS1 acts as a clutch, disengaging the ATPase activity of the AAA+ ring from the MTBD (Huang *et al.* 2012). Then, a cryo-EM structure of LIS1 bound to the motor domain shed light on the mechanism, which is essentially due to LIS1 binding to the AAA+ ring and sterically blocking the linker domain from completing its powerstroke (Toropova *et al.* 2014).

Finally, some results suggest an interesting interplay between different cofactors. When I began my PhD, the consensus model was that dynactin binds to the N-terminus of the DIC through its own p150$^{\text{Glued}}$ subunit (Schroer 2004). This same region of the DIC was also thought to bind to the LIS1/NUDE complex. One study suggested that the dynactin and LIS1/NUDE bind to the DIC in a mutually exclusive way, resulting in a reciprocal regulation between the two co-factors (McKenney *et al.* 2011). While we have no evidence to exclude the possibility that the DIC also binds to dynactin, some of the work I describe in Chapter 4 suggests that there is another, very prominent dynein/dynactin interface between the Arp1 filament of dynactin and the DHC N-terminus of dynein (Urnavicius *et al.* 2015).

Clearly, even at the time that this thesis is being written, the functions of the dynein tail domain remain poorly understood. Some previously controversial issues have been resolved, but as is often the case, this only led to more questions. However, many of the tools that were not available just a few years ago are finally in hand and ready to be put to use. Therefore it is an exciting time in this important field as many of the longstanding questions about dynein function are set to be resolved in the near future.

1.4   Experimental aims of this thesis

The last section of this introduction has provided just a glimpse of the complicated body of knowledge about the dynein tail, resulting from efforts made using a combination of genetics, biochemistry, structural biology and cell biology. Having witnessed first-hand how using a recombinant construct of the dynein motor domain brought such clarity, and ultimately a structure, to a field that had been clouded by contradictory results, I was inspired to help accelerate the same process in the dynein tail.

When I began my PhD, there were still no reports of the expression of a recombinant minimal dynein tail complex. So my first aim was to develop a strategy

for the expression and purification of such a complex. My next aim was to characterise this minimal tail complex biochemically, as having an *in vitro* sample of a minimal dynein tail would make certain experiments possible that had never been done before. For example characterising its native phosphorylation state, testing its stabilty in different buffer conditions and understanding the function of the different subunits. My final aim was to characterise the minimal tail complex structurally, using whatever methods would be most suitable. In practice, after trying crystallography on the entire tail complex, I shifted to a negative stain EM approach for that particular construct. However, I managed to solve the crystal structure of a smaller piece of the dynein tail. This was the first structure of any part of the DHC tail, and it played an important role in understanding how dynein binds dynactin.

Chapter 2

The expression of a minimal dynein tail complex

2.1 Introduction

The dynein tail is the site of cargo binding and is involved in regulation (Kardon & Vale 2009). Previously, there was no high-resolution structural information on the dynein tail. In order to shed light on dynein regulation and cargo binding, my initial goal was to solve this structure using X-ray crystallography. The primary prerequisite for this techniques is a large amount of a homogeneous, stable sample without intrinsic flexibility. This chapter describes my efforts to establish a protocol for producing such a sample.

Before I began, there were no published methods for purifying the dynein tail complex separate from the motor domains. Therefore, my first goal was to purify a minimal dynein tail complex from *S. cerevisiae* (yeast). I chose yeast because it had already been established as a useful model system for dynein for the following reasons. First, genomic modification via homologous recombination is both simple to execute and extremely reliable. Furthermore, yeast dynein is simpler than mammalian dynein both structurally and functionally. In fact, yeast dynein only has one function, for which it is not essential (Eshel *et al.* 1993). Therefore, the endogenous dynein genes of yeast are free to be modified using homologous recombination without causing a deleterious phenotype.

The first tail truncation construct I attempted was unsuccessful, so I took a step back and first tried to overexpress the full-length dynein complex. Once I accomplished this, I began truncating the dynein heavy chain (DHC) in search of a minimal tail complex that would still bind all accessory chains while eliminating the flexible linker and the motor domains. After finding a suitable truncation of the DHC, I explored different expression systems, in hopes of increasing yield and purity. Finally, I screened crystallisation conditions and determined a low-resolution initial 3D model of the tail complex using negative stain EM.

Some of the initial work on expressing and purifying the dynein complex from yeast was done in collaboration with Max Schlager (Section 2.2.2). Andrew Carter constructed the plasmid that expresses the NH truncation in four separate cassettes (Section 2.2.6). EM data were collected and processed with the help of Linas Urnavicius, Andrew Carter helped with manual particle picking (Section 2.2.8) The remainder of the experiments described in this chapter are my own work.

2.2 Results

2.2.1 Expression of the dynein tail complex as a polyprotein

I first attempted to express the dynein tail complex using a polyprotein approach since this technique had proved advantageous for the recombinant expression of several other complexes (de Felipe *et al.* 2006, Vijayachandran *et al.* 2011). The defining feature of this approach is that the genes coding for all the components of a complex are fused together to create a single open reading frame (ORF). Individual genes are separated by short sequences that code for a flexible linker containing the recognition site for a highly specific protease. Including the gene for the protease itself at the 5′ end of the ORF ensures that that protease levels are high enough to cleave the subunits apart either during, or shortly after polyprotein expression. The primary benefit of using a polyprotein, which was my motivation for using it to express the dynein tail, is that the stoichiometric expression of all subunits prevents one subunit from becoming limiting (Chaplin *et al.* 1999, Lorens *et al.* 2004). Furthermore, the subunits are translated consecutively by a single ribosome. I speculated that the resulting temporal and spacial proximity might aid in proper complex formation, especially when subunits contain inherently disordered regions that require a binding partner to fold properly.

As this was my first attempt to express the dynein tail recombinantly, I had little information about where to truncate the DHC. I decided to include the N-terminal 1363 residues since the construct used in the crystal structure of the motor domain began at residue 1364 (A. P. Carter *et al.* 2011). Furthermore, an *in vivo* study had previously shown that this truncation was viable in yeast cells (Markus *et al.* 2009). For the other three subunits, I included the full-length proteins.

I amplified the first 4089bp of *DYN1* (which codes for the first 1363aa of the DHC) and the full-length genes of *PAC11*, *DYN3* and *DYN2* (which code for the DIC, DLIC and DLC, respectively) from genomic DNA of the wild type W303 yeast strain (CY176, Fig. A.4). As *DYN2* contains two introns, I designed primers to amplify each exon and then stitched them together using a PCR-based approach to create my own intron-free version of this gene. I also amplified the TEV protease gene from the Carter lab plasmid p107 (Fig. A.5) and a ZZ-tag from Carter lab plasmid p61 (Fig. A.5). I then used PCR stitching to link together all of these genes in frame, with the appropriate linkers, in order to create an 8kb insert. The insert began with the TEV protease, linked to the ZZ-tag by a TEV cleavage site. The ZZ-tag was linked to the first 4089bp of *DYN1* by a PreScission cleavage site and then the following three genes were all linked by additional TEV cleavage sites. This insert was cloned into the pYES2 yeast expression vector (Thermo Fisher Scientific), which places the inserted cassette under control of the *GAL1* promoter (Fig. 2.1).

The polyprotein plasmid was transformed into two yeast strains for expression: DSY-5 (MoBiTec, CY223 in Fig. A.4) and BCY123 (Nagai lab, MRC-LMB, CY282 in Fig. A.4) and 12L of each strain was grown. The DSY-5 strain grew much slower than the BCY123 one, taking 48 hours longer to reach the density required for harvesting. Furthermore, both resulting pellets were much smaller than expected from 12L of yeast. The DSY-5 pellet weighed $\sim$ 60g and the BCY123 pellet weighed $\sim$ 30g, compared to a typical yield of $\sim$ 150g of yeast from 12L of culture. The combination of slow growth rates and small pellets suggest that overexpression of this dynein tail construct may limit cell growth. Nonetheless, a test purification
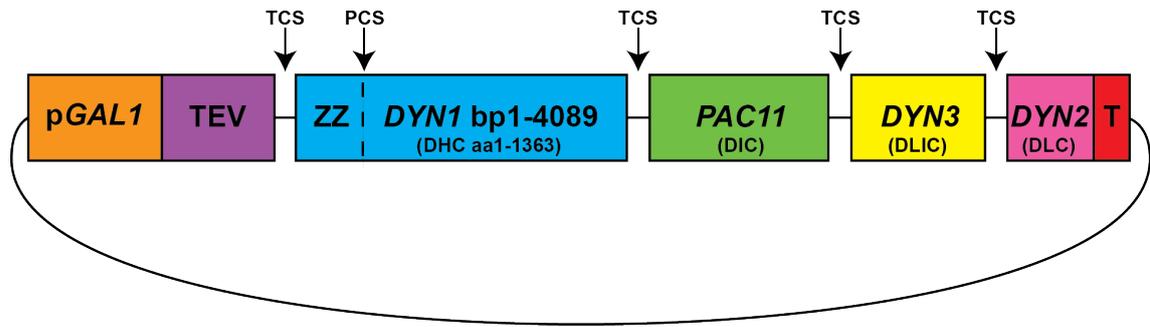
Figure 2.1: **Schematic diagram of the dynein tail polyprotein plasmid.**
The four subunits of the dynein complex were expressed from a single open reading frame and then cleaved apart by the TEV protease that was expressed from the same mRNA transcript. The open reading frame was preceded by the *GAL1* promoter (p*GAL1*) for high level inducible expression and followed by the *CYC1* transcriptional terminator (T) for efficient termination of mRNA. The genes were separated by linker peptides containing TEV cleavage cites (TCS). An N-terminal ZZ-tag was fused to the DHC by a PreScission cleavage site (PCS).

was performed on each pellet to test whether the complex actually expressed and if so, whether it could be purified successfully. For this test purification, I followed a protocol similar to the one used for purification of the yeast dynein motor domain (Reck-Peterson *et al.* 2006). Cell lysate was incubated with IgG beads (GE Healthcare), which are designed to bind with high affinity to the N-terminal ZZ-tag of the DHC. After washing away unbound protein, bound protein was released by PreScission cleavage and the supernatant was collected.

When the samples were concentrated and analysed by SDS-PAGE (Fig. 2.2) it appeared that the expression and purification were at least partially successful. Both strains produced a protein that ran at the expected height of the truncated DHC, but in BCY123 it appears to have remained bound to the IgG beads even after PreScission cleavage. While there were some very faint bands near where the DIC and DLIC were expected to run on the gel, there were also several other bands of similar heights present. Therefore, the identities of these lower-MW bands were not clear.

From this experiment it wasn't possible to determine whether the limited, slow cell growth and low protein yield were due to the polyprotein approach or the somewhat arbitrarily chosen DHC truncation at residue 1363. So I decided to take a step back and attempt to purify full-length dynein complex before testing DHC truncations. However, making the polyprotein with a full-length DHC would have required making the already unwieldy 15kb polyprotein plasmid 8kb larger. So for the expression of the full-length dynein complex and identification of a suitable DHC truncation I turned to a different strategy, which avoids the problem of making a polyprotein on a plasmid altogether by overexpressing the genomic copies of all four dynein genes.
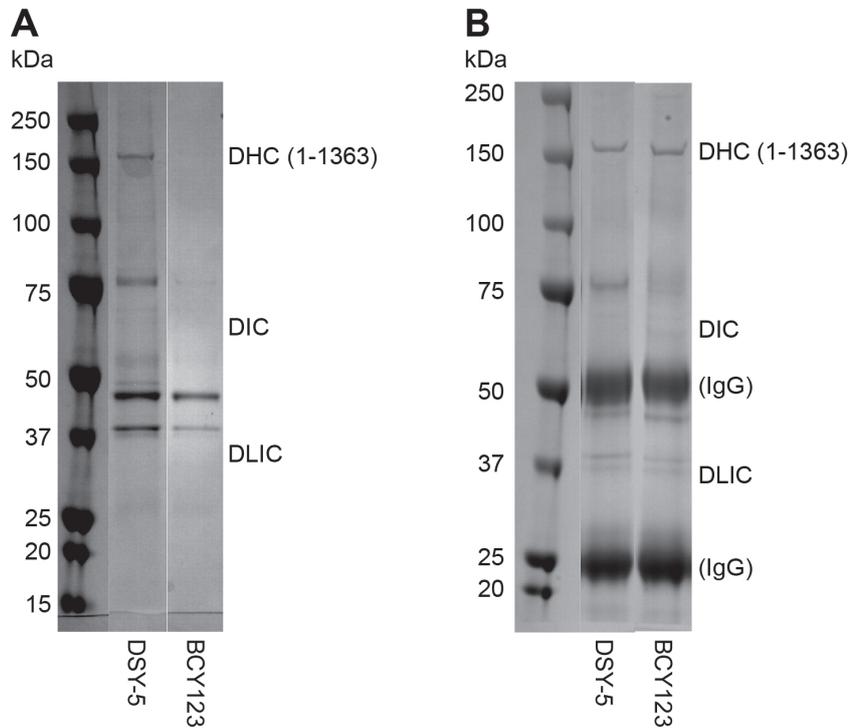
Figure 2.2: **Gels showing the dynein tail polyprotein purified from both DSY-5 and BCY123. A)** A polyprotein dynein tail was expressed in two yeast strains, DSY-5 and BCY123. 12L of each strain was grown and the protein was purified using affinity chromatography and then concentrated and analysed by SDS-PAGE. The predicted heights for the truncated DHC, DIC and DLIC are indicated to the right of the gel. The DSY-5 contains a band at the correct height for the truncated DHC, but its level was low. **B)** A slurry of the IgG sepharose beads used in the purification described in **A** was boiled and the supernatant was analysed by SDS-PAGE. In addition to the predicted heights for the dynein subunits, the expected heights for the two IgG subunits are indicated to the right of the gel. Both samples contain bands at the correct height for the truncated DHC.

30

2.2.2 Overexpression of endogenous dynein genes using multiple *GAL1* promoters

Using the commercial yeast expression strain, DSY-5 (MoBiTec, CY223 in Fig. A.4), as a background strain we replaced the endogenous promoter at the 5′ end of each dynein gene with a copy of the strong, inducible *GAL1* promoter. The *GAL1* promoter is convenient because it is inhibited by glucose, and induced by galactose, both of which support the growth of yeast. DSY-5 was selected as the background strain because it lacks two major vacuolar proteases, Pep4 and Prb1, which increases yield and limits non-specific degradation. It also contains mutations in four different biosynthetic pathways, *leu2*, *trp1*, *ura3-52* and *his3*. This allowed us to perform homologous recombination in four consecutive steps, using a different selectable marker for each one. While inserting a *GAL1* promoter at the 5' end of each dynein gene, we simultaneously reverted one of the biosynthetic pathways back to wildtype. This provided the selection pressure necessary for picking colonies with the desired genotype. The *GAL1* promoter inserted in front of *DYN1* (the DHC gene) was followed by ZZ-tag, so that it was fused to the DHC N-terminus by a TEV cleavage site. Each intermediate strain was verified by colony PCR and the final strain was verified using colony PCR and Sanger sequencing. The resulting strain is referred to as CY286 (Fig. A.4)(Fig. 2.3).
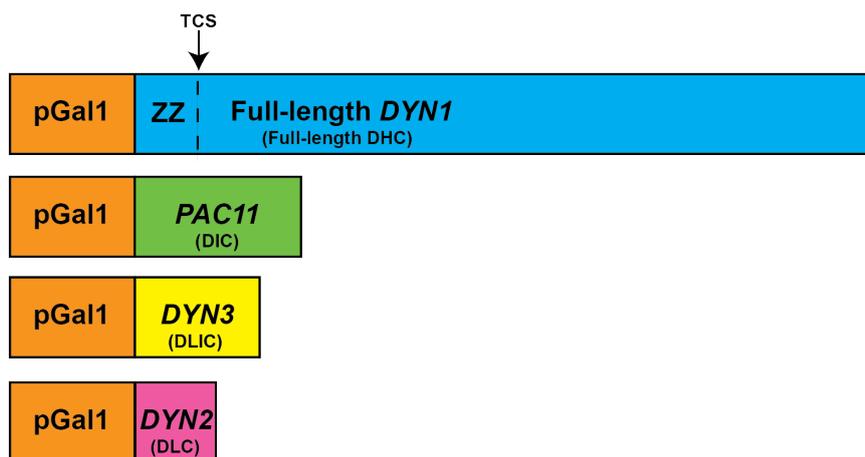
Figure 2.3: **Schematic diagram of CY286, a yeast strain engineered to overexpress the full-length dynein complex using the multiple *Gal1* promoter approach.** Using homologous recombination, the promoter for each of the four dynein subunit genes was replaced by the *GAL1* promoter (p*GAL1*) for high level inducible expression. An N-terminal ZZ-tag was fused to the DHC by a TEV cleavage site (TCS).

We performed an initial test expression and purification from 2L of culture following the same protocol used above for the dynein tail polyprotein, except we used TEV instead of PreScission to cleave the protein off of the IgG sepharose beads. I analysed the purified CY286 sample and a boiled slurry of the IgG sepharose beads using SDS-PAGE and found that the full length DHC was present in both, but the other subunits were not clearly visible in this first experiment (Fig. 2.4). The low level of full-length DHC suggested that even if the other subunits were present in the sample, they would not be visible since their MWs are several-fold less than
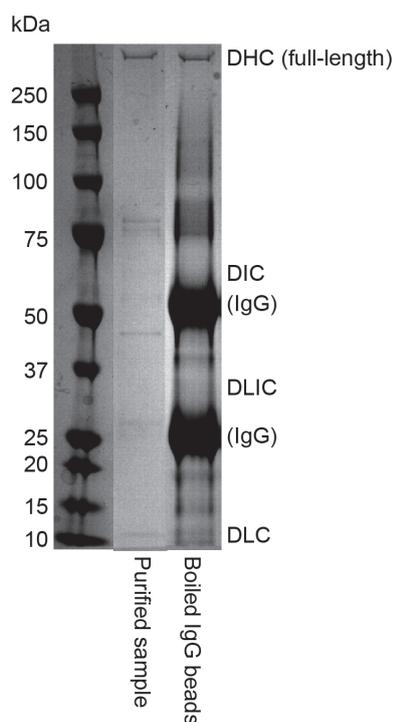
Figure 2.4: **Gel showing a test purification of the full-length dynein complex from 2L of CY286.** 2L of CY286 were grown and a small-scale purification was performed. After TEV cleavage and concentration, the entire sample was analysed by SDS-PAGE. The purified sample showed a clear band at the expected height of the full-length DHC, but there were no visible bands at the expected heights for the other three subunits. After TEV cleavage, a slurry of the IgG sepharose beads was boiled and the supernatant was analysed by SDS-PAGE. This sample also has a band at the height expected for the full-length DHC. In addition to the predicted heights for the dynein subunits, the expected heights for the two IgG subunits are indicated to the right of the gel.

that of the full-length DHC. Therefore, it was not possible to conclude whether the entire complex was co-purifying or not.

In hopes of purifying enough protein to see all the subunits clearly on a gel I grew another 36L of CY286 culture and purified the full-length dynein complex from it. After TEV cleavage and concentration, there was $\sim 300 \mu$L of sample, with a protein content of ~ 5mg/mL. I loaded 20μL of this sample into both a $4-12\%$ gradient gel as well as a 12% gel in order to ensure adequate band separation for identifying the dynein subunits (Fig. 2.5). Both gels showed bands at the expected heights for the DHC, DLIC and DLC. While there was a band at the height expected for the DIC, it was much less prominent than the other three. The DIC band was also more broad than the DLIC band, possibly indicating degradation of the DIC. As the DLCs only bind the dynein complex through the DIC, it was puzzling how there could be so much DLC present without a stoichiometric amount of DIC. To confirm that the bands observed correspond to the dynein subunits, I needed to excise the bands and submit them for MS analysis. However, as these gels showed that the sample had several non-specific proteins co-purifying with the dynein, I decided to first purify the sample further.

In order to achieve the level of purity necessary for MS analysis and to test
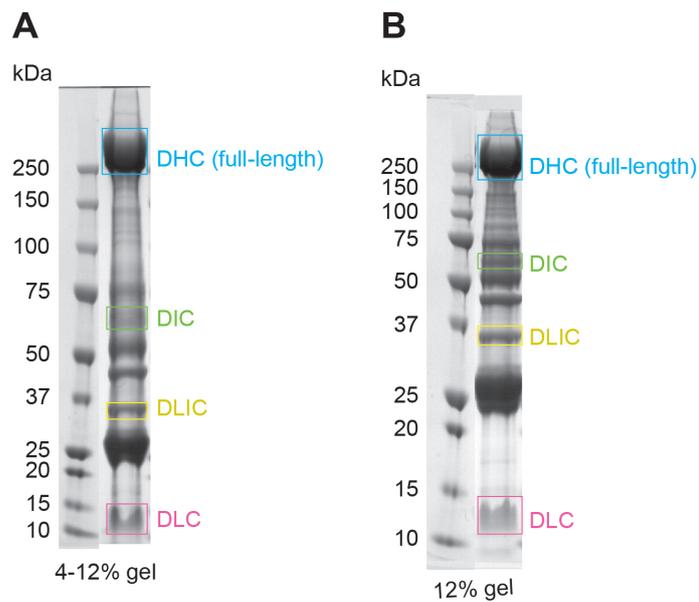
Figure 2.5: **Gel showing a large-scale purification of the full-length dynein complex from 36L of CY286.** 36L of CY286 were grown and a large-scale purification was performed. After TEV cleavage and concentration, there was $\sim 300\mu$L of sample, with a protein content of ~ 5mg/mL. **A)** SDS-PAGE analysis of 20μL of the sample using a $4-12\%$ gradient gel. **B)** SDS-PAGE analysis of 20μL of the sample using a 12% gel. In both **A** and **B**, boxes surround bands that appeared at the predicted heights of the DHC (cyan), DIC (green), DLIC (yellow) and DLC (magenta).
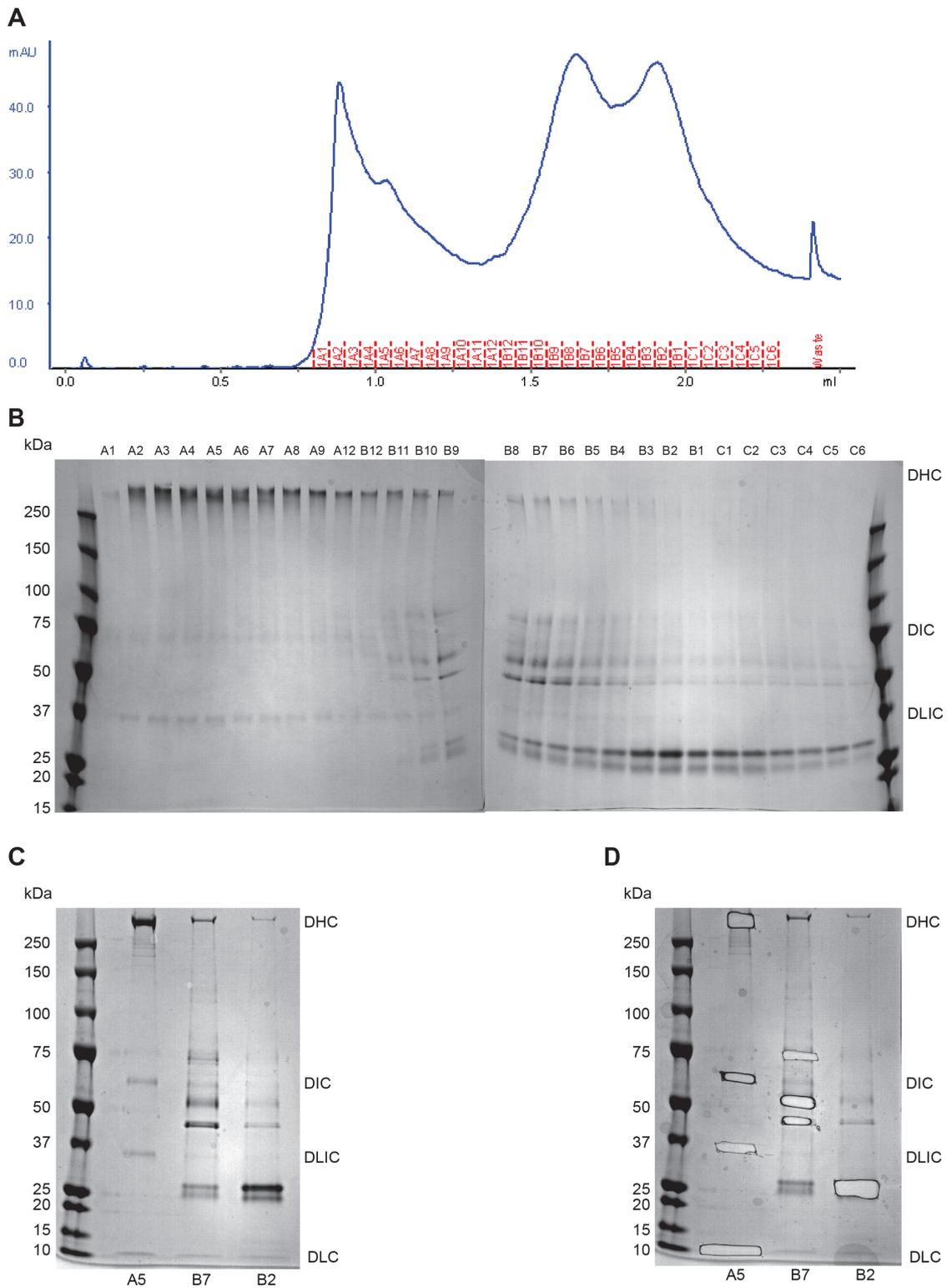
Figure 2.6: **The purification of full-length dynein complex from CY286 and identification of subunits by mass spectrometry. A)** Full-length dynein was purified from 36L of CY286. 50μL of the 5mg/mL sample was loaded on a Superose6 column and 50μL fractions were collected. **B)** 20μL of each fraction (except for A10 and A11) were analysed by SDS-PAGE. **C)** The remainder of fractions A5, B7 and B2 were concentrated and analysed by SDS-PAGE. **D)** Relevant bands from **C** were excised and analysed by MS, confirming the identity of the four dynein subunits.

whether the dynein complex remains intact after purification, I performed size exclusion chromatography (SEC) on the sample. Using an ÄKTAmicro system, I loaded 50μL of the 5mg/mL sample on a Superose6 column (GE Healthcare), collected 50μL fractions and then analysed 20μL of each fraction using SDS-PAGE (Fig. 2.6). The chromatogram shows that in addition to a large aggregate peak eluting at the void volume of the column (~ 0.85mL), three other peaks are present, centred in fractions A5, B7 and B2 (Fig. 2.6A). The A5 peak eluted at ~ 1.00mL and overlaps significantly with the aggregate peak. The B7 peak eluted at ~ 1.65mL and the B2 peak eluted at ~ 1.90mL.

I analysed all of the fractions collected except for A10 and A11 using SDS-PAGE (Fig. 2.6B). It is clear from the gels that the A5 peak contains the full-length dynein complex, whereas the B7 and B2 peaks consist of various contaminants. However, the Superose6 column was not particularly well-suited for purification of this ~ 1.2MDa complex, as the dynein peak was very close to the void volume. This means that the fractions containing the full-length dynein complex are likely a mixture of well-behaved, monodisperse protein and aggregate. Nevertheless, the column successfully separated the dynein complex from the lower molecular weight contaminants, which allowed me to identify the components of the complex as well as the contaminants by MS.

In order to ensure a clean sample, I ran the remainder of the protein from fractions A5, B7 and B2 on a fresh gel (Fig. 2.6C). The four bands that correspond to the MWs of the dynein subunits were all clearly visible in the A5 fraction. Therefore, I excised these bands and submitted them for MS analysis (Fig. 2.6D). I also excised the most prominent bands in the B7 and B2 fractions and identified them by MS in order to determine whether they are non-specific contaminants or dynein-related proteins that co-purified due to a specific interaction. The four bands in the A5 fraction were positively identified as Dyn1 (DHC), Pac11 (DIC), Dyn3 (DLIC) and Dyn2 (DLC). The bands in B7 and B2 were all identified as other yeast proteins with no apparent connection to dynein.

Finally, I calculated the yield of full-length dynein from the 36L culture of CY286. As noted above, the total amount of protein after the TEV cleavage step was $\sim 300\mu$L at ~ 5mg/mL. However, integration of the peaks in Fig. 2.6 showed that only about 1/3 of the total protein is dynein. Therefore, the final yeild of full-length dynein from 36L of culture can be estimated as ~ 0.5mg.

2.2.3   First series of DHC truncations

With a means for expressing and purifying full-length dynein complex, I turned my attention back toward identifying a DHC truncation that would yield a minimal tail complex. In the CY286 background, I made a series of C-terminal DHC truncations by inserting a stop codon into various positions within the *DYN1* gene (Fig. 2.7). In order to accomplish this, I first had to revert CY286 into a uracil auxotroph so that I could apply a selection pressure to favour clones that successfully recombined. I removed the functional *URA3* gene that we had previously inserted and selected for clones lacking this gene using plates containing 5-Fluoroorotic acid (5-FOA). The Ura3 enzyme metabolises 5-FOA into 5-fluorouracil, which is toxic, therefore growing yeast in the presence of 5-FOA is a way to select for clones that lack the *URA3* gene. Once CY286 was auxotrophic for uracil, I transformed the functional
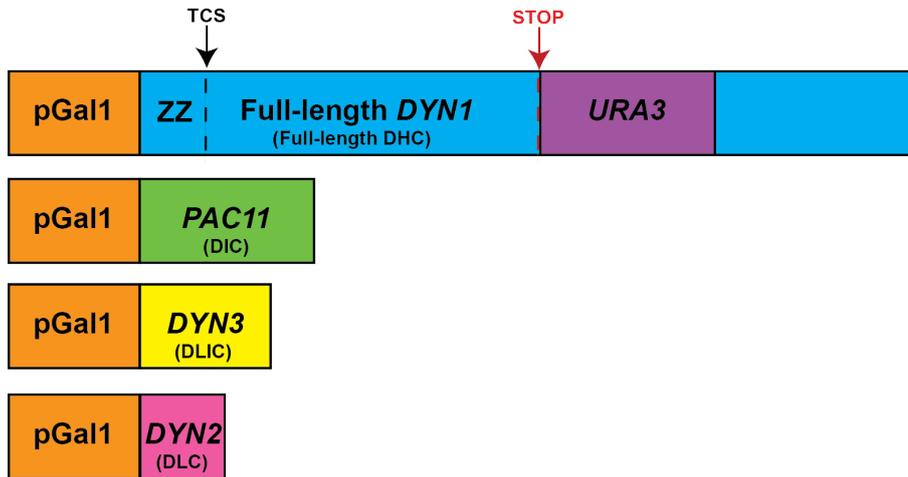
Figure 2.7: **Schematic diagram of the introduction of DHC truncations into CY286.** Using homologous recombination, CY286 was reverted back into a uracil auxotroph by removing the functional *URA3* gene that had been inserted when it was first created. Then, a functional *URA3* gene with a stop codon 5' to its promoter was targeted to various points throughout the *DYN1* (DHC) gene. In effect, this produced C-terminal truncations of the DHC.

URA3 gene back into the strain, targeting it to various points throughout the *DYN1* (DHC) gene. By including a stop codon at the 5' end of the *URA3* promoter, I was able to effectively truncate the DHC C-terminally at specific positions.

As the literature did not have any reliable descriptions of the DIC and DLIC binding sites on the DHC, I designed 27 different truncations, spanning a very large region within the DHC (Table 2.1). The shortest truncation (DHC-A) terminated after residue 279, which was designed to be too short to bind any other dynein subunits. The longest truncation (DHC-Ω) terminated after residue 1477, which is well into the linker domain as shown by the previous motor domain crystal structure (A. P. Carter *et al.* 2011). I also specifically included a truncation that terminated at residue 1363 (DHC-Y), which is equivalent to the truncation I used for the polyprotein in Section 2.2.1. The remaining truncations were roughly evenly distributed over the remaining sequence between residues 279 and 1477.

While designing the truncations, I referred to a document I created in order to summarise several predictions about this region of the yeast DHC (Figs. A.1 and A.2) including secondary structure predictions (McGuffin *et al.* 2000), coiled-coil predictions (Lupas *et al.* 1991) and domain boundary predictions (Finn *et al.* 2013). Starting just after the beginning of the predicted DHC N1 domain, I introduced a pair of truncations every few α-helices. The first truncation in each pair was near the middle of the predicted α-helix and the second was in the predicted loop following the α-helix. This was done to reduce the likelihood that a single bad truncation due to local secondary structure would prevent me from finding a more global domain boundary.

The yeast strain for expressing truncation DHC-C was not completed due to cloning difficulties, but the other 26 strains were created and verified. I then grew 1L of each of the 26 DHC truncations and performed a small-scale purification for each pellet. I also grew 1L of CY286 and purified it alongside the truncations as a

Name	Length (aa)	MW (kDa)	Results
DHC-A	279	31.6	X
DHC-B	288	32.7	
DHC-C	466	53.1	X
DHC-D	477	54.5	
DHC-E	557	63.9	*
DHC-F	576	66.0	
DHC-G	666	76.3	
DHC-H	743	85.4	
DHC-I	760	87.5	
DHC-J	791	90.9	
DHC-K	856	98.2	
DHC-L	870	99.9	*
DHC-M	954	109.4	*
DHC-N	962	110.5	X
DHC-O	1048	120.7	**
DHC-P	1065	122.7	
DHC-Q	1148	132.4	
DHC-R	1175	135.4	
DHC-S	1239	142.7	
DHC-T	1254	144.6	
DHC-U	1300	150.0	
DHC-V	1310	151.0	
DHC-W	1340	154.8	**
DHC-X	1346	155.4	
DHC-Y	1363	157.4	**
DHC-Z	1459	169.0	X
DHC-Ω	1477	171.1	**

Table 2.1: **List of the first set of 27 DHC truncations.** The length in amino acids and predicted MW after TEV cleavage is shown for each construct. In the results column, an X indicates a a failed experiment either due to unsuccessful cloning (DHC-C) or contamination of the truncated construct with full-length DHC (DHC-A, DHC-N and DHC-Z). * indicates truncations that expressed well enough to produce a visible DHC band on the gels in Fig. 2.8. ** indicates truncations that successfully pulled down the DIC, DLIC and DLC.
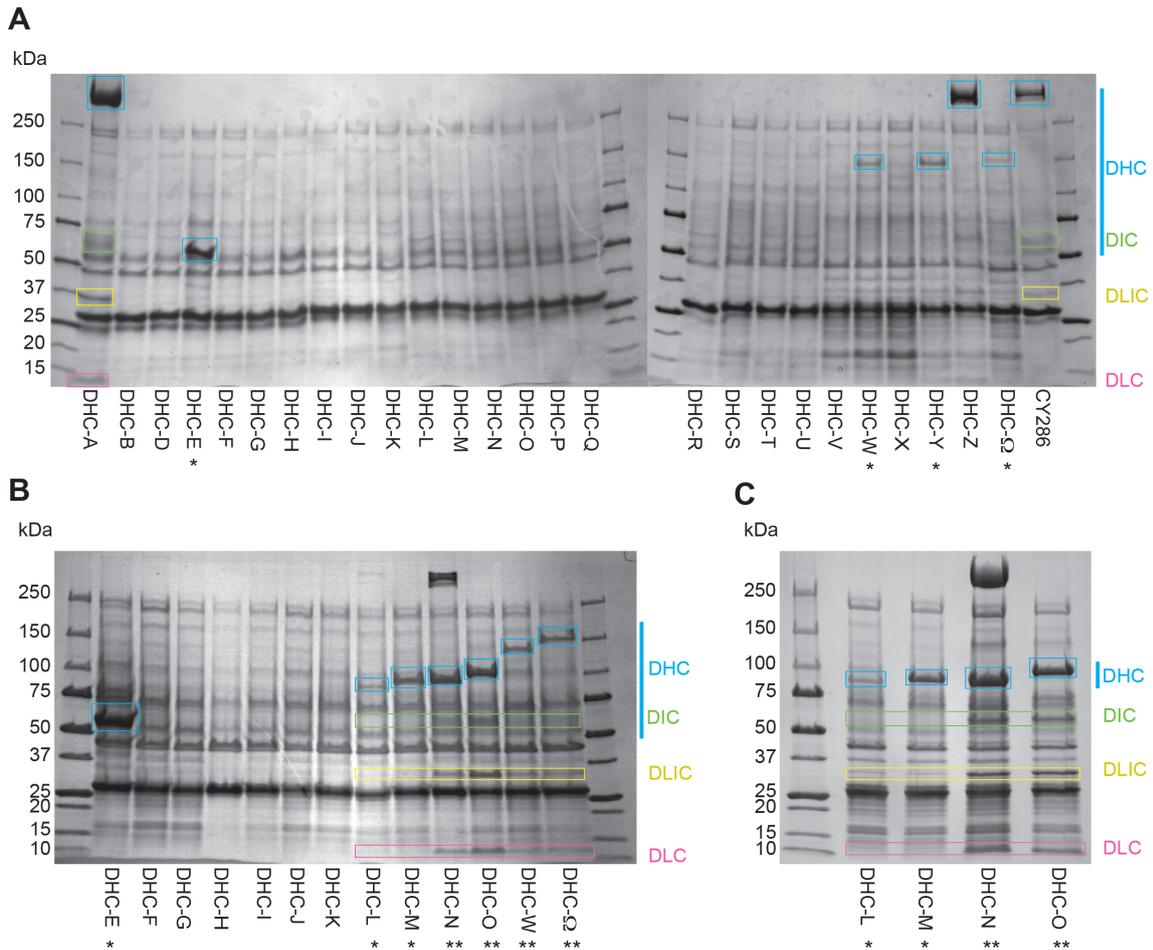
Figure 2.8: **Test purifications of the first series of DHC truncations. A)** 1L of CY286 and each of the 26 successfully cloned DHC truncations was grown. Small-scale purifications were performed and the samples were analysed by SDS-PAGE. DHC-A and DHC-Z were contaminated with full-length DHC, but served as full-length controls along with the CY286 lane. DHC-E expressed at high levels, but did not bind any other subunits. DHC-W, DHC-Y and DHC-Ω all produced faint bands for the truncated DHC, but no other subunits were clearly present. Boxes surround bands that appeared at the predicted heights of the DHC (cyan), DIC (green), DLIC (yellow) and DLC (magenta). **B)** The experiment shown in **A** was repeated, but with fewer constructs. DHC-L, DHC-M, DHC-N and DHC-O all produced bands for the truncated DHC. Only DHC-N and DHC-O produced bands for the accessory subunits. **C)** The experiments done in **A** and **B** were repeated, but with only four constructs and 3L of yeast culture for each. DHC-N was contaminated with full-length DHC. All three accessory subunits co-purified with DHC-O, while none of them appeared to co-purify with DHC-L or DHC-M.

full-length control. After TEV cleavage was complete I concentrated each sample down to $\sim 20\mu$L and loaded the entire sample on a gel for SDS-PAGE analysis (Fig. 2.8A).

Two of the strains (truncations DHC-A and DHC-Z) appear to have been contaminated with full-length DHC, since their band pattern looks similar to that of CY286. The majority of the other constructs did not produce visible bands for any of the dynein subunits, indicating a failure either to express or to purify. However, a few constructs did show bands for dynein subunits. DHC-E (557aa) was the shortest truncation to produce a visible band, and it was almost as prominent as the CY286 DHC band. As the full length DHC is >7 times the length of DHC-E, this indicates that the yield of DHC-E was roughly sevenfold the yield of CY286. However, DHC-E did not appear to bind any of the other subunits.

The only other truncations to show DHC bands for this initial experiment were DHC-W, DHC-Y and DHC-Ω. Each of these three constructs produced weak DHC bands, but it was difficult to determine whether the other three subunits were present as the protein levels were low and this high-throughput method of purification is not stringent enough to remove all background on the gel. In order to reproduce the original results and improve the signal on the gel, I repeated the experiment using only half as many constructs. For this second attempt, I included DHC-E, DHC-W and DHC-Ω, which all produced visible DHC bands previously. I also included DHC-F to DHC-O as these ten constructs were purified in parallel the first time and I suspected that the fact none of the ten showed a single dynein band may have been due to experimental error.

This second batch of constructs reproduced the results described above for DHC-E, DHC-W and DHC-Ω. More importantly, it revealed four new DHC truncations that expressed, DHC-L, DHC-M, DHC-N and DHC-O (Fig. 2.8). DHC-L (870aa) had a very weak band for the truncated DHC and no bands for the other subunits. DHC-M (954aa) showed a more prominent DHC band, but still no accessory subunits were visible. DHC-N (962aa) displayed all four dynein subunits prominently, however this strain also produced a large amount of full-length DHC, so it is possible that the accessory subunits were pulled down by this contaminant rather than the truncated DHC itself. DHC-O (1048aa) was not contaminated by full-length DHC and still pulled down all of the accessory subunits.

To reproduce these new results, I grew another 3L each of DHC-L, DHC-M, DHC-N and DHC-O and purified them again. These purifications showed the same results as above and the gel is more clear (Fig. 2.8C). Therefore, I was able to conclude that somewhere between residues 954 and 1048 there is a position where I can truncate the DHC as much as possible while retaining the ability to bind all accessory subunits. This narrowed down the region where an ideal DHC truncation for a minimal tail complex might exist from a range of 1198aa to one that is only 94aa.

2.2.4 Second series of DHC truncations

The 94aa region between DHC-M and DHC-O (residues 954 and 1048) is still fairly large and includes four predicted α-helices and one β-strand (Fig. 2.9). Therefore, I performed a finer screen of 12 new C-terminal truncations within this region (hereafter referred to as the "N region"). Using the CY286 background again and the

```
                                                                                         M    NA
             890       900       910       920       930       940       950       960
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
881   NEGRITSSPSLAAMKRSLLKDIEALLNKVVLINFLHDPDHPLSTTLTFNSLVIKLKDDIQNCIEQVQNLHCKINSYVKEW
626   E--EEEEE--HHHHHHHHHHHHHHHHHH--------------------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
                 H25                                          H26

      N/NB     NC          ND  NE        NF    NG        NH        NI          NJ        NK
             970       980       990       1000      1010      1020      1030      1040
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
961   QKMEFLWQITEEAFLEVVDNSTQRCFGILKGLLDSQSKFDLIISRNNFSKNLVLHTEDAQRHIRSKMDSWILYVSKHLLT
682   H-HHHHH---HHHHHHHH----HHHHHHHHHHHHHHHHH-----E---EEEEE-HHHHHHHHHHHHHHHHHHHHHHHHHHH
                 H27             H28                               H29

              Q              NL  P
             1050      1060      1070      1080      1090      1100      1110      1120
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1041  IYERDARKLHEDMNRDREAVEDMDINFTSLKNITVIIEAVNVNKRHLTERDIQIKLLGSVMRALTKLKVRFPSHFVYIDQ
745   HHHHHHHHHHHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH------HHHHHHH
                          H30                                              H31
```

Figure 2.9: **Secondary structure prediction of N region of the DHC with the second series of truncations highlighted.** E=predicted β-strand, H=predicted α-helix, turquoise=predicted coiled-coil, red=final residues from the first series of DHC truncations in this region: DHC-M, DHC-N, DHC-O and DHC-P, green=final residues from twelve new N region DHC truncations. Note that the new DHC-NB is equivalent to DHC-N from the first series.

Figure 2.10: **Schematic diagram of the fusion of GFP onto the C-terminus of DHC truncations.** Into a strain of CY286 that was auxotrophic for uracil, I transformed a DNA fragment that contained a GFP followed by a stop codon and a functional *URA3* gene. This fragment was targeted to various points throughout the *DYN1* (DHC) gene, producing truncations of the DHC with a GFP fused to the artificial C-terminus.

Name	Length (aa)	MW (kDa)
DHC-NA	958	109.9
DHC-NB	962	110.5
DHC-NC	968	111.3
DHC-ND	979	112.5
DHC-NE	981	112.7
DHC-NF	990	113.8
DHC-NG	995	114.2
DHC-NH	1004	115.3
DHC-NI	1016	116.7
DHC-NJ	1025	117.8
DHC-NK	1035	119.0
DHC-NL	1063	122.5
DHC-NA-GFP	958	136.9
DHC-NB-GFP	962	137.5
DHC-NC-GFP	968	138.3
DHC-ND-GFP	979	139.5
DHC-NE-GFP	981	139.7
DHC-NF-GFP	990	140.8
DHC-NG-GFP	995	141.2
DHC-NH-GFP	1004	142.3
DHC-NI-GFP	1016	143.7
DHC-NJ-GFP	1025	144.8
DHC-NK-GFP	1035	146.0
DHC-NL-GFP	1063	149.5

Table 2.2: **List of the second set of 24 DHC constructs.** The length of the DHC truncation in amino acids and predicted MW of the entire DHC (including GFP where applicable) after TEV cleavage is shown for each construct.

same strategy used for the first series, I inserted a stop codon into various positions within the *DYN1* gene (Table 2.2). The shortest (DHC-NA) terminated after residue 958, which lies halfway between the original DHC-M and DHC-N. DHC-N (terminating after residue 962) was reproduced in order to eliminate the full-length DHC contamination in the first series and was renamed DHC-NB. The longest truncation terminated after residue 1063 (DHC-NL) in order to include the entire predicted α-helix that runs from residue 1017 to residue 1061 (Fig. 2.9).

I also designed an equivalent construct for each of these truncations that fused a green fluorescent protein (GFP) C-terminal to the site of truncation. This was to ensure that if the DHC in a particular truncation was the right length for a minimal tail complex, any disorder introduced by the truncation itself would be followed by an ordered protein to help stabilise the protein and prevent it from being targeted for degradation. In order to accomplish this, after the intended final residue in each truncation I inserted a GFP gene in frame with *DYN1*, followed by a stop codon (Fig. 2.10). These truncations are named DHC-NA-GFP, DHC-NB-GFP, etc.

The yeast strains for expressing DHC-NG and DHC-NF-GFP were not completed due to cloning difficulties, but the other 22 strains were created and verified. I then
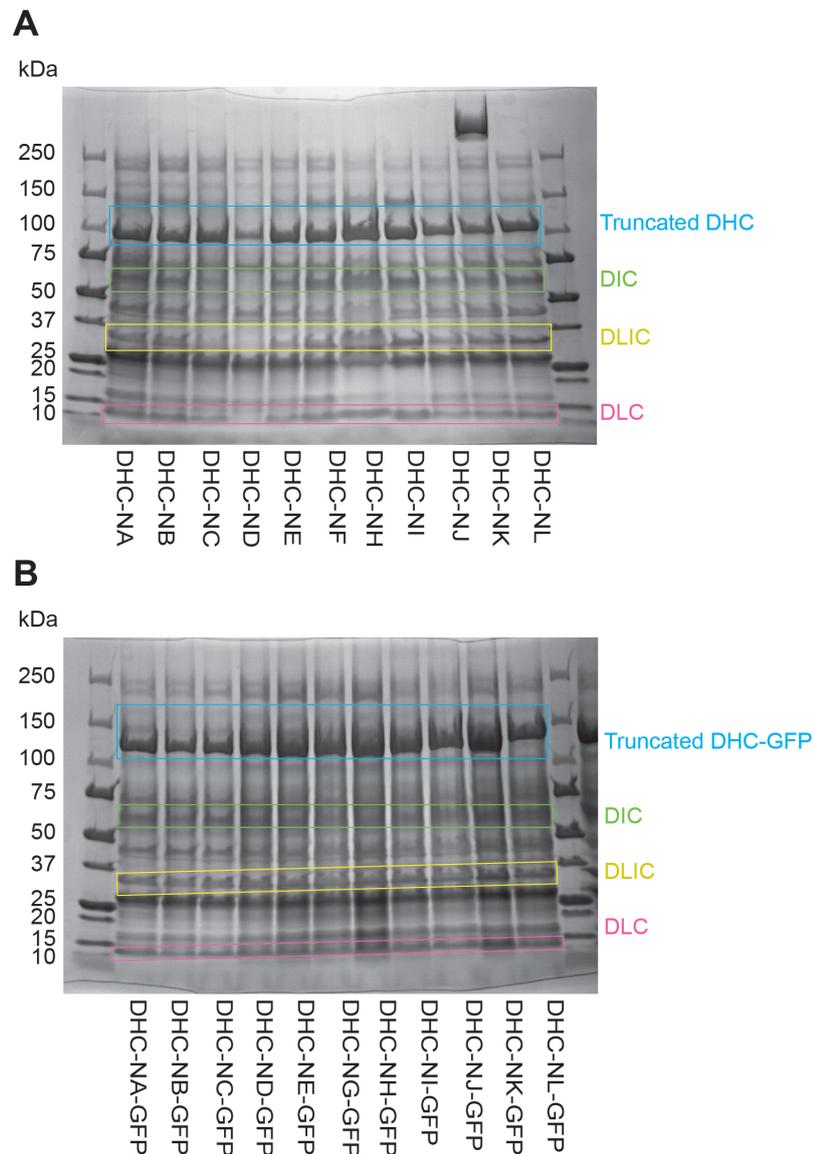
Figure 2.11: **Test purifications of the second series of DHC truncations.**
A) 1L of each N region DHC truncation was grown. Small-scale purifications were
performed and the samples were analysed by SDS-PAGE. DHC-NK was contami-
nated with full-length DHC. All of the N region truncations produced bands for the
truncated DHCs. Some had higher expression levels (*e.g.*, DHC-NH) while others
had lower expression levels (*e.g.*, DHC-ND). The expression levels of the DHC trun-
cation tended to correlate with the amount of accessory subunits that were pulled
down with it. Boxes surround bands that appeared at the predicted heights of the
DHC (cyan), DIC (green), DLIC (yellow) and DLC (magenta). **B)** As in **A**, but
with the set of N region truncations that included a C-terminal GFP.

grew 1L of each N region construct and performed a small-scale purification for each pellet. After TEV cleavage was complete I concentrated each sample down to $\sim 20\mu$L and loaded the entire sample on a gel for SDS-PAGE analysis. Contrary to the first round of heavy chain truncations, every one of the 22 new constructs tested expressed the DHC truncation well enough to see a clear band on the gel (Fig. 2.11). DHC-NK was contaminated by full-length dynein and was excluded from further analysis.

My next goal was to test some of the N region constructs using SEC to eliminate nonspecific proteins, to test the stability each construct and to determine whether the sample was homogeneous. As 1L of culture did not produce enough protein for SEC, I began with only 6 constructs, and scaled up the expression of each to 4L. In order to test the effect of the C-terminal GFP, I used constructs that expressed well both with and without the GFP. To begin with the three most different N-region constructs I selected the shortest (DHC-NA), the longest (DHC-NL) and one in the middle (DHC-NH). For the one in the middle, I chose DHC-NH because it had the highest expression levels for all subunits as well as lower levels of the background contaminants (Fig. 2.11A).

I grew 4L of each construct and the proteins were purified as before, except after TEV cleavage the samples were concentrated to $\sim 50\mu$L and loaded on a Superose6 column using an ÄKTAmicro system (both GE Healthcare). For each construct, I looked for a peak that eluted between 1.2mL and 1.5mL since this is where a complex the size of the dynein tail is expected to elute (for clarity, in this section I refer to any peaks in this range as "dynein peaks"). I then selected the fractions that span the dynein peak, pooled them and concentrated them to $\sim 50\mu$L. I did the same for any other peaks of interest. These concentrated samples were then analysed using SDS-PAGE and compared to the original samples that were loaded for each SEC run (labelled as "Input" in Figs. 2.12, 2.13 and 2.14).

DHC-NA and DHC-NA-GFP

Both DHC-NA and DHC-NA-GFP failed to produce clear dynein peaks. The chromatogram for DHC-NA was almost completely flat in the 1.2mL to 1.5mL range (Fig. 2.12A), whereas DHC-NA-GFP had an increase in UV absorbance, but not in the form of a prominent, symmetrical peak (Fig. 2.12B). Both constructs had a very small peak at the void volume and a small, yet distinct peak at ~ 1.7mL. To isolate the indistinct dynein peak in the DHC-NA-GFP sample, I pooled fractions A10-B2. I also pooled fractions B4-B8 in order to isolate the the peak that eluted at ~ 1.7mL. Despite DHC-NA's lack of a distinct dynein peak, I pooled the same fractions from this sample for a direct comparison with the GFP-tagged version. SDS-PAGE analysis revealed that both DHC-NA and DHC-NA-GFP retained the other three dynein subunits at least through the TEV cleavage and concentration steps prior to SEC (Input, Fig. 2.12). Because the input contained all four dynein subunits, but fractions A10-B2 contained only DHC, both constructs appear incapable of maintaining an intact complex when subjected to the low concentrations experienced during gel filtration. The A10-B2 lane in both gels shows a distinct band for the truncated DHC, but the DIC, DLIC and DLC bands are not present. The B4-B8 lane in both gels shows that SEC is successful in separating the most abundant contaminants from the DHC.
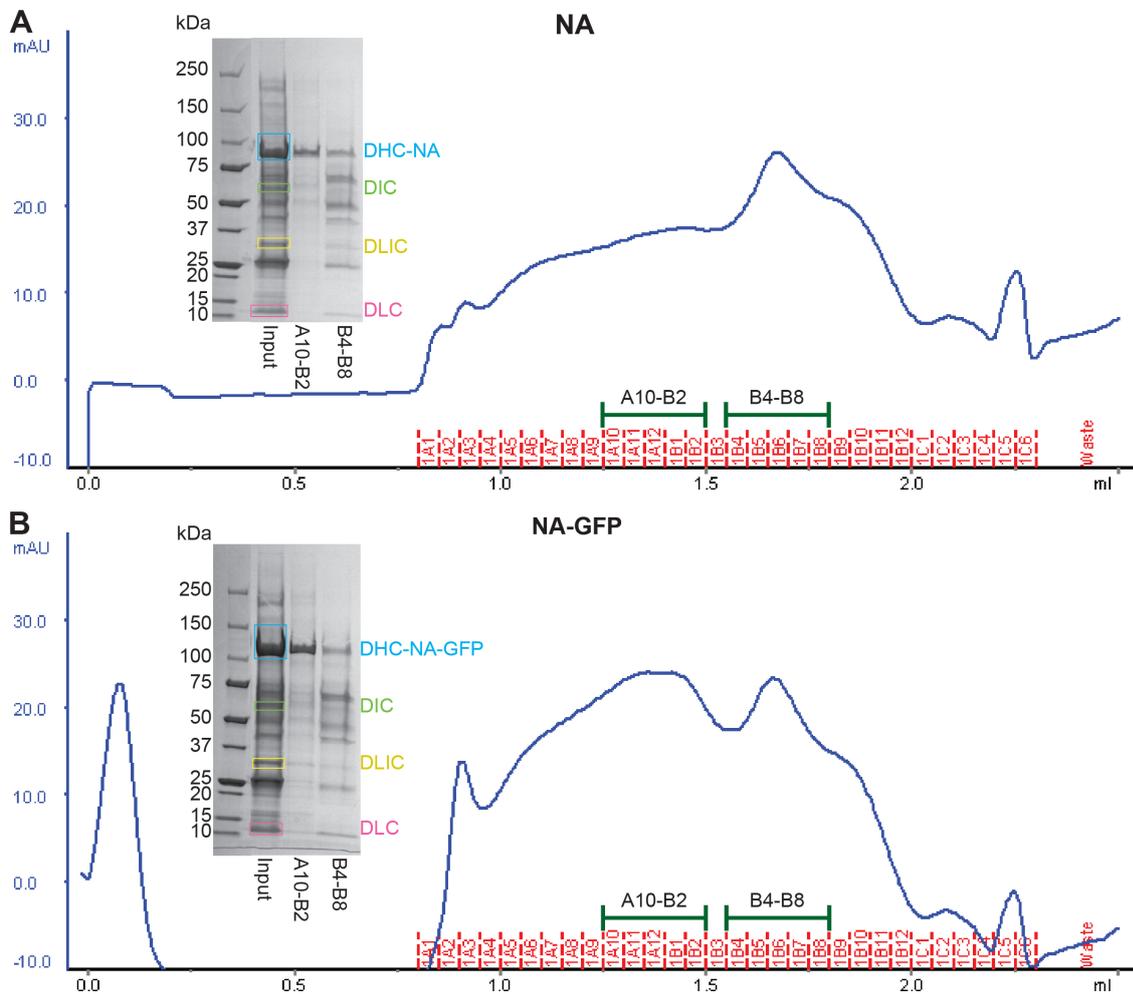
Figure 2.12: **Purification of DHC-NA and DHC-NA-GFP by SEC.** DHC-NA (**A**) and DHC-NA-GFP (**B**) were each purified from 4L of yeast culture. Both samples were concentrated to $\sim 50\mu$L. 10μL were reserved as the "input" sample and the remainder was loaded on a Superose6 column using an ÄKTAmicro system (chromatograms). Fractions of interest were pooled (green bars) and concentrated to $\sim 50\mu$L. Then 10μL of each sample was analysed by SDS-PAGE (gel insets). Dynein subunits are highlighted in colour-coded boxes.
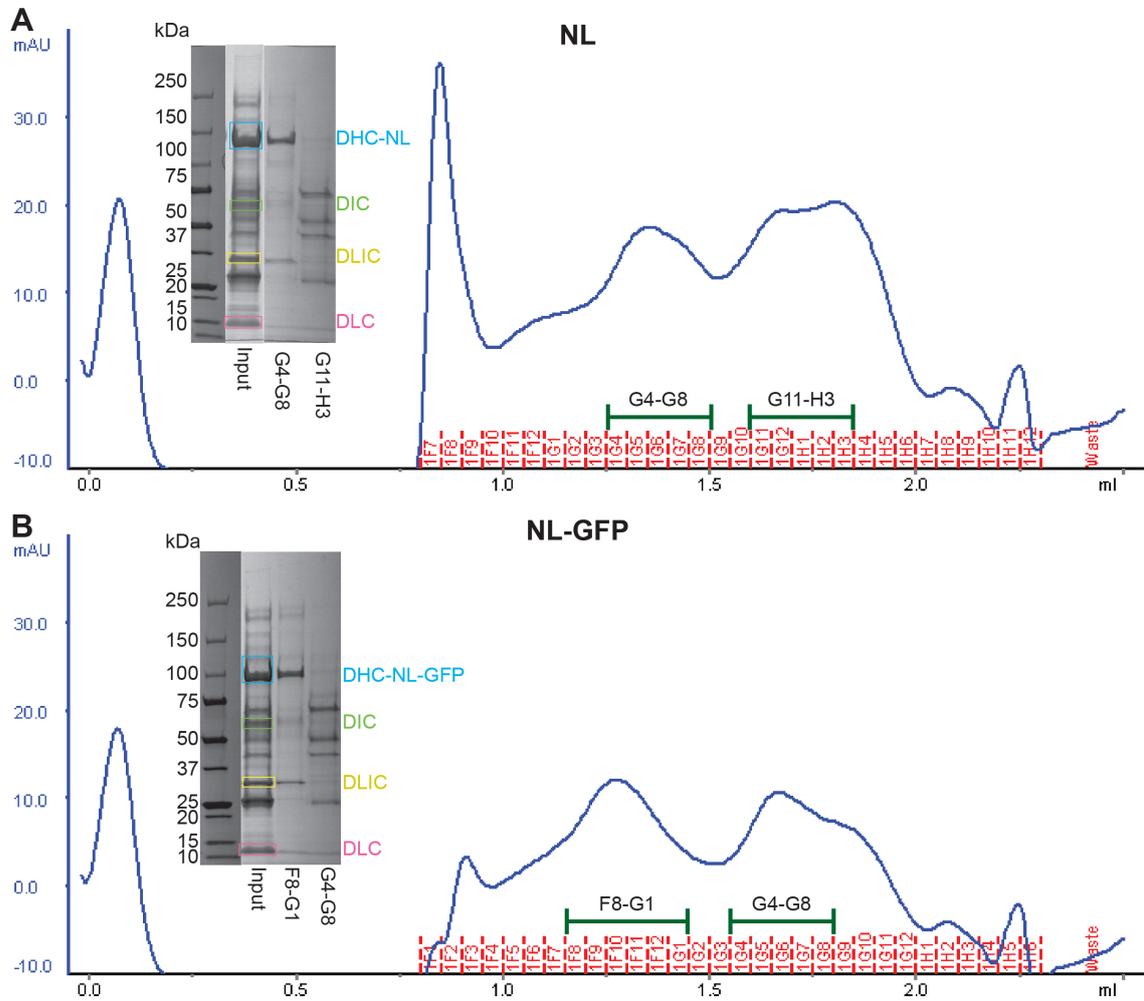
Figure 2.13: **Purification of DHC-NL and DHC-NL-GFP by SEC.** DHC-NL (**A**) and DHC-NL-GFP (**B**) were each purified from 4L of yeast culture. Both samples were concentrated to $\sim 50\mu$L. 10μL were reserved as the "input" sample and the remainder was loaded on a Superose6 column using an ÄKTAmicro system (chromatograms). Fractions of interest were pooled (green bars) and concentrated to $\sim 50\mu$L. Then 10μL of each sample was analysed by SDS-PAGE (gel insets). Dynein subunits are highlighted in colour-coded boxes.

DHC-NL and DHC-NL-GFP

The chromatograms for DHC-NL and DHC-NL-GFP are similar overall with the primary difference being that DHC-NL has a sharp, prominent peak at the void volume whereas DHC-NL-GFP has a very small one (Fig. 2.13). This may indicate that in the context of the DHC-NL truncation, the C-terminal GFP is effective in preventing aggregation. Both constructs produced a small, yet fairly symmetrical dynein peak and an equally small double-peak in the 1.6mL to 2.0mL range. The dynein peak was isolated by pooling fractions G4-G8 from the DHC-NL run and fractions F8-G1 from the DHC-NL-GFP run. The double-peak was isolated by pooling fractions G11-H3 from the DHC-NL run and fractions G4-G8 from the DHC-NL-GFP run. SDS-PAGE analysis showed that just like DHC-NA and DHC-NA-GFP, both DHC-NL and DHC-NL-GFP contained all four dynein subunits through the TEV cleavage and concentration steps (Input, 2.13). In contrast to DHC-NA and DHC-NA-GFP, the DHC and DLIC are both present in the dynein peak, indicating the formation of a subcomplex consisting of these two subunits. However, the DIC and DLC bands are still largely absent from these samples, which may indicate that the DHC-NL truncation introduces instability into the complex. The double peaks from both DHC-NL and DHC-NL-GFP contain similar contaminants as the 1.7mL peak in DHC-NA and DHC-NA-GFP and similarly demonstrates that it is possible to separate these from the DHC/DLIC subcomplex using SEC.

DHC-NH and DHC-NH-GFP

The chromatograms for DHC-NH and DHC-NH-GFP both contain a sharp, prominent dynein peak and a smaller double-peak that elutes at ~ 1.7mL (Fig. 2.14). In a reversal from the trend seen with DHC-NL, DHC-NH-GFP had a large aggregate peak elute at the void volume, whereas the bare DHC-NH had a very small aggregate peak. The dynein peaks were isolated by pooling fractions D9-E3 and D3-E8 from the DHC-NH and DHC-NH-GFP runs, respectively. The double peaks were isolated by pooling fractions E5-E8 and D11-E1 from the DHC-NH and DHC-NH-GFP runs, respectively. SDS-PAGE analysis showed that DHC-NH and DHC-NH-GFP also contained all four dynein subunits after the TEV cleavage and concentration steps (Input, 2.14). In contrast to the other two truncations tested, the dynein peaks from both DHC-NH and DHC-NH-GFP retained all four dynein subunits, indicating the successful purification of a complete dynein tail complex. The double peaks from both DHC-NH and DHC-NH-GFP contain the same contaminants as observed in DHC-NA and DHC-NL, demonstrating that SEC is a useful technique for separating these proteins from the minimal tail complex.

2.2.5 Large-scale production of DHC-NH from CY339 using a 120L fermentor

As DHC-NH expressed, co-purified with the DIC, DLIC and DLC, and was fairly pure even after a simple SEC run, it seemed a reasonable target for a first attempt at solving a structure of the dynein tail complex. However, if I was going to use X-ray crystallography, the low expression level needed to be addressed. I made several attempts to improve the expression of DHC-NH from the same yeast strain (hereafter referred to as CY339, Fig. A.4) by only modifying growth conditions,
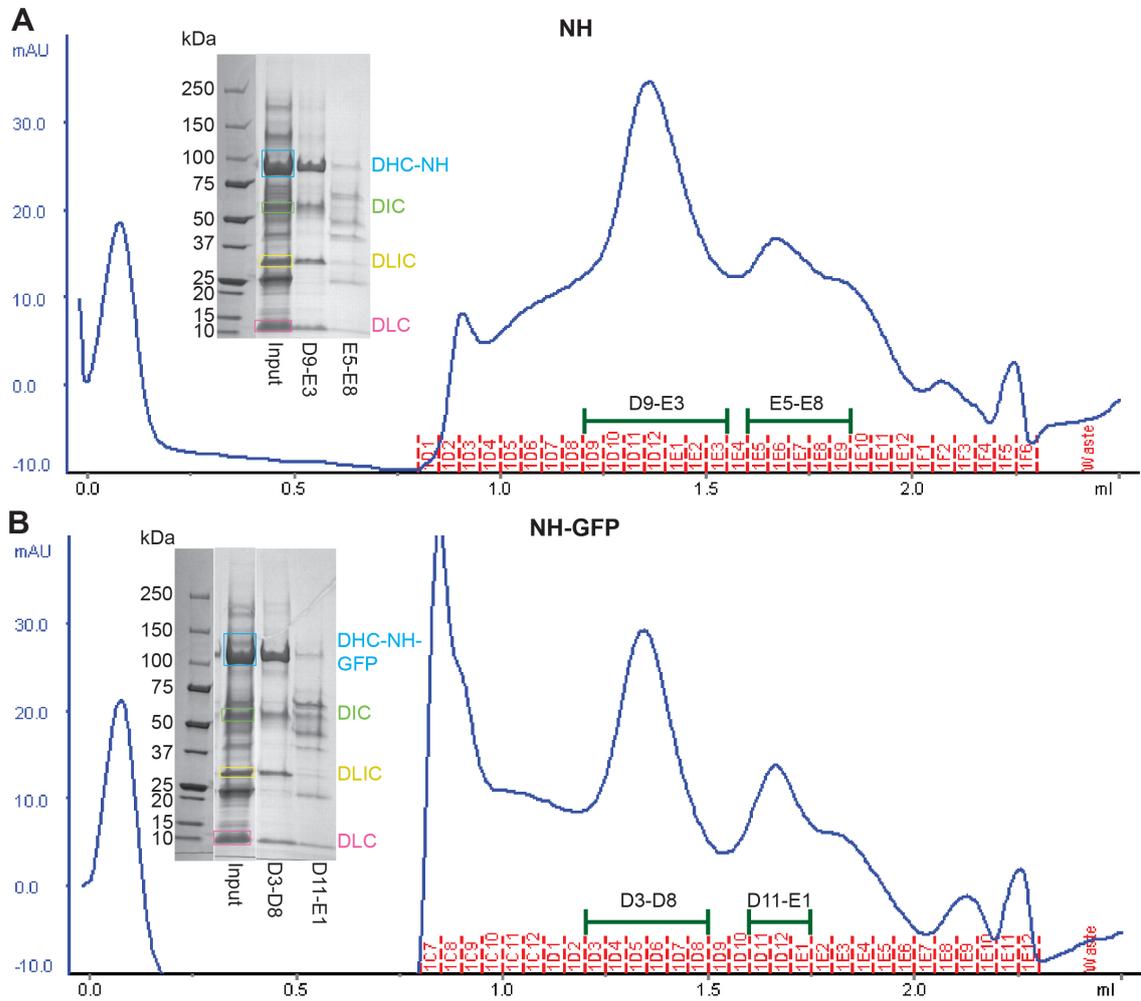
Figure 2.14: **Purification of DHC-NH and DHC-NH-GFP by SEC.** DHC-NH (**A**) and DHC-NH-GFP (**B**) were each purified from 4L of yeast culture. Both samples were concentrated to $\sim 50\mu$L. 10μL were reserved as the "input" sample and the remainder was loaded on a Superose6 column using an ÄKTAmicro system (chromatograms). Fractions of interest were pooled (green bars) and concentrated to $\sim 50\mu$L. Then 10μL of each sample was analysed by SDS-PAGE (gel insets). Dynein subunits are highlighted in colour-coded boxes.

such as media composition, growth temperature, density at the point of induction and length of induction before harvesting. None of these experiments produced any significant increase in the amount of protein produced per gram of CY339 harvested. As a short-term solution, I decided to simply grow more yeast, which I hoped would provide me with enough protein to decide whether this construct was really a good crystallography target before investing time in more involved strategies for boosting expression.

Logistically, growing more than 36L of CY339 in shaker flasks was not feasible, so I used a large fermentor to grow 120L at a time, which simplified the growth and harvest while quadrupling the amount of yeast per batch. A typical yield from a 36L shaker flask growth was $\sim$ 1kg wet mass of yeast, while the typical yield from a 120L fermentor growth was $\sim$ 4.5kg wet mass.

I performed a large-scale purification of DHC-NH from $\sim$ 1kg of CY339 grown in the fermentor. The increase in total protein after the TEV cleavage step allowed me to introduce an anion exchange chromatography step using a MonoQ column (GE Healthcare) before the final SEC step. This was not only beneficial for producing a more pure sample, but also because it allowed me to avoid concentrating the sample while several contaminants were still present. This second benefit arose because I could load a large volume of dilute sample onto the MonoQ column, rather than immediately concentrating it to a volume suitable for SEC. In this way, the dynein tail was only concentrated once it was fairly clean.

The MonoQ run consisted of a shallow gradient from 20mM KCl to 500mM KCl over 60 column volumes and produced a prominent peak at a conductance of $\sim$ 35mS/cm (Fig. 2.15A,B). The peak began sharply, but had a shoulder that eluted slightly after and then it tailed off gradually over several column volumes. I analysed all the fractions that eluted around this peak using SDS-PAGE in order to determine whether it contained the dynein tail complex, how clean it was, and whether it remained intact (Fig. 2.15C). The prominent peak contained the dynein tail complex, and aside from two bands in the 100kDa to 150kDa range and one band around 75kDa, the sample was already very clean. While the majority of the complex eluted in the sharp peak, there was still a considerable amount of the complex eluting as the peak tailed off. The peak's shoulder was created by the presence of a $\sim$ 250kDa contaminant, preventing the inclusion of these later fractions in the final sample. The complex also appeared robust at 170mM KCl, the concentration required for elution, since all four subunits eluted together.

I pooled fractions E2-E7 from the MonoQ run, concentrated the sample down to $\sim$ 500μL and then loaded it on a Superose6 column for SEC (Fig. 2.16). The chromatogram revealed a single, prominent, symmetrical peak, which eluted at $\sim$ 12mL (Fig. 2.16A). SDS-PAGE analysis confirmed that the peak consisted of a very pure, complete dynein tail complex (Fig. 2.16B,C). I pooled fractions B11-B8 and concentrated the sample to $\sim$ 75μL. This sample had a final concentration of $\sim$ 5mg/mL, meaning the final protein yield from $\sim$ 1kg of yeast was $\sim$ 0.375mg.

I used 10μL of the final sample for SDS-PAGE, in order to confirm the purity at a higher concentration (Fig. 2.16C). With the remaining 65μL, I set up a simple PEG3350/pH crystal screen by hand, in order to get a first impression of how this complex behaves in droplets. The majority of these droplets developed a heavy brown precipitate, but those with a higher pH and/or lower PEG3350 concentration remained clear. This indicated that this complex may be a suitable target for
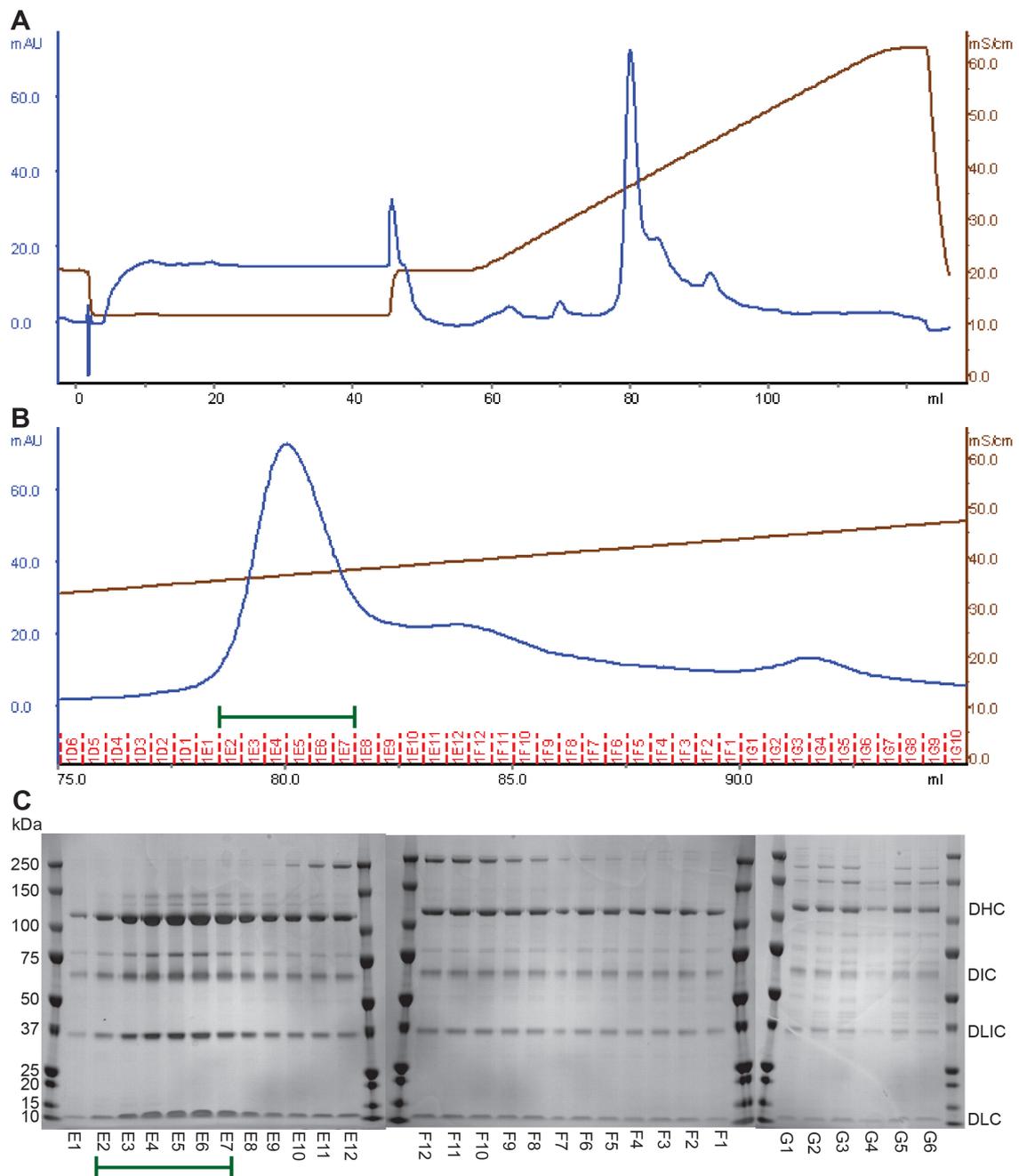
Figure 2.15: **Large-scale purification of DHC-NH using anion exchange chromatography.** DHC-NH was purified from $\sim$ 1kg of CY339 grown in a 120L fermentor. After TEV cleavage, the sample was bound to a MonoQ column and then eluted using a salt gradient. **A)** The MonoQ chromatogram. UV absorbance is shown in blue and conductivity is shown in brown, A prominent peak eluted just before the mid-point of the gradient. **B)** A closer look at the prominent peak from the chromatogram in **A**. The green bar indicates fractions that were pooled for SEC. **C)** SDS-PAGE analysis of fractions around the most prominent peaks in **A**. The green bar indicates fractions that were pooled for SEC.
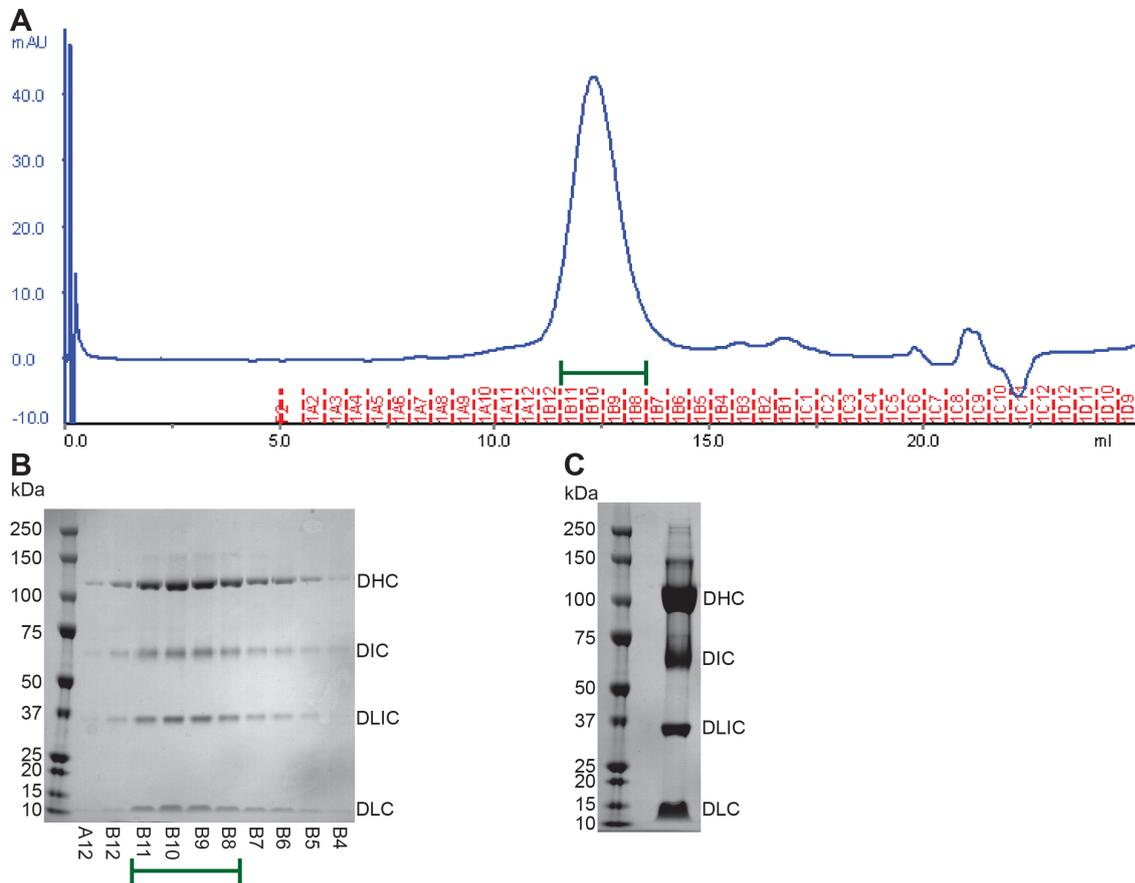
Figure 2.16: **Large-scale purification of DHC-NH using size exclusion chromatography.** DHC-NH was purified from $\sim$ 1kg of CY339 grown in a 120L fermentor. After TEV cleavage, the sample was bound to a MonoQ column and then eluted using a salt gradient. **A)** The MonoQ chromatogram. UV absorbance is shown in blue and conductivity is shown in brown, A prominent peak eluted just before the mid-point of the gradient. **B)** A closer look at the prominent peak from the chromatogram in **A**. The green bar indicates fractions that were pooled for SEC. **C)** SDS-PAGE analysis of fractions around the most prominent peaks in **A**. The green bar indicates fractions that were pooled for SEC.

crystallography, but the yield remained problematic if I hoped to perform a large number of crystal screens.

I continued to grow CY339 in the 120L fermentor and performed several large-scale purifications of the tail complex as described in Section 2.2.5. After several such iterations, I was able to set up 15-20 crystal screens, none of which produced any hits. Typically, I was only able to set up three 96-well plates for every kg of yeast, using 200nL of protein at 5mg/mL per droplet. This made it clear that any attempt to produce crystals of the dynein tail complex might require a long and drawn out screening process, requiring much larger amounts of protein than I could produce using this method. Therefore, while I continued performing these large-scale purifications from the fermentor growths, I also began cloning new plasmids, in hopes of finally increasing the yield of protein to a level suitable for crystallography.

2.2.6 Expression of DHC-NH from a plasmid

One possible explanation for the low yield observed when using CY339 is that even though each of the dynein genes has a strong promoter at its 5′ end, there is only one copy of each gene per cell. As plasmids provide a means of having multiple copies of each gene per cell, I decided to express the NH construct using two different plasmid systems. The first is a return to the polyprotein system described in Section 2.2.1, but using the DHC-NH truncation. The other was a more conventional approach, which includes placing each dynein gene in its own expression cassette (complete with both a promoter and terminator sequence) and then expressing all four cassettes from a single plasmid.

I also had to consider the different media necessary for these two expression strategies. CY339 was created using selectable markers, but as the genome itself was modified, it can be considered a stable strain that requires no selection pressure. On the other hand, strains expressing plasmids are traditionally grown using a selection pressure that encourages plasmid retention. However, the media used to exert this selection pressure are defined, and therefore not as rich in nutrients. This means there is a possible trade-off between ensuring a high percentage of cells contain the plasmid and providing the maximal amount of nutrients to all cells. To investigate this balance, when I first compared plasmid-based expression to CY339, I tested each strain in both a defined media (Ura-) and a rich media (2xYP).

Polyprotein expression of DHC-NH

To create the NH polyprotein plasmid I first used PCR to linearise the polyprotein plasmid described in Section 2.2.1. In the process of doing so, I also removed the necessary base pairs from *DYN1* so that only the first 1004 residues of the DHC would be expressed. Then I ligated this linear fragment back into a circular plasmid using Gibson assembly (Fig. 2.17).

The polyprotein plasmid was transformed into BCY123 and expressed in shaker flasks. 1L was grown in Ura- media and 1L was grown in 2xYP. For a direct comparison, I also grew 1L of CY339 in each type of media as well. I performed a small-scale purification on each of the four cell pellets and analysed the samples using SDS-PAGE directly after TEV cleavage and concentration (Fig. 2.18).

For CY339, 2xYP media resulted in a much higher expression level than Ura-media. This was expected as there is no need for a selective pressure in this strain,
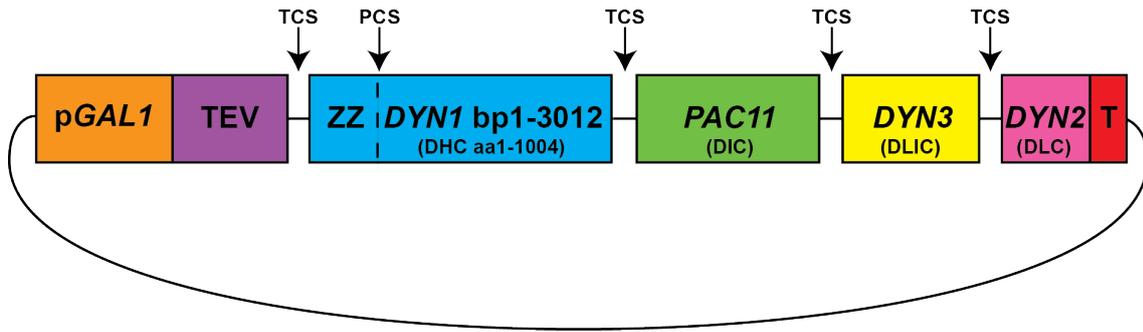
Figure 2.17: **Schematic diagram of the DHC-NH polyprotein plasmid** The four subunits of the dynein complex were expressed from a single open reading frame and then cleaved apart by the TEV protease that was expressed from the same mRNA transcript. Only the part of *DYN1* that codes for the first 1004 amino acids was included. The open reading frame was preceded by the *GAL1* promoter (p*GAL1*) for high level inducible expression and followed by the *CYC1* transcriptional terminator (T) for efficient termination of mRNA. The genes were separated by linker peptides containing TEV cleavage cites (TCS). An N-terminal ZZ-tag was fused to the DHC by a PreScission cleavage site (PCS).
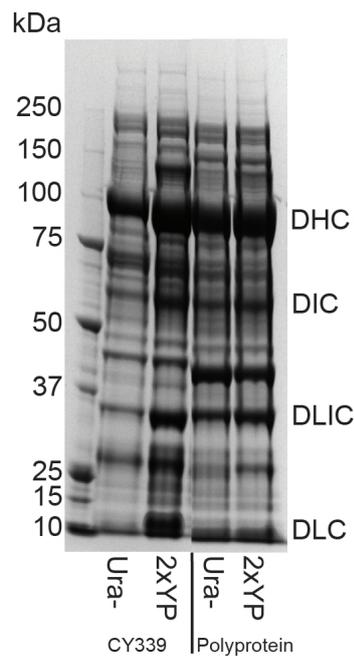
Figure 2.18: **Gels comparing the expression of the NH tail construct from both CY339 and a polyprotein plasmid in two different types of media.** Both CY339 and BCY123 transformed with the NH polyprotein were grown in 1L of Ura- and 1L of 2xYP media. Small-scale purifications performed and the concentrated samples were analysed by SDS-PAGE. 2xYP provided a much higher yield than Ura- when using CY339. However, the yield was roughly equivalent in both media when using the polyprotein. When comparing CY339 to the polyprotein, the amount of DHC is comparable, but there is a markedly lower amount of DLC in the polyprotein samples.

and 2xYP is much more nutrient-rich than defined media. For the polyprotein plasmid, the expression is comparable in both media, suggesting a balance between selective pressure and nutrient availability. As 2xYP showed a clear advantage for CY339, which does not require any selective pressure, it is likely that a certain percentage of cells grown in 2xYP did lose the plasmid. But the smaller number of yeast cells that contained the plasmid in 2xYP must have expressed enough protein to compensate, as there was as much total protein produced as in Ura-, where it is assumed that the majority of the cells contain the plasmid. Therefore, at least for this plasmid system, the media used does not have a large effect on the final protein yield.

The CY339 and NH polyprotein samples grown in 2xYP appear similar in most respects. The DHC, DIC and DLIC are present at similar levels in both samples, with slightly more in the genomic sample. However, there is a clear decrease in the DLC level when using the polyprotein. This is most likely due to the fact that there should be two DLCs binding to each DIC and the polyprotein only has one copy of the DLC gene, enforcing a strict 1:1:1:1 ratio between all four subunits. In conclusion, it seemed that the polyprotein method works for expressing the dynein tail, but not at the correct stoichiometry. Therefore, I tested another type of plasmid-based expression strategy to see if this could overcome the problem of stoichiometry.

Expression of truncation NH using multiple cassettes

A plasmid was constructed by Andrew Carter using PCR and Gibson Assembly, which contained four separate expression cassettes, each containing a different dynein gene preceded by its own *GAL1* promoter and followed by its own *CYC1* terminator sequence (Fig. 2.19). The cassettes were linked together by 40bp stretches of unique sequence. Using these unique sequences as primer annealing sites for PCR, individual cassettes can be amplified, altered, and reintegrated into the plasmid. This was included in the design in order to make future modification of individual subunits within the dynein complex more efficient. This plasmid is hereafter referred to as the NH-MC plasmid.

In order to compare this new NH-MC plasmid with CY339 and the NH polyprotein, I grew 1L of each. The two plasmids were transformed into BCY123 for this experiment. Since the polyprotein plasmid seemed to yield roughly the same amount of protein regardless of the media (Fig. 2.18), I used 2xYP for all three strains in this experiment. I performed a small-scale purification on each of the three cell pellets and analysed the samples using SDS-PAGE directly after TEV cleavage and concentration (Fig. 2.20).

The yeast strain containing the NH-MC plasmid expressed all four dynein subunits at a noticeably higher level than either the genomic expression strain or the polyprotein plasmid strain (Fig. 2.20). From this point forward, I began using this plasmid for expressing the tail complex from yeast. I also used it as the background in which I created further dynein tail constructs, including phosphomutants and truncations of the DIC and DLIC, as described in Chapter 3.

2.2.7 Expression of truncation NH in Sf9 cells

While the NH-MC plasmid did provide a roughly two-fold increase in expression levels, this was not a significant enough difference to make crystallography trials
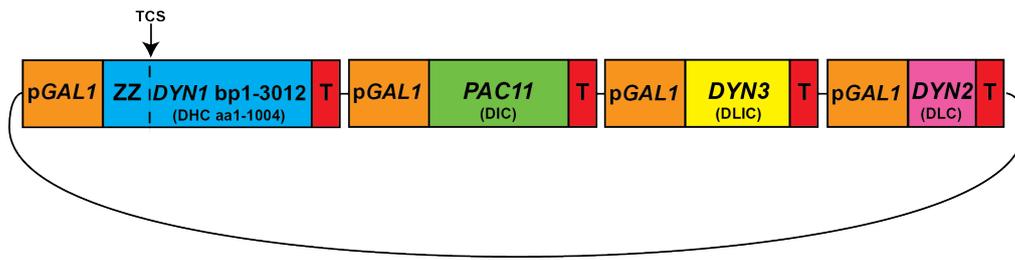
Figure 2.19: **Schematic diagram of the NH-MC plasmid.** Each of the four subunits of the dynein complex was cloned into its own expression cassette, including both a *GAL1* promoter (p*GAL1*) and *CYC1* transcriptional terminator (T). Only the part of *DYN1* that codes for the first 1004 amino acids was included. An N-terminal ZZ-tag (ZZ) was fused to the DHC by a TEV cleavage site (TCS).
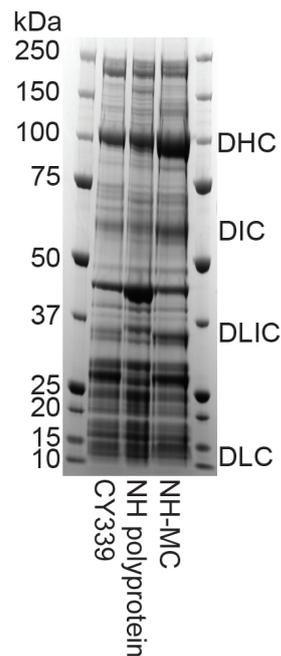
Figure 2.20: **SDS-PAGE analysis showing the protein yield from CY339, the NH polyprotein plasmid and the NH-MC plasmid.** 1L of CY339, 1L of BCY123 transformed with the NH polyprotein plasmid and 1L of BCY123 transformed with the NH-MC plasmid were grown in 2xYP. A small-scale purification was performed on each pellet and the concentrated samples were analysed by SDS-PAGE. The NH-MC plasmid has a much higher protein yield than the other two expression strategies.

any less labour-intensive. Around this time, my colleagues in the Carter lab were establishing a baculovirus expression facility and beginning to express the dynein motor domain using this system. They succeeded in obtaining much higher yields of purified protein per litre of culture than possible with yeast. So I designed a plasmid that would allow me to express the DHC-NH construct of the yeast tail complex using the baculovirus system.

The specific expression system I used is called MultiBac (Bieniossek *et al.* 2008). This system allows for the expression of multi-subunit protein complexes from a single baculovirus, which is used to infect insect cells cultured in suspension. All four dynein genes were synthesised by Epoch Life Science in order to optimise codon usage for Sf9 cells. As in previous constructs, the DHC had a ZZ tag linked to its N-terminus for purification. Then, similar to the multiple cassette approach I used in yeast (Section 2.2.6), each gene was cloned into its own expression cassette, where it was preceded by a polyhedrin promoter and followed by a SV40 terminator sequence, and then inserted into the pACEBac1 vector. The polyhedrin promoter is used because it only becomes active at a very late stage of viral infection, ensuring that high cell densities can be reached before recombinant expression begins. The complete plasmid was then inserted into the EMBacY baculovirus genome, ultimately producing a functional baculovirus that was capable of infecting Sf9 cells.

When amplifying the virus for the first time, I collected the cells used to produce the P2 generation of virus and performed an expression test on them. As a positive control, I also generated a P2 virus using a motor domain construct (CY184, Fig. A.4), which was known to express well in Sf9 cells. I briefly spun down 100μL of each cell suspension and removed the supernatant. Then I resuspended the pellet in 1X LDS sample buffer, boiled for 5 min and loaded 10μL of this crude lysate on
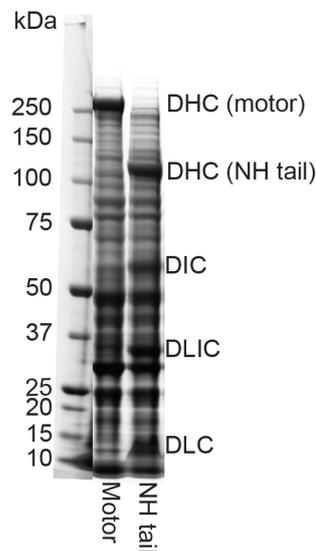
Figure 2.21: **Gel showing expression of NH truncation in crude lysate of Sf9 cells.** A virus known to successfully experess the dynein motor domain and a virus created for the expression of the NH tail construct were used to infect Sf9 cells. The pellets from the 50mL P2 cultures were collected and lysed and 10μL of crude lysate was loaded for each sample. Each subunit of the tail complex expressed above background, suggesting that similar levels of the tail were produced as the motor domain construct.

a gel for SDS-PAGE analysis (Fig. 2.21).

As expected, the motor domain construct produced a single band near the 250kDa mark, which was clearly visible above the background proteins in the lysate. The DHC band from the tail construct is as prominent as that of the motor domain construct, which suggests that it expressed very well. The DIC, DLIC and DLC bands are also all clearly visible above the background when compared to the motor domain lane of the gel. From $\sim 100\mu$L of Sf9 cell culture, there was almost as much tail complex present in the lysate as there typically is from 1L of yeast culture, indicating that expression is a few orders of magnitude higher in insect cells. This large increase in the relative amount of dynein tail in the crude lysate not only suggested that I might be able to get a much higher yield overall, but also that it would be significantly easier to separate the dynein tail from impurities. Therefore, I began to scale up the expression of this construct using Sf9 cells and performed a test large-scale purification from 3L of cell culture.

This large-scale purification was done using the same buffers and protocol as the large-scale purification from yeast, except cell lysis was performed manually using a tissue homogeniser. The MonoQ chromatogram showed a similar pattern as the sample purified from large amounts of yeast except for one clear difference. The Sf9 sample showed a large increase in UV absorbance immediately after sample loading was complete and the column was washed in low salt buffer (Fig. 2.22). By comparing the ratio of absorbance at 280nm and 260nm, I was able to determine that this peak is due to nucleic acids (most likely RNA) eluting off of the column rather than protein. I confirmed this by analysing a sample from this peak using SDS-PAGE. The fractions from this RNA peak produced no bands on the gel despite having a UV absorbance at 280nm several times higher than that of some other fractions that produced clear protein bands on the gel. The dynein peak, which elutes at ~ 30mS/cm looked very similar in shape to the one observed when purifying the tail complex from 1kg of CY339 (Fig. 2.15, but had a slightly larger amplitude. Therefore, even at this stage of the purification, it was clear that from just 3L of Sf9 cells (which produced a ~ 45g pellet) it was possible to produce more protein than I was able to get from a 1kg yeast pellet.

SDS-PAGE analysis of the fractions that eluted during the dynein peak revealed that the dynein tail complex was present, intact and relatively pure. I pooled fractions D2-E4, concentrated them to $\sim 500\mu$L and ran the sample over a Superose6 column. Once again, the chromatogram was very similar to the one produced from the yeast purification (Fig. 2.23A), producing a single prominent and symmetrical dynein peak. The input for this SEC run and the individual fractions from the peak were analysed by SDS-PAGE, which showed a very pure and complete dynein tail complex (Fig. 2.23B).

The final yield of dynein tail from 3L of Sf9 cells was ~ 0.5mg, which is comparable to the typical yield from 1kg wet mass of yeast. As expression in Sf9 cells finally provided a much higher yield of the dynein tail, I continued crystal screening using this sample. By performing dynein tail purifications from 6L of Sf9 cells at a time, I was routinely able to obtain 2mg to 3mg of sample after the MonoQ step. However, when concentrating this sample for the subsesquent SEC step, the protein often precipitated out of solution and formed a slimy, white solid, which could not dissolve back into solution. This seemed to occur whenever the concentration of dynein tail exceeded ~ 5mg/mL. SDS-PAGE analysis of the precipitate confirmed
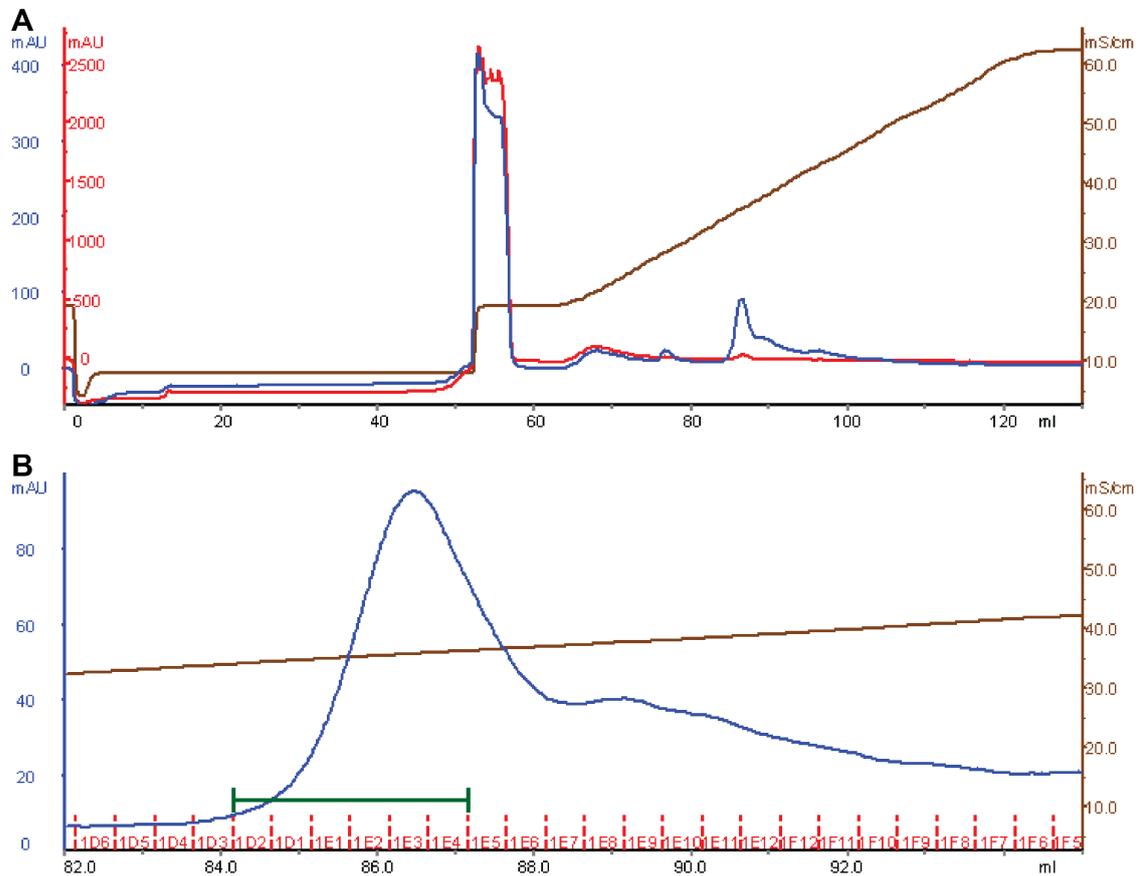
Figure 2.22: **Anion exchange chromatography of the dynein tail complex from Sf9 cells.** 3L of Sf9 cells infected with the NH tail virus were harvested and a large-scale purification was performed using a similar protocol to the one used for CY339. **A)** The anion exchange chromatography chromatogram shows a similar profile as in Fig. 2.15, suggesting the sample behaved similarly to the equivalent construct expressed from yeast. One clear difference is the large amount of nculeic acid that eluted off the column during the pre-gradient wash. **B)** A closer look at the dynein peak. The green bar indicates the fractions that were pooled and concentrated for SEC.
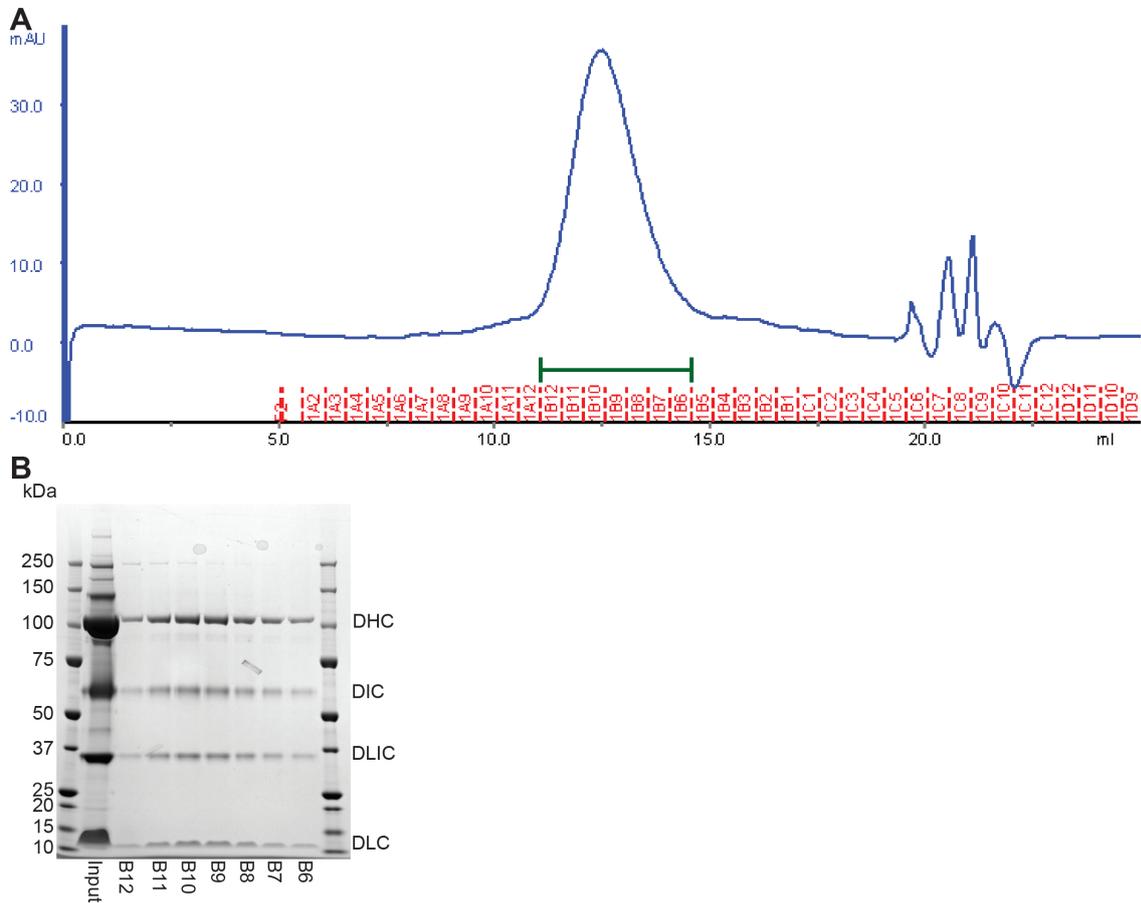
Figure 2.23: **Size exclusion chromatography of the dynein tail complex from Sf9 cells.** The dynein tail complex was purified from 3L of Sf9 cells. After anion exchange chromatography, the dynein peak was pooled, concentrated and loaded on a Superose6 column. **A)** The chromatogram shows that a single, symmetrical peak eluted at a similar volume as when the dynein tail was purified from yeast. The green bar indicates fractions that were analsyed by SDS-PAGE. **B)** SDS-PAGE analysis comparing the input to the peak fractions for the SEC run. The gel revealed that the dynein complex was intact and that SEC successfully removed any contaminants that remained after the anion exchange step.

that it contained the dynein tail. As this is not typically a characteristic associated with proteins that are likely to crystallise (indeed, none of the screens I managed to set up produced any promising hits), I spent a considerable amount of time and effort trying to understand the cause of this precipitation, and to remedy it. Those attempts to improve the behaviour of this protein sample are the focus of Chapter 3. However, as single particle EM requires only a dilute sample, I was already able to begin exploring this other powerful structural technique immediately.

2.2.8 Negative stain EM of truncation NH

To produce the sample used for EM, I repeated the purification described in Section 2.2.7 using 6L of Sf9 cells. The SEC fraction with the most pure sample (B10, Fig. 2.24) was concentrated to ~ 0.16mg/mL, which is equivalent to ~ 350nM for this construct. I made a series of dilutions, to produce samples at 50nM, 20nM and 10nM. Linas and I made several negative stain grids using all three of these sample concentrations. We then screened them in the microscope and determined that the optimal concentration, which gives a good distribution of single particles on the grid was 20nM.

Using one of these grids, I collected 200 micrographs, one of which can be seen in Fig. 2.25A. Then Andrew and I picked $\sim 10,400$ particles manually. A 2D classification was used to sort the particles into 50 2D classes (Fig. 2.25B). Even a cursory glance at the 50 classes is enough to get a sense of the tail structure (Fig. 2.25C,D,E,F). In many classes, the complex is clearly divided into two equal halves, which is expected since the tail complex contains a dimer of each subunit (Fig. 2.25C). Going along the long axis of each monomer, there are three distinct domains (Fig. 2.25C,D,E,F). Two of the three appear quite rigidly attached to each other, and to their dimer partners. In some classes, the third domain is tightly dimerised (Fig. 2.25C), while in other it appears flexible enough that the two monomers are completely separate (Fig. 2.25F), suggesting that in this part of the complex there is some inherent flexibility. The 50 2D classes offer several different views of the tail complex, which allowed us to calculate an initial 3D model. We refined the initial 3D model and performed a 3D classification to separate the particles into 5 distinct 3D classes, one of which is shown in Fig. 2.25G.
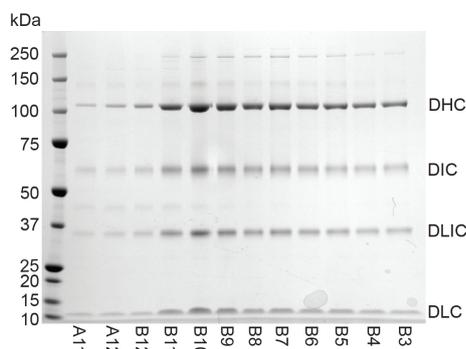
Figure 2.24: **Gel showing sample used to make EM grids.** The NH tail construct was purified from 6L of Sf9 cells. After SEC, the peak fractions were analysed by SDS-PAGE and fraction B10 was selected for further study by negative stain EM.
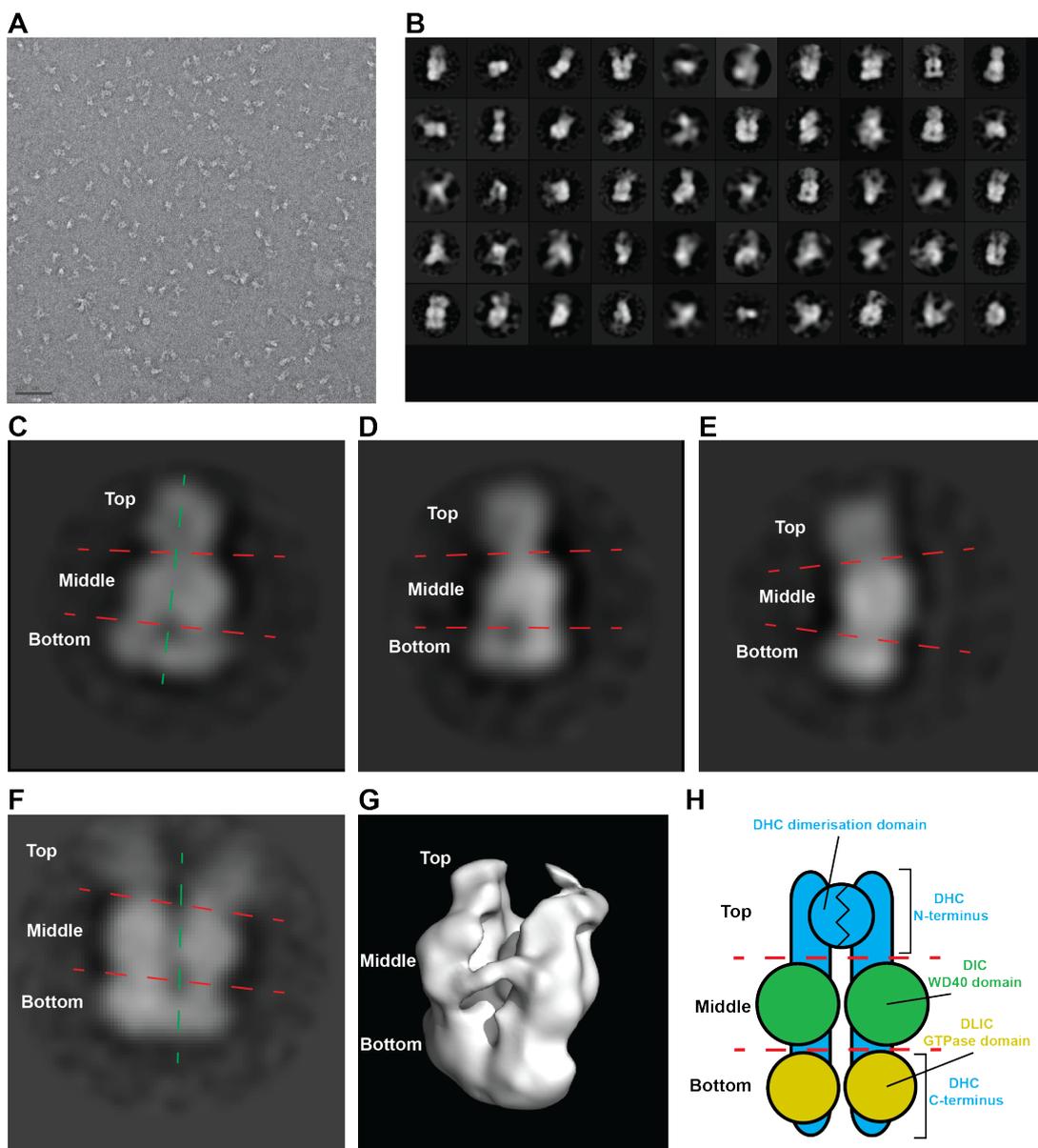
Figure 2.25: **A low-resolution structure of the dynein tail as determined by negative stain EM.** Negative stain grids were made using the dynein tail complex purified from Sf9 cells. 200 micrographs were collected, and $\sim 10,400$ particles were picked manually. The particles were sorted into 50 2D classes, which were used to calculate an initial 3D model. **A)** Representative micrograph. **B)** 2D classes. Most appear to be various views of the same structure, but some may be contaminants or disordered complexes. **C,D and E)** Three 2D classes that offer different views of the same structure. **C** is a front view, **E** is a side view and **D** is somewhere in between. In all three classes, the dotted red lines separate the three distinct domains, which are labelled Top, Middle and Bottom. In **C**, a green dotted line divides the two halves that appear roughly equal in size. **F)**A view similar to **C**, but with the top domain dissociated. Several 2D classes in **B** show the top domain in a tightly dimerised conformation, while others show it in this open conformation. **G)** One 3D class from the refined 3D reconstruction. The three domains are labelled, but the top domain is missing due to its flexibility. **H)** Schematic diagram of the dynein tail based on the negative stain model as well as data described in chapter 4.

2.3 Discussion

2.3.1 Overexpression of a recombinant full-length dynein complex

The earliest *in vitro* studies of cytoplasmic dynein relied on a microtubule affinity purification (Paschal *et al.* 1987). Several years later, it became possible to overexpress mammalian DHCs recombinantly (Mazumdar *et al.* 1996). This eventually led to the introduction of modifications, such as those used to create minimal motor constructs of the *S. cerevisiae* DHC used in motility assays (Reck-Peterson *et al.* 2006) and structural studies (A. P. Carter *et al.* 2011). However, studies of the full dynein complex remained constrained by the endogenous expression levels of the dynein subunits. The yield from these techniques was enough for western blots and motility assays (Kardon *et al.* 2009), but nowhere near abundant or pure enough for structural studies. Here, I have presented the first successful strategy for co-overexpressing all known subunits of a recombinant full-length dynein complex. This is useful in its own right, as a means of producing larger amounts of high-purity dynein for motility assays, biochemical studies and even structural studies. Furthermore, it provided a background in which many other dynein constructs have been created.

2.3.2 Description of a minimal tail complex: truncation NH

Taking advantage of the ability to overexpress the full dynein complex, I made several DHC truncations and ultimately found one that eliminates the flexible linker while retaining all four dynein subunits. This was the first time a dynein tail complex was overexpressed recombinantly and the first report of a pure *in vitro* sample containing the isolated tail region of dynein. The ability to purify the dynein tail complex, along with the freedom to define and modify the components of the complex was an important step in understanding its function and structure. In addition to providing the requisite sample for structural studies it opened the door to myriad biochemical experiments, such as well-controlled binding assays with known interactors, discovery of unknown interactors via co-immunoprecipitation, and characterisation and manipulation of post-translational modifications. I took advantage of this myself and some of these biochemical experiments are described in Chapter 3. Furthermore, the NH truncation I discovered in yeast has already been used by Max Schlager to guide the creation of a similar human tail construct (discussed further in Chapter 4). Since the alignment between these two species is fairly clear in the N region of the DHC, he designed a human version of the dynein tail complex without the need to tediously screen dozens of DHC truncations (Fig. A.3). This human version of the NH truncation ultimately proved useful for structural studies that are described in Chapter 4.

Some surprising observations

Even the process of defining this truncation resulted in some unexpected observations that are interesting in themselves. The first of these arose when making the first series of DHC truncations (Section 2.2.3). Previous studies had suggested that the DIC binds between residues 446 and 701 of the *R. norvegicus* DHC and residues

629 and 730 of the *D. discoideum* DHC (Habura *et al.* 1999, Tynan *et al.* 2000) and that the DLIC binds between residues 649 and 800 of the *R. norvegicus* DHC (Tynan *et al.* 2000). Given this information, I expected to see some shorter DHC truncations that bind the DIC and not the DLIC. Strikingly, there was a difference of only 94aa between the longest *S. cerevisiae* DHC truncation to bind no accessory proteins (truncation M, 954aa) and the next truncation (truncation O, 1048aa), which bound all accessory proteins (Fig. 2.8C). One possibility is that the DIC and DLIC binding sites are spatially separated from each other, but the entire region of the DHC needed for binding both can be destabilised by truncations. If truncating the DHC within this region disrupts it to the point that the protein is targeted for degradation, this would also explain the observation that between truncation E (557aa) and truncation L (870aa) no truncations expressed well enough to be visible on a gel (Fig. 2.8B). Some evidence in support of this model arose from a very recent study, which used negative stain EM to suggest that the DIC and DLIC are not adjacent when bound to the DHC (Chowdhury *et al.* 2015). However, this study was done using bovine dynein, and it is also possible that there are substantial differences between mammalian and yeast dynein.

The second unexpected observation was made when performing SEC on selected constructs from the second series of DHC truncations (Section 2.2.4). It was already known that treating the dynein complex with the chaotropic salt potassium iodide dissociates it into two distinct subcomplexes (S. Gill *et al.* 1994). One subcomplex consists of the DLICs and DHCs while the second contains the DLCs and DICs. Furthermore, it was shown that the DLIC is very tightly bound to the DHC and that this subcomplex is highly unstable due to the absence of IC. (S. J. King *et al.* 2002). In light of all this, the way that truncations NA, NL and NH behaved over gel filtration was intriguing.

First, it is worth noting that NA, NL and NH (as well as the equivalent truncations fused to a C-terminal GFP) all retained the DIC, DLIC and DLC through TEV cleavage and concentration, but some of these constructs lost subunits when run over a SEC column (Figs. 2.12, 2.13, and 2.14). For example, both NA and NA-GFP dissociated from all accessory chains completely (2.12). This suggests that truncation NA (residues 1-958) may contain some of the residues involved in the DIC and DLIC interfaces, but not all of them. As expected based on previous studies, the bare NA truncation was not a stable protein, which may be due to its lack of the DIC's stabilising influence as reported by S. J. King *et al.* 2002.

A comparison of the NL and NH truncations is where things become much less clear. The NH truncation (residues 1-1004) is long enough to bind all accessory chains and form a stable complex (Fig. 2.14). However, truncation NL (residues 1-1063) only binds the DLIC despite being 59aa longer than truncation NH (Fig. 2.13). It seems that the additional 59aa actually have a destablising effect on the entire complex and somehow prevent the DIC from binding. In agreement with the previous subcomplex studies (S. Gill *et al.* 1994, S. J. King *et al.* 2002), the complex formed between truncation NL and the DLIC is more stable than the bare NA truncation, but still not a very well-behaved protein. Finally, the C-terminal GFP actually had an opposite effect on truncations NL and NH; its presence resulted in less aggregation of truncation NL, but more aggregation of truncation NH (Figs. 2.13 and 2.14). A possible explanation is that the additional 59aa on the NL C-terminus are inherently disordered and the addition of a GFP serves to decrease

aggregation, whereas the NH truncation is already stable, so fusing on a GFP gets in the way and induces aggregation. This is why for all future work, I used the NH construct that lacked the C-terminal GFP.

Suitability of truncation NH for crystallisation trials

When I was finally successful in increasing the yield of truncation NH by expressing it in Sf9 cells, its suitability as a crystallisation target came under question. The first warning sign was that the sample was very sensitive and would often form a white, goopy precipitate if there were air bubbles present during the TEV cleavage step or when it was concentrated to more than 5mg/mL. Initially, I resolved to simply be very careful while purifying the delicate sample and set crystal screens at 5mg/mL, However, when these screens produced no viable hits I began to think about possible explanations for this problem that might also prevent crystallisation. The most simple explanation is that the NH truncation is good enough to make a fairly stable complex, but its C-terminus is still somehow sub-optimal. Furthermore, I had not yet tried making truncations of the accessory chains, so it is possible that the DHC truncation is optimal, but the DIC, DLIC or DLC contain inherently disordered regions that cause the entire complex to aggregate. Truncations of the DLIC and DIC are described in Sections 3.2.3 and 3.2.4, respectively. Finally, whenever using a heterologous expression system, it is possible that improper post-translational modifications (PTMs) will arise, which can have a negative effect on protein stability. Sf9 cells are particularly notorious for hyperphosphorylation. I briefly investigated the native phosphorylation state of the dynein complex, which is described in Chapter 3.

2.3.3 An initial 3D model of the tail complex

Although the NH truncation did not appear suitable for X-ray crystallography, it did prove a useful sample for negative stain EM. I was able to collect enough single-particle electron micrographs of the yeast dynein tail to calculate the initial 3D model presented in Section 2.2.8. This model provides the first 3D structural information on the tail region of cytoplasmic dynein. Despite being low-resolution, it already provides some insights into the overall architecture of the dynein tail.

The general shape approximated a rectangular prism with side-lengths that were roughly 1:2:3 (10nm x 18 nm x 35nm) (Fig. 2.25). The complex was clearly organised as a dimer of two equal halves, but the resolution was too low to see whether this is a rotational or translational symmetry. Each monomer could be divided into three globular domains, with the middle one being the largest of the three. The "middle" and "bottom" domains (as labelled in (Fig. 2.25C-G) appear tightly bound together and have a rigid dimer interface with no conformational changes between these domains visible in the 2D class averages. Strikingly, the "top" domain does adopt a variety of conformations, which are visible in different 2D class averages (Fig. 2.25C,F). In some classes, the two small domains form a tight dimer, appearing more as one larger globular domain centred on the axis of symmetry. In other classes, they stretch away from the axis of symmetry, oriented in opposite directions, giving the complex the appearance that it is "fraying" at the end.

Without a higher resolution structure or labelling experiments, I was not able to assign the electron density in the 3D models to specific structures within the tail

complex. Based on further work carried out by my colleagues and myself (described in Chapter 4), it later became possible to assign the density. Fig. 2.25H provides a schematic diagram of the dynein tail organisation, which takes into account data from Chapters 2 and 4 of this thesis, as well as a recently published paper by Chowdhury *et al.* 2015. The crystal structure and cryo-EM maps used to help determine this model and a further discussion of the organisation of the dynein tail are included in Chapter 4.

2.3.4   Future directions

While the NH truncation did not prove a promising crystallography target, many other strategies for getting a useful crystal structure of the dynein tail remained. The most obvious first step would be to go back and test some more N region truncations using SEC. I only tested NA, NH and NL, but it is possible that the most similar truncations to NH (for example NG and NI) would have a slightly more stable C-terminus and behave better overall in solution. It might also be worth trying to fuse proteins other than GFP onto the C-terminus and to try including a short flexible linker between the GFP and the DHC.

Next, as until this point I had only modified one of the four dynein subunits, I would see if modification of the DIC, DLIC and/or DLC might improve overall complex stability. The DIC is a particularly good candidate for N-terminal truncations as this region is thought to be important for binding many of dynein's interaction partners, and it may be disordered in their absence. Another angle to solve the same problem would be to co-express some of these known interactors with the dynein tail, so that they might bind and lock any inherently disordered regions within the DIC into a more rigid conformation.

Furthermore, I would try to describe the PTMs present in the dynein tail, expressed from both yeast and Sf9 cells. This would serve two purposes. First describing the native state of yeast dynein would contribute to a better biochemical understanding of the dynein tail's functions. Phosphorylation of both the DIC and DLIC is thought to be an important part of how mammalian dynein is regulated, so it would be interesting to see if it plays a significant role in yeast despite the more limited role of dynein in this organism. And a description of the yeast dynein's ubiquitination may shed light on how dynein levels are regulated throughout the cell cycle, not to mention the potential for finally understanding why it is so difficult to express dynein at high levels from yeast. Second, it is possible that the dynein expressed in Sf9 cells has PTMs that are not present in native dynein, or lacks PTMs which are present in native dynein. This alone may be causing the solubility problems I observed with the dynein tail.

Finally, after successfully creating negative stain grids using the NH truncation, a logical next step would be to attempt imaging the same sample using cryo-EM. Even a modest improvement in the resolution of the 3D structure of the dynein tail may prove invaluable in understanding the function of this complex better. At the very least, it might provide useful information for designing future constructs that could be used for x-ray crystallography.

Together, the NH-MC plasmid that I used to express the dynein tail from yeast and the plasmid I used for baculovirus expression provide the basic foundation necessary for all of the above experiments. Chapter 3 describes my attempts to carry

out some of the experiments described above using these two plasmids.

2.3.5   Conclusions

In this chapter I have presented the design, purification and initial characterisation of a dynein tail complex. This is the first description of a successful attempt to isolate the dynein tail in vitro. The results presented here have opened the door to further biochemical and structural studies on this important part of the dynein complex, which may contribute to a better understanding of both cargo binding and regulation of dynein. I have also presented the first structural insights into the overall architecture of the dynein tail, providing a foundation upon which future structural studies may continue to build.

Chapter 3

Further biochemical characterisation of the minimal dynein tail complex

3.1 Introduction

Although expressing the yeast dynein tail complex in Sf9 cells increased the yield significantly, the sample often precipitated out of solution during concentration, making it ill-suited for crystallography experiments. This chapter describes some of my efforts to improve the stability of the sample using a variety of techniques. It also describes the tools I developed in order to perform these experiments, which may be a useful resource for future studies in their own right.

When initially searching for the optimal dynein tail complex in Chapter 2, I varied only one characteristic of a single dynein subunit, namely the length of the DHC. Having optimised that one parameter, I broadened my focus to include the other three subunits, as well as the complex as a whole. These experiments, which were always done in the context of the minimal tail complex (truncation DHC-NH) described in Chapter 2, are the topic of the current chapter. First, I addressed the stability of the complex as a whole by testing various buffer conditions. When this failed to improve the behaviour of the complex, I addressed each subunit one by one. I attempted to find another yeast DLC and truncate both the DLIC and DIC to eliminate inherently disordered regions. I also considered the possibility that adding yet another subunit, or at least a fragment of one, might improve the stability of the DIC, and consequently the entire tail complex. I also investigated the phosphorylation state of the complex as a whole.

Andrew Carter proposed the idea that Tda2 may be a yeast homologue of Tctex-1 (Section 3.2.2). Carina Motz helped purify the DIC truncations (Seciton 3.2.4). The remainder of the experiments described in this chapter were my own work.

3.2 Results

3.2.1 Stability of the dynein tail in various buffers

One simple explanation for why a protein might aggregate at high concentrations is that the buffer conditions are not favourable. In particular, salt concentrations often

67

require tweaking as some proteins are most soluble at very low concentrations while other require a much higher salt level. Until this point, I had always used 150mM KCl since this is near physiological conditions and worked well for the dynein motor domain, so I set out to optimise the salt concentration for the dynein tail. One complication of studying a protein complex is that higher salt concentrations may disrupt the protein-protein interactions required for complex formation, so I first performed a simple experiment to determine the salt tolerance of the dynein tail complex.

I began by performing a standard large-scale purification of the dynein tail from $\sim$ 1kg of yeast that had been grown in the fermentor. Once the sample was bound to IgG beads and had been washed in lysis buffer, I divided the beads into 12 equal parts and put them into separate columns. I then washed each sample with 50mL of modified TEV buffers, each of which had the same composition except for the concentration of KCl, which varied from 100mM to 3M. Then I washed each sample back into standard TEV buffer (150mM KCl), cleaved the samples off of the IgG beads, concentrated and compared the samples using SDS-PAGE (Fig. 3.1).
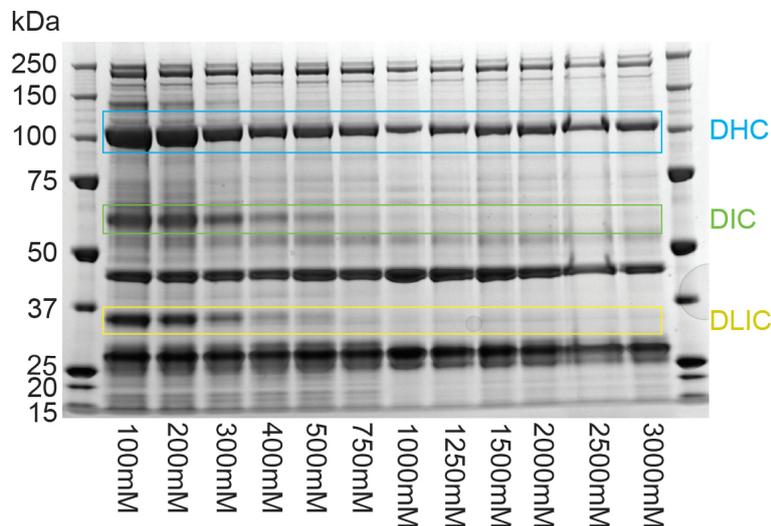
Figure 3.1: **Stability of the dynein tail complex in various salt conditions.** 12 samples of the dynein tail were subjected to various concentrations of KCl and then analysed by SDS-PAGE. Boxes surround bands that appeared at the predicted heights of the DHC (cyan), DIC (green) and DLIC (yellow).

There is a clear trend that as the concentration of KCl increases, the amount of DHC in the sample decreases. This continues up to somewhere between 500mM and 1M. At this point, the amount of DHC remains fairly constant all the way up to 3M, but there is much less present than at lower salt concentrations. Furthermore, the amounts of DIC and DLIC rapidly diminish within the 100mM to 500mM range. Above 500mM there is virtually no DIC or DLIC visible in any of the samples. Even the 200mM KCl sample appears to have slightly less of the dynein tail complex than the 100mM one. The effect is not large enough to exclude experimental or gel-loading error, but it at least suggsets that 150mM KCl may already be approaching the maximum salt concentration that the dynein tail can tolerate.

The only other major component of TEV buffer is 20mM Tris-HCl (pH = 8.0). Once I had established the useful range of KCl concentrations, I wanted to test

whether the Tris itself might be problematic. First, I purified some dynein tail from yeast and set up a solubility screen using a 96-well plate with various buffers titrated to different pH ranges. Using a robot to set up droplets with 200nL of sample plus 200nL of buffer, I was able to see how the dynein tail responds to different buffers and different pHs. The sample was most soluble in Tris-HCl (pH = 8.0) and ADA (pH = 7.0). So I performed another simple experiment, exactly as described above except the KCl ranged from 0mM to 500mM and each salt concentration was tested with both 20mM Tris-HCl (pH = 8.0) and 20mM ADA (pH = 7.0). The samples were analysed using SDS-PAGE (Fig. 3.2).
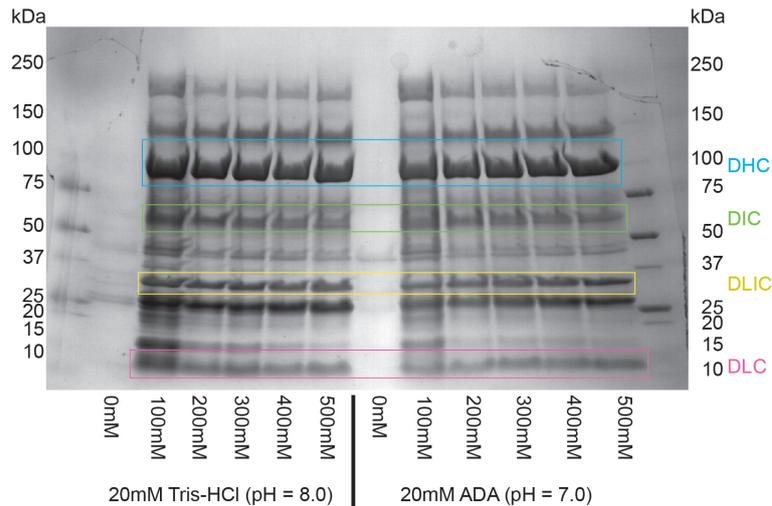
Figure 3.2: **Stability of the dynein tail complex in varoius salt concentrations and pH values.** 12 samples of the dynein tail were subjected to various concentration of KCl and two different buffers titrated to different pH values before being analysed by SDS-PAGE.

The most striking result in this experiment is that 0mM KCl results in almost no protein in the final sample, including contaminants. This is expected as with no salt in the buffer, the proteins would tend to shield themselves by aggregating and binding non-specifically to the sepharose resin in the IgG beads. There is no clear difference between the Tris-HCl sample and ADA sample at any single salt concentration, suggesting that Tris-HCl is a suitable buffer for the dynein tail. Interestingly, this experiment did not show as clear of a trend for the DIC and DLIC to dissociate at higher salt conentrations, although they do still seem to decrease slightly as the concentration of KCl is increased. One last observation is that the 100mM sample for both buffers has a noticeably higher background of non-specific proteins. Therefore, after these two experiments, I was unable to find any convincing reasons to change the TEV buffer from the 20mM Tris-HCl (pH = 8.0) and 150mM KCl composition I had been using all along.

3.2.2 Tda2: a potential yeast homologue of human Tctex-1

As discussed in Chapter 1 (Section 1.3.2), human cytoplasmic dynein has at least three distinct DLCs (LC8, Tctex-1 and Roadblock/LC7) bound to each DIC, while only one DLC is known to exist in yeast (Dyn2). However, there are two DLC
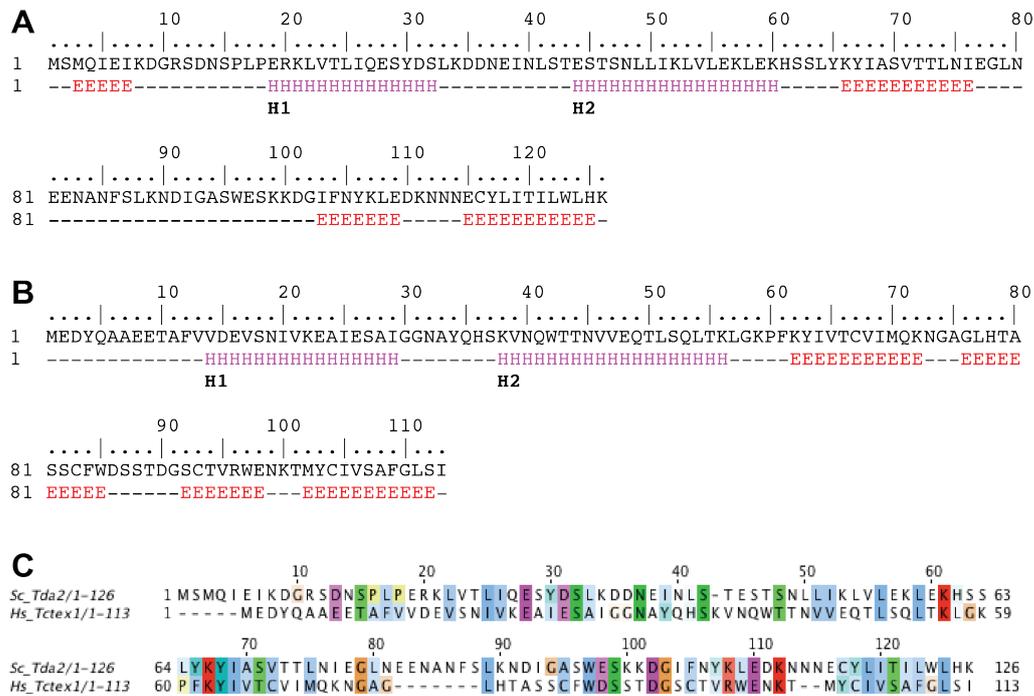
Figure 3.3: **Secondary structure predictions and alignment of Tda2 and Tctex-1.** **A)** Secondary structure prediction of the yeast protein of unknown function, Tda2. E=predicted β-strand, H=predicted α-helix. **B)** Secondary structure prediction of the human dynein light chain, Tctex-1. Key as in **A**. **C)** Alignment between Tda2 and Tctex-1 performed in ClustalX and visualised in Jalview 2. Coloured by sidechain character of conserved and similar residues.
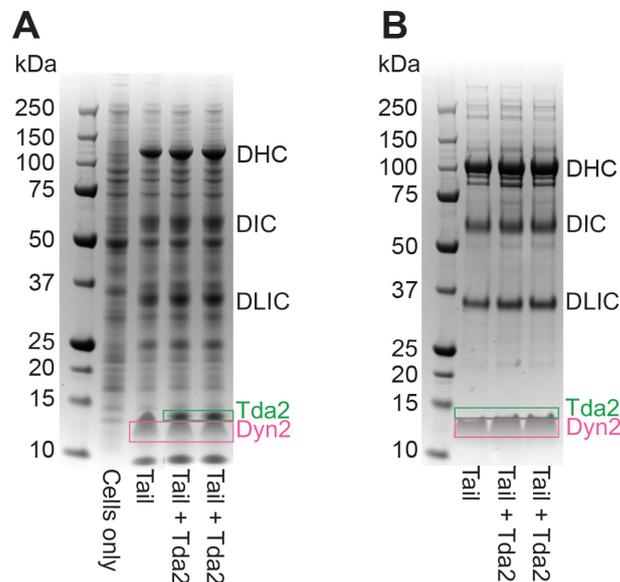
Figure 3.4: **Gels showing co-expression and purification of the dynein tail complex with Tda2.** **A)** Gel comparing the crude lysate of Sf9 cells that were not infected, cells that were infected with a virus encoding the dynein tail complex and two samples of cells that were infected with different viruses encoding the dynein tail complex and Tda2. **B)** Gel comparing the purified dynein tail complex from cells that did not overexpress Tda2 to two samples that did.

binding sites present on the yeast DIC, suggesting that there are two DLCs present for each DIC (Stuchell-Brereton *et al.* 2011). These two binding sites are found in a region of the yeast DIC that aligns with the LC8 and Tctex-1 binding sites of the human DIC. Furthermore, in human cytoplasmic dynein, LC8 and Tctex-1 are structurally similar and bind the DIC through a similar mechanism, despite being very different in amino acid sequence (Williams *et al.* 2007). Previously, these data were understood by a model that describes two copies of Dyn2 per DIC (Stuchell-Brereton *et al.* 2011). However, there were no studies that directly tested this model, and so it remained possible that there are two distinct DLCs in yeast. This provided another possible explanation for the aggregation problem observed with the dynein tail complex described in Chapter 2. Therefore, it seemed worthwhile to search the yeast genome for any proteins of unknown function that are similar to either the human Tctex-1 or the yeast Dyn2 in sequence and/or structure.

Andrew Carter performed a search of the *S. cerevisiae* genome database and found one promising candidate for a yeast homologue of Tctex-1. The function of this protein, called Tda2, remains unknown, but its predicted size and pI (14.5kDa and 4.93, respectively) are similar to those of Tctex-1 (12.5kDa and 5.00, respectively)(Gasteiger *et al.* 2005). Furthermore, the sequence is predicted to contain a Tctex-1 domain from residues 42-124 (Finn *et al.* 2013) and is predicted to have a secondary structure (McGuffin *et al.* 2000, Fig. 3.3A,B) and have a 3D fold similar to Tctex-1 (Kelley & Sternberg 2009). I performed an alignment between yeast Tda2 and human Tctex-1 using ClustalX (Larkin *et al.* 2007) and visualised the alignment using Jalview 2 (Waterhouse *et al.* 2009, Fig. 3.3C). Since all these bioinformatics techniques indicated a possible relationship between Tda2 and Tctex-1, I cloned a Tda2 cassette into the plasmid used for baculovirus expression (Section 2.2.6).

After verifying the sequence and amplifying virus, I infected a 50mL culture of Sf9 cells and co-expressed Tda2 with the dynein tail complex. By analysing a sample of crude lysate with SDS-PAGE, I was able to confirm the presence of a new protein being overexpressed, which ran at a height of $\sim$ 14kDa (Fig. 3.4A). However, after performing a small-scale purification and analysing the final sample by SDS-PAGE, the extra band is missing, suggesting that Tda2 does not interact with the dynein complex (Fig. 3.4B) However, the two bands ran very close together in Fig. 3.4A, showing some overlap. Furthermore, the DLC bands in Fig. 3.4B ran slightly higher in the two lanes with Tda2 in the sample. So to test in a more quantitative way, I determined the ratio of DLC to DLIC in each of the three lanes of 3.4B using Image Lab (Bio-Rad). The tail only sample had a ratio of 0.93, and the two +Tda2 samples had ratios of 0.99 (left) and 1.08 (right). The fact that the difference between the two +Tda2 samples is larger than the difference between the tail only sample and one of the +Tda2 samples suggests that the variation between all three samples is within the range of experimental error, and that there is no Tda2 present in any of the samples.

3.2.3   DLIC truncations

Since the DLCs did not seem to be the subunit responsible for the instability of the dynein tail complex, I next tried to modify the DLIC. The DLIC was predicted to consist of a N-terminal G-protein domain followed by an inherently disordered C-terminus (Perrone *et al.* 2003 and Finn *et al.* 2013). Therefore, in the context

of the DHC-NH minimal tail complex, I attempted to find a C-terminal truncation of the DLIC that might eliminate its flexible C-terminus while preserving an intact dynein tail complex.

First, I used a variety of bioinformatics techniques to guide construct design. To begin, I performed a secondary structure prediction (McGuffin *et al.* 2000) and created a document to summarise all my findings (Fig. 3.5). This secondary structure prediction already shows that the C-terminal 100 residues are predicted to contain long stretches of sequence with no predicted secondary structure. Next, I predicted disordered regions (Prilusky *et al.* 2005), which provided a more specific region within the C-terminus that is predicted to be disordered (Fig. 3.6). Then I confirmed the location of the predicted G-protein domain (Kelley & Sternberg 2009), which begins within 30 residues of the N-terminus and ends within the predicted disordered region. Having narrowed down the region of interest to be between residues 150 and 312 (the C-terminus) I analysed all possible C-terminal truncations within this region using PPCPred (Mizianty & Kurgan 2011) to see if any potential truncations were predicted to be particularly crystallisable.

Using all of the bioinformatics predictions, I designed a series of 20 C-terminal DLIC truncations (Table 3.1). The shortest (DLIC-A) would terminate after residue 167 and the longest (DLIC-T) would terminate after residue 301. Once I had designed the truncations themselves, I modified the multiple cassette plasmid described in Section 2.2.6 so that they would each be expressed in the context of the full tail complex in yeast.

Name	Length (aa)	MW (kDa)
DLIC-A	167	19.3
DLIC-B	181	21.0
DLIC-C	187	21.8
DLIC-D	196	22.9
DLIC-E	200	23.4
DLIC-F	204	23.9
DLIC-G	206	24.2
DLIC-H	210	24.7
DLIC-I	214	25.8
DLIC-J	222	26.2
DLIC-K	231	27.2
DLIC-L	237	27.8
DLIC-M	242	28.3
DLIC-N	251	29.4
DLIC-O	258	30.3
DLIC-P	268	31.5
DLIC-Q	275	32.4
DLIC-R	282	33.1
DLIC-S	292	34.4
DLIC-T	301	35.4

Table 3.1: **List of 20 DLIC truncations.** The name, length and predicted MW after TEV cleavage is shown for each DLIC construct.

```
              10        20        30        40        50        60        70        80
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
   1    MDNCNAWDKLLSQNESTINSTETATITAIIYSPSSKTLHQFINICFPEGSNSILDTTLINYATIGWTNDLKENYSVDVYT
   1    ---HHHHHHHHHHHH-----EEEEEEEEE----HHHHHHHHHH---------------EEEEE-------EEEEEEE
           H1                         H2

              90        100       110       120       130       140       150       160
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
  81    LIRNTDDALDLLKPFLQEHSSKVRWLILLDWTLNDQKLWLNELSYAFNKIKQLNDDNEFSVWCLNSGEILNLQRHTTVWQ
  46    E------HHHHH---------EEEEEEE-------HHHHHHHHHHHHHH------EEEEEE----HHHHH------
           H3                        H4                          DLIC→              H5

              170       180       190       200       210       220       230       240
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
 161    SVHIDFILQTLRSFCYFNDSLFYICDHTEEKREAQRLKYQLLKHFCEDRMKDHIEMVKRSEILIPKGCDSIGLIK
  88    HHHHHHHHHHHHHHHHH---EEEEE------HHHHHHHHHHHHH-----------------EE---------E
          H6                        H7

              250       260       270       280       290       300       310
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
 241    TVDERFEPTEVKEQHFLLRYMDFIPTIDKIREDRKTTSGIDKDKLYPLEVFKVNIQEELGKMFAKYRENSRI
 128    EE--------------------------------------------------HHHHHHHHHHHHHHH-----
                                                                      H8
                            ←DLIC
```

Figure 3.5: **Secondary structure prediction of DLIC.** E=predicted β-strand, H=predicted α-helix, green=predicted G-protein domain, yellow=predicted disordered regions, red=last residue in truncations, turquoise=potential phosphorylation site, blue text=DLIC domain predicted by Pfam.

```
Summary:
Number Disordered Regions:   3
Longest Disordered Region:   35
Number Disordered Residues: 54
Predicted disorder segment: [185]-[219] length: 35 score: -0.13 ± 0.07
Predicted disorder segment: [224]-[236] length: 13 score: -0.03 ± 0.01
Predicted disorder segment: [263]-[268] length:  6 score: -0.04 ± 0.02


   1 MDNCNAWDKL LSQNESTINS TETATITAII YSPSSKTLHQ FINICFPEGS
  51 NSILDTTLIN YATIGWTNDL KENYSVDVYT LIRNTDDALD LLKPFLQEHS
 101 SKVRWLILLD WTLNDQKLWL NELSYAFNKI KQLNDDNEFS VWCLNSGEIL
 151 NLQRHTTVWQ SVHIDFILQT LRSFCYFNDS SLFYICEDHT EEKREEAQRL
 201 KYQELLKHFC EDRDMKDHIE MVKRSEILIP KGCDSIGLIK TVDERFEPTE
 251 VKEQHFLARY MDFIPTIDKI REDRKTTSGI DLDKLYPLEV FKVNIQEELG
 301 KMFAKYRENS R
```


Figure 3.6: **Disorder prediction of DLIC.** Red=predicted disordered regions, green=predicted ordered regions.

73

I grew a 1L culture of yeast expressing each of the 20 DLIC truncation constructs and performed a small-scale purification on each one. Unfortunately, I unknowingly used a new batch of TEV protease for these purifications, which had not been tested and was not functional. Since I performed all 20 purifications in parallel, none of the final samples showed any dynein bands when analysed using SDS-PAGE. Any constructs that expressed and purified successfully would have remained bound to the IgG beads as a result of the faulty TEV protease. Due to time constraints, I never repeated this experiment with functional TEV protease, but these constructs remain in the Carter lab plasmid database and may be of interest for future studies.

3.2.4 DIC truncations

Similar to the DLIC, the DIC is composed of a highly structured domain and an inherently disordered region. The DIC is thought to contain a WD40 domain at the C-terminus (Finn *et al.* 2013), while the N-terminus was predicted to be largely unstructured (Barbar 2012). The N-terminus also contains binding sites for other proteins, including the DLCs and some of dynein's regulators, such as dynactin and LIS1/NUDE (McKenney *et al.* 2011 and Barbar 2012), which may induce a more structured conformation upon binding. Therefore, while it was predicted to be unstructured in isolation, it is possible that in cells, the DIC spends the majority of its time in highly structured complexes with other proteins. At least one example of this has been clearly shown by solving a crystal structure of two DLCs bound to a peptide fragment of the DIC, which becomes more structured upon binding the
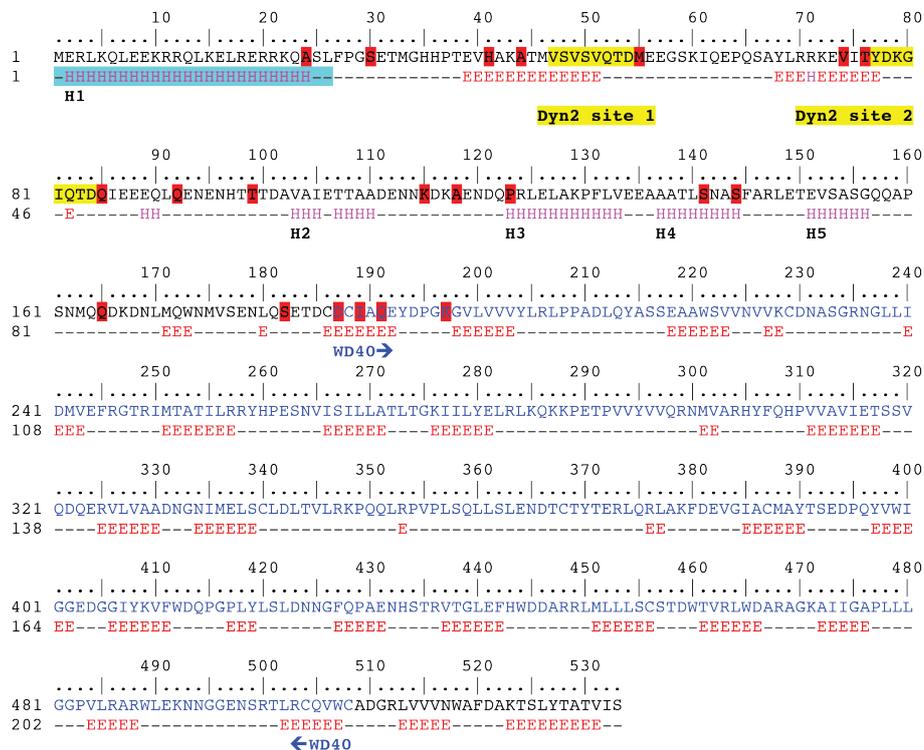
Figure 3.7: **Secondary structure prediction of DIC.** E=predicted β-strand, H=predicted α-helix, yellow=DLC binding sites, turquoise=predicted coiled-coil and Nip100 binding site, red=first residue in truncations, blue text=predicted WD40 domain.

DLCs (Williams *et al.* 2007). Nevertheless, for the purpose of solving a structure of the minimal dynein tail, a simple solution would be to simply eliminate any flexible regions entirely. Therefore, at the same time that I was designing the DLIC truncations, I applied a similar method as the one described above (Section 3.2.3) to design a series of 21 truncations that, within the context of the minimal tail complex (DHC-NH) eliminate the N-terminus of the DIC.

Again, I began by performing a secondary structure prediction (McGuffin *et al.* 2000) and created a document to summarise all my findings (Fig. 3.7). The secondary structure prediction shows that residues 170-533 contain only β-strands, while the majority of the N-terminus is α-helical in structure. The only β-strands predicted within the N-terminus are the two DLC binding sites, which had already been mapped to residues 47-54 and residues 77-84 (Stuchell-Brereton *et al.* 2011). The secondary structure prediction is in agreement with the Phyre2 alignment, which predicts that residues 187-507 constitute a WD40 domain (Kelley & Sternberg 2009). The disorder prediction (Prilusky *et al.* 2005) further strengthens the
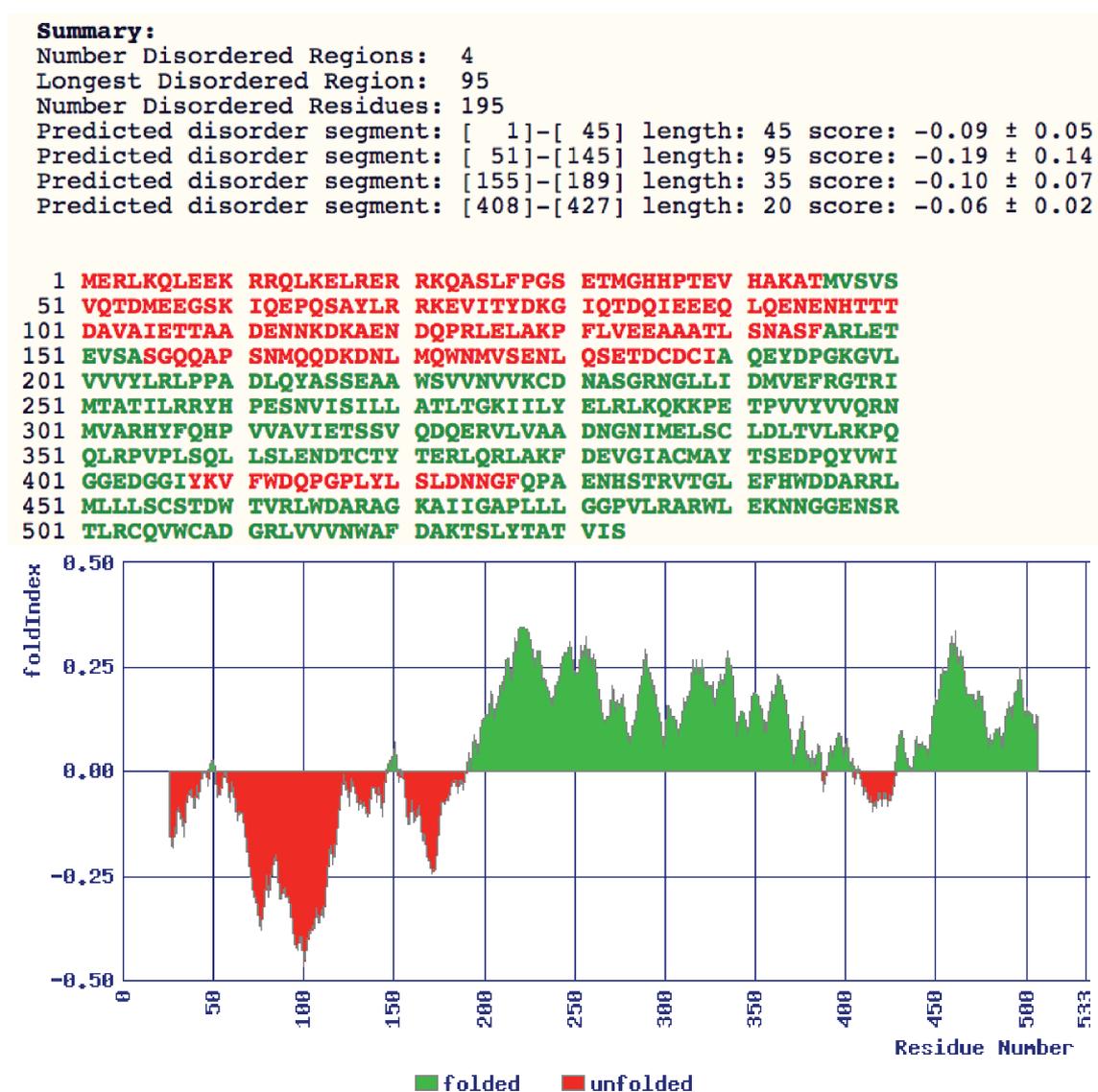
Figure 3.8: **Disorder prediction of DIC.** Red=predicted disordered regions, green=predicted ordered regions.

Name	First residue	MW (kDa)	Results
DIC-A	24	57.1	
DIC-B	30	56.5	
DIC-C	41	55.3	*
DIC-D	44	55.0	*
DIC-E	55	53.8	
DIC-F	74	51.6	*
DIC-G	76	51.3	*
DIC-H	85	50.3	*
DIC-I	92	49.5	
DIC-J	99	48.6	*
DIC-K	115	47.0	
DIC-L	118	46.6	
DIC-M	123	46.1	
DIC-N	141	44.1	
DIC-O	144	43.8	*
DIC-P	165	41.6	
DIC-Q	182	39.6	
DIC-R	187	39.0	
DIC-S	189	38.8	
DIC-T	191	38.6	
DIC-U	197	37.9	

Table 3.2: **List of 21 DIC truncations.** The N-terminal residue and predicted MW after TEV cleavage is shown for each DIC construct. In the results column, *=truncations that expressed and purified well enough to see a clear peak using SEC.

argument as residues 1-189 are predicted to be largely disordered (Fig. 3.8), whereas the C-terminal WD40 domain is well structured. A coiled-coil prediction (Lupas *et al.* 1991) calculated a very high score for the first 25 residues, which are also thought to bind to Nip100 (the yeast homologue of mammalian p150[Glued]), a major dynactin component (McKenney *et al.* 2011). Having narrowed down the region of interest to the N-terminal 200 residues, I analysed all possible N-terminal truncations within the first 200 residues using PPCPred (Mizianty & Kurgan 2011) to see if any potential truncations were predicted to be particularly crystallisable.

Using the bioinformatics-based predictions in combination with the previously mapped binding sites for Nip100 and the DLCs, I designed a series of 21 C-terminal DIC truncations (3.2). In addition to simply eliminating the flexible N-terminus, I also placed truncations to eliminate the binding sites to Nip100 and/or the DLCs. The longest truncation (DIC-A) would begin with residue 24 and the shortest (DIC-U) would begin with residue 197. Once I had designed the truncations themselves, I modified the multiple cassette plasmid described in Section 2.2.6 so that they would each be expressed in the context of the full tail complex in yeast.

I grew a 1L culture of yeast expressing each of the 21 DIC truncation constructs and performed a small-scale purification on each one. Rather than perform SDS-PAGE immediately after the TEV cleavage step, we first loaded each sample on a
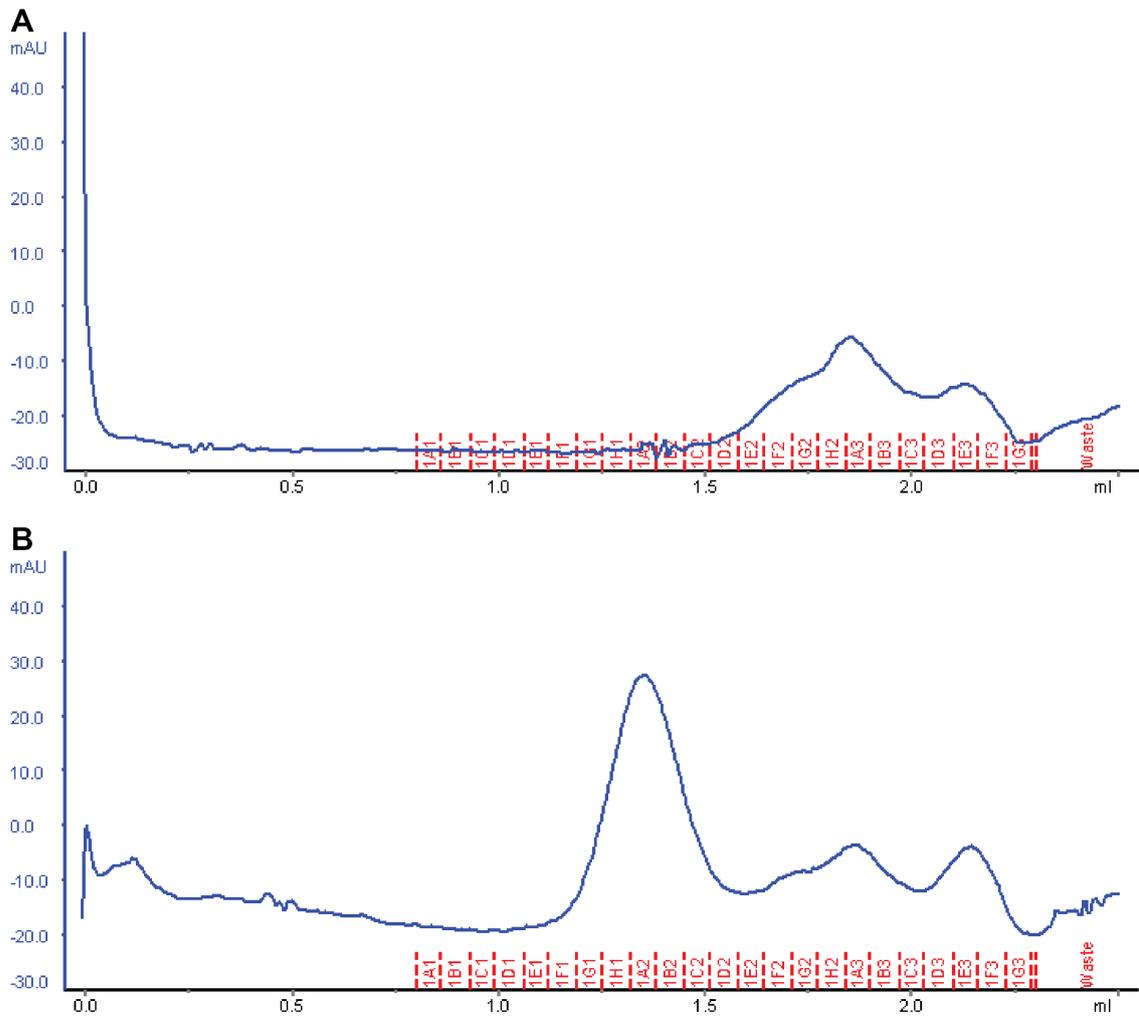
Figure 3.9: **Representative chromatograms of two DIC truncations. A)** The chromatogram for the DIC-A construct, which is an example of a truncation that failed to produce a peak. **B)** The chromatogramp for the DIC-C construct, which is an example of a truncation that produced a peak.

Superose6 column and performed SEC. Out of the 21 samples tested, 7 produced peaks in their chromatograms: DIC-C, DIC-D, DIC-F, DIC-G, DIC-H, DIC-J and DIC-O (Table 3.2, Fig. 3.9). We then concentrated the peak fraction from each of those 7 truncations and analysed using SDS-PAGE (Fig. 3.10). Truncations DIC-C and DIC-D, which are only three residues apart, appeared very similar. Aside from the slightly lower MW DIC, they looked identical to the standard tail complex. This result is in line with the fact that these two truncations only eliminate the N-terminal helix, which is thought to bind Nip100. Truncation DIC-F, eliminates the first of the two DLC binding sites as well, and just as one might predict, the amount of DLC seen on the gel is reduced significantly. Interestingly, truncation DIC-G, which is only two residues shorter, seems to lose all but a trace of DLC binding capacity despite including the entire second DLC binding site. Truncations DIC-H and DIC-J are visibly shorter and lack both DLC binding sites, and no DLC is visible on the gel whatsoever. DIC-O looks similar except the DIC is much smaller as it is 45 residues shorter than DIC-J.

From an analysis of the SEC chromatograms alone, all of the DIC truncations that produced peaks appeared to be stable, homogeneous samples. Furthermore, even the shortest of those truncations, DIC-O, eluted where a dimer would be expected. Although SEC-MALS analysis would provide more confidence in this observation, it seemed that even those complexes that lacked the DLCs completely formed stable dimers. The implications of this unexpected observation are discussed further in the discussion of both Chapters 3 and 4.
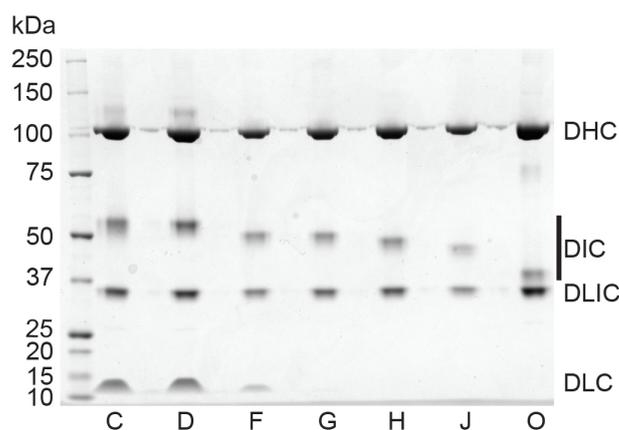
Figure 3.10: **SDS-PAGE of DIC truncations after SEC.** The seven DIC truncations which produced peaks in SEC were compared using SDS-PAGE. The bar on the right indicates the range over which the various DIC truncations appear. The DLC disappears as more of the DIC N-terminus is eliminated.

3.2.5   The interaction between dynein and dynactin

Rather than removing it completely, another potential way to stabilise the flexible N-terminus of the DIC is to bind it to one of dynein's binding partners. The N-terminal 25 residues of the yeast DIC were thought to bind Nip100 (McKenney *et al.* 2011), the largest subunit of the dynactin complex and yeast homologue of the mammalian p150[Glued]. So, I set out to purify yeast dynactin in order to bind it to and stabilise the dynein tail. This would have also been a more biologically

relevant solution than removing the DIC N-terminus since dynactin is known to be necessary for virtually all of dynein's cellular functions, but the mechanism by which dynactin binds dynein was largely unknown. Therefore, any biochemical or structural information on this interface may have been useful and interesting in itself.

To obtain dynactin, I used a previously described yeast strain that has a ZZ-tag fused onto the N-terminus of the Arp10 subunit by a linker containing a TEV cleavage site (Kardon *et al.* 2009). This allows endogenous dynactin to be immuno-precipitated and purified, which was demonstrated to be functional in motility assays (Kardon *et al.* 2009). I grew 24L of this strain and harvested when the OD_{600} was between 0.8 and 1.0. I performed a large-scale purification using the same buffers and protocol I use for dynein. After TEV cleavage, I concentrated the entire sample down to $\sim 20\mu L$ and analysed by SDS-PAGE (Fig. 3.11). There were prominent bands at the expected heights for Nip100, Arp1 and Arp10, as well as a lot of background on the gel. I cut the entire lane into 30 slices and had each sample analysed by MS. Each of the known subunits of yeast dynactin was positively identified at approximately the correct height on the gel for its size. Interestingly, some full-length DHC was also identified in this sample, indicating that some of the dynactin remained bound to dynein throughout the purification process.
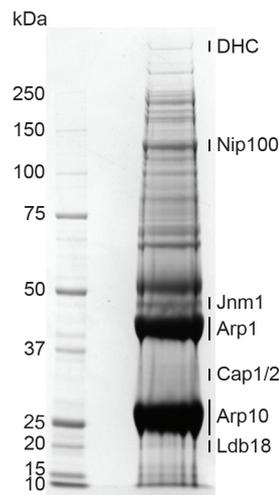
Figure 3.11: **Endogenous yeast dynactin sample used for MS analysis.** Endogenous dynactin was purified from 24L of yeast and analysed by SDS-PAGE. The entire gel lane was separated into 30 bands for MS analysis. Black bars to the right of the gel indicate the approximate width of the bands excised that contained each dynactin subunit. Even when bands were not visible by eye, the subunit was detected by MS.

While this seemed a promising way to produce yeast dynactin capable of binding dynein, the yield was far too low to perform any assays that could detect the interaction. One option would have been to insert *GAL1* promoters in front of each of the dynactin genes in this strain, theoretically increasing the yield in the same way the dynein yield was originally increased. Unfortunately, with seven known dynactin genes and the possibility that there are others that remain unknown, this seemed both time-consuming and less than likely to be successful. However, right around this time a study was published that described a recombinant complex formed be-

tween small fragments of rat DIC and the rat homologue of Nip100, which is called p150 (Siglin *et al.* 2013). Since this interaction was described as robust and of high enough affinity to be studied *in vitro*, I decided to make a yeast equivalent of the p150 fragment.

```
                          *
Rat     404   VRQQRERLQEELSQAESTIDELKEQVDAALGAEEMVEMLTDRNLNLEEKVRELRETVGDL   463
Yeast   249   -----------LQQTNNTIGDLALQIEAQSKSSNIVDKLTNDNILLTSNIKALNNELEEL   297


Rat     464   EAMNEMNDELQENARETELELREQL-----------DMAGARVREAQKRVEAAQETV   509
Yeast   298   QAKEKLDENLRITYEQLEQELRLQLSNLQSALENEKEIAGTYIEE-NSRLKATLESI   353
```

Figure 3.12: **Alignment of rat p150 and yeast Nip100.** A somewhat conserved region of p150 from *R. norvegicus* and Nip100 from *S. cerevisiae* was aligned using Blast. Red=conserved residues (26% within this region), yellow=similar residues (also 26%), *=indicates the conserved N-terminus of a fragment of p150 that binds DIC.

I performed an alignment between rat p150 and yeast Nip100 in order to ensure that I could make an equivalent construct to the fragment used by Siglin *et al.* 2013 (residues 415-530 of p150). Overall, the fragments contained 52% similarity, and 26% of the sequence which was fully conserved. The N-terminus of this fragment aligned well to the yeast sequence and it was easy to identify residue 249 of Nip100 as the equivalent of residue 415 of p150 (Fig. 3.12). However, the C-terminus of the fragment was not as clear, so I made five different fragments, all beginning at residue 249, but each ending at different points near residue 343 since this was my best guess as the equivalent of residue 530 in p150 (Table 3.3).

Name	Residues	MW (kDa)	Concentration (mg/mL)
Nip100-A	249-323	8.7	6.6
Nip100-B	249-336	10.2	29.1
Nip100-C	249-343	11.0	4.9
Nip100-D	249-352	12.0	9.7
Nip100-E	249-355	14.6	45.7

Table 3.3: **List of five Nip100 fragments.** The range of residues, predicted MW after TEV cleavage and final concentration after purification is shown for each Nip100 fragment.

Since Siglin *et al.* 2013 successfully expressed their p150 fragment in *E. coli*, I cloned my five Nip100 fragments into a bacterial expression vector. I grew 1L of each construct and purified all five using affinity chromatography followed by SEC. After SEC I pooled each peak and concentrated all samples to a volume between 1mL and 2mL. The final concentration of each construct is listed in Table 3.3. Meanwhile, I performed a standard dynein tail purification from 1L of Sf9 cells up until the point where the tail was bound to the 2mL of IgG beads in TEV buffer. I divided the 2mL of dynein-bound beads into 6 tubes, one for each Nip100 construct and incubated with ~ 1mg of Nip100 fragment. I also did the same with IgG beads that had been washed into TEV buffer, but never had dynein tail bound as a control for non-specific binding to the sepharose resin. With all twelve
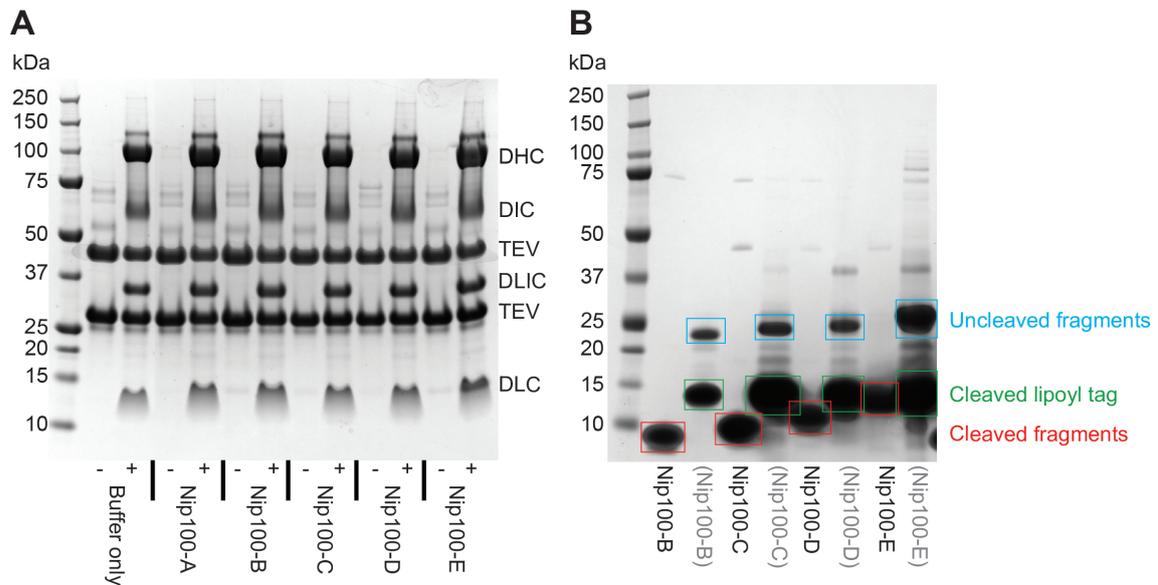
Figure 3.13: **SDS-PAGE analysis shows that five Nip100 fragments fail to bind to the dynein tail complex. A)** Five different Nip100 fragments and a buffer control were incubated with bare IgG beads (-) or IgG beads that already had dynein tail bound (+). After washing the beads, TEV cleavage and concentration, the samples were analysed by SDS-PAGE. None of the five Nip100 fragments co-purified with the dynein tail complex. The two prominent bands in every lane are self-cleaving TEV protease. **B)** SDS-PAGE analysis performed during the purification of the Nip1000 fragments. The lanes labelled in black are the final products of Nip100 B-E (A was not analysed). The lanes with grey labels show the uncleaved protein and cleaved lipoyl tag that bound to the Ni-NTA column after TEV cleavage. This was eluted and analysed on the gel as well to compare the ratio of cleaved to uncleaved protein. In the red boxes, Nip100B-E have purified and run at the expected heights on the gel. Therefore, the failure to see them in **A** is not due to a failure to purify the fragments, but rather a failure for them to bind the dynein tail.

samples, I washed the beads to remove unbound protein and then performed TEV cleavage, concentrated the samples and analysed by SDS-PAGE. The gel shows that the dynein tail purifications were successful, but none of the Nip100 fragments were pulled down by the dynein tail or the bare IgG beads (Fig. 3.13).

In conclusion, I was not able to reproduce the results reported by Siglin *et al.* 2013 in the context of the entire dynein tail complex. There are many reasons why this might have been the case, including a failure to reproduce the exact constructs, too low of concentrations to observe a stable complex, inherent differences between yeast dynein/dynactin and mammalian dynein/dynactin, and steric clashes resulting from the presence of the entire dynein tail complex. These possibilities are explored in more detail in the Chapter 3 discussion.

3.2.6 Phosphorylation of the dynein tail

Each time I attempted a new strategy for the expression of the dynein complex, I verified the presence of all four dynein subunits by MS. There were often peptides with modifications that are typically introduced during the MS analysis, for example

carbamylation, acetylation and methylation. I never observed any of these modifications consistently across different samples, and so I dismissed them as artifacts. However, every construct that I expressed in yeast produced the same DLIC peptide showing phosphorylation of residue S278. Since in humans, the C-terminus of the DLIC is thought to have phosphorylation sites (Dell *et al.* 2000), I set out to test whether this might be a relevant phosphorylation site in the yeast dynein tail. My goals were to confirm a previously unknown phosphorylation site in the yeast tail and to see whether the phosphorylation state of this site would have an effect on the stability of the behaviour of the complex *in vitro*.

I performed two small-scale purifications of the dynein tail complex from yeast grown in the 120L fermentor. The yeast used for this experiment expressed the dynein tail using the genomic expression strategy (Section 2.2.2). One purification was performed as usual. For the other purification, I included Halt Phosphatase Inhibitor Cocktail (Thermo Fisher Scientific) in the lysis buffer to prevent the dephosphorylation of cytoplasmic dynein by phosphatases released from other cellular compartments during and after cell disruption. After the lysis step, both purifications were carried out as usual, without Halt included in any other buffers for either sample. After TEV cleavage and concentration, the +Halt sample was $\sim$ 1mg/mL and the -Halt sample was $\sim$ 1.5mg/mL. Next, I incubated equal amounts of each sample with Lambda Protein Phosphatase (λPP, New England BioLabs) or ddH$_2$O to test whether any native phosphorylation that remained could be removed. I then analysed all four samples using SDS-PAGE. Phosphorylated proteins were imaged using Pro-Q Diamond Phosphorylation Gel Stain and total protein was imaged using SYPRO Ruby (both Thermo Fisher Scientific).

The total protein stain revealed that all four samples contained the full dynein complex and that an equal amount was loaded in each lane of the gel (Fig. 3.14A). The phosphorylation stain revealed that the Halt cocktail had no effect, since each of the bands that showed up in + Halt lanes, showed up with an equal intensity in the - Halt lanes (Fig. 3.14B). However, there were clear differences between samples treated with λPP and those that were only treated with ddH$_2$O. The most obvious difference was in the DLIC, which was not unexpected as this is the subunit that contained the phosphorylated residue in the MS results. The DLIC is clearly phosphorylated in the water control samples, and the amount of phosphorylation is markedly decreased in the samples treated with λPP. Although it is less obvious, the DHC also appears to have some amount of phosphorylation in the water control samples, which is mostly removed by λPP treatment. The DIC and DLC do not appear to be phosphorylated at all in any of the samples.

As the DHC showed some evidence of being phosphorylated, I looked back through all my old MS data, but could not find any evidence for phosphorylation sites within the DHC. Since the DHC bands in the phosphorylation stained gel are much weaker than the DLIC bands, it is possible that only a small population of dynein is phosphorylated on the DHC, and the MS that was performed simply for identification of proteins was not sensitive enough to detect low levels of phosphorylation. However, I did find a previous high-throughput study that identified thousands of phosphorylation sites within the yeast proteome, including three phosphorylation sites in the DHC tail region: S620, T623 and S890 (Holt *et al.* 2009). In order to confirm the presence of these specific phosphorylation sites, as well as the S278 site within the DLIC, I created 12 phosphomutant constructs. I mutated each

of the four potential phosphorylation sites to alanine (to eliminate phosphorylation completely) as well as the phosphomimetic amino acids aspartate and glutamate (Table 3.4).

The mutations were introduced in the same background in which the DLIC and DIC truncations were made (Sections 3.2.3 and 3.2.4). I verified the sequence and grew a 1L culture of each. Due to time constraints I was not yet able to perform a small-scale purification of these constructs, but they remain in the Carter lab plasmid database and may be of interest for future studies. For example, if a structure can be determined of the entire dynein tail, it may be interesting to compare the phosphomutants to test whether they have an observable effect on the overall conformation. Furthermore, if it becomes possible to reproducibly bind dynein to various binding partners (*e.g.*, dynactin, Num1, Pac1) it would be interesting to test whether any of these phospho-sites act as a regulatory switch that prevents or allows this binding to occur.

3.3 Discussion

In this chapter, I describe several different approaches to improving the stability and solubility of the dynein tail complex. While some were inconclusive, and others showed no clear improvement, the experiments were included because they each contribute either a little more knowledge about dynein or describe constructs that I created that may be useful to others in the future. Some of the information in this chapter also becomes more interesting in light of the results I describe in Chapter 4.

3.3.1 The DIC/DLC subcomplex

Perhaps the most intriguing results in this chapter came from the series of DIC truncations. The fact that several of these constructs were expressed and behaved well in solution is interesting in itself. I successfully purified dynein complexes that lack what was thought to be the primary binding site for the dynactin complex. This makes possible many interesting experiments concerning the interaction between dynein and its most important regulator.

By removing even more of the DIC N-terminus, I was also able to purify dynein complexes that have only one of the two DLCs bound, as well as complexes that lack the DLCs completely. This is relevant for several reasons. First, the fact that the intensity of the single DLC band in Fig. 3.10 is halved when the first DLC binding site is removed corroborates the results of the Tda2 experiment and supports the model that both DLCs are the same protein. Furthermore, a study was published shortly after I completed this experiment, which presented a crystal structure of a fragment of the yeast DIC bound to Dyn2 (Rao *et al.* 2013). However, the DIC peptide used in the crystal structure corresponds to the C-terminal DLC binding site (the one that binds LC8 in mammalian dynein). An equivalent peptide for the N-terminal DLC binding site (the one that binds Tctex-1 in mammalian dynein) was designed and synthesised, but proved insoluble due to its hydrophobicity (Rao *et al.* 2013). Therefore, even after this study, there is no conclusive evidence that there are two copies of the same DLC (Dyn2) bound to each DIC in yeast dynein.

The Rao *et al.* 2013 paper also argues that deletion of the DLCs inhibits homodimerisation of the dynein complex and results in a slower motor capable only
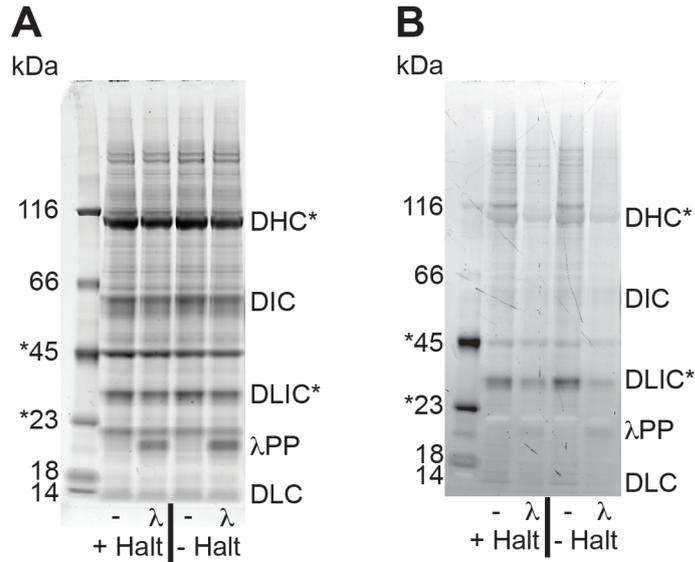
Figure 3.14: **Phosphorylation of the dynein tail complex expressed in *S. cerevisiae*.** Two small scale purifications of the dynein tail complex expressed in yeast were performed, one with Halt Phosphatase Inhibitor Cocktail in the lysis buffer and the other without. One sample from each purification was incubated with buffer (-) and the other was incubated with λ protein phosphatase (λ). All samples were analysed by SDS-PAGE. * indicates phosphoproteins. **A)** The gel is stained with SYPRO Ruby to reveal total protein in the sample. This gel showed that an equal amount of tail complex had been loaded into each lane. **B)** The same gel is stained with Pro-Q Diamond Phosphorylation Gel Stain to reveal phosphorylated proteins only. This suggested that DLIC is phosphorylated, and that this phosphorylation can be partially removed by λ treatment. The DHC may also show a low level of phosphorylation, which is also decreased after λ treatment, but it is less clear.

Subunit	Mutation
DLIC	S278A
DLIC	S278D
DLIC	S278E
DHC	S620A
DHC	S620D
DHC	S620E
DHC	T623A
DHC	T623D
DHC	T623E
DHC	S890A
DHC	S890D
DHC	S890E

Table 3.4: **List of 12 dynein tail phosphomutants.** The subunit and specific point mutation are listed for each phosphomutant.

of short run lengths. I did not subject any of my constructs to *in vitro* motility assays. However, the SEC chromatograms show that the elution volumes of DIC-C (which retains both DLC sites), DIC-F (which lacks the N-terminal DLC site) and DIC-G (which lacks both DLC sites) are very similar, suggesting that all of these constructs homodimerise. So, while I cannot say whether the truncations that lack the DLCs would have similar motile deficiencies, it seems unlikely that the effects described by Rao *et al.* 2013 were due to a failure to homodimerise. One key difference between their study and mine is that they use full-length DIC and delete the DLC gene, while I have actually removed the region of the DIC that binds the DLCs altogether. Therefore, it is possible that the motility defects they observe are caused by the disordered DIC N-terminus without the stabilising influence of the DLCs.

The observation that dynein complexes completely lacking DLCs still dimerise is unexpected as the consensus within the dynein field for the past several years has been that the primary role of the DLCs is to drive homodimerisation of the dynein motor (Barbar 2008). While it remains possible that the DLCs play some role in this process, my results suggest that they are not essential for homodimerisation. Other experiments I performed, which I describe in Chapter 4, provide further evidence that the DLCs are not the only part of the dynein complex driving homodimerisation.

3.3.2 The dynein/dynactin supercomplex

There are several possible explanations for why my attempt to bind Nip100 fragments to the dynein tail failed. The simplest is simply that none of the five fragments contained the DIC binding site used by Siglin *et al.* 2013, but this seems unlikely as the alignment between the rat p150 fragment and yeast Nip100 showed 52% similarity and 26% identity. It could also have been that the 10:1 ratio of Nip100 fragment:dynein tail that I used was not enough to reach the concentration necessary for binding. Finally, the fragments may have bound initially only to be removed by the subsequent washes due to a weak or transient interaction. However, the fact that some endogenous DHC co-purified with the dynactin complex described above (Fig. 3.11) suggests that the interaction between these two complexes is strong enough to survive several washes with TEV buffer.

So, how is it possible that the two fragments bind *in vitro*, but the Nip100 fragment doesn't bind the whole dynein complex? It could be that a DIC fragment would have bound the Nip100 fragment, but in the context of the whole dynein complex the DIC is sterically hindered from interacting unless some other interaction with dynactin occurs first, *i.e.* the minimal fragments can bind each other, as can the full complexes, but the Nip100 fragment cannot bind the full dynein complex. At this point I felt that any further studies investigating the interaction between dynein and dynactin would best be done using the full dynactin complex. Since I was not able to produce this in sufficient amounts, I set this project to the side until a better way of producing dynactin was available.

Incidentally, some colleagues in the Carter lab successfully purified large quantities of endogenous dynactin from pig brains, which they used to make a super processive dynein motor (Schlager *et al.* 2014). They also showed that recombinantly expressed human dynein is not processive, neither on its own, nor after the

addition of pig brain dynactin. The complex only formed upon the addition of a third component, a fragment of a cargo adaptor called BICD2. This was quite surprising given the model that had been established in yeast a few years earlier. A slightly older study showed that endogenous yeast dynein is processive on its own, but the addition of endogenous dynactin increases the processivity roughly two-fold (Kardon *et al.* 2009). These two results seem to show a clear difference between how the dynein/dynactin supercomplex forms in yeast and mammals. However, the fact that the yeast components were all purified from endogenous levels of yeast means that it is possible some third component, like a yeast homologue of BICD2 was co-purified with either the dynein and dynactin. Therefore the mechanism for how yeast dynein/dynactin is formed, and if it is fundamentally different from the equivalent mechanism in mammals remains unknown. As a result of further work on the triple complex formed by pig brain dynactin, recombinant human dynein and BICD2, we now have a clear picture of how this complex comes together (Urnavicius *et al.* 2015). This will be discussed in more detail in Chapter 4.

3.3.3 Future directions

A source of the whole dynactin complex

As the primary obstacle in the way of studying the dynein/dynactin complex was the limiting amount of dynactin, I would consider producing the full yeast dynactin complex a priority. Although it has more subunits and a much more skewed stoichiometry than dynein due to the presence of an actin-like filament, I would first try to apply the same strategies I successfully used for expressing dynein in Chapter 2 to the expression of dynactin. If those methods are unsuccessful, it may also be possible to develop a higher-throughput protocol for purifying endogenous dynactin from large amounts of yeast. Since yeast is cheap to grow and the IgG beads used for affinity purification are the most expensive aspect of purifying recombinant dynein, it might be possible to use an adapted MT-affinity purification, similar to those commonly used for purifying dynein and dynactin from mammalian tissue. As I mentioned above, some of my colleagues have already established a protocol for purifying dynactin from pig brains. So as a last resort, I might test whether pig dynactin can bind to yeast dynein and see if this helps elucidate any fundamental differences between the mammalian and yeast dynein/dynactin complexes.

Biochemical and structural characterisation of the DIC truncations

Having only purified each of the DIC truncations once, I would scale up the seven constructs that produced SEC peaks and characterise them further. The first thing I would do is perform SEC-MALS to get a more accurate estimate of the MW of each construct. This would confirm that the constructs that lack the DLCs are still homodimerising. I would also try to prepare negative stain grids, at least with the shortest of the constructs that expressed, DIC-O. If I could collect enough images to create a 3D reconstruction similar to the one I presented in Section 2.2.8, I might be able to compare the two structures and locate the DIC N-terminus and DLCs within the dynein tail. Furthermore, in light of the interplay between my data and the data reported by Rao *et al.* 2013, I would also be interested in making these truncations in the background of the full-length dynein so that I could perform

motility assays with them. Finally, if the DLIC truncation series was tested again, and some truncations appeared useful for eliminating disordered regions, I would make a construct that combines the best of the DIC truncations with the best of the DLIC truncations and screen it for crystallisation.

3.3.4 Conclusions

In this chapter, I successfully identified several DIC truncations that are useful due to their ability to selectively prevent the DIC from binding to dynactin, or one or both DLCs. These truncations are important because they shed new light on the role of DLCs within the dynein complex by providing evidence that the complex can homodimerise even in the complete absence of DLCs. They may also be useful in further dissecting the role of the DIC/Nip100 interface in forming the dynein/dynactin supercomplex. I also provided evidence for at least one phosphorylation site within the yeast DLIC at residue S278. At this stage, I began to doubt that the dynein tail complex was a good crystallography target. So I tried to crystallise one of the other original DHC truncations, DHC-E, which expressed at a very high level (Section 2.2.3). This is the topic of Chapter 4.

Chapter 4

The crystal structure of the dynein heavy chain N-terminus

4.1 Introduction

While attempting to make the full dynein tail complex into a suitable crystallography candidate, I decided to revisit one of the other DHC truncations described in Chapter 2. This construct, DHC-E, contained only the N-terminal 557 of the DHC, did not bind any of the accessory subunits and expressed at very high levels. I selected this construct as a potential crystallography target and ultimately solved its structure. This chapter describes the process of solving this crystal structure, the first of any part of the DHC tail region.

The crystals from initial screening were very small and did not diffract to high resolution, but through optimisation I was able to increase the crystal size and diffraction power. Since no homology models existed, I screened heavy atoms for experimental phasing, and ultimately got initial phases from a selenomethionine derivative. From the initial phases, I obtained an electron density map good enough to build an initial model, which I used to design new constructs in an effort to further increase the resolution.

This crystallography project was done in collaboration with Minmin Yu. All of the construct design, protein expression, protein purification, crystallisation and model building were my own work. The crystal harvesting, data collection and data processing were done with Minmin's guidance.

The crystal structure of DHC-E has been deposited in the PDB with the accession code 5AFR in connection to the following article:

Urnavicius, L.*, Zhang, K.*, Diamant, A. G.*, Motz, C., Schlager, M. A., Yu, M., Patel, N. A., Robinson, C. V. and Carter, A. P. The structure of the dynactin complex and its interaction with dynein. *Science* **347**, 1441-1446 (2015).

Much of the work presented in Section 4.2.8 was carried out by my colleagues, who were co-authors of the above article. In particular, Linas Urnavicius and Kai Zhang performed the EM work I describe. My contribution was essential to the paper as it allowed for the interpretation of a cryo-EM structure that was otherwise difficult to interpret. I also include figures adapted from those that Linas and Kai generated, which I indicated by citing Urnavicius *et al.* 2015 in the figure legends.

4.2 Results

4.2.1 Purification of DHC-E from Sf9 cells

At the time I decided to revisit expression of the DHC-E construct, I had switched from yeast to baculovirus expression due to the large increase in expression. So I had the DHC-E truncation of the yeast DYN1 gene codon optimised for Sf9 cells and synthesised, and I cloned it into a baculovirus. When generating P2 virus, I harvested the cell pellet and analysed the crude lysate by SDS-PAGE. The gel showed a prominent band at $\sim$ 75kDa, which is close to the predicted MW of DHC-E when when still fused to its ZZ-tag (Fig. 4.1A). The 75kDa band was much larger and more intense than any other bands in the lysate of uninfected Sf9 cells.
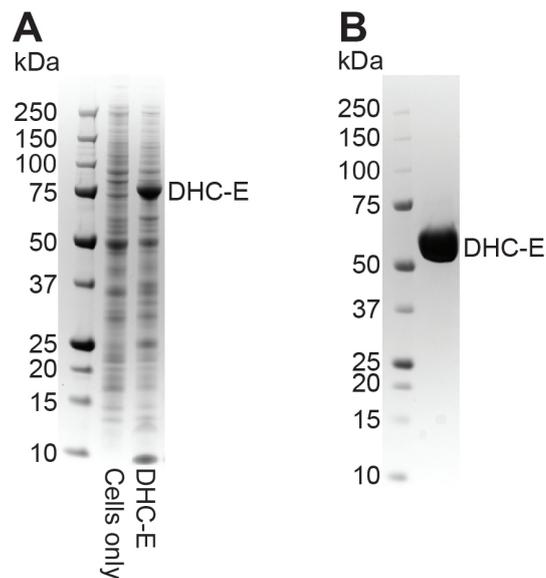
Figure 4.1: **SDS-PAGE analysis of DHC-E expressed in and purified from Sf9 cells.** DHC-E was cloned into a baculovirus and expressed in Sf9, then purified by affinity chromatography and SEC. **A)** A crude lysate sample from the p2 generation, which shows a high level of DHC-E expression above that of all background proteins in a sample of uninfected Sf9 cells. **B)** The purified DHC-E after affinity chromatography, TEV cleavage and SEC. The difference in MW between A and B is accounted for by the ZZ-tag, which is removed before SEC.

After this indication of high expression levels, I grew 8L of Sf9 cells infected with the DHC-E baculovirus. The DHC-E protein was easily purified, requiring only affinity chromatography (with the ZZ-tag), TEV cleavage and SEC to produce a large amount of highly pure sample (Fig. 4.1B). From 8L of culture, there was 55mg of total protein after TEV cleavage. One caveat is that despite being easy to purify, only $\sim$ 10% of the protein was usable after SEC; the other 90% formed either aggregates or higher-order structures and eluted far earlier than a dimer of DHC-E would be expected to (Fig. 4.2A). Nevertheless, I pooled and concentrated the smaller, symmetrical peak that did elute where a dimer was expected, and then ran it back through the column. There was $\sim$ 5mg of dimer after the first run, and I was able to recover nearly 100% of it after the second run (Fig. 4.2B). This demonstrated that the dimer form of DHC-E is stable in solution and that it is not

A

B
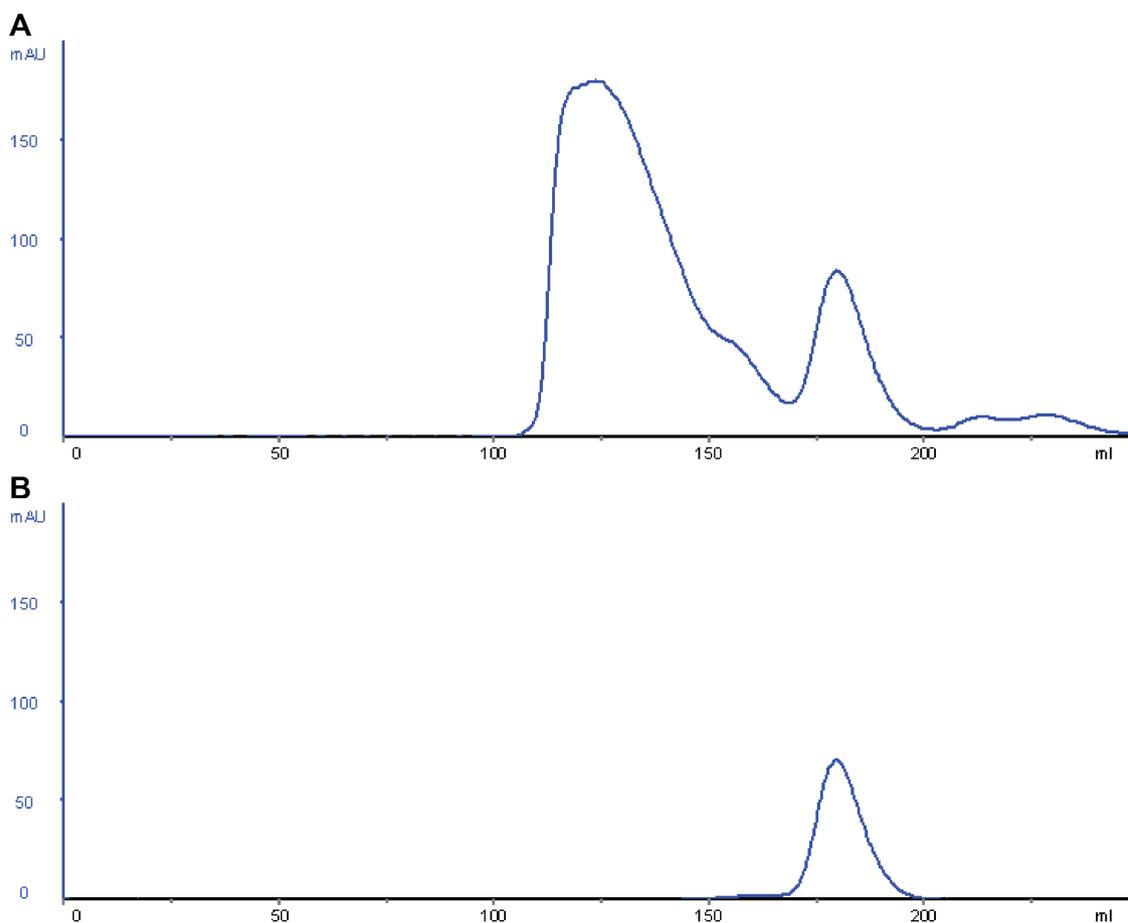
Figure 4.2: **SEC of DHC-E purified from 8L of Sf9 cells.** DHC-E was purified from 8L of Sf9 cells and loaded on a Superdex200 for two consecutive runs. **A)** Immediately after TEV cleavage, there was $\sim$ 55mg of protein. After SEC, the majority of it eluted sa higher-order structures or aggregates. A small peak eluted at $\sim$ 180mL, the expected volume for a dimer of DHC-E. **B)** The dimer peak from A was pooled, concentrated and loaded on the same column under the same conditions. There was only $\sim$ 5 mg of protein left, but the second run had a very high recovery rate and all 5mg eluted as a dimer again. There was no sign of aggregation or higher-order structures forming.

in equilibrium with the higher-order species, which most likely form in the cells or an earlier stage of purification. I concentrated the final product down to 5.3mg/mL and checked the purity by SDS-PAGE (Fig. 4.1B). The final purity and yield were high enough that I proceeded to set up some initial crystal screens using this sample.

4.2.2   Initial crystal screens

The MRC-LMB has a dedicated X-ray crystallography suite, which provides more than 20 distinct screens for protein crystals. These screens are assembled from commercial screens produced by Hampton, Emerald, Jena Bioscience and Molecular Dimensions. There is a certain amount of overlap between conditions, with several appearing in more than one LMB screen. However, if all 20 LMB screens are set, this inherent redundancy can be advantageous as comparing crystals produced in iden-

Screen/Well	Buffer (pH)	Precipitant(s)	Morphology	Diffraction (Å)
LMB3/A10	0.1M HEPES (7.0)	1.6M $(NH_4)_2SO_4$	fine needles	none
LMB3/B3	0.1M MES (6.0)	2.4M $(NH_4)_2SO_4$	fine needles	37
LMB3/C12	0.1M Bicine (9.0)	1M LiCl, 10% PEG (MW=6,000)	fine needles	9.7
LMB5/G4	0.05M HEPES (7.0)	1.6M Li_2SO_4, 0.05M $MgSO_4$	thick needles	8.0
LMB9/A9	0.1M Tris-HCl (8.5)	0.2M Li_2SO_4, 18% PEG (MW=8,000)	small nuggets	7.9

Table 4.1: **Initial DHC-E crystal hits.** DHC-E was purified from 8L of Sf9 cells and concentrated to 5.3mg/mL. Several 96-well plate LMB crystal screens were set with 200nL of protein and 200nL of reservoir per droplet. The crystal screens were set and incubated at 4°C. After 24 hours, the five conditions listed in the table produced crystal hits. Some crystals from each condition were tested for diffraction using synchrotron radiation. The conditions, crystal morphology and diffraction resolution limit are listed for each of the initial crystal hits.

tical or similar conditions can often provide useful insights about which conditions are truly favourable for a particular protein to crystallise.

Using the DHC-E sample purified from 8L of Sf9 cells, I was able to set up several of these LMB screens. I used a Mosquito robot (TTP Labtech) at 4°C to dispense 200nL of protein and 200nL of reservoir solution for each drop. Then I sealed the plates and incubated at 4°C. Within 24 hours, there were already at least three different conditions with crystal growth (Table 4.1). The crystals grew over a wide range of pH values, from 6.0 to 9.0. Most of the conditions had a high salt concentration as the precipitant, but a couple also used PEG. I allowed the crystals to grow for several more days to ensure they had equilibrated fully and to give the crystals time to reach their maximum size. Then I harvested some crystals from each condition, washed them into a cryo-solution and flash-cooled them in LN_2. I screened the crystals for diffraction using beamline I03 at Diamond Light Source. The best crystals diffracted to a maximum resolution of 7.9Å (Table 4.1). In conclusion, even from the very first large-scale purification of the DHC-E construct, I was able to produce diffracting protein crystals. As there were so many variables in the expression, purification and crystallisation of DHC-E left to optimise, I focused my attention and efforts toward this project and began by attempting to optimise the expression.

4.2.3   The expression of DHC-E from *E. coli*

Although Sf9 cells produced much more protein than yeast, the fact that only 10% of the DHC-E construct was suitable for crystallisation trials meant that producing enough sample for crystal optimisation would have been quite labour-intensive and time-consuming. I was already familiar with a modified pRSET(A) plasmid (Dodd *et al.* 2004) that I used to express fragments of Nip100 and other proteins in *E. coli* (Section 3.2.5). This plasmid was capable of producing as much as 50mg of protein per litre of culture for some constructs. Furthermore, the DHC-E construct contained only a single subunit, which is 557aa in length, so it was well within the range of what *E. coli* are typically capable of producing. Therefore, I cloned the DHC-E fragment into this bacterial expression vector and attempted to express it in *E. coli*.

I purified DHC-E from 6L of *E. coli* using affinity chromatography and SEC. From 6L of *E. coli*, I was able to purify nearly 50mg of pure DHC-E, whereas from 8L of Sf9 cells, the yield was only 5mg. The expression of DHC-E in *E. coli* avoided the aggregation problem described in Section 4.2.1. Figs. 4.3A and 4.3B provide a direct comparison of the two expression systems. The total amount of protein present in the Sf9 sample (4.3A) is not significantly less than the protein present in the *E. coli* sample (4.3B). However, the distribution shifted dramatically so that there is only a negligible aggregate peak in the *E. coli* sample, with the majority of the protein eluting as a dimer. This accounts for the dramatic increase in the final yield of dimeric DHC-E from *E. coli*. Figs. 4.3C and 4.2B provide a direct comparison of the final products after the second SEC run (note the difference in scale). One possible explanation for the aggregate formation occurring in Sf9 cells and not in bacteria is that the Sf9 cells may have been putting on unnecessary post-translational modifications that interfere with protein stability.

With a glut of DHC-E on hand, I not only set the same LMB screens that produced the initial hits, but all 20 of the LMB screens in duplicate. The first set was under the same conditions as the initial hits: I used a protein concentration of 5.3mg/mL, 200nL of protein and 200nL of reservoir per drop, and the plates were set at 4°C. The second set was under the same conditions except at room temperature (RT, $\sim$ 20°C). After 24 hours, 58 of the 1920 conditions screened at RT and 11 of the 1920 conditions screened at 4°C already had signs of crystalline growth. The majority of the RT hits and all of the 4°C hits were clusters of fine needles. However, four of the RT hits produced more three-dimensional, nugget-shaped crystals that looked similar to the LMB9/A9 initial hits that diffracted to 7.9Å (Table 4.1). After one week, there were no visible changes, indicating that these droplets reached equilibrium within the first 24 hours.

The inherent redundancy within the 20 LMB screens was useful in this case because the four most promising hits shared certain characteristics, which I was able to use for crystal optimisation (Section 4.2.4). The most striking similarity between these conditions was that they all contained 10-20% PEG (MW=8,000). Many of them also had a salt present at a concentration of 0.2M, although the specific salt varied from condition to condition. Finally, the formation of crystals did not appear to be very sensitive to the pH of the reservoir solution as two of the conditions contained no buffer at all. However, the protein solution itself contained 20mM Tris-HCl (pH=8.0), which may have still buffered the final concentration to a pH near 8.0. Taking this into account and looking at the buffers present in the other conditions that produced similar crystals suggested that crystal formation occured anywhere between a pH of 8.0 and 10.5.

Overall, not only did the expression of DHC-E in *E. coli* provide a ten-fold increase in yield while simultaneously reducing the amount of time it takes to express the construct (24 hours instead of 72), but it unexpectedly solved the aggregation problem observed when using the baculovirus system. Having optimised both the expression and purification in one step, I was now able to make large amounts of DHC-E with ease. By reproducing the initial crystal hits that formed from DHC-E expressed in Sf9 cells, I showed that the DHC-E expressed in *E. coli* behaves similarly. This provided me with the material necessary to attempt the dozens of different techniques I employed for crystal optimisation. Hereafter, unless otherwise noted, all DHC-E samples were purified from *E. coli*.
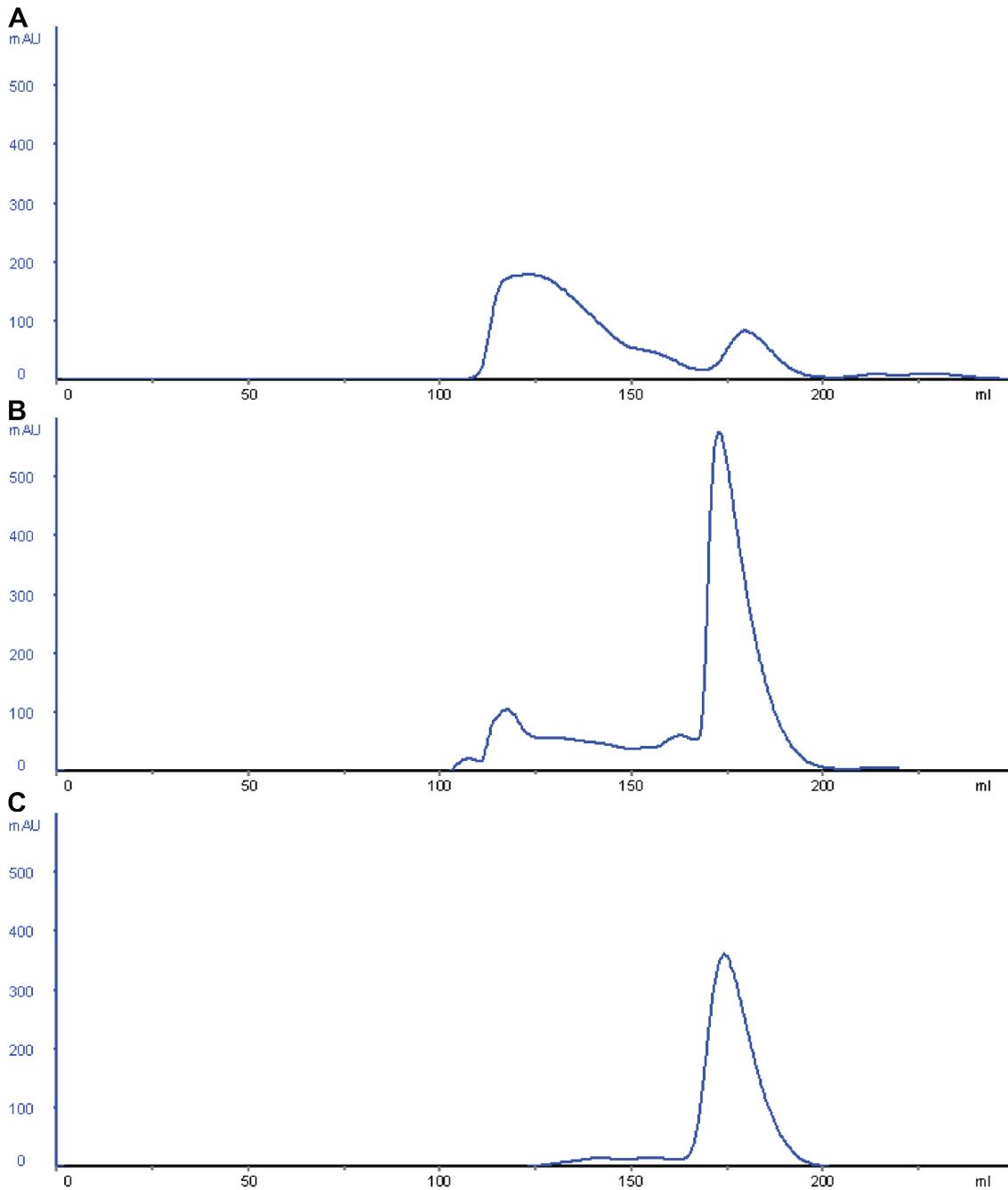
Figure 4.3: **A comparison of DHC-E purified from 8L of Sf9 cells vs. 6L of _E. coli_.** DHC-E was purified from 8L of Sf9 cells and 6L of _E. coli_, then each was loaded on the same Superdex200 column. **A)** The same chromatogram from Fig. 4.2A re-scaled for a direct comparison with B. **B)** 6L of _E. coli_ produced much more DHC-E total. Also, the percentage of total protein that remained in the dimer form was much higher. **C)** The dimer peak from B was pooled, concentrated and run over the column again. There is much more protein in this peak than the one in Fig. 4.2B.

Screen/Well	Buffer (pH)	Precipitant(s)
LMB2/H3	0.1M CAPS (10.5)	0.2M NaCl, 20% PEG (MW=8,000)
LMB8/H10	none	0.2M 0.2M $Mg(CH_3CH_2)_2$, 10% PEG (MW=8,000)
LMB9/A6	none	0.05M $(NH_4)_2SO_4$, 0.1M Na Citrate, 15% PEG (MW=8,000)
LMB11/F2	0.1M Tris-HCl (8.5)	0.2M Li_2SO_4, 10% PEG (MW=8,000), 10% PEG (MW=1,000)

Table 4.2: **Initial crystal hits reproduced using DHC-E expressed in *E.coli.*** DHC-E was purified from 6L of *E. coli* and concentrated to 5.3mg/mL. 20 different 96-well plate LMB crystal screens were set in duplicate with 200nL of protein and 200nL of reservoir per droplet. Each LMB screen was set and incubated once at 4°C and once at 20°C. After 24 hours, 58 conditions produced crystals at 20°C and 11 produced crystals at 4°C. Of the 20°C hits, four conditions produced large, nugget-shaped crystals, similar in appearance to those that diffracted to 7.9Å. These four conditions are listed in the table, including the composition of each one. One ingredient that is common to all four conditions is 10-20% PEG (MW=8,000).

Initial condition	Across variable	Down variable	Buffer (pH)
LMB3/C12	0-2.0M LiCl	0-20% PEG (MW=6,000)	0.1M Bicine (9.0)
LMB5/G4	0-2.0M Li_2SO_4	0-0.5M $MgSO_4$	0.05M HEPES (7.0)
LMB9/A9	0-2.0M Li_2SO_4	0-30% PEG (MW=8,000)	0.1M Tris-HCl (8.5)

Table 4.3: **Optimisation matrices for three initial crystal conditions.** A 48-well plate optimisation matrix was designed for each of the three initial crystal hits that diffracted with sub-nanometer resolution. All three matrices were created using a Dragonfly robot so that the across variable would increase in even steps across the six columns and the down variable would increase in even steps across all eight rows. The buffer was held constant in all 48 conditions.

4.2.4 Crystal optimisation

48-well plate optimisation matrices

With large amounts of highly pure DHC-E finally available, I began a systematic attempt to improve the diffraction resolution limit of the crystals I had obtained in the initial screens. Of the three initial hits that were capable of sub-nanometer diffraction (Table 4.1), one had a high salt concentration (LMB5/G4), one had a high PEG concentration (LMB9/A9) and one had moderate amounts of both salt and PEG (LMB3/C12). Using a Dragonfly robot (TTP Labtech) I created 48-well plate optimisation screens from four corner solutions that centred on each of the three initial hits. Each 48-well plate had eight rows and six columns, so I varied the concentration of the salt across the columns and the concentration of the precipitant/second salt down the rows (Table 4.3). All 48 conditions for each plate had the buffer concentration and pH that was present in the initial hit condition being optimised since crystals were not pH sensitive (Table 4.3).

I set up the three optimisation matrices in duplicate, once at 4°C and once at RT, using a Dragonfly robot. Each droplet was set with 500nL of DHC-E (5.3mg/mL)
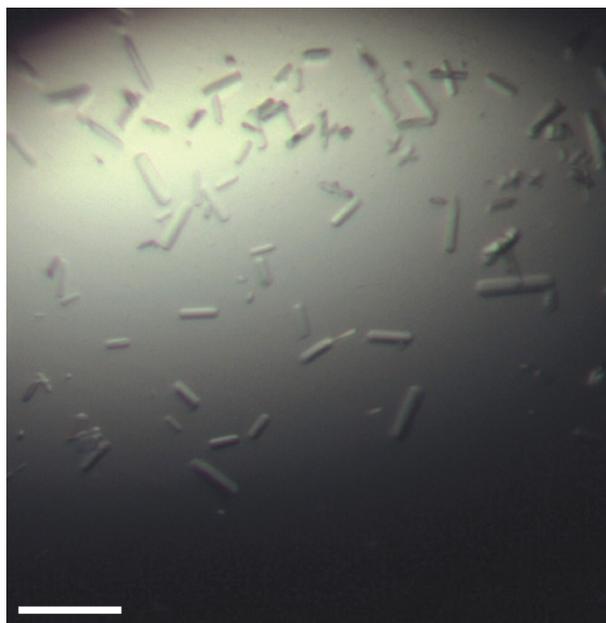
Figure 4.4: **Preliminary optimisation of DHC-E crystals.** DHC-E was purified from *E. coli* and concentrated to 5.3mg/mL. The sample was used to screen for crystals by adding 500nL protein to 500nL reservoir solution from several 48-well plate optimisiation matrices that were based on initial crystal hits. The crystals shown were grown in a condition that was similar to the one that produced many small crystals (0.1M Tris-HCl (pH = 8.5), 0.2M Li_2SO_4, 6.3% PEG (MW = 8,000, v/v)). When screened with synchrotron radiation, these crystals diffracted to 6.4Å. Measure bar = 100μm.

and 500nL of reservoir solution. The plates were checked for crystal growth after 24 hours and again after one week. The LMB3/C12 and LMB5/G4 screens only produced needles that were visually indistinguishable from the initial hits. A selection of needles from both screens were screened for diffraction using the I03 beamline at Diamond Light Source, but none of them diffracted beyond 8Å. The 4°C LMB9/A9 screen only produced nugget-shaped crystals that were similar in appearance to the initial crystal hit and failed to diffract beyond 8Å. However, the RT screen had a few conditions that produced crystals that were rod-shaped and more translucent in appearance than the initial nugget-shaped crystals (Fig. 4.4). When these crystals were screened using synchrotron radiation, they diffracted to 6.4Å.

The best diffracting crystals grew in a condition that was identical to LMB9/A9, except the concentration of PEG (MW=8,000) was reduced from 18% to 6.3% (0.1M Tris-HCl (pH = 8.5), 0.2M Li_2SO_4, 6.3% PEG (MW = 8,000)). Interestingly, the Li_2SO_4 concentration was already optimal, which is in agreement with the previous result that almost every hit in the initial screens had a salt concentration of 0.2M, regardless of the specific salt (Table 4.1). The PEG concentration however, needed to be much lower, which may have resulted in fewer nucleation events and favoured crystal growth instead. This would explain why crystals in this optimised condition were fewer in number, but larger in size than the initial hit ($\sim 50\mu$m in the longest direction compared to $\sim 10\mu$m in all directions). Despite being larger than the initial hit, the crystals produced in this condition were still less than 100μm in their longest direction, and less than $10~\mu$m in their shortest direction (Fig. 4.4). So, I aimed to

96

increase the size even further in hopes that this would also lead to another increase in diffraction power. From this point on, I focused on the LMB9/A9 condition and applied a variety of crystallisation techniques in order to decrease the number of crystals per droplet and increase their size.

Crystal seeding

Phase diagrams are a useful way to illustrate the process of protein crystallisation by vapour-diffusion and provide a useful background for the theory behind crystal seeding (Fig. 4.5). When a droplet is first set, if protein and/or precipitant concentrations are low enough, the sample exists in a soluble zone (S in Fig. 4.5) below the saturation point and the protein remains in solution. As the droplet equilibrates with the higher concentration of precipitant in the reservoir, the concentration of protein and precipitant both increase, bringing the sample across the saturation point and into the metastable zone (M in Fig. 4.5). In this zone, the addition of soluble protein onto existing crystals is favourable, but the more energetically costly nucleation of new crystals is not yet spontaneous. If the protein and precipitant concentrations increase further, the sample moves into the labile zone (L in Fig. 4.5), where nucleation of new crystals in spontaneous. Finally, if the protien and/or precipitant concentrations become too high, the sample moves into the precipitation zone (P in Fig. 4.5). Here, the protein becomes insoluble too quickly for the formation of ordered crystals and forms amorphous precipitate instead.

Therefore, a common approach for producing larger crystals when small crystals are readily available is to break up a mature crystal into very small crystal fragments that are still capable of growing (called seeds), and then transfer them into a new solution that is within the metastable zone (Asherie 2004, Fig. 4.5C). I already had the crystals necessary for making seeds, but had not identified a condition suitable for seeding into, *i.e.*, a condition in the metastable zone. So I set up screens using conditions similar to LMB9/A9, but with various PEG concentrations, in order to identify the threshold below which the drops would be in the metastable zone at equilibrium. Since I planned to use hanging-drop vapour diffusion for seeding, I identified the threshold using this format. Using 15-well EasyXtal hanging-drop plates (QIAGEN) I set up a screen where every condition had 0.1M Tris-HCl (pH = 8.5) and 0.2M Li_2SO_4 and the concentration of PEG (MW=8,000) ranged from 1.4% to 21% in 1.4% increments (Table 4.4).

Using this screen I set up three droplets for each condition, all on the same screw cap. I set up each droplet with 1μL of DHC-E (at 5.3mg/mL) and 1μL of reservoir solution. The plate was set and left to equilibrate at room temperature ($\sim$ 20°C) for 24 hours and then I examined each droplet for crystal growth and/or precipitate formation. The droplets remained clear for all PEG concentrations from 1.4% to 8.4%. Several clusters of crystals formed in the droplets that had between 9.8% and 12.6% PEG. PEG concentrations from 14% to 21% produced only amorphous precipitate (Table 4.4). The crystals in the hanging-drop format grew larger than those in the sitting-drop format. Therefore, from this point on I primarily used the hanging-drop format.

After identifying 8.4% PEG (MW=8,000) as a condition that was potentially in the metastable zone, I used a reservoir solution with this PEG concentration for two different seeding techniques. The first technique, called streak seeding, is based on the direct transfer of very small crystal fragments from a pulverised donor crystal to
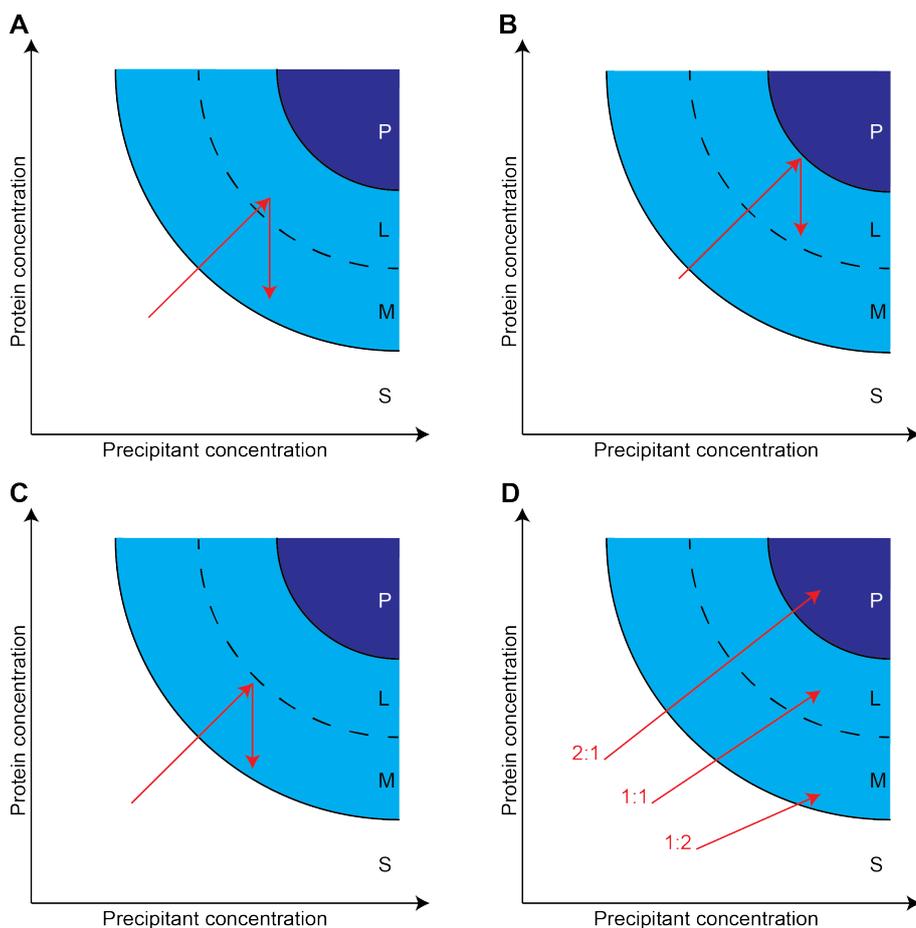
Figure 4.5: **Theoretical phase diagrams comparing two crystallisation scenarios.** S=soluble zone, where protein and/or precipitant concentrations are low enough that protein remains completely soluble and droplets appear clear. M=metastable zone, where protein is supersaturated enough that crystal growth is supported, but not nucleation of new crystals. L=labile zone, where protein is supersaturated enough to favour crystal nucleation. P=precipitation zone, where protein is so supersaturated that it comes out of solution too quickly to form ordered crystals and forms amorphous precipitate instead. **A)** An ideal scenario for crystal growth. The red arrows show the droplet becoming supersaturated as it approaches equilibrium just across the boundary between the metastable and labile zones. At equilibrium, a few crystal nuclei form, reducing the protein concentration enough that the droplet passes into the metastable zone where only growth of the existing crystals can occur. This results in few, large crystals per droplet. **B)** A less ideal scenario for crystal growth. As in **A**, the droplet becoming supersaturated as it approaches equilibrium, which lies much deeper within the labile zone. As a result, new crystals will continue to nucleate even as the protein concentration gradually decreases, resulting in crystals that are greater in number and smaller in size. **C)** The idea scenario for an acceptor droplet for seeding. The droplet reaches equilibrium in the metastable zone, meaning no new nucleation will occur and all crystal growth will take place on seeds. **D)** A demonstration of the effects of changing protein:reservoir ratios when setting up droplets. Each ratio reaches equilibrium at the same precipitant concentration, but different protein concentrations, which can cause dramatic changes in crystal nucleation and growth.

% PEG (v/v)	Result
1.4	clear drop
2.8	clear drop
4.2	clear drop
5.6	clear drop
7.0	clear drop
8.4	clear drop
9.8	crystals
11.2	crystals
12.6	crystals
14.0	precipitate
15.4	precipitate
16.8	precipitate
18.2	precipitate
19.6	precipitate
21.0	precipitate

Table 4.4: **Screening for the nucleation threshold of DHC-E crystals in PEG (MW=8,000).** I created a screen in a 15-well EasyXtal hanging drop plate that had 0.1M Tris-HCl (pH = 8.5) and 0.2M Li_2SO_4 in every condition, but concentrations of PEG (MW=8,000) ranging from 1.4% to 21.0%. For each condition, three droplets were set. Each droplet was set with $1\mu L$ of DHC-E (at 5.3mg/mL and $1\mu L$ of reservoir solution. The plate was set and left to equilibrate at room temperature ($\sim 20°C$) for 24 hours and then each droplet was examined for crystal growth and/or precipitate formation. The highest PEG concentration tested to maintain a clear drop was 8.4%. Crystals formed between 9.8% and 12.6% PEG. Higher PEG concentrations produced only amorphous precipitate.

a droplet that is already in the metastable zone via a cat whisker or similar fibre. I set up a 15-well EasyXtal hanging-drop plate with all of the wells containing 500μL of the same reservoir solution, which included 0.1M Tris-HCl (pH = 8.5), 0.2M Li_2SO_4 and 8.4% PEG (MW=8,000). Then I pipetted 1μL of DHC-E and 1μL of reservoir solution onto the screw caps and sealed them so that the droplets could equilibrate at RT for 24 hours and reach the metastable zone. After the droplets equilibrated I smashed some fully-grown crystals using a streak seeding tool (Hampton Research) and then dragged the fibre once through each equilibrated droplet. I allowed the plate to incubate at RT for 24 hours and then checked for crystal growth. All of the droplets contained showers of tiny crystals along the path that the fibre took through the droplet. Although the amount of seeds on the tool diluted with each successive droplet, there were never any droplets with large single crystals. The crystals always grew as clusters of small crystals that were impossible to separate without damaging them, thus rendering them less useful for X-ray crystallography.

In order to increase control over the number of crystal seeds delivered to the metastable droplets, I employed another technique that makes use of a tool called a Seed Bead (Hampton Research). First, I made a 1X seed stock by smashing the crystals from a single droplet using a a mesh LithoLoop (Molecular Dimensions). Then I pipetted the entire 2μL drop into an Eppendorf tube containing 98μL of reservoir solution (0.1M Tris-HCl (pH = 8.5), 0.2M Li_2SO_4 and 8.4% PEG (MW=8,000)) and a Seed Bead. Before doing this, I checked that crystal were stable in reservoir solution overnight. By vortexing this solution for 30 seconds, the bead pulverised the crystals into tiny seeds. Then I made serial dilutions by pipetting 10μL of the 1X seed stock into 90μL of reservoir solution, and repeating until I had seed stocks all the way down to a 10^{-7} dilution.

I set up a 15-well EasyXtal plate with every well containing the same 500μL 8.4% PEG reservoir solution used to make the seed stocks. Then I pipetted 1μL of DHC-E and 1μL of the various seed stock solutions onto the screw caps. I set up at least three droplets for each dilution of the seed stock and then sealed the plate so that the droplets could equilibrate at RT for 24 hours before checking them for any crystal growth. The 10^{-6} and 10^{-7} seed stocks failed to produce any crystals, but all of the others produced large amounts of very irregular looking crystals in each droplet. These crystals did not look as ordered as the crystals that grow *de novo* and when screened for diffraction using synchrotron radiation, they did not diffract to better than 15Å resolution.

I therefore concluded that seeding was not an effective way to improve the diffraction resolution limit of DHC-E crystals and focused instead on improving the *de novo* crystals. However, in the process of attempting these seeding techniques I did discover that using the hanging-drop vapour diffusion method produced slightly larger crystals than the ones I grew using the sitting-drop method and that it was easier to harvest intact crystals from the 15-well EasyXtal plates than from sitting-drop plates. I also found a very narrow range of PEG concentrations that promote crystal growth using this method (9.8% to 12.6%). I incorporated these findings into the next set of strategies I used to improve the diffraction power of the DHC-E crystals.

Oil vapour-diffusion barriers and protein:reservoir ratios

I attempted several other strategies for improving the quality of the DHC-E crystals. Some of these strategies include adding various concentrations of glycerol to the

	10%	30%	50%	70%	90%
0.5 : 1.0	O	O	O	X	X
1.0 : 1.0	*	*	*	X	X
1.0 : 0.5	*	*	**	X	X

Table 4.5: **Oil barriers and protein:reservoir ratios.** I created a screen in 15-well EasyXtal hanging-drop plates that had 0.1M Tris-HCl (pH = 8.5), 0.2M Li_2SO_4 and either 9.5%, 11% or 12.5% PEG (MW=8,000) in all 15 conditions. Two plates were created for each PEG concentration. Three droplets were set in each screw cap, following the protein:reservoir ratios indicated on the left. For example, in the top row, $0.5\mu L$ of DHC-E was mixed with $1\mu L$ of reservoir solution. One plate for each PEG concentration was left without adding any oil to the reservoir. The other plate had $500\mu L$ of oil added on top of the reservoir before re-sealing the screw cap. The oil for each column contained the indicated percentage of paraffin oil, with the remainder being silicone oil. The plates was set and left to equilibrate at room temperature ($\sim 20°C$) for 24 hours and then each droplet was examined for crystal growth and/or precipitate formation. All 90% droplets and most 70% droplets dried out completely (X). All 0.5:1.0 droplets remained clear (O). The remaining conditions all had crystals of some sort (*). The largest crystals were observed in the 11% and 12.5% PEG plates when a 1.0:0.5 drop ratio was used in combination with 50% paraffin oil (**).

reservoir solution, Silver Bullets Bio and HT additive screens (Hampton Research) and beginning with lower or higher concentrations of DHC-E itself. All of these were attempts to change the crystallisation dynamics slightly in order to favour fewer, larger crystals. None of those strategies led to an improvement in DHC-E crystals. However, a significant increase in both crystal size and diffraction resolution limit came from an experiment that combined two different strategies.

The first of these strategies was a variation of a popular method for slowing down droplet equilibration, and therefore crystal growth, known as under oil crystallisation. Typically this strategy is performed by covering the droplet itself with an oil that slows down the rate of vapour-diffusion between the droplet and the reservoir (Chayen *et al.* 1992). However, as I was using a hanging-drop approach, I opted to use a slightly different technique, which involves covering the reservoir with a layer of oil instead of the droplet (Chayen 1997). This also provides the benefit that harvesting crystals is no different from normal, whereas in traditional under oil experiments one must carefully remove the oil from the droplet before crystals can be harvested. I tested various mixtures of two oils to see which would provide the most effective barrier for slowing crystal growth without preventing it entirely. 100% paraffin oil (mineral oil) allows very little vapour diffusion to occur, while 100% silicone oil allows vapour diffusion to occur at approximately the same rate as when there is no oil present at all. By mixing the two oils in various proportions, it is possible to fine tune the rate of vapour diffusion for optimal crystal growth. I prepared five different mixtures: 10%, 30%, 50%, 70% and 90% paraffin oil, with the remainder of each mixture being silicone oil (Table 4.5).

The second strategy that I used in combination with the oil vapour-diffusion

P:R	[PEG]$_i$ (% v/v)	[DHC-E]$_i$ (mg/mL)	[PEG]$_f$ (% v/v)	[DHC-E]$_f$ (mg/mL)
0.5:1.0	7.33	1.77	11.0	2.65
1.0:1.0	5.50	2.65	11.0	5.30
1.0:0.75	4.71	3.03	11.0	7.07
1.0:0.5	3.67	3.53	11.0	10.6

Table 4.6: **Various protein:reservoir drop ratios and the effects on protein concentration before and after reaching equilibrium.** All calculations were done assuming a 5.3mg/mL DHC-E sample and 11% PEG in the reservoir solution. If 1μL of 5.3mg/mL DHC-E is added to 1μL of 11% PEG, the initial protein concentration in the droplet is 2.65mg/mL and the initial PEG concentration is 5.5%. In order for the PEG to return to equilibrium with the 11% reservoir, the droplet must shrink to half its original volume through vapour diffusion, doubling the protein concentration back up to 5.3mg/mL. If only 0.5μL of 11% PEG is added to 1μL of 5.3mg/mL DHC-E, the initial PEG concentration is 3.67mg/mL and the initial protein concentration is 3.53mg/mL. In order for the PEG to return to equilibrium with the reservoir solution, the droplet must shrink to one third of its original volume through vapour diffusion, tripling the protein concentration to 10.6mg/mL. The other two ratios were calculated using similar logic.

barriers was to change the ratio of protein sample to reservoir solution added when setting droplets. Crystal growth dynamics are not only determined by where on a phase diagram the droplet ends up after reaching equilibrium, but also where it begins and the path that it takes to get there (Luft *et al.* 2007). Therefore starting with different ratios of protein and reservoir solution in the droplet may have dramatic effects on the number and size of crystals formed. Figure 4.5D qualitatively illustrates the effect of different drop ratios on the initial condition, equilibrium condition and the path between the two. Table 4.6 demonstrates the theoretical initial and final DHC-E concentrations calculated for four different protein:reservoir ratios when 5.3mg/mL DHC-E is mixed with 11% PEG. The predicted values do not change for different PEG concentrations because my calculations oversimplify the system and don't account for the DHC-E and other solutes present in the protein sample, but they provide an overall trend for what happens to the protein concentration when different drop ratios are used.

I combined both the oil barrier and drop ratio experiments into a single 15-well EasyXtal hanging-drop screen, which is represented by Table 4.5. As I mentioned above, when I was determining the PEG concentration at which a droplet would be in the metastable zone at equilibrium for hanging-drop experiments, I discovered that 9.8% to 12.6% yielded crystals (Table 4.4). Therefore, when creating the oil barrier and drop ratio screens I used three new PEG concentrations that spanned this narrow range: 9.5%, 11% and 12.5%. I made the screen in Table 4.5 twice for each of these three PEG concentrations. All six plates were set up using 5.3mg/mL DHC-E, with three droplets set per condition. The drop ratios described in Table 4.5 were used for each row. One of the two plates for each PEG concentration was left without oil, while the other had 500μL of oil added to each reservoir before re-sealing the screw caps. Each column of the plate had a different percentage of paraffin oil (Table 4.5) with the remainder being silicone oil. The plates were set up
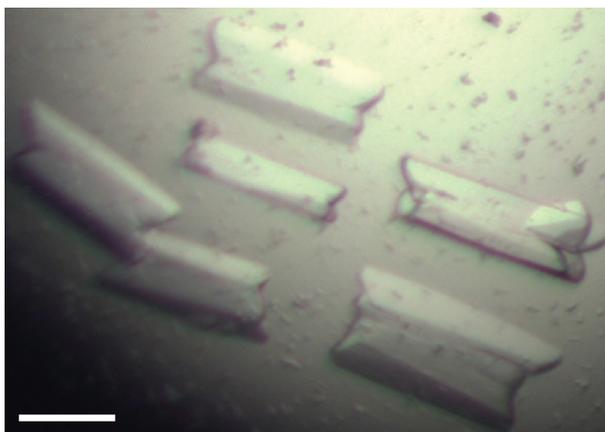
Figure 4.6: **Larger, optimised crystals of DHC-E.** DHC-E was purified and concentrated to 5.3mg/mL. $1\mu L$ of protein was added to to 500nL reservoir solution (0.1M Tris-HCl (pH $= 8.5$), 0.2M Li_2SO_4, 11% PEG (MW $= 8,000$, v/v)) on the screw cap of a 15-well EasyXtal hanging-drop plate. The $500\mu L$ reservoir solution was covered with $500\mu L$ of an oil mixture (50% paraffin oil, 50% silicone oil) before re-sealing. After 24 hours, the largest crystals had grown up to $\sim 200\mu m$ by $100\mu m$. When screened with synchrotron radiation, these crystals diffracted to 6.4Å. Measure bar $= 100\mu m$.

and left to incubate at RT ($\sim$ 20°C for 24 hours, then examined for crystal growth.

All of the conditions that had 90% paraffin oil, and most of the ones that had 70% paraffin oil on the reservoir dried out completely. All of the conditions that were set with a drop ratio of 0.5:1.0 remained clear. The remaining conditions contained crystals of various sizes and shapes. Most of them contained crystals similar in size and appearance to those already obtained by hanging-drop, described in Table 4.4. However, the 11% and 12.5% PEG plates had a few conditions that contained far fewer and much larger crystals. The largest of these crystals ($\sim 200\mu m$ in one direction and $\sim 100\mu m$ in another) were found in conditions that were set up with a 1.0:0.5 drop ratio and had 50% paraffin oil covering the reservoir (Fig. 4.6).

I harvested some of these optimised crystals and screened them for diffraction using the I03 beamline at Diamond Light Source. The best-diffracting crystals I tested had spots out to 4.2Å in the best direction, but were fairly anisotropic with a resolution of only 5.5Å in the worst direction. However, as this resolution is often high enough to at least see α-helices, I decided to try to solve a low-resolution structure using these crystals. I also thought that even a low-resolution structure might guide the design of future constructs that would form crystals with a higher resolution diffraction limit. Therefore, I collected several native datasets using these crystals and began setting up experiments to begin the phasing process.

4.2.5 Phasing the DHC-E crystals

I first searched for a homologous protein with a known structure that I could use as a search model for molecular replacement. However, when the amino acid sequence of DHC-E was used as the query sequence in a PSI-BLAST search (Altschul *et al.* 1997), the only hits were DHCs from other species. Since none of these have known structures for the N-terminal region, molecular replacement was not an option. I

instead tried two different strategies for calculating experimental phases, heavy atom soaks of native DHC-E crystals and crystallisation of a selenomethione (SeMet) derivative of DHC-E.

Heavy atom derivatives of native DHC-E crystals

I attempted to produce heavy atom derivatives of the native DHC-E crystals described in Section 4.2.4 using two different strategies. The first, was to soak native DHC-E crystals in various heavy atom solutions before freezing. The second, was to derivatise DHC-E with heavy atoms by co-crystallisation of the protein with the heavy atom compounds already in the reservoir solution used to set the droplet.

I first tried soaking DHC-E crystals in a variety of heavy atom compounds from the commercial Heavy Atom Screens M1 and M2 (Hampton Research). To accomplish this, I determined a stabilising solution into which I could transfer DHC-E crsytals for several hours without them dissolving or cracking. The stabilising solution I used was identical to the cryo-protectant solution I use for freezing crystals (0.1M Tris-HCl (pH = 8.5), 0.2M Li_2SO_4, 11% PEG (MW = 8,000, v/v) and 30% ethylene glycol (v/v)). For each of the heavy atom compounds tested (Table 4.7), I prepared a 10mM solution in the stabilising solution. Then I placed $1\mu L$ droplets of these solutions on the screw caps of a hanging-drop plate that contained stabilising solution in the reservoirs. A few crystals were also soaked in stabilising solution without any heavy atom compound as a control to ensure the treatment itself was not detrimental to the crystals. I transferred freshly grown native DHC-E crystals into each heavy atom solution and allowed the crystals to incubate for two hours. After the soak was complete I transferred the crystals to fresh stabilising solution and washed them for one minute before mounting them in cryoloops and freezing them in LN_2.

Although the majority of the heavy atom soaks did not visibly damage the crystals when observed under a dissecting microscope, none of the crystals diffracted as well post-treatment as native crystals that were soaked in stabilising solution without a heavy atom compound (4.5Å, Table 4.7). Some compounds caused the crystals to lose the ability to diffract completely, whereas others only reduced the resolution to $\sim$ 8Å.

In addition to the Heavy Atom Screens made by Hampton Research, I tried derivatising DHC-E with two compounds that were specially designed for experimental phasing: a tantalum bromide cluster, $[Ta_6Br_{12}]^{2+} \cdot 2Br^-$ (Jena Bioscience), and the highly substituted benzene ring 5-Amino-2,4,6-triiodoisophthalic acid (I3C, Hampton Research). $[Ta_6Br_{12}]^{2+} \cdot 2Br^-$ has proven useful for experimental phasing due to its large size, high electron density, and the presence of two different anomalous scatterers, Ta and Br (Knäblein *et al.* 1997). Furthermore, $[Ta_6Br_{12}]^{2+} \cdot 2Br^-$ is green in solution and turns crystals green upon binding, providing a convenient way to visually determine whether a soak has been successful. I3C, binds protein through its amino group and carboxyl groups more readily than traditional heavy-metal ions. Furthermore, it contains a 6Å equilateral triangle of iodine atoms, which simplifies recognition of compound within electron density (Beck *et al.* 2008).

I prepared a 10mM $[Ta_6Br_{12}]^{2+} \cdot 2Br^-$ solution in stabilising solution and set up drops in a hanging-drop plate as described above. Then I soaked 10 native DHC-E crystals in this solution for various lengths of time, ranging from 10 minutes to two hours. Roughly every ten minutes, I transferred one crystal to a fresh droplet

Heavy atom compound	Result
Thallium(I) chloride	9Å
Thallium(III) acetate hydrate	11Å
Cadmium chloride hydrate	21Å
Sodium hexachloroiridate(III)	38Å
Ammonium hexabromoosmate(IV)	8Å
Ammonium tetrathiotungstate(VI)	12Å
Samarium(III) acetate hydrate	13Å
Lanthanum(III) nitrate hexahydrate	N/A
Europium(III) nitrate hexahydrate	40Å
Gadolinium(III) chloride hydrate	N/A
Lutetium(III) acetate hydrate	12Å
Ytterbium(III) chloride hydrate	19Å
Neodymium(III) chloride hydrate	13Å
Control (no heavy atom compound)	4.5Å

Table 4.7: **Heavy atom soaks of DHC-E crystals in various compounds and subsequent diffraction resolution limits.** Native DHC-E crystals were soaked in a 10mM solution of various heavy atom compounds for two hours. After soaking the crystals were washed briefly and screened for diffraction. The results column reports the resolution limit for the best-diffracting crystal from each compound. The crystals lost diffraction power after soaking with all of the heavy atom compounds. The most detrimental soaks resulted in no diffraction at all (N/A), whereas crystals soaked in at least two compounds were still able to diffract to sub-nanometer resolutions. Crystals soaked in stabilising solution alone diffracted to 4.5Å.

of stabilising solution, allowed it to wash for 1 minute and then snap froze it in LN$_2$. The crystals were tinged with green even after 10 and 20 minutes of soaking, but the amount of green colour increased steadily until about one hour of soaking. From this point on all the crystals were deeply stained green and showed no signs of physical damage. When screened for diffraction with synchrotron radiation, most of the crystals diffracted to between 7Å and 8Å, with the best (the crystal that had been soaked for 30 minutes) reaching 6.5Å. I performed a fluorescence scan on this crystal and found the K edge for bromine. Using this scan I determined the peak wavelength to be 0.91882Å and the inflection wavelength to be 0.91956Å. I collected a dataset at each of these wavelengths, but subsequent attempts to calculate initial phases using single-wavelength anomalous dispersion (SAD) and multi-wavelength anomalous dispersion (MAD) failed due to a lack of anomalous signal. This was determined by looking at the CC$_{anom}$=0.5 when scaling and merging the datasets in Aimless (Evans & Murshudov 2013). None of the datasets had a value below 10Å, indicating that any [Ta$_6$Br$_{12}$]$^{2+}$ · 2Br$^-$ binding may have been non-specific.

I repeated this process using I3C, except I began with a 500mM I3C solution for the crystal soaks. Since I could see that the crystals were dissolving over time in this solution, I diluted the I3C to 100mM and repeated the experiment with new crystals. At this concentration I did not observe any physical defects in the crystals, even after one hour. The best diffraction of these crystals came from a crystal that had been soaked in 100mM I3C for only four minutes and diffracted to 4.5Å. Using a fluorescence scan, I found the L_I edge of iodine and determined the peak wavelength to be 2.3879Å. I collected a dataset at this wavelength and attempted to calculate initial phases using SAD, but as with the [Ta$_6$Br$_{12}$]$^{2+}$ · 2Br$^-$ datasets, there was no anomalous signal and so I could not use the dataset for experimental phasing.

I also attempted to co-crystallise DHC-E with I3C by setting up hanging-drop trays using a reservoir solution that contained 10mM I3C. Crystals formed in the same way as native DHC-E crystals and were visually indistinguishable. Of the 16 that I screened for diffraction using synchrotron radiation, the best one diffracted to 5.6Å. However, a fluorescence edge scan of this crystal suggested that insufficient I3C was bound to produce a measurable signal.

In conclusion, none of the heavy atom derivitisation strategies I attempted, whether through soaking or co-crystallisation, were successful in obtaining experimental phases for the DHC-E crystals. Since I could express DHC-E from *E. coli* and there were several well-established protocols for producing SeMet derivatives of proteins using bacteria, I decided to try this approach before investing more time in heavy atom soaks and co-crystallisation screens.

A selenomethionine derivative of DHC-E

To express the SeMet derivative of DHC-E, I transformed the plasmid described in Section 4.2.3 into a a methionine auxotrophic strain of *E. coli* called B834(DE3) (Novagen). I grew these cells in M9 media supplemented with selenomethionine and the other 19 standard amino acids, as well as all the other nutrients and vitamins required for optimal bacterial growth (Ramakrishnan *et al.* 1993). The remainder of the bacterial growth, protein expression and purification of the SeMet DHC-E were performed as with native DHC-E, except all buffers were thoroughly degassed and made with double the normal amount of DTT or BME in order to minimise SeMet
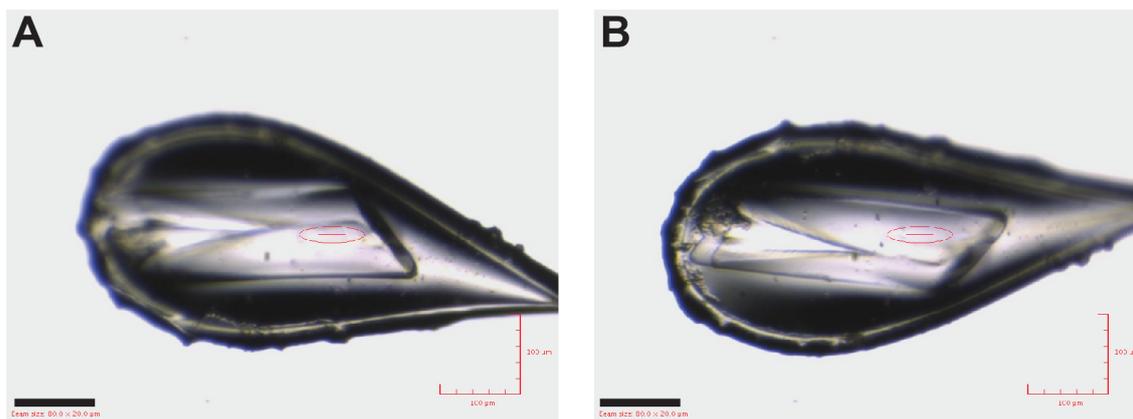
Figure 4.7: **SeMet DHC-E crystal in cryoloop, aligned for data collection.**
SeMet DHC-E was purified and concentrated to 5.3mg/mL. 1μL of protein was
added to to 500nL reservoir solution (0.1M Tris-HCl (pH = 8.5), 0.2M Li$_2$SO$_4$,
12.5% PEG (MW = 8,000, v/v)) on the screw cap of a 15-well EasyXtal hanging-
drop plate. The 500μL reservoir solution was covered with 500μL of an oil mixture
(50% paraffin oil, 50% silicone oil) before re-sealing. Crystals grew overnight, were
washed in cryoprotectant and flash frozen the next morning in LN$_2$. When screened
with synchrotron radiation, this crystal diffracted out to 3.7Å. Measure bars =
100μm. **A** and **B** are two views of the same crystal, rotated 180°around the long
axis of the cryoloop.

oxidation. SeMet DHC-E crystallised in the same conditions as native DHC-E to
yield crystals that were similar in size and appearance (Fig. 4.7).

32 SeMet DHC-E crystals were screened for diffraction using the I03 beamline at
Diamond Light Source. The majority of the crystals diffracted to between 5Å and
8Å, but several diffracted to 4Å and the best crystal had spots visible out to 3.7Å.
While screening, the auto processing software provided at the beamline suggested
that the crystals belonged to the *P222* point group and had unit cell parameters of
A=72.13Å, B=149.28Å , C=179.56Å , and $\alpha=\beta=\gamma=90°$.

Using one of the poorly diffracting crystals, I performed a fluorescence scan and
found the K edge for selenium. I determined the peak wavelength to be 0.97952Å,
the inflection wavelength to be 0.97970Å, and a high-energy remote wavelength to be
0.91841Å. There were two crystals that showed better diffraction in screening than
the rest. These two crystals were large enough that I could collect multiple datasets
off of each one without any signs of radiation damage by moving to a new region
of the crystal each time. For each anomalous dataset, I used only 20% transmission
and collected between 300°and 360°of data even though *P222* crystals require only
90°of data collection for a complete dataset. These measures were taken in order
to minimise radiation damage and collect high-redundancy datasets, in hopes of
maximising the anomalous signal necessary for phasing. Table 4.8 provides a full
summary of all datasets collected from these two crystals . After collecting at least
one dataset at each anomalous wavelength from each crystal, I collected two "native"
datasets using a wavelength of 0.97626Å. These datasets were of a lower redundancy
(180°) and were collected using 100% transmission. These were collected to allow
for future phase extension after a successful initial phasing experiment.

Crystal	Dataset	λ (Å)	Ω oscillation (°)	Exposure (s)	No. images	Transmission	Resolution (Å)
1	Peak$_A$	0.9795	0.2	0.25	1500	20%	3.5
1	Inflection$_A$	0.9797	0.2	0.25	1500	20%	3.5
1	Remote$_A$	0.9393	0.2	0.25	1500	20%	3.5
2	Peak$_B$	0.9795	0.2	0.17	1500	20%	4.0
2	Inflection$_B$	0.9797	0.2	0.14	1800	20%	4.0
2	Remote$_B$	0.9184	0.2	0.15	1500	20%	4.0
2	Peak$_C$	0.9795	0.2	0.13	1800	20%	4.0
2	Native$_A$	0.9763	0.2	0.20	900	100%	3.0
2	Native$_B$	0.9763	0.2	0.19	900	100%	3.0

Table 4.8: **Summary of the datasets collected from two SeMet DHC-E crystals.** The parameters used to collect nine datasets from two crystals, including wavelength, Ω oscillation, exposure time, number of images, transmission and the maximum resolution used to define the distance of the detector from the crystal.

Calculation of initial phases

I indexed and integrated the nine datasets described in Table 4.8 using iMosflm (Battye *et al.* 2011), which agreed with the auto processing results that both crystals belonged to the *P222* point group and, on average, had unit cell parameters of A∼ 72Å, B∼ 150Å, C∼ 180Å and $\alpha=\beta=\gamma=90°$. Each dataset was then scaled and merged using Aimless (Evans & Murshudov 2013) from within the CCP4 package (Winn *et al.* 2011). A summary of the diffraction statistics for datasets that were used in further steps can be found in Table 4.9. For crystal 1, all three datasets were scaled together, but merged independently so that one merged MTZ file was produced for each wavelength (Peak$_A$, Inflection$_A$ and Remote$_A$ in Table 4.9). For crystal 2, all six datasets were scaled together and one merged MTZ was produced for each wavelength. This resulted in the six original datasets being reduced to four since Peak$_B$ and Peak$_C$ were merged into a single dataset as well as Native$_A$ and Native$_B$. The only dataset used from this crystal in future steps was the one that resulted from merging Native$_A$ and Native$_B$ together (Native$_{comb}$ in Table 4.9). I tried phasing using many different combinations of datasets and empirically determined that combining datasets in this way and omitting the others gave the best initial phases.

With these datasets, I began trying to calculate initial phases using a variety of techniques and software packages. I first tried the SAD approach using the Peak$_A$ dataset, but this failed to produce any useful solution. I also tried the MAD approach using the Peak$_A$, Inflection$_A$ and Remote$_A$ datasets, but this also failed to produce credible electron density maps. It was only by also incorporating the Native$_{comb}$ dataset and using the multiple isomorphous replacement with anomalous scattering (MIRAS) approach that I successfully calculated experimental phases. This was accomplished by running AutoSHARP (Vonrhein *et al.* 2007) in MIRAS mode using SHELXD (Sheldrick 2010) for heavy atom detection with all four datasets described in Table 4.9. After initial phasing and automated density modification by solvent flattening, the AutoSHARP run produced an electron density map with a clear solvent boundary and features that were consistent with α-helices (Fig. 4.8A). There was also one region, which appeared to be a β-sheet and showed some signs of separation between the individual β-strands (Fig. 4.8B). The observed electron density

of the initial map suggested that the majority of the density could be accounted for by α-helices rather than β-strands. This observation was in line with the predicted secondary structure of DHC-E, so I used the map to build an initial model.

	Native$_{comb}$	**Peak$_A$**	**Inflection$_A$**	**Remote$_A$**
Diffraction data				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	$a = 71.9$	$a = 72.0$	$a = 72.1$	$a = 72.1$
	$b = 148.9$	$b = 149.0$	$b = 148.9$	$b = 149.2$
	$c = 179.6$	$c = 179.4$	$c = 179.5$	$c = 178.8$
Unit cell (°)	$\alpha = 90.0$	$\alpha = 90.0$	$\alpha = 90.0$	$\alpha = 90.0$
	$\beta = 90.0$	$\beta = 90.0$	$\beta = 90.0$	$\beta = 90.0$
	$\gamma = 90.0$	$\gamma = 90.0$	$\gamma = 90.0$	$\gamma = 90.0$
Resolution (Å)	52.6 - 4.0	64.9 - 4.0	61.0 - 4.0	64.9 - 4.0
R_{Sym}[1]	10.6 (80.2)*	19.6 (126.9)	14.0 (96.8)	13.0 (85.6)
$<I>/<\sigma I>$	10.4 (3.9)	7.9 (2.3)	8.5 (2.4)	8.9 (2.7)
Completeness (%)	100.0 (100.0)	99.8 (99.3)	99.8 (99.3)	99.8 (99.3)
Redundancy	11.6 (11.7)	10.0 (10.1)	10.0 (10.1)	9.7 (9.8)
No. reflections	196,661 (55,179)	169,522 (47,577)	169,334 (47,518)	163,986 (45,845)
No. unique reflections	16,985 (4,732)	16,985 (4,699)	16,984 (4,699)	16,986 (4,699)
Wavelength (Å)	0.9763	0.9795	0.9797	0.9393
CC_{anom}=0.5 cutoff (Å)	-	6.7	-	-
Phasing statistics				
FOM[2]	0.31	-	-	-

Table 4.9: **Data collection and phasing statistics for the four datasets used to solve the SeMet DHC-E structure.** *Values in parentheses refer to the highest resolution shell. [1]R_{Sym} (I) = $(\Sigma_{hkl}\Sigma_i |I_{hkl,i} - <I_{hkl}>|) / \Sigma_{hkl}\Sigma_i I_{hkl,i}$ for n independent reflections and i observations of a given reflection. [2]FOM = mean figure of merit.

4.2.6 Initial model building and density modification

I began building the DHC-E model by manually placing ideal α-helices into each tube of density that resembled an α-helix using *Coot* (Emsley *et al.* 2010). By visualising crystallographic symmetry mates, I ensured that all the α-helices I built belonged to the same asymmetric unit and were not related to each other by crystallographic symmetry. Once all of the obvious helical density had been filled in this manner, I also modelled eight β strands into the β-sheet (Fig. 4.8B). The resulting model consisted of a globular domain framed on either side by elongated bundles of α-helices (Fig. 4.9A). The majority of the globular domain was situated in a different plane than the elongated domains with only one α-helix per monomer appearing close enough to make a contact with the elongated domains (Fig. 4.9B). The globular domain was made up of a β-sheet sandwiched between α-helices (Fig. 4.9C). The elongated domains consisted only of α-helices, which could be divided roughly into two bundles, one slightly longer than the other (Fig. 4.9D). The overall
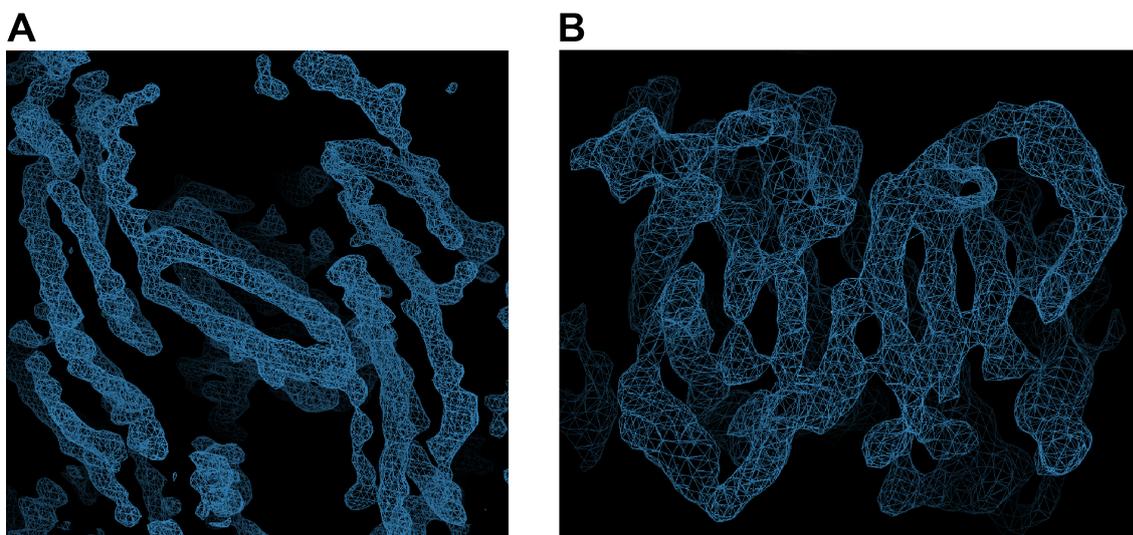
Figure 4.8: **Initial electron density map of DHC-E after experimental phasing by MIRAS and solvent flattening.** Three anomalous datasets and one "native" dataset collected from SeMet DHC-E crystals, were used to calculate initial phases. Phases were calculated using AutoSHARP in MIRAS mode followed by solvent flattening in order to produce an initial map. Electron density was contoured to 1.5σ. **A)** A global view of α-helical bundles. **B)** A close-up view of a β-sheet region. Some strand separation was already visible in the initial map.

structure revealed an obvious two-fold symmetry, centred on the globular domain, which displayed internal symmetry within itself (Fig. 4.9A). These observations led me to conclude that the globular domain was a previously unknown dimerisation domain within the DHC. I expected the DHC-E might crystallise as a dimer, since this is how it eluted during SEC, but the extensive dimerisation interface observed in the crystal structure confirmed that this dimer is biologically relevant and not just an artifact of purification.

I also combined the anomalous difference amplitudes from the Peak$_A$ dataset with the experimental phases in order to create an anomalous difference density map using FFT (Read & Schierbeek 1988). Although there are eight methionine residues for each copy of DHC-E (not including the N-terminal methionine), only five SeMet peaks per monomer where visible in the anomalous difference map. However, the five SeMet peaks that were visible were useful in beginning to assign specific residues to the initial model. For example, the fact that M358 and M366 are only separated by eight residues and were predicted to be within a single α-helix allowed me to assign that α-helix with great confidence. Furthermore, M394 was predicted to be in the middle of the next α-helix and M216 was predicted to be at the very N-terminus of a 24-amino acid α-helix, which made assigning those α-helices possible, as well as determining their orientations. Finally, M186 was predicted to be in a loop region, just after a short β-strand. This corresponded to the location of the final anomalous peak, which could be found within an extended loop located at the joint between the dimerisation domain and the elongated domains.

Once I assigned these key regions using the SeMet peaks, I was able to identify the majority of the α-helices by referring to the secondary structure predictions (Fig. A.1). The prediction began with two α-helices followed by four consecutive
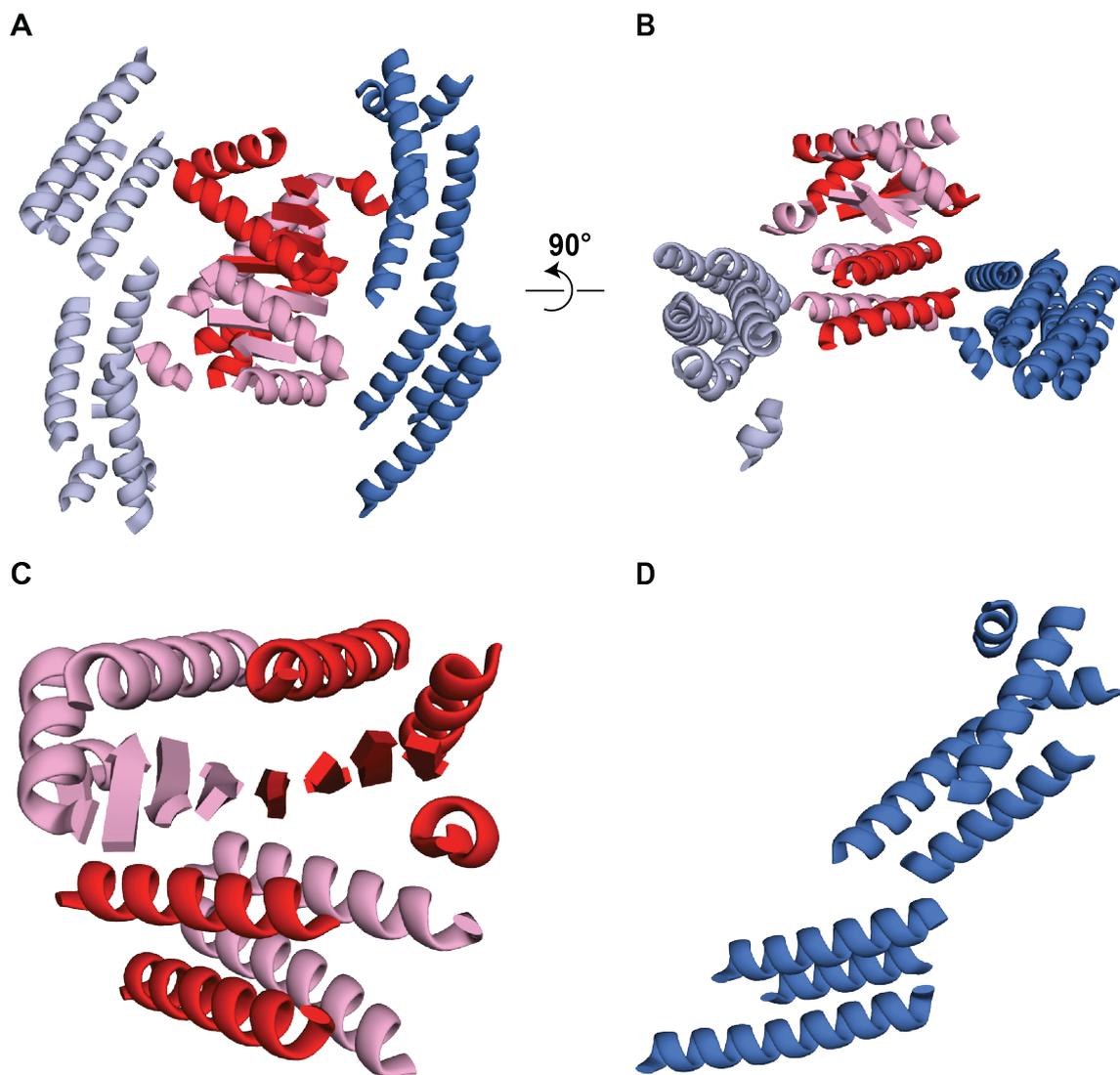
Figure 4.9: **An overview of the initial DHC-E model.** An initial DHC-E model was built by placing α-helices and β-strands in in tubes of electron density that were of the correct diameter for each type of secondary structure motif. **A)** A view of the entire model that highlights the 2-fold non-crystallographic symmetry of the DHC-E dimer. The axis of symmetry comes out of the page, directly in the centre of this image. The dimerisation domain is coloured red (Chain A) and pink (Chain B). The elongated domains are coloured dark blue (Chain A) and light blue (Chain B). **B)** A view of the entire model after rotating **A** 90°about a horizontal axis through the centre of the molecule. This view highlights the fact that the dimerisation domain sits in a different plane than the elongated domains, with the exception of the bottom two α helices, which appear to make bridge the two domains together. **C)** A closer look at the dimerisation domain, which reveals the intricate dimer interface. A single β-sheet is formed by eight β-strands, with each monomer contributing four. Above the β-sheet, there are two anti-parallel α-helices that appear to contact each other. Below the β-sheet, there are four more α-helices, two from each monomer, which appear to form a tight bundle. **D)** A closer look at one of the elongated domains, which demonstrates that it is made up of two smaller bundles of α-helices.

β-strands and then two more α-helices, which corresponds well with the structure of the dimerisation domain. After the dimerisation domain, the SeMet positions, as well as the variation in the length of predicted α-helices allowed me to trace the carbon backbone through each α-helix that I had built. However, I could not always determine the orientation due to a lack of density for the connections between α-helices. In this way I assigned all the α-helices until residue 412. At this point, five α-helices remained in the secondary structure prediction, but no more α-helices could be modelled due to a lack of clear electron density. There were also three more methionine residues in the part of the DHC-E sequence that I could not build. Since this density was missing from both the full map, as well as the anomalous difference map, I concluded that the final five α-helices were disordered in the crystal. There was enough empty space within the crystal structure to accommodate the missing α-helices without disrupting the crystal packing. With the initial model complete, I was also able to determine that the asymmetric unit consisted of a single DHC-E dimer, meaning that the only non-crystallographic symmetry (NCS) I would be able to take advantage of when improving the model and/or map would be the symmetry that existed between the two monomers.

I next attempted to improve the map quality using various techniques. First, I re-processed all of the original datasets used for phasing with Diffraction Anisotropy Server in order to perform anisotropic scaling and ellipsoidal truncation (Strong *et al.* 2006). This was to correct for any negative effects of the inherent anisotropy in the DHC-E crystals. Next, I created a NCS mask for each subdomain of each monomer of the initial model and determined the superposition matrix necessary to transform the subdomains from one monomer onto the other. I used the anisotropy corrected amplitudes, experimental phases, NCS masks and superposition matrix to run DM (K. Cowtan 1994). In the DM job, I performed 200 cycles of phase extension to 3.25Å with NCS averaging, solvent flattening and histogram matching. The resulting map showed a dramatic improvement in some regions of the map, but other regions were relatively unaffected. In the most improved regions, the tubes of density for α-helices were covered with bumps that corresponded to the side chains of larger amino acids and more connections between α-helices were visible (Fig. 4.10). Despite these improvements in the parts of the map I had already modelled, the density for the C-terminal five α-helices did not appear. In order to make sure this was not because the mask used for NCS averaging only covered the region of the assymetric unit I had already modelled, I added several α-helices in the open space where the C-terminus must be, made new NCS masks and repeated the DM run. This still didn't result in any new density becoming visible in this region. So I concluded that the C-terminus is disordered and would not become visible without changing the crystals themselves.

Using the new electron density map, I assigned all residues and modelled the majority of the connections between α-helices. I was not able to confidently assign the individual strands within the β-sheet, nor determine their orientation, since the connections between them were not clear enough. I also modelled larger side chains when there were bumps in the density at the expected places. For α-helices with multiple large side chains visible, I was able to rotate the α-helix and check that the registry was correct, but this was not possible for all of them. While building, I alternated between the two copies of the DHC-E dimer in order to take advantage of the regions of the map that were best in each copy. Once I had modelled
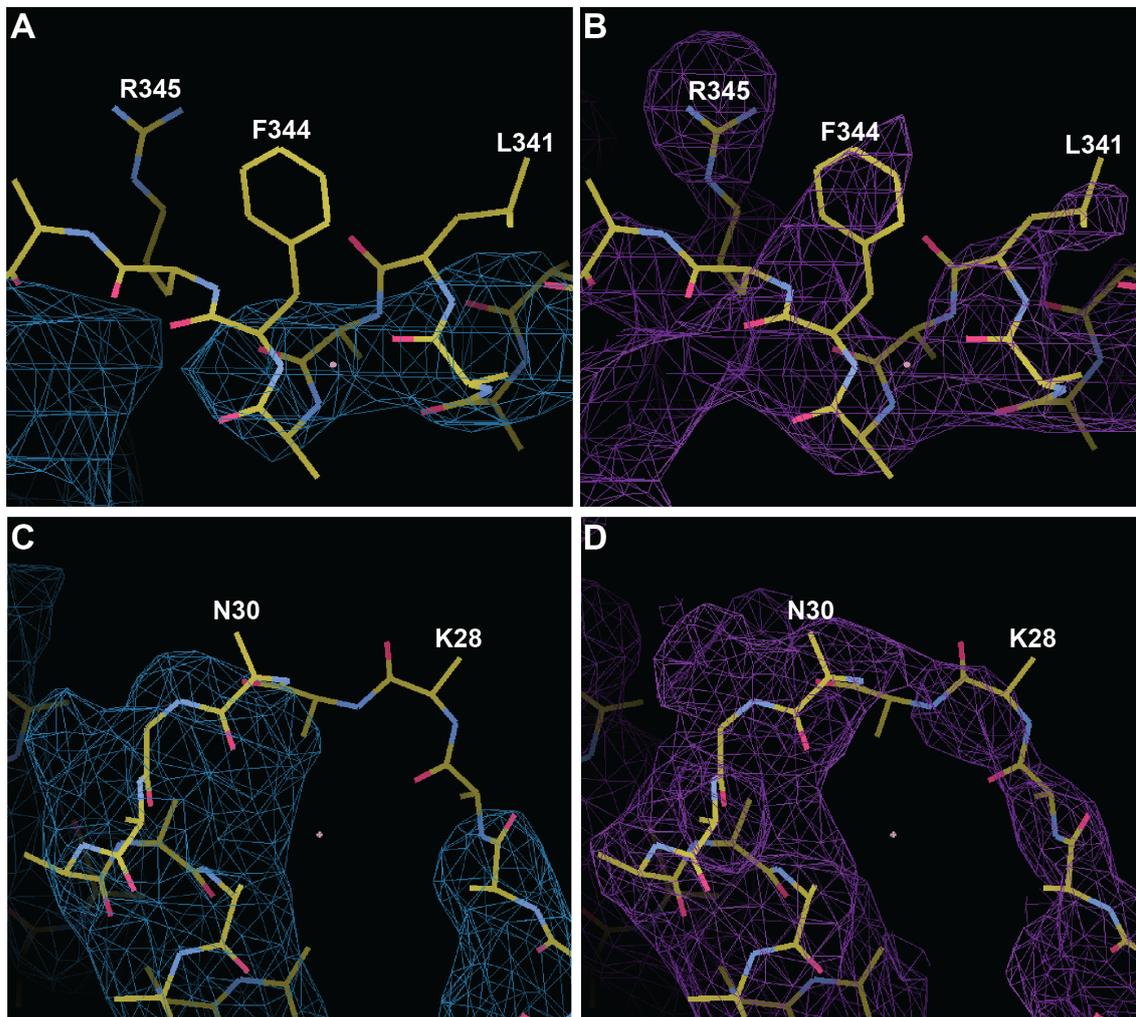
Figure 4.10: **A comparison of the electron density map before and after DM.** All datasets used for experimental phasing were corrected for anisotropy. After anisoptropic scaling and ellipsoidal truncation, the amplitudes from these datasets were combined with the experimental phases. Phases were extended and NCS averaging was performed between the two monomers within the asymmetric unit. **A and B)** A region within one of the elongated domains of Chain A where the map improved dramatically after running DM. Three large sidechains were able to be built into electron density after DM, whereas only the carbon backbone could be built before. **C and D)** The loop that connects the two first α-helices at the N-terminus was not visible in the initial map, preventing correct assignment and orientation of these two α-helices. After DM, the connection became visible, allowing the orientation of the two α-helices to be determined. Both maps were contoured to 1.5σ.
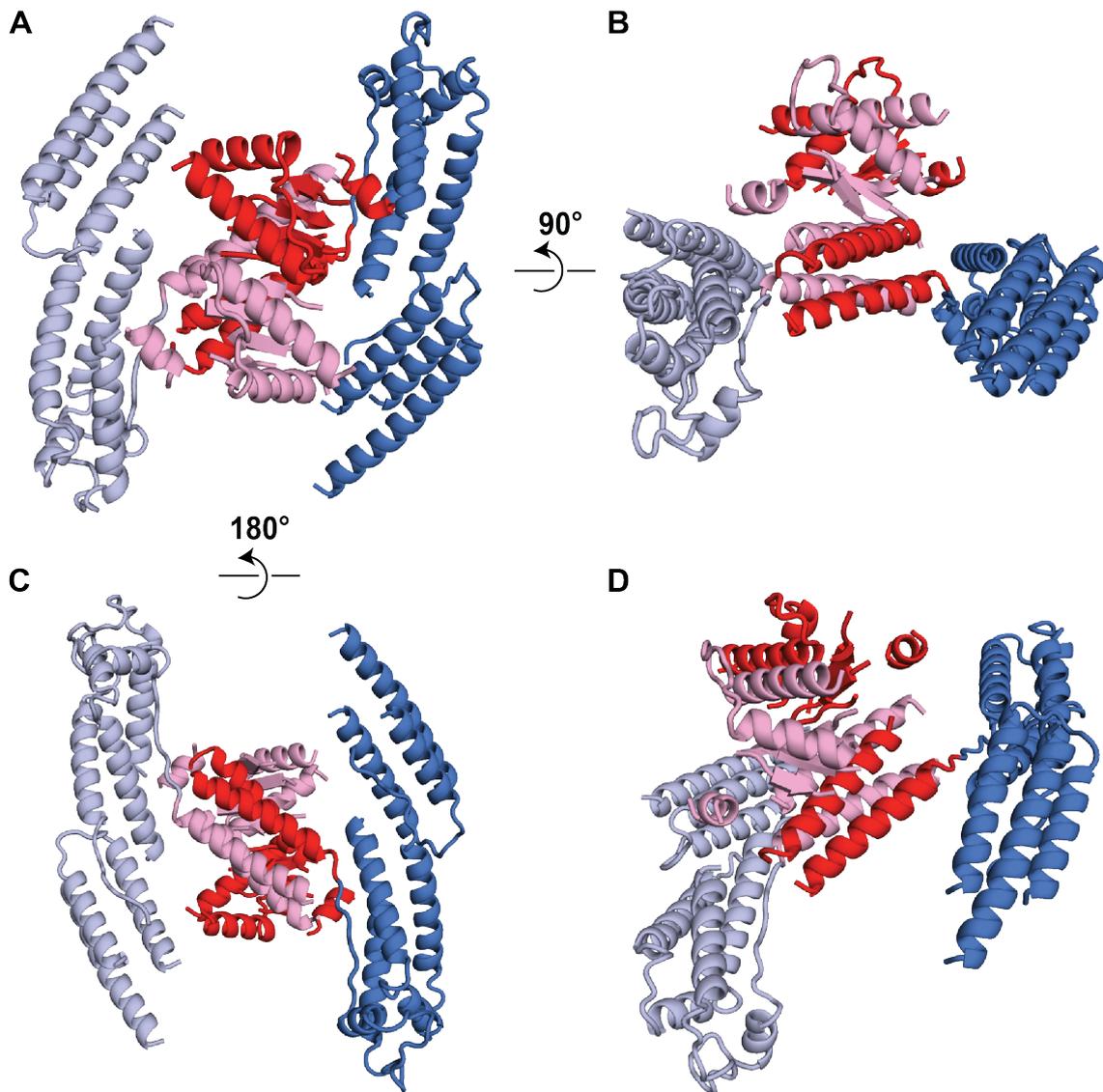
Figure 4.11: **An overview of the final DHC-E model.** The initial DHC-E model was updated after improving the electron density map (PDB accession code 5AFR). Loops connecting some of the α-helices were built in, including the one that connects the dimerisation domain to the elongated domain. **A)** A view of the entire model equivalent to the one used in Fig. 4.9A. The dimerisation domain is coloured red (Chain A) and pink (Chain B). The elongated domains are coloured dark blue (Chain A) and light blue (Chain B). **B)** A view of the entire model after rotating **A** 90°about a horizontal axis through the centre of the molecule. This view highlights the new loops that have been built in to the dimerisation domain as well as at the near end of the elongated domain. **C)** A view of the entire model after rotating **A** 180°about a horizontal axis through the centre of the molecule. This view highlights the underside of the dimerisation domain and provides a clearer view of the connections between the dimerisation domain and the elongated domains. **D)** A view that emphasises the distance between the dimerisation domain and the elongated domain. With the exception of the peptide backbone that connects these two domains, at this resolution there are no visible contacts between the two domains, suggesting that they may rotate freely with respect to each other.

as much as possible into the post-DM map (Fig. 4.11), I carried out refinement of the model against the experimental data using REFMAC5 (Murshudov *et al.* 2011). Despite attempting many different types of refinement (including restrained refinement, and rigid body refinement), and changing many of the parameters (with and without experimental phases, with and without local NCS, with and without jelly-body refinement), the best R-factor and R-free I obtained were 0.39 and 0.43, respectively. I also tried running the refinement after deleting various α-helices and loops, to check whether I had over-interpreted the electron density. In every case, the difference density always suggested that what I had built belonged there. This most likely indicates that there is some larger problem with the model, though I was never able to determine what this might be. Working with such a low-resolution density map, it is possible that the α-helices are slightly out of position. If enough α-helices are off by a half a rotation, this might be enough to prevent the R-factors from decreasing.

Since it seemed the DHC-E crystals I had already obtained would not provide me with the quality of data I needed to obtain a higher-resolution density map, nor a complete model of the construct, I designed several new constructs. Many of these were designed rationally using the initial model to make specific predictions about how to improve the crystals, but others still relied on a trial and error approach. As the low quality of the phases seemed to be the limiting factor in improving the original density map, my goal was to collect better phase information or to find a new crystal form, which would allow me to take multi-crystal averaging.

4.2.7 Creating and screening new crystal constructs of the DHC N-terminus

Expressing DHC-E in *E. coli* not only increased the yield and solubility of the sample, but also made the screening of new constructs much higher-throughput. For most constructs, 1L of *E. coli* culture produced enough protein to set up all 20 of the LMB initial crystal screens at $\sim$ 5mg/mL. So once the plasmids were created and transformed into the correct expression strain, it was convenient to express 12 or 24 constructs in parallel. Furthermore, the purification protocol, which relied only on an affinity chromatography step followed by SEC, was compatible with performing up to six purifications in parallel. With this high-throughput system for screening constructs and an initial model on which I could base specific predictions about obtaining better crystals, I designed and screened several new constructs. My two goals were to improve the diffraction resolution limit of the crystals and/or the completeness of the model, which still lacked the C-terminus.

DHC-E equivalent constructs in 12 new species

One well known strategy for improving crystal quality is to crystallise an equivalent construct from a different species. The aim is to find a species that has enough differences in the protein structure to subtly affect the formation of crystals, while remaining similar enough to the original construct that it will still crystallise and provide a relevant structure. As this strategy relies on a "trial and error" approach and I had established the high-throughput system for screening constructs, I selected twelve different species to compare. I included a diverse assortment of species across

Species	Length	MW (kDa)	Results
S. cerevisiae	557	60.1	E,P,C,D (Sections 4.2.3-6)
A. gossypii	540	61.3	E,P
C. elegans	597	67.9	E,P
C. thermophilum	646	71.1	E,P
D. discoideum	656	73.8	E,P,C
D. melanogaster	609	69.9	X
D. rerio	614	69.9	X
G. gallus	582	66.9	E
H. sapiens	615	70.3	E,P
K. lactis	542	62.5	E,P
M. musculus	613	70.1	E,P
N. crassa	640	71.1	E,P
T. thermophila	581	67.3	E,P

Table 4.10: **List of equivalent constructs of DHC-E from 12 different species.** The C-terminal residue and predicted MW after TEV cleavage are shown for each DHC-E equivalent. In the results column, X=expression level was too low for purification, E=expression level was high enough to complete purification through SEC step, P=SEC produced a peak that was clean enough to set up crystal screens, C=crystal screens produced hits that were optimised and screened for diffraction, D=crystals diffracted.

eukaryotic taxa in order to increase the chances of finding one species that would have the optimal amount of divergence with *S. cerevisiae*. I included some fungal species, which align more closely with *S. cerivisiae*, as well as several more distantly related metazoans. I performed an alignment between the full-length cytoplasmic DHCs from all twelve species plus *S. cerevisiae* using ClustalX (Larkin *et al.* 2007). A visual inspection of the alignment in Jalview 2 (Waterhouse *et al.* 2009) revealed that all thirteen sequences clearly aligned in the region surrounding the *S. cerevisiae* DHC-E truncation (Fig. 4.12). So I designed the twelve equivalent constructs (Table 4.10), ending each with the residue that aligned with D557 of *S. cerevisiae* (red arrow, Fig. 4.12). I had each fragment synthesised and codon optimised for expression in *E. coli* and then I cloned the fragments into the same bacterial expression vector I had used previously Section 4.2.3).

I grew 1L of *E. coli* for each of the twelve DHC-E equivalents from different species. Two of the constructs (*D. melanogaster* and *D. rerio*) did not express at high enough levels to purify the protein (Fig. 4.13). Of the remaining constructs, all but the *G. gallus* construct were clean enough after SEC to set up crystal screens. However, the *D. discoideum* construct was the only one to produce any crystal hits. I attempted to optimise these crystals using many of the same techniques described in Section 4.2.4, but I was unable to improve any of them significantly. I screened some of the initial crystals for diffraction using the I03 beamline at Diamond Light Source, but they did not diffract, indicating that they may have been very disordered.

Although none of these constructs proved useful for crystallography, at least two of those that purified well may be of use for future experiments. At 613 amino acids in length, the *M. musculus* construct includes residue F580, which is the site of a well
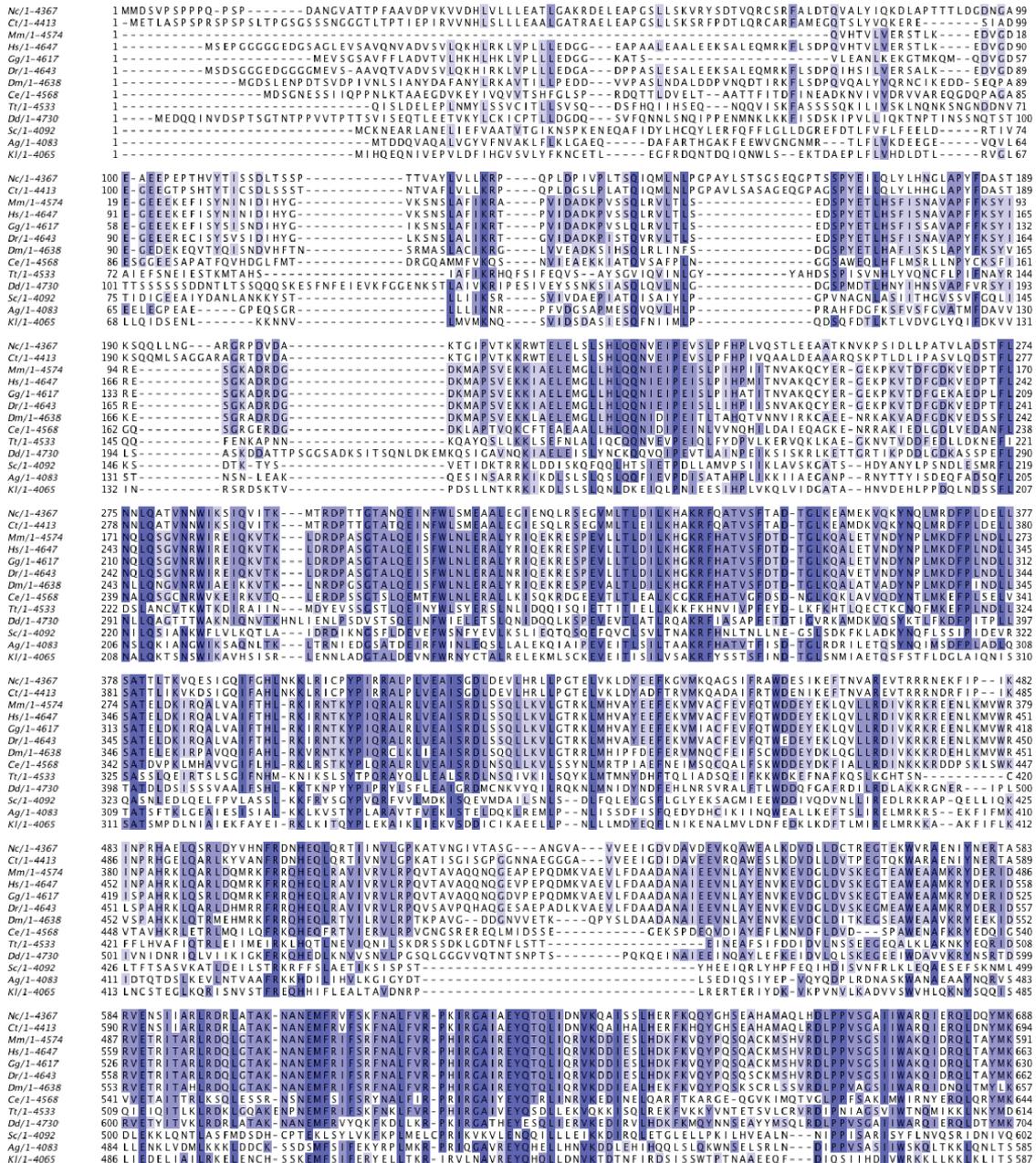
Figure 4.12: **An alignment of the DHC N-terminus from thirteen different species.** Full-length DHCs from thirteen different species were aligned using ClustalX. The alignment was visualised with Jalview 2 using the "percentage identity" colour scheme (darker purple indicates higher conservation). Only the N-terminal ∼ 15% is shown. DHC-E equivalent constructs were designed in twelve other species by truncating at the residues above the red arrow, which aligned with the C-terminal residue of the *S. cerevisiae* construct. Ag=*Ashbya gossypii*, Ce=*Caenorhabditis elegans*, Ct=*Chaetomium thermophilum*, Dd=*Dictyostelium discoideum*, Dm=*Drosophila melanogaster*, Dr=*Danio rerio*, Gg=*Gallus gallus*, Hs=*Homo sapiens*, Kl=*Kluyveromyces lactis*, Mm=*Mus musculus*, Nc=*Neurospora crassa*, Sc=*Saccharomyces cerevisiae*, Tt=*Tetrahymena thermophila*.
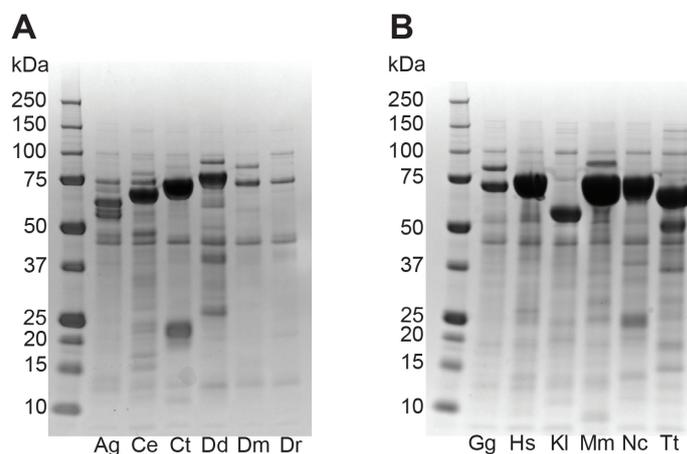
Figure 4.13: **SDS-PAGE analysis of equivalent constructs of DHC-E from twelve different species.** 1L of *E. coli* was grown for each equivalent construct of DHC-E from twelve different species. Protein was purified using a Ni-NTA column before removing the His-tag by TEV cleavage. The gels presented show the samples after TEV cleavage and just before the SEC polishing step. Ag=*Ashbya gossypii*, Ce=*Caenorhabditis elegans*, Ct=*Chaetomium thermophilum*, Dd=*Dictyostelium discoideum*, Dm=*Drosophila melanogaster*, Dr=*Danio rerio*, Gg=*Gallus gallus*, Hs=*Homo sapiens*, Kl=*Kluyveromyces lactis*, Mm=*Mus musculus*, Nc=*Neurospora crassa*, Sc=*Saccharomyces cerevisiae*, Tt=*Tetrahymena thermophila*.

described point mutation called leg at odd angles (Loa). Despite resulting from the addition of a single hydroxyl group (F580Y), the Loa mutation has dramatic effects on development in mice, which have been linked to defects in neuronal migration Ori-McKenney & Vallee 2011). It might be interesting to express both the wild-type and mutant versions of this construct and then characterise them biochemically to see if the mutant version folds properly or if it adopts an entirely different conformation. Additionally, in light of some of my colleagues' recent work in the Carter lab, which will be discussed in detail in Section 4.2.8, the human construct may be useful in further dissecting the interraction between dynein and dynactin.

A finer screen of C-terminal truncations near DHC-E

The next set of constructs consisted of a finer screen of C-terminal truncations in the region surrounding the DHC-E truncation (hereafter referred to as the E region). I based these truncations on the secondary structure prediction (Fig. 4.14), the same way I designed the first and second of DHC truncations in Sections 2.2.3 and 2.2.4. I was not able to use the crystal structure to guide the design, since all of the E region truncations fall within the region of the DHC-E construct that was not visible in the electron density. However, I made a fine screen in this region in hopes of altering or introducing new crystal contacts that might force the C-terminus into a more ordered state. The shortest constructs (DHC-EA, DHC-EB and DHC-EC) removed the last few α-helices from the DHC-E construct in order to test whether removing the predicted coiled-coil would produce higher quality crystals. I thought that disruption of one of the α-helices that forms the coiled-coil might have caused the entire C-terminus to be disordered in the crystal lattice. I also made a few

```
                                              EA      CEB     EC      D
           410       420       430       440       450       460       470       480
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
401   IVQDVNLLIREDLRKRAPQELLIQKLTFTSASVKATLDEILSTRKRFFSLAETIKSISPSTYHEEICRLYHPFEQIIDIS
262   HHHHHHHHHHH----------------HHHHHHHHHHHHHHHHHHHHHHHHHH-----HHHHHHHHHHHH--------
                         H13                                    H14

                ED                  EE        EF        EG                EH    E/EI  EJ
           490       500       510       520       530       540       550       560
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
481   VNFRLKLEQAESEFSKNMLDLEKKLQNTLASFMDSDHCPTEKLSYLVKFKPLMELCPRIKVKVLENQQILLLEIKKDIRQ
312   ----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-----HHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHH
          H15                           H16                   H17

                EK        ELF
           570       580       590       600       610       620       630       640
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
561   LETGLELLPKILHVEALNNIPPISARISYFLNVQSRIDNIVQYLEALFGSNWNDTLEGRSISTSIVQLRKETNPHDVFLH
377   HHHHHHHH-----------HHHHHHHHHHHHHHHHHHHHHHH------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
              H18                            H19
```

Figure 4.14: **Secondary structure prediction of E region of the DHC with the new series of truncations highlighted.** H=predicted α-helix, turquoise=predicted coiled-coil, red=final residues from the first series of DHC truncations in this region: DHC-C, DHC-D, DHC-E and DHC-F, green=final residues from twelve new E region DHC truncations. Note that the new DHC-EI is equivalent to the original DHC-E truncation.

truncations that were longer than the original DHC-E construct. I designed DHC-EK and DHC-EL to include the remainder of the predicted coiled-coil that DHC-E falls halfway through. Finally, I reproduced the original DHC-E construct as a control, DHC-EI. This was necessary becasuse although the constructs shorter than DHC-E were made by modifying the original DHC-E plasmid, the longer ones had to be made by amplifying the gene from a different template. So to ensure this difference in cloning procedure for the longer constructs did not lead to differences in expression, I re-cloned the DHC-E construct in the same way as DHC-J, DHC-K and DHC-L. One of the PCR reactions necessary for the creation of DHC-EE failed, so this construct was not completed.

I grew 1L of each construct, plus one extra litre of the *E. coli* that had been transformed with the DHC-EL construct, which I did not induce wth IPTG. This uninduced culture served as a control to test for leaky expression and to identify non-specific proteins that are present in the cell lysate in the absence of overexpression. I performed a purification on all twelve pellets by lysing the cells, incubating with Ni-NTA agarose beads, and then washing. After eluting bound protein, I performed a TEV digest and analysed all samples by SDS-PAGE (Fig. 4.15). Several constructs expressed just as well as the original DHC-E construct (equivalent to then new DHC-EI in Fig. 4.15). However the shortest four constructs and the longest one expressed only at very low levels. Therefore, I proceeded to complete the purification only for the constructs that expressed well: DHC-EF, DHC-EG, DHC-EH, DHC-EI, DHC-EJ and DHC-EK.

I ran the samples of the six highly expressed constructs back through a Ni-NTA column in order to re-bind all of the non-specific proteins, TEV protease and the His-tagged lipoyl domain, which had been cleaved off. Then I concentrated each sample in order to perform SEC. DHC-EF and DHC-EG produced complex chromatograms, with several peaks, and no prominent one at the expected elution

119
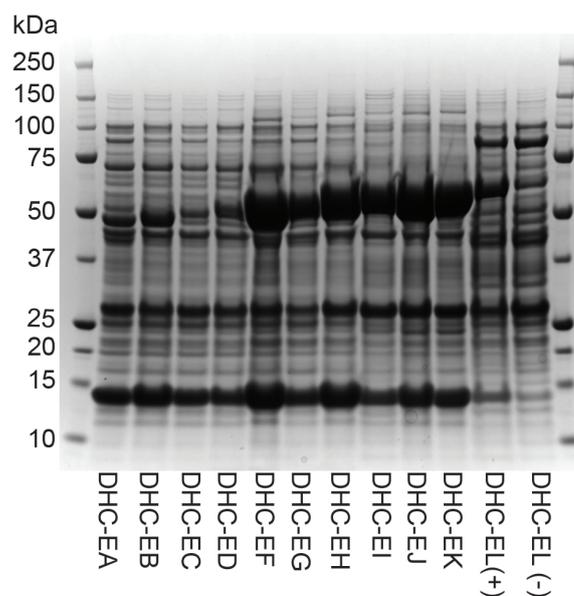
Figure 4.15: **SDS-PAGE analysis during test purifications of 11 E region truncations from 1L of culture each.** 1L of *E. coli* was grown for each E region construct except DHC-EE, which was not completed due to difficulties with cloning. 1L of the *E. coli* strain that had been transformed with the DHC-EL plasmid was also grown without inducing expression with IPTG (-). This control served to check for leaky expression and to identify non-specific proteins in the *E. coli* lysate by direct comparison with the 1L culture that was induced for DHC-EL expression (+). Protein was purified from lysates using a Ni-NTA column before removing the His-tag by TEV cleavage. The gels presented show the samples after TEV cleavage but before running the eluate back through the Ni-NTA column to re-bind non-specific proteins, TEV protease and the His-tagged lipoyl domain.

Name	Last residue	MW (kDa)	Results
DHC-EA	460	52.4	X
DHC-EB	467	53.3	X
DHC-EC	472	54.0	X
DHC-ED	495	56.7	X
DHC-EE	515	59.0	PCR failed
DHC-EF	524	60.0	E
DHC-EG	531	60.9	E
DHC-EH	548	62.9	E,P
DHC-EI	557	63.9	E,P,C,D
DHC-EJ	560	64.3	E,P,C
DHC-EK	569	65.2	E,P,C
DHC-EL	575	66.0	X

Table 4.11: **List of 12 E-region truncations.** The C-terminal residue and predicted MW after TEV cleavage are shown for each construct. In the results column, X=expression level was too low for purification, E=expression level was high enough to complete purification through SEC step, P=SEC produced a peak that was clean enough to set up crystal screens, C=crystal screens produced hits that were optimised and screened for diffraction, D=crystals diffracted.

volume for a dimer. The remaining four constructs (DHC-EH, DHC-EI, DHC-EJ an DHC-EK) all produced chromatograms very similar to the original DHC-E construct. Therefore, I pooled the prominent dimer peaks and concentrated each sample to 5.3mg/mL for crystal screening. I set up all 20 LMB crystal screens using each of these four samples with droplets composed of 200nL protein plus 200nL reservoir. The plates were set up and incubated at $\sim 20°C$.

After 24 hours of incubation, the DHC-EI construct had already produced dozens of crystal hits in dozens of conditions, similar to the original DHC-E construct Sections 4.2.2 and 4.2.3, confirming that the template used for PCR had no effect and that the high-throughput purification techniques I was using were not detrimental to crystal quality. DHC-EJ and DHC-EK each produced a single crystal hit, both in the same condition. This condition, LMB2/H3, has a reservoir solution composed of composed of 0.1M CAPS (pH = 10.5), 0.2M NaCl and 20% PEG (MW=8,000, w/v). This condition is strikingly similar to the initial hit that led to successful optimisation of the DHC-E crystals, LMB9/A9. Both had 0.2M of a salt, and both had 18% or 20% PEG (MW=8,000). I harvested some of the initial crystals from both DHC-EJ and DHC-EK and screened them for diffraction using the I03 beamline at Diamond Light Source. Neither showed any diffraction, but as these crystals were smaller and much more fragile than DHC-E crystals, this might have been due to difficulties in crystal harvesting.

Interestingly, these results suggest that the construct is most stable when α-helices H14, H15 and H16 (Fig. 4.14) are included, since constructs DHC-EA through DHC-ED, which eliminated these α-helices all failed to express at high levels. This was somewhat unexpected as these α-helices are apparently not ordered enough to be visible in the original DHC-E crystals. Furthermore, the crystallisation of DHC-EJ showed what a large effect including even three more amino acids can
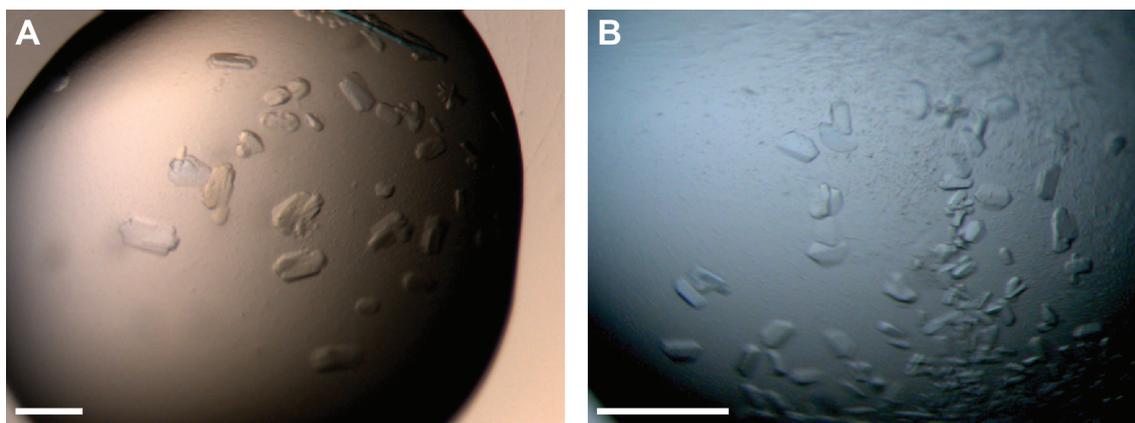
Figure 4.16: **Crystal hits from two new E region truncations, DHC-EJ and DHC-EK.** Each construct was purified from 1L of *E.coli*, concentrated to 5.3mg/mL and used to set up all 20 96-well plate LMB crystal screens. Each droplet was set with 200nL of protein and 200nL of reservoir solution. Plates were incubated at $\sim 20°$C and examined after 24 hours and one week for crystal growth. **A)** Crystals observed from DHC-EJ in a condition with a reservoir solution composed of 0.1M CAPS (pH = 10.5), 0.2M NaCl and 20% PEG (MW=8,000, w/v). **B)** Crystals observed from DHC-EK. These crystals grew in the same condition as those shown in **A**. Measure bars = 100μm.

have on crystallisation. DHC-EJ is just three residues longer than DHC-EI, yet it produced a single initial hit out of 20 LMB screens, which did not diffract, whereas DHC-EI produced dozens of initial hits, several of which diffract with sub-nanometer resolution.

Model-based DHC-E constructs

In an attempt to take advantage of the new information revealed by the DHC-E crystal structure, I made one last set of rationally designed constructs, which were based on observations made while building the model. The first observation is that the four and a half α-helices that are predicted to be at the C-terminus of DHC-E (H13-H17 in Fig. 4.14) were not ordered enough to be found within the electron density map. Although some of the shorter constructs in Table 4.11 eliminated three of these αhelices, none of the truncations went far enough to eliminate all of them. Therefore, I designed two new C-terminal truncations, both of which would terminate shortly after H12, the final α-helix that I was able to build in the DHC-E model. I called these constructs C-terminal A and C-terminalB (Table 4.12).

The second observation I used was that the DHC-E construct is sub-divided into two distinct domains. The N-terminal ~ 180 residues constitute a globular dimerisation domain (Fig. 4.9C), whereas the remainder forms an α-helical elongated domain (Fig. 4.9D). Furthermore, in the crystal structure, there are no apparent contacts between these two domains other than the protein backbone that joins them together. I hypothesised that the connection between these two domains may be flexible and hinder crystal formation. I also hoped that by crystallising even one of the two domains on its own, I would be able to build a higher resolution model for that domain, and that by incorporated this into the complete DHC-E model, the refinement would finally succeed for the entire DHC-E structure. Therefore I

Name	Last residue	MW (kDa)	Results
C-terminalA	1-411	47	X
C-terminalB	1-418	48	E
DimerA	1-177	20	E
ElongatedA	178-557	45	E,P
DimerB	1-181	21	E
ElongatedB	182-557	44	E,P,C,D

Table 4.12: **List of six model-based constructs.** The residues included and predicted MW after TEV cleavage are shown for each construct. In the results column, X=expression level was too low for purification, E=expression level was high enough to complete purification through SEC step, P=SEC produced a peak that was clean enough to set up crystal screens, C=crystal screens produced hits that were optimised and screened for diffraction, D=crystals diffracted.

designed two pairs of constructs that would express each domain separately. The first pair separated the two domains between residues 179 and 180 (DimerA and ElongatedA) and the second pair separated the two domains between residues 181 and 182 (DimerB and ElongatedB) (Table 4.12).

I grew 1L cultures for each of these six model-based constructs and purified the proteins. C-terminalA did not express at high enough levels to complete the purification. DimerA and Dimer B, as well as C-terminalB expressed fairly well, but did not perform well during SEC. Both elongated domain constructs expressed very well and produced prominent, single, symmetrical peaks during SEC. Therefore, I concentrated both samples to 5.3mg/mL and set up all 20 LMB crystal screens with each sample, using the same conditions described above for the E region truncations.

ElongatedA failed to produce any crystal hits, but ElongatedB produced one promising hit (Fig. 4.17). The condition, LMB9/B10 had a reservoir solution composed of 0.1M MES (pH = 6.5), and 10% PEG (MW=20,000, w/v). Using the in-house X-ray generator I confirmed that the crystals were protein and produced some reflections at $\sim$ 9Å. Using the same strategies I described in Section 4.2.4, I optimised the ElongatedB crystals. The initial hit produced extremely thin, 2-dimensional plates, which were difficult to handle. After optimisation, I found a condition (0.1M Tris (pH = 8.5) and 15% PEG (MW=10,000) that produced slightly thicker, more durable crystals. When screened for diffraction using synchrotron radiation, these crystals produced spots as far as 2.8Å, but the highest resolution I ever observed after processing a complete dataset was 3.3Å.

I truncated the model from Section 4.2.6 to include only the residues that constitute the ElongatedB construct. Using this as a search model, I performed molecular replacement with the 3.3Å ElongatedB dataset. However, all attempts to phase the ElongatedB crystals failed. I used several different datasets and processed the datasets in various ways using different programs. For example, I tried integrating with both XDS and iMOSFLM. I also tried imposing different sized unit cells, as XDS sometimes calculated a unit cell that was twice the size of the one that iMOS-FLM calculated, possibly indicating non-crystallographic pseudosymmetry. I also made several variations of the search model by including more or fewer α-helices, side-chains, loops, etc. I put all of these different datasets and search models into
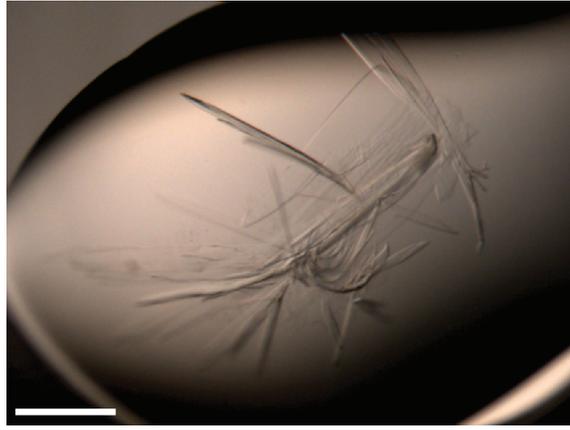
Figure 4.17: **Crystal hits of the ElongatedB construct, which is the equivalent of DHC-E without the dimerisation domain.** The construct was purified from 1L of *E.coli*, concentrated to 5.3mg/mL and used to set up all 20 96-well plate LMB crystal screens. Each droplet was set with 200nL of protein and 200nL of reservoir solution. Plates were incubated at ∼ 20°C and examined after 24 hours and one week for crystal growth. Crystals formed in a condition with a reservoir solution composed of 0.1M MES (pH = 6.5), and 10% PEG (MW=20,000, w/v). Measure bar = 100μm.

various molecular replacement programs, experimenting with different parameters, but I never obtained a clear solution. I looked at the models and electron density produced by some of the highest scoring solutions, but there was always a problem either with the crystal packing or the density itself was not realistic. This led me to believe that there may be a significant conformational change between the elongated domain in the DHC-E construct and the ElongatedB construct. Therefore, I decided to phase the ElongatedB crystals experimentally, using the same SeMet approach that was successful for the entire DHC-E construct. I successfully purified and crystallised a SeMet derivative of ElongatedB. A fluorescence scan of the crystals showed that SeMet was successfully incorporated, but none of the SeMet crystals diffracted beyond 6Å. As I have only purified the SeMet derivative of ElongatedB and crystallised it once, this could be due to day-to-day variations in the protocol and may not indicate an inability of the SeMet protein to crystallise as well. However, due to time constraints I have not been able to reproduce the experiment and/or optimise the crystallisation of the SeMet sample in order to collect higher resolution anomalous datasets.

Another possible explanation for why molecular replacement failed, which I could not exclude at that point in time, was that the DHC-E model itself had serious errors. Based on the experiments using REFMAC5 that I described in Section 4.2.6, this did not seem likely, but the low resolution of the map and the very high R-factors left it an open question. However, at the same time I was solving the structure of the DHC N-terminus using X-ray crystallography, my colleagues in the Carter lab were solving the structure of dynactin as well as a dynein tail/dynactin complex using cryo-EM. This was relevant to my work for two reasons. First, their cryo-EM maps of the dynein tail agreed with the model I had built using X-ray crystallography, which provided some validation for the DHC-E model I had built. Second, the cryo-EM structure of the dynein/dynactin complex revealed the interface between these

124

two large complexes for the first time, and in doing so, shed light on why the DHC N-terminus is required to take the 3-dimensional shape I had observed.

4.2.8 The DHC N-terminus at the dynein/dynactin interface

Around the same time that I was building the initial model of the DHC-E construct, my colleagues in the Carter group were solving a cryo-EM structure of the dynactin complex purified from pig brains (Fig. 4.18). They showed that dynactin is made up of 23 subunits, most of which are involved with forming and capping an actin-like filament. The remaining subunits, including p150[Glued], form the shoulder, a subcomplex that wraps around part of the filament. They ultimately achieved an overall resolution of 4.0Å, reaching 3.5Å in the interior of the rigid filament.

Before the dynactin structure was solved, two other studies had already shown that dynein and dynactin only form a stable complex in the presence of an adaptor, like BICD2N. Furthermore, only after forming this dynein/dynactin/BICD2N triple-complex (DDB), does dynein become highly processive (Schlager *et al.* 2014, McKenney *et al.* 2014). After the success my colleagues had in solving the dynactin structure by cryo-EM, they set out to solve the structure of the DDB complex as well. DDB proved an unsuitable sample for single-particle cryo-EM due to the inherent flexibility that the linker provides between the dynein motor domains and the tail. So they focused instead on the triple-complex of a recombinant human dynein tail, pig brain dynactin and a recombinantly mouse BICD2N (TDB). In fact, the human tail complex, which was expressed in Sf9 cells, was designed using a DHC truncation based on the one I had previously discovered in yeast when making the DHC-NH tail construct Section 2.2.4).

The TDB complex was sufficiently stable that my colleagues were able to solve the structure by cryo-EM (Fig. 4.19), although the overall resolution of 8.2Å was lower than that of dynactin alone (4.0Å). Nevertheless, the TDB structure provided the first glimpse of the dynein/dynactin interface, and immediately provided new insights into DDB function. For example, BICD2N runs along the entire length of the dynactin filament, and makes extensive contacts with both dynein and dynactin. This explains why dynein and dynactin fail to assemble *in vitro* in the absence of
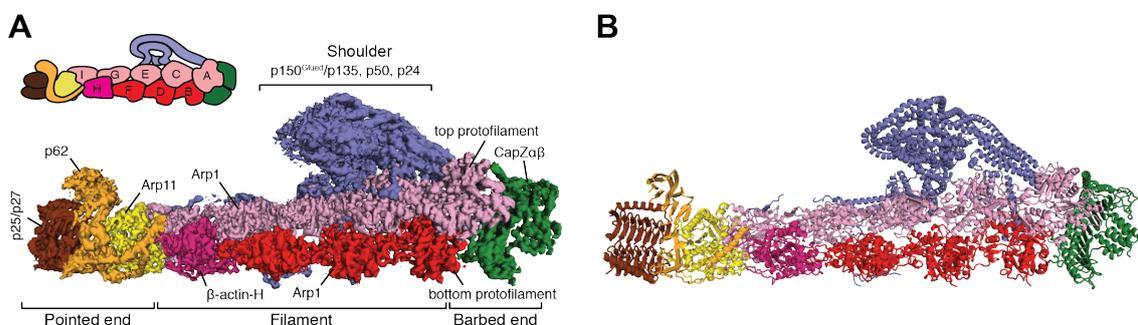
Figure 4.18: **Cyro-EM structure of the dynactin complex. A)** A 4.0Å cryo-EM map of dynactin segmented and coloured according to its components (EMDB accession code 2857). **B)** A molecular model of dynactin (PDB accession code 5AFT). Figure adapted from Urnavicius *et al.* 2015 with permission of the author.
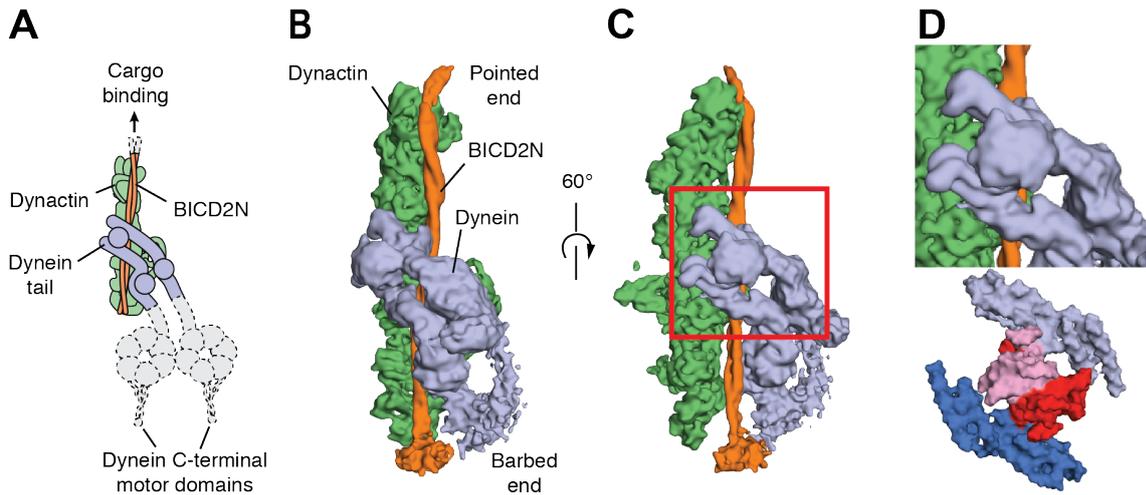
Figure 4.19: **Cyro-EM structure of the dynein tail/dynactin/BICD2N complex (TDB). A)** A cartoon model of the TDB complex. **B)** An 8.2Å cryo-EM structure of TDB (EMDB accession code 2860). **C)** Another view of the cryo-EM structure of TDB. The red box outlines a region of the dynein tail that resembles the DHC-E crystal structure. **D)** A closer look at the boxed region from **C** and a surface representation of the DHC-E crystal structure (PDB accession code 5AFR), showing the similarity between the two structures. Figure adapted from Urnavicius *et al.* 2015 with permission of the author.

BICD2N or other adaptors (Schlager *et al.* 2014, McKenney *et al.* 2014). However, the limited resolution of the TDB structure made it difficult to interpret the dynein tail region, the majority of which was even lower than 8.2Å resolution. The low resolution cryo-EM map wasn't sufficient to positively identify the C-terminus or N-terminus of the dynein tail, as no labels had been included. Nor was it possible to assign density to the different subunits within the tail complex with any amount of certainty. For example, there were three globular domains visible, but it was not clear whether they were made up of the DHC itself or the accessory subunits (DIC, DLIC and DLCs) (Fig. 4.19).

Before the TDB structure was solved, I had already built the DHC-E model based purely on the X-ray density map. But once the TDB structure was solved, even a cursory glance at the region of the dynein tail closest to dynactin and BICD2N was enough to see that it bore a resemblance to the DHC-E crystal structure (Fig. 4.19D). There were two elongated domains bridged together by a globular domain and the dimensions of and spacings between these domains were similar to those of the DHC-E model. By concluding that this part of the EM structure corresponds to the DHC-E model, we identified one of the three globular domains (shown in red in Fig. 4.20A) as the dimerisation domain, and therefore the N-terminus of the DHC. Consequently, we concluded that the other end of the dynein tail was formed by the C-terminus. By modelling the DIC WD40 domains using Phyre2 (Kelley & Sternberg 2009), we were able to show that the DIC WD40 domains are roughly the correct size and shape to fit into the other two globular domains at this end of the dynein tail (Fig. 4.20B).

However, there was one clear difference between the DHC-E crystal structure and the cryo-EM map of the dynein tail. In the crystal structure, the dimer contains a
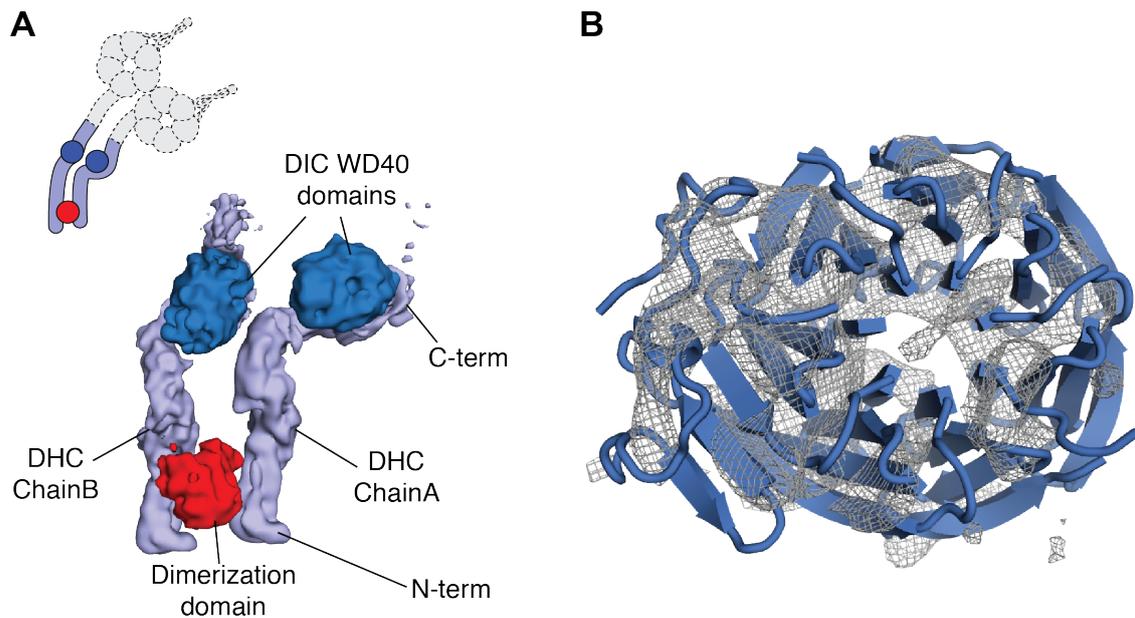
Figure 4.20: **Assignment of density within the cryo-EM structure of the dynein tail. A)** The dynein tail density from has been segmented out of the TDB structure (EMDB accession code 2860) and coloured to highlight different domains. The similarity between the cryo-EM map and DHC-E crystal structure shown in Fig. 4.19D allowed the assignment of the dynein tail N-terminus and C-terminus as well as the dimerisation domain. **B)** The density containing the DIC WD40 domains were assigned by showing that a homology model could fit into the cryo-EM density. Figure adapted from Urnavicius *et al.* 2015 with permission of the author.
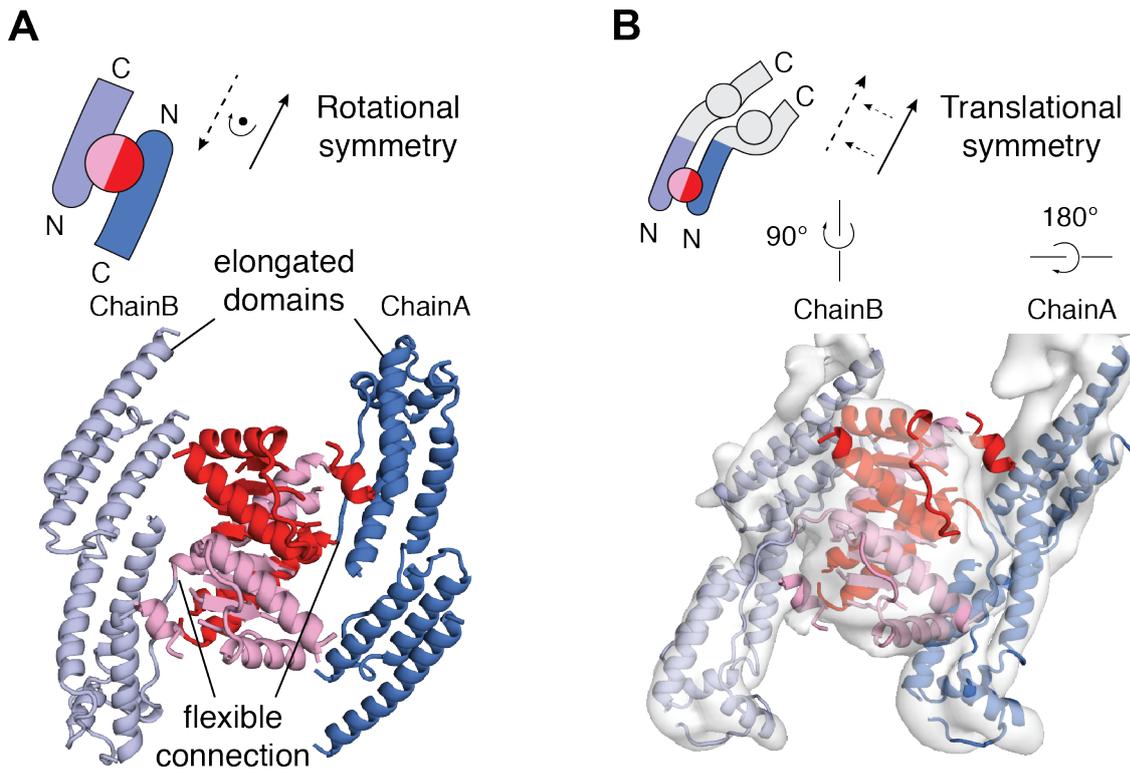
Figure 4.21: **The DHC-E crystal structure modelled into the TDB cryo-EM electron density map. A)** The DHC-E crystal structure (PDB accession code 5AFR), which clearly displays a two-fold rotational symmetry around the dimerisation domain. The cartoon model shows that in ChainA, the N-terminal part of the elongated domain is oriented toward the top of the figure, while in Chain B, it is oriented toward the bottom of the figure. **B)** The DHC-E crystal structure after each elongated domain has been separated from the dimerisation domain and rotated in order to fit into the TDB cryo-EM density (EMDB accession code 2860). The cartoon model shows that now the N-termini of both elongated domains are oriented toward the bottom of the figure and both C-termini are oriented toward the top. Figure adapted from Urnavicius *et al.* 2015 with permission of the author.

2-fold rotational symmetry, with the two elongated domains positioned anti-parallel to each other (Fig. 4.21A). On the other hand, the dynein tail observed by cryo-EM displays a translational symmetry, with the two DHCs running parallel to each other (Fig. 4.20A). Even when building the DHC-E model for the first time, it seemed likely that the rotational symmetry I observed was a crystal-packing artifact; in order for dynein to function, the two motor domains must be oriented in the same direction so that both microtubule-binding domains can reach the same microtubule. However, with the crystal structure alone, I could not exclude the possibility that the DHCs might turn toward each other after the first 400 residues that I could see in the crystal structure, providing the translational symmetry required for dynein motility.

A comparison between the DHC-E crystal structure and the TDB cryo-EM structure revealed how the elongated domains are arranged within the context of a processive motor. Attempts to fit the DHC-E crystal structure into the cryo-EM density were only successful once one of the elongated domains was rotated 180°, so that it
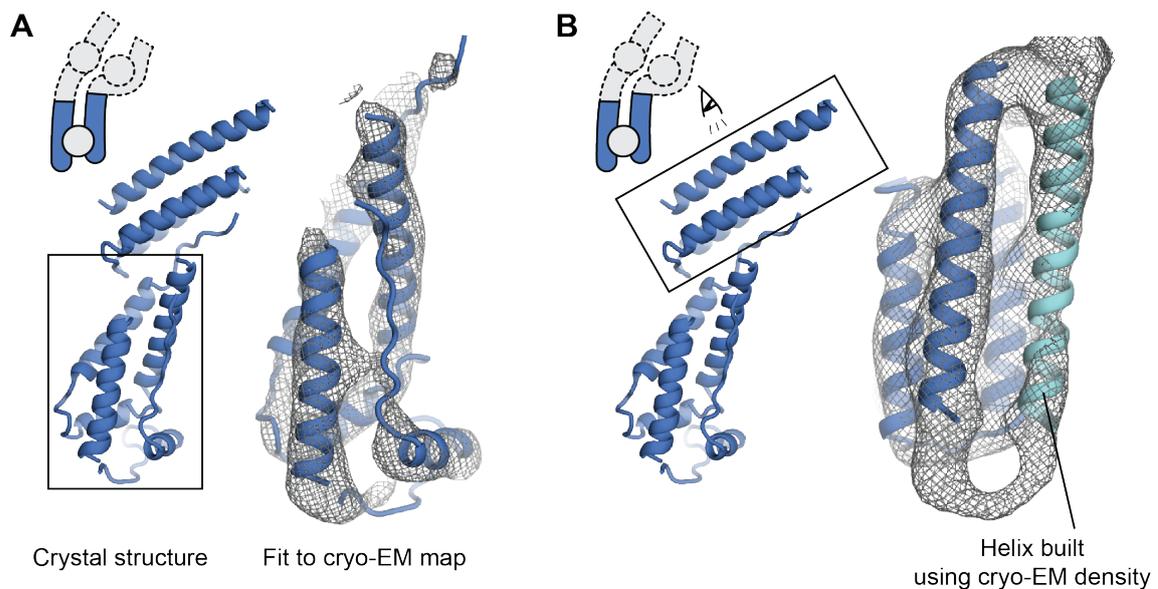
Figure 4.22: **The elongated domains of the DHC-E crystal structure after being fitted into the TDB cryo-EM electron density map. A)** One of the α-helical bundles from the elongated domain of the DHC-E crystal structure (PDB accession code 5AFR) after fitting into TDB cryo-EM density (EMDB accession code 2860). **B)** The other α-helcial bundle from the elongated domain after fitting into TDB cryo-EM density. In this case, the cryo-EM map even contained clear density for one more α-helix, which only showed weak density in the X-ray map. Figure adapted from Urnavicius *et al.* 2015 with permission of the author.
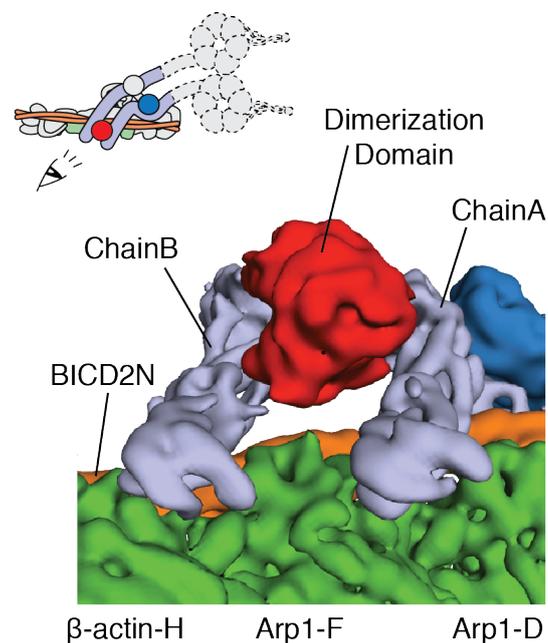
Figure 4.23: **The interface between the DHC N-terminus and the Arp1 filament of dynactin.** The elongated domains of the DHC N-terminus slot into adjacent grooves within the Arp1 filament of dynactin. The translational symmetry and distance between the elongated domains of the DHC correlates precisely with the translational symmetry and distance between Arp1 monomers within the dynactin filament. Figure adapted from Urnavicius *et al.* 2015 with permission of the author.

129

would run parallel to the other one (Fig. 4.21B). Once this rotation was performed, the cryo-EM density fit the α-helices of the DHC-E elongated domains extremely well (Fig. 4.21B, Fig. 4.22A). This provided further evidence for the idea that the link between the dimerisation domain and the elongated domains is flexible, and that the elongated domains can rotate freely with respect to the dimerisation domain. The electron density in this portion of the cryo-EM map was so clear that we were able to model in one α-helix that I had not yet modelled due to weak density in the X-ray maps (Fig. 4.22B). While the dimerisation domain roughly matched the size and shape of the globular domain observed the in the cryo-EM map, the density was not clear enough to see individual secondary structure elements.

The most unexpected discovery we made by combining the DHC-E crystal structure and TDB cryo-EM structures was the extensive interface that exists between the DHC N-terminus and dynactin. Before these two structures were solved, the most important dynein/dynactin interaction was thought to occur between the DIC and p150[Glued] (McKenney *et al.* 2011). No interactions had been mapped between the N-terminal 500 residues of the DHC and dynactin, yet this is where we observed the majority of the contacts between dynein and dynactin and BICD2N (Fig. 4.23).

In conclusion, the agreement between the DHC-E crystal structure and the TDB cryo-EM structure provided a means of validating the accuracy of both structures and provided insights that could not be made from either structure alone. It was only by taking advantage of the DHC-E crystal structure that it was possible to assign specific dynein subunits within the TDB cryo-EM structure. Likewise, the cryo-EM map confirmed that the rotational symmetry observed in the DHC-E crystal structure is most likely a crystal artifact, as dynein must adopt a translational symmetry in order to bind dynactin and BICD2N. Finally, all of this taken together led to the unexpected revelation that the DHC N-terminus is responsible for a large number of the contacts between dynein, dynactin and BICD2N in the highly processive DDB complex.

4.3 Discussion

4.3.1 The first crystal structure within the DHC tail region

The crystal structure of DHC-E was the first to be solved for any part of the DHC N-terminus. Previously, residues 1363-4092 in *S. cerevisiae* and their equivalent from *D. discoideum* had been crystallised (Schmidt *et al.* 2012, Kon *et al.* 2012). But none of the N-terminal 1362 residues (about one third) had been crystallised, leaving the cargo-binding domain unknown. The DHC-E crystal structure only allowed me to build a model for $\sim$ 400 of these residues, but even from this relatively small fraction of the DHC some interesting features could be seen for the first time.

The most prominent feature of the DHC-E crystal structure is the intricate $\sim$ 180 amino acid dimerisation domain that forms at the very N-terminus of the DHC (Fig. 4.9A,B,C). This dimerisation domain was previously unknown and completely unexpected. Several years ago, at least one study attempted to map a homodimerisation domain within the DHC (Habura *et al.* 1999). However, their chemical cross-linking and *in vitro* affinity methods failed to detect any homodimerisation activity at the very N-terminus of the DHC. They instead mapped the key homodimerisation site to residues 627-780 in *D. discoideum*. According to my alignment of 13 different

cytoplasmic DHCs (Fig. 4.12), this region is equivalent to residues 527-681 in *S. cerevisiae*. This corresponds to the region of the DHC just C-terminal to what I was able to build in the crystal structure (residues 1-412). The cryo-EM TDB structure includes the part of the DHC that is equivalent to this reported homodimerisation site, which may overlap with the DIC binding sites (Fig. 4.20A). From the EM structure, the DHCs don't appear to contact each other at all after the N-terminal dimerisation domain, but we can't exclude the possibility that this is only true within the TDB complex. So the dimerisation site reported by Habura *et al.* 1999 may exist only when dynein is on its own.

The presence of this unexpected dimerisation domain has complicated the already poorly understood question of how dynein homodimerises, . As I discussed in Section 3.3.1, one of the most popular models in recent years has been that the DLCs dimerise while also binding to the DICs, which themselves are bound to the DHCs, thus the DLCs drive the homodimerisation of the entire dynein complex. However, as I have shown in Section 3.2.4, even complexes that lack the region of the DIC that binds the DLCs, and therefore lack the DLCs completely, purify as soluble and well-behaved homodimers when analysed by SEC. The presence of the DHC dimerisation domain may answer the question of how dynein was able to homodimerise in the absence of DLCs, but it raises further questions. For example, what is the importance of the DHC dimerisation domain relative to the DLCs? What happens when only one or the other is present? If dynein lacks both of these mechanisms, will it still dimerise? Later in the discussion, I propose some future experiments to try to answer these questions.

Another prominent feature that came from the crystal structure is the clear domain boundary between the dimerisation domain and the elongated domains. The domains appear to not make any contacts, other than through the peptide backbone that links them together into a single peptide (Fig. 4.11D), which suggested that these two domains may be able to rotate freely with respect to one another. This idea became even more plausible when we realised that the elongated domains must rotate around this flexible connection in order to fit into the cryo-EM density of the TDB map (Fig. 4.21). It is also worth noting that the elongated domain, which has a much higher amount of sequence conservation than the dimerisation domain (Fig. 4.12) corresponds well with the DHC-N1 domain family predicted by the Pfam database (Finn *et al.* 2013).

4.3.2 The synergistic effect of combining the DHC-E crystal structure with the TDB cryo-EM structure

This project was yet another example of how approaching the same complex using multiple structural techniques can have a synergistic effect, which occurs when the strengths from one technique fill in the gaps left by another. In this case, the crystal structure was slightly higher resolution than the cryo-EM map. But perhaps even more important in this case was the fact that the crystal structure included a smaller fragment of a single, subunit. This helped in the assignment of density within the cryo-EM map, since at 8.2Å, it was impossible to be sure where specific subunits of the full dynein complex should have been placed in the density (Fig. 4.19). At the same time, the cryo-EM structure helped resolve the question of rotational symmetry versus translatinoal symmetry between the two DHC elongated domains

(Fig. 4.21). It also provided a biological context for the DHC-E construct, and quite unexpectedly showed that this part of the dynein complex plays a key role in the dynein/dynactin interface.

The hybrid crystal/cryo-EM TDB structure also answered some questions that first arose when I calculated an initial 3D model of the dynein tail complex using negative stain EM (Section 2.2.8). First, it established the orientation of the dynein tail more conclusively, which is actually the opposite from what I first supposed, based on the negative stain model (Fig. 2.25). It also allowed us to assign the correct subunits to the three distinct domains that were visible in the negative stain 2D class averages.

However, this assignment added another twist to the already perplexing DHC dimerisation domain. As I described in Section 2.2.8, one of the three domains is only an intact dimer in some class averages, and in others, it is separated, giving the complex the appearance of "fraying" at one end. I initially speculated that this might be the C-terminus of the DHC, which was flexible due to the truncation. However, we now know that this density corresponds to the DHC N-terminus. So in some classes, the DHC dimerisation domain appears not to be intact, even when the rest of the tail remains tightly associated. This may mean that neither the DHC dimersiation domain nor the DLCs are necessary for homodimerisation and that the dimer interface is extensive. However, the cryo-EM structure does not seem to support this, as the two dynein monomers appear fairly separate aside from the DHC dimerisation domain (Fig. 4.20A). In agreement with the TDB cryo-EM map, but in contradiction to my own negative stain data, a study that was only just published shows many negative stain 2D class averages of the full-length dynein complex with the two DHCs completely splayed apart (Chowdhury *et al.* 2015). In this conformation, the only points of contact between the two dynein monomers appear to be at the DHC N-terminus and one of the three DLCs (LC7/Roadblock). For now, it remains something of a mystery how all of these results can be accounted for by a single model. It is also worth noting that these studies all used different constructs of dynein purified from different species, and analysed them using different techniques.

For years, one of the big questions about dynactin has been why the primary activator of dynein, a microtubule motor, is built around an actin-like filament. The combined knowledge gleaned from both structures finally suggests an explanation. As the rotation of the DHC elongated domains relative to the dimerisation domain demonstrated, the dynein tail on its own is quite flexible and capable of adopting many different conformations. Further evidence of this was very recently provided by the negative stain EM work done by Chowdhury *et al.* 2015. However, as microtubules exhibit translational symmetry both along their length and around their circumference, it follows that the two microtubule binding domains within a single dynein motor must also exhibit translational symmetry in order to take steps in a single direction and make forward progress. It seems that the best mechanism the cell has for imposing translational symmetry on a molecule is simply to bind it to another filament. So in a sense, the requirement that dynein has translation symmetry in order to walk along a microtubule is satisfied by binding it to the actin-like filament within dynactin.

Finally, the TDB structure structure shed some light on why some of the experiments I describe in Chapter 3 didn't work. In Section 3.2.5 I present my efforts

to bind a fragment of Nip100 (the yeast homologue of p150$^{\text{Glued}}$) to the dynein tail complex. This experiment was based on the consensus at the time, which was that the primary interface between dynein and dynactin occurs between the N-terminus of the DIC and a coiled-coil region of p150$^{\text{Glued}}$. However, as the TDB structure revealed, there are extensive contacts between the DHC and the Arp1 filament (Fig. 4.23). Neither the p150$^{\text{Glued}}$ coiled-coil domains nor the DIC N-terminus were visible in the TDB structure, so it is still possible that the two do interact. But even if they do, that interaction now seems less likely to have an effect on the overall stability of the dynein tail than binding the entire dynactin complex would have. Therefore, any future experiments attempting to stabilise the dynein tail with dynactin should probably make use of the entire complex.

4.3.3 Future directions

The most obvious next step would be to continue my efforts to improve the resolution of the DHC-E crystal structure. Many of the strategies I described in Section 4.2.7 showed promise and a more focused effort on any one of these strategies may provide the quality of data necessary to build a complete atomic structure. Perhaps the first priority would be to continue crystallising the SeMet derivative of the ElongatedB construct, as the native crystals did show some diffraction spots as far as 2.8Å. I could also try phasing these crystals using heavy atom soaks, or even by performing molecular replacement using the appropriate segment of the cryo-EM electron density map as search model. Improving the model of the elongated domains to atomic resolution would improve the refinement of the entire DHC-E structure, which might even be enough for the density in the dimerisation domain to reach atomic resolution as well.

It would also be worthwhile to revisit some of the other constructs I tested, or even to design new ones. For example, the DHC-EJ and DHC-EK constructs were capable of crystallising, but I only had time for a brief attempt at optimising the initial crystals. Investing a little more time and effort in optimising these crystals may yield even crystals that are even better than the original DHC-E ones. Furthermore, an even finer screen around the DHC-E truncation may be useful. I have demonstrated on two separate occasions that changing a construct by less than 5 amino acids can be the difference between crystals that diffract to better than 3.5Å and no diffraction (DHC-EI vs. DHC-EJ) or even no crystals (ElongatedA vs. ElongatedB) at all. Therefore, a screen that tests, for example, every possible C-terminal truncation of the DHC between amino acids 550 and 560 might produce a construct with a significantly higher diffraction power than the original DHC-E.

Another possibility is to revisit the initial goal I set in Section 2.1 of solving the structure of the entire yeast tail complex. The crystal structure of DHC-E (residues 1-557 of the DHC) only accounts for a part of the DHC-NH tail construct (residues 1-1004) described in Chapter 2. So it might be worth trying to solve the remainder of this complex by crystallography. As the DHC-E truncation seems to have found a clear boundary domain, it might be possible to make the inverse truncation, expressing only residues 558-1004 of the DHC along with the full length accessory subunits. One explanation for why the DHC-E construct crystallised while the whole tail complex did not is that there is a flexible connection between the DHC-E construct and the remainder of the tail complex. If separating the N-

terminal half from the rest is what allowed the DHC-E construct to crystallise, it is conceivable that the C-terminal half might crystallise on its own as well.

One simple, yet interesting experiment would be to try and form the TDB triple complex using only the human equivalent of the DHC-E construct rather than the whole tail complex. I have already purified large amounts of this construct when I attempted to crystallise it. Since the cryo-EM structure of the TDB complex suggests that the majority of the contacts that dynein makes with both dynactin and BICD2N occur through the elongated domains at the DHC N-terminus, this construct may still be able to form the TDB complex, even without any of the dynein accessory chains. Not only would this test whether the rest of the tail complex is necessary for complex formation, but if not, the resulting complex might be more rigid, and therefore a more suitable sample for cryo-EM. Since the DHC C-termini and all of the accessory chains were very flexible in the TDB map, a complex that includes only the DHC-E construct of dynein might result in more uniform particles and a higher resolution for the overall map.

Now that we have a better understanding of how mammalian dynein binds dynactin and BICD2N to become processive, the question of how this process occurs in yeast becomes even more complicated. There aren't any known BICD2 homologues in yeast, but the one known cargo, Num1, is predicted to contain a long coiled-coil motif in its N-terminus, the same part that is responsible for dynein binding (X. Tang *et al.* 2012). Therefore, one could imagine that in yeast Num1, which is found at the cell cortex, recruits dynein and dynactin and allows them to assemble into a triple complex similar to TDB. The situation is not a perfect parallel because unlike mammalian dynein, yeast dynein is processive on its own, although dynactin has been shown to have the effect of increasing processivity (Kardon *et al.* 2009). This study used endogenously expressed dynein and dynactin, purified from yeast, meaning it is possible there was also an endogenous cargo adaptor, like Num1, present when extra-processive dynein/dynactin complexes were formed. So it might be revealing to reproduce these results using recombinant full-length dynein, recombinant dynactin, both in the presence and absence of recombinant Num1. If a yeast dynein/dynactin complex can be formed without any adaptor, it would also be interesting to solve its structure by cryo-EM for a comparison with the mammalian version. Any differences between the two structures might answer the question of why yeast dynein is processive on its own, whereas mammalian dynein requires BICD2N and dynactin.

Finally, one way to test the relative importance of the DHC dimerisation domain versus the DLCs would be to make a mutant yeast strain that lacks the first 181 residues of the DHC, and compare it to the DIC-H truncation described in Section 3.2.4, which lacks the DLC binding sites of the DIC. I would also make a strain that combines these two mutations, to see if this prevents homodimerisation completely or if there are even more mechanisms in place. Rescue strains could be created by replacing the DHC dimerisation domain and the DIC N-terminus with a glutathione S-transferase (GST) domain, or other known homodimer domain. As a background for each of these mutant strains I would use CY286, the strain I described in Section 2.2.2, which overexpresses the full-length dynein complex. That way I would be able to purify the mutant complexes and characterise them biochemically. After checking whether each mutant forms a dimer or monomer by SEC-MALS, their motility could be measured using an *in vitro* motility assay. The strains could also be analysed *in*

vivo by testing each mutant in a spindle osciliation assay. This is a sensitive way to detect a defective dynein phenotype by measuring how rapidly the spindle oscillates through the bud neck of hydroxyurea arrested yeast cells (Toropova *et al.* 2014).

4.3.4 Conclusions

In this chapter I have presented the crystal structure of the DHC-E construct, which constitutes the DHC N-terminus. I described each step of the structure solving process, including purification, initial crystal screening, crystal optimisation, data collection, experimental phasing and model building. This was the first crystal structure of any part of the DHC tail domain. It revealed that the N-terminal ~ 180 residues of the DHC form a previously unknown dimerisation domain. Furthermore, this crystal structure was fit into and allowed the interpretation of a cryo-EM structure of the dynein tail/dynactin/BICD2N complex. This hybrid structure provided the unexpected insight that dynein interacts with dynactin primarily through the DHC N-terminus. The combination of these two structures has provided a first glimpse at how dynein adopts its processive conformation by binding dynactin and one known adaptor, thus paving the way for future studies to probe how cytoplasmic dynein interacts with diverse cargoes in order to carry out its myriad roles within the cell.

Chapter 5

Methods

5.1 Molecular Biology

5.1.1 General techniques

Polymerase chain reaction (PCR)

Genes were amplified using the Phusion High-Fidelity DNA Polymerase kit (New England BioLabs) according to the manufacturer's protocol. A standard 50μL reaction volume contained 10μL of 5X Phusion HF buffer, 5μL of 2mM dNTPs, 0.25μL forward primer, 0.25μL reverse primer, < 250ng of template DNA, 0.5μL of Phusion polymerase and enough ddH$_2$O to bring the total volume to 50μL. The mixture was vortexed, centrifuged briefly and put in a thermocycler set to run the following program with 35 cycles of denaturation, annealing and extension:

Step	Temperature (°C)	Time
Initial denaturation	98	30s
Denaturation	98	10s
Annealing	68	30s
Extension	72	30s/kb
Final extension	72	10min

PCR stitching was performed similarly, but with an equal amount of two templates that overlapped by $\sim$ 20bp.

Verification and gel extraction of PCR products

Each 50μL PCR product was mixed with 5μL of 10X DNA loading buffer (3.7M bromphenol blue, 50mM Tris-HCl (pH = 7.6) and 60% glycerol (v/v)). Samples were then run in a 1% agarose gel containing SYBR Safe DNA gel stain (Thermo Fisher Scientific) at 120V for 30 to 60min. At least one lane was loaded with 10μL of the 1Kb+ DNA Ladder (Thermo Fisher Scientific) in order to confirm the approximate length of PCR products. A blue light was used to visualise DNA bands, which were excised using a scalpel. DNA was purified using QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's protocol. Each excised sample was added to three volumes of QG buffer and incubated at 50°C until agarose was fully dissolved. DNA was bound to a QIAquick spin column by applying and

centrifuging at $17,900g$ for 60s. The column was washed by adding $750\mu L$ of PE buffer and centrifuging twice more at $17,900g$ for 60s each time, removing flow-through between spins. DNA was eluted by adding $30\mu L$ of EB buffer or ddH$_2$O to the spin column, incubating for 60s, then placing the column in a 1.5mL centrifuge tube and centrifuging at $17,900g$ for 60s. Final DNA concentration was determined from a $2\mu L$ sample using a NanoDrop.

Plasmid purification from bacterial stocks

Plasmids were purified from bacterial stocks using QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's protocol. A single colony of bacteria was used to inoculate 10mL of 2X TYE medium containing the antibiotic for which the plasmid of interest provides resistance. This culture was incubated at 37°C with 210RPM shaking overnight. 2mL of the overnight culture was decanted into a 2mL centrifuge tube and centrifuged at $8,000g$ for 3min. The supernatant was discarded and the pellet was resuspended in $250\mu L$ buffer P1. After adding $250\mu L$ buffer P2, the tube was gently inverted 6 times, or until the entire solution was a homogeneous blue. Then $350\mu L$ of buffer N3 was added and the tube was gently inverted another 6times, or until all the blue colour disappeared. The tube was centrifuged at $17,900g$ for 10min and the supernatant was recovered and applied to a QIAprep spin column, which was then centrifuged at $17,900g$ for 1min. The flow-through was discarded, and the column was washed by adding $750\mu L$ of PE buffer and centrifuging twice more at $17,900g$ for 60s each time, removing flow-through between spins. DNA was eluted by adding $30\mu L$ of EB buffer or ddH$_2$O to the spin column, incubating for 60s, then placing the column in a 1.5mL centrifuge tube and centrifuging at $17,900g$ for 60s. Final DNA concentration was determined from a $2\mu L$ sample using a NanoDrop.

Preparation of Gibson assembly mix (GA)

50mM NAD (AppliChem) was prepared by dissolving 33.2mg in 1mL MPW. Then 3mL of 1M Tris-HCl (pH $= 7.4$), $300\mu L$ of 1M MgCl$_2$, $60\mu L$ of $100\mu M$ dNTPs (Fermentas), $300\mu L$ of 1M DTT (Fermentas), 1.5g of PEG (MW $= 8,000$, v/v), $600\mu L$ of 50mM NAD and enough ddH$_2$O to bring the total volume to 6mL were mixed together. The resulting isothermal assembly buffer (IAB) was divided into $320\mu L$ aliquots and stored at -20°C until ready for the next step. To each thawed $320\mu L$ aliquot of IAB, 1.2μ T5 exonuclease (New England BioLabs), $20\mu L$ Phusion polymerase (New England BioLabs), $160\mu L$ Taq ligase (New England BioLabs) and $700\mu L$ MPW were added. The resulting Gibson assembly mix (GA) was divided into $15\mu L$ aliquots and stored at -20°C.

Ligation of PCR products into vectors via Gibson assembly

Acceptor vectors were linearised either by a double digest with two site-specific restriction enzymes or by PCR amplification. The approximate size of the linear vector was confirmed by gel electrophoresis and purified by gel extraction as described above. Inserts were designed to include 15 to 25bp overlap with each end of the linear vector and created using PCR as described above. In a total volume $< 5\mu L$, 100ng of insert and 100ng of linear vector were added to $15\mu L$ of GA mix.

The final mixture was vortexed, centrifuged and incubated at 50°C for 30min. The ligated product was either frozen at −20°C for storage or transformed into bacteria immediately.

Preparation of competent bacterial cells

A single colony of the required strain was used to inoculate a 3mL culture, which was incubated at 37°C with 210RPM shaking overnight. The overnight culture was used to inoculate 25mL of SOB (VWR) in a 250mL polycarbonate Erlenmeyer flask, which was incubated at 37°C with 210RPM shaking for 12 to 14hr. 2mL of this culture was used to inoculate 250mL of SOB++ (SOB plus 10mM $MgCl_2$ and 10mM $MgSO_4$) in a 1L polycarbonate Erlenmeyer flask. This culture was incubated at 20°C with 210RPM shaking until the OD_{600} was between 0.3 and 0.6. The flask was then put on ice for 60min and the cells were pelleted by centrifuging in detergent free bottles at $4,200g$ and 4°C for 10min. The pellets were gently resuspended in 80mL CCMB80 buffer (10mM KOAc (pH = 7.0), 80mM $CaCl_2$, 20mM $MnCl_2$, 10mM $MgCl_2$ and 10% glycerol (v/v)) and incubated on ice for 20min. The cells were pelleted again by centrifuging at $4,200g$ and 4°C for 10min and resuspended gently in 10mL of CCMB80 buffer. The OD_{600} of this suspension was adjusted to 5.0 and the cells were divided into 100μL aliquots and stored at −80°C.

Bacterial transformation of GA reactions or purified plasmids

A 30μL aliquot of competent cells was thawed on ice. When the incubation was complete, 1μL of ligated GA reaction or purified plasmid was added to the competent cells. The tube was gently flicked and placed back on ice for 20min. Then the bacteria were heatshocked by placing the tube in a 42°C water bath for 45s and quickly moving it back on ice for another 2min. If the antibiotic marker being used required a recovery step, 900μL SOC medium (2% tryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_2$ and 20mM glucose) was added and the culture was allowed to recover at 37°C for 1 to 4hr. After recovery, the culture was briefly centrifuged to pellet cells, the supernatant was discarded, and the pellet was resuspended in 30μL sterile ddH_2O. The cells were then spread on TYE agar plates containing the appropriate antibiotic for selection of transformed bacteria.

Verification of new plasmids and storage as glycerol stocks

A sample of each plasmid constructed using the methods described above was sent for Sanger sequencing at Source Bioscience (Cambridge, UK) in order to confirm the sequence. Both forward and reverse primers were used to ensure sufficient redundancy between reads. Once a plasmid's sequence was confirmed, the remaining 8mL of the overnight culture used in the miniprep was centrifuged at $4,000g$ for 5min. The supernatant was discarded and the pellet was resuspended in 900μL of 20% glycerol (v/v). This suspension was stored at −80°C and used to inoculate a TYE agar plate for future experiments.

Transformation of PCR products or plasmids into yeast

A single colony of the acceptor strain of yeast was used to inoculate 10mL of YEPD media. The culture was incubated at 30°C with 200RPM shaking overnight. The OD_{600} of the overnight culture was determined and the required volume was used to inoculate 100mL of YEPD to an initial OD_{600} of 0.2. This culture was incubated at 30°C with 200RPM shaking until the OD_{600} was between 0.5 and 0.7. The cells were pelleted by centrifuging in two 50mL tubes at 1,000g for 5min and then resuspended in 20mL sterile ddH_2O and combined. They were pelleted again and resuspended in 1mL of LiAc TE buffer (100mM LiAc, 10mM Tris-HCl (pH = 7.5) and 1mM EDTA). For each transformation, $100\mu L$ of cell suspension was moved to a 2mL centrifuge tube. Either 20 to $50\mu L$ of PCR product or $1\mu L$ of purified plasmid was added to the cells. Then $5\mu L$ of pre-boiled salmon sperm DNA (ssDNA) and $600\mu L$ of PEG LiAc TE (LiAc TE with 40% PEG (MW = 3,350, v/v) were added. The tube was gently inverted several times and the cells were incubated in a 30°C water bath for 30min. $70\mu L$ of sterile DMSO was added and the tube was mixed again by gentle inversion and incubated in a 42°C water bath for 15min. The cells were pelleted by centrifuging at 4,000g for 5min and resuspended in $100\mu L$ sterile ddH_2O. The suspension was then spread on the appropriate selective plates and incubated at 30°C for 3 days. Single colonies were restreaked on selective plates and incubated at 30°C for another 3 days.

Colony PCR for verification of yeast strains and storage as glycerol stocks

After restreaking newly constructed yeast strains, single colonies were picked and resuspended in $10\mu L$ of 10mM NaOH. The suspension was incubated at 95°C for 30min and centrifuged briefly in a tabletop centrifuge to pellet cell debris. The supernatant was reserved as a genomic DNA template for colony PCR. Forward and reverse primers were designed to bind $\sim$ 250bp to each side of successful homologous recombination sites. In this way, a $\sim$ 500bp fragment would be amplified only if the transformation was successful. $25\mu L$ PCR reactions were set up using the Quick-Load *Taq* 2X Master Mix kit (New England BioLabs) according to the manufacturer's protocol. $1\mu L$ of genomic DNA template, $0.25\mu L$ of each primer (at $100\mu M$) and $11\mu L$ of ddH_2O were added to $12.5\mu L$ of Quick-Load *Taq* 2X Master Mix. The mixture was vortexed, centrifuged briefly and put in a thermocycler set to run the following program with 30 cycles of denaturation, annealing and extension:

Step	Temperature (°C)	Time
Initial denaturation	95	30s
Denaturation	95	15s
Annealing	60	30s
Extension	68	1min/kb
Final extension	68	5min

$10\mu L$ samples from each PCR reaction were analysed via gel electrophoresis as described above. A colony from any strains that were positive for all intended homologous recombination events was used to inoculate a 10mL culture, which was incubated at 30°C with 200RPM shaking overnight. The culture was then pelleted by centrifuging at $4,000g$ for 3min and the supernatant was discarded. The pellet was

resuspended in 900μL of 20% glycerol (v/v). This suspension was stored at -80°C and used to inoculate the appropriate selective agar plate for future experiments.

5.1.2 Cloning for expression in *S. cerevisiae*

Cloning for expression from pYES2 plasmids

Dynein genes were amplified by PCR using genomic DNA from the wild type yeast strain W303 (CY176, Fig. A.4) as a template. Inserts were assembled by combining fragments via PCR stitching and then ligated into the linearised pYES2 vector by Gibson assembly. The Gibson reaction products were transformed into *E. coli* and plated on ampicilin containing TYE plates for selection of successfully ligated plasmids. Bacterial colonies were used to inoculate 10mL overnight cultures, and then minipreps were performed to isolate plasmid DNA. Plasmid sequences were confirmed by Sanger sequencing. Successful clones were added to the Carter lab plasmid library as glycerol stocks and documented in the database. A list of all plasmids described in this thesis can be found in the Appendix (Figs. A.5 and A.6).

Plasmids were transformed into expression strains such as DSY-5 (CY223, Fig. A.4) and BCY123 (CY282, Fig. A.4). Yeast was plated on Ura- selective plates and incubated at 30°C for 3 days. Single colonies were restreaked on Ura- plates and incubated for another 3 days at 30°C. Single colonies from this plate were used to inoculate 10mL overnight cultures, either to make a glycerol stock of the strain containing the plasmid, or to begin an expression immediately.

Cloning for expression from endogenous genes

The commercial strain, DSY-5 (CY223, Fig. A.4) was used as a background strain. Linear DNA fragments were assembled using PCR to amplify individual segments and PCR stitching to join multiple fragments together. For homologous recombination, linear DNA fragments were designed to include 500bp of overlap with the genomic sequence on both sides of the site of recombination. Auxotrophic marker genes used for selection were amplified from closely related fungal species other than *S. cerevisiae*. This was to ensure that the gene product would be functional within the *S. cerevisiae* biosynthetic pathway, but would not have a similar enough DNA sequence to favour homologous recombination with the endogenous auxotrophic marker gene. For example, when using a background strain with the common *ura3-52* point mutation for selection, the functional *URA3* gene from *K. lactis* was included in the linear DNA fragment.

Once constructed and purified, linear DNA fragments were transformed into the background strain and yeast was plated on auxotrophic selective plates. After three days, plates were checked for colonies and successful clones were restreaked on a fresh selective plate. After three more days, a single colony was picked and tested by colony PCR and/or Sanger sequencing to verify the presence of the intended genomic modification. Successful clones were added to the Carter lab yeast library as glycerol stocks and documented in the database. A list of all yeast strains described in this thesis can be found in the Appendix (Fig. A.4).

5.1.3 Cloning for expression in Sf9 cells

Transformation of plasmids into EMBacY cells to generate bacmid

All plasmids were assembled using PCR and Gibson assembly and then verified by Sanger sequencing. The pACEBAC1 vector was used for all baculovirus expression plasmids (Geneva Biotech). To prepare bacmids, verified plasmids were diluted to $25\text{ng}/\mu\text{L}$ and $1\mu\text{L}$ of dilute plasmid was transformed into a $30\mu\text{L}$ aliquot of EMBacY competent cells. Cells were recovered without antibiotics for 4 hours and then plated on selective plates containing kanamycin ($50\mu\text{g/mL}$), gentamycin ($7\mu\text{g/mL}$), tetracylcine ($10\mu\text{g/mL}$), Blue-Gal ($100\mu\text{g/mL}$) and IPTG ($40\mu\text{g/mL}$). After 48 hours, white colonies were selected and grown in 3mL 2xTY media plus kanamycin ($50\mu\text{g/mL}$), gentamycin ($7\mu\text{g/mL}$) and tetracylcine ($10\mu\text{g/mL}$) overnight at 37°C. The next morning, an inoculating loop was used to streak a new selective plate to ensure that only white colonies would grow before proceeding with the bacmid.

Bacmid preparation

The bacmid prep was begun by centrifuging an overnight culture at 4000g for 5min. The resulting pellet was resuspended in $300\mu\text{L}$ P1 buffer (Qiagen) and transferred to a 2mL eppendorf tube. After adding $300\mu\text{L}$ P2 buffer (Qiagen), the tube was mixed by gentle inversion and incubated at RT for 5min. $400\mu\text{L}$ P3 buffer (Qiagen) was added and the tube was mixed again by gentle inversion and left on ice for 6min. The tube was centrifuged at 13,000RPM for 10min to pellet the white precipitate that formed. The supernatant was transferred to a new tube and centrifuged again to remove residual chromatin flakes. The supernatant was moved again to a new tube that contained $800\mu\text{L}$ isopropanol at RT. The tube was mixed by gentle inversion and placed on ice for 10min. The tube was centrifuged at 13,000RPM for 10min at RT before removing the supernatant and adding $500\mu\text{L}$ of RT 70% ethanol. The tube was inverted several times to wash the DNA pellet and then centrifuged at 13,000RPM for 5min at RT. This ethanol was was repeated two more times for a total of three washes. Then the supernatant was removed and the DNA pellet was allowed to airdry for no longer than 1min before gently dissolving it in $50\mu\text{L}$ sterile-filtered EB buffer (Qiagen) on ice. DNA concentration was determined by Nanodrop and bacmids were stored at 4°C.

Transfection of bacmids into Sf9 cells to generate baculovirus

2mL of Sf9 cell suspension that was already at $5\text{x}10^5$ cells/mL was placed into a 6-well plate. The cells were left to attach to the plate surface while bacmid DNA mix was prepared. To prepare bacmid DNA mix, $198\mu\text{L}$ medium (SF900-II, Life Technologies) was gently mixed with $2\mu\text{L}$ bacmid DNA (prepared at $1\mu\text{g}/\mu\text{L}$). $6\mu\text{L}$ of FuGENE HD Trasnfection Reagent (Promega) was added directly to the bacmid DNA mixture while ensuring that no undiluted FuGENE came in direct contact with the plastic tube. The bacmid DNA mixture was incubated at RT for 15min. 10min into these 15min, media was removed from the attached Sf9 cells. $800\mu\text{L}$ fresh media was added to the incubated bacmid DNA mixture, which was then added directly to the attached Sf9 cells. The plate was incubated for 3 days at 27°C in a plastic dish that also contained damp paper towels to maintain a high humidity. After the

first day, 1mL fresh media was added to the cells. At the end of the three days, 2mL of media was removed, which served as the P1 virus stock. Virus stocks were stored at 4°C.

5.1.4 Cloning for expression in *E. coli*

Cloning of DHC-E for expression in *E. coli*

A gene encoding the N-terminal 557 amino acids of the dynein heavy chain (DHC-E) of *S. cerevisiae* was commercially synthesised (GeneArt, Thermo Fisher Scientific). This synthetic gene was codon optimised for Sf9 cells, although it was ultimately expressed in *E. coli*. The gene was cloned into a modified pRSET(A) plasmid, which links a His-lipoyl domain fusion tag to the N-terminus via a TEV cleavage site (Dodd *et al.* 2004). This plasmid was a generous gift from Mark Allen. The final plasmid was verified by Sanger sequencing.

All other bacterial constructs were inserted into the same pRSET(A) vector. Inserts used to express further E-region truncations were amplified off of existing template sequences using PCR. Inserts used to express the DHC-E construct from 12 different species were synthesised (GeneArt, Thermo Fisher Scientific) and codon optimised for *E. coli*.

5.2 Protein expression and purification

5.2.1 General techniques

SDS-PAGE analysis of protein samples

Dilute protein samples (< 3mg/mL) were prepared for SDS-PAGE by directly adding 15μL of protein to 5μL of 4X NuPAGE LDS sample Buffer (Thermo Fisher Scientific). More concentrated samples were first diluted with ddH$_2$O and then added to NuPAGE LDS sample buffer in the same way. Samples were vortexed and centrifuged briefly in a tabletop centrifuge, then placed in a 95°C water bath for 5min. 10 to 20μL of sample were loaded into each well of precast NuPAGE Novex $4-12\%$ Bis-Tris Protein Gels (Thermo Fisher Scientific). Gels were run in either 1X NuPAGE MOPS SDS Running Buffer or 1X NuPAGE MES SDS Running Buffer (both Thermo Fisher Scientific) depending on the expected distribution of protein molecular weights in the sample. A voltage of 200V was applied for between 30 and 45 minutes, then the gels were rinsed, submerged in InstantBlue (Expedeon) and gently shaken for > 1hr. Gels were destained in ddH$_2$O for > 1hr and imaged using a Gel Doc XR system with a white light conversion screen controlled by Image Lab software (all Bio-Rad).

5.2.2 Expression in *S. cerevisae*

Shaker flask expression

For each litre grown, a 10mL YEPD starter culture was inoculated from a single colony and grown overnight at 30°C with 200RPM shaking. The next evening, the 10mL culture was transferred to a 2L flask that contained 1L of 2xYEPD. This was

grown overnight at 30°C with 200RPM shaking. The following afternoon, the culture was monitored to determine when all the glucose had been consumed. This was initially done by measuring the OD_{600} periodically to see when it no longer increased. Later, it became possible to simply dip a urinalysis test strip (Diastix, Bayer) in a 1mL sample of culture to determine the remaining glucose concentration. Once glucose had been depleted, 100mL of 20% galactose was added, to induce expression. Cells were induced overnight at 30°C and 200RPM. The following morning cells were harvested by centrifuging in 1L bottles at 4000RPM for 5min. The supernatant was discarded and the pellet was resuspended in ddH_2O. The cells were centrifuged again at 4000RPM for 5min and the supernatant was discarded. The cells were resuspended in as little ddH_2O as possible and then flash frozen in LN_2. The frozen pellet was ground into a fine powder using a waring blender, pre-cooled with LN_2 and stored at -80°C.

120L fermentor expression

Fermentor expression was similar as in shaker flasks, aside from the scale and equipment used. 6 10mL starter cultures were grown and used to inoculate 6L of YEPD. These were grown overnight and then used to inoculate $\sim$ 108L of 2xYEPD. The fermentor was kept at 30°C and the agitator was kept at 300RPM throughout the growth. Once the glucose had been fully consumed, 6L of 40% galactose was added, and cells were induced overnight. The cells were harvested in the same way as from shaker flasks.

5.2.3 Purification from *S. cerevisae*

Small-scale protein purification from yeast

A stock of 4X lysis buffer (120mM HEPES (pH = 7.2), 200mM $KC_2H_3O_2$, 8mM $MG(C_2H_3O_2)_2$, 4mM EGTA, 40% glycerol (v/v)) was prepared in advance. The day of the purification, 100mL of 4X lysis buffer was made up to 1mM DTT, 0.1mM MgATP, 1mM PMSF and 0.8% Triton X-100. Lysis was achieved by adding 6mL of this fresh 4X lysis buffer to 30mL of frozen yeast powder in a 50mL Falcon tube. The tube was placed in a 37°C water bath until all the powder had just thawed, at which point it was returned to ice. The lysate was centrifuged in a Type 70 ti rotor at 70,000RPM and 4°C for 30min. While the lysate was spinning, 0.2mL of packed IgG Sepharose beads (GE Healthcare) were washed into 1X lysis buffer. The lysate supernatant was collected in a 50mL falcon tube and then the washed IgG beads were added. The lysate was nutated at 4°C for 1hr. The lysate was poured into a disposable column (BioRad) and the beads were washed twice with 1X lysis buffer. The beads were then washed three times with 1X TEV buffer (20mM Tris-HCl (pH = 8.0), 150mM KCl, 10% glycerol (v/v), 1mM DTT and 0.1mM MgATP) and resuspended in $\sim$ 500μL of TEV buffer. The beads were moved to a 2mL Eppendorf tube and the slurry was allowed to settle. Buffer was removed by pipette until the remaining total volume was $\sim$ 300μL. 3μL of TEV protease was added and the tube was gently shaken at 16°C for 1hr. The tube was spun to pellet the beads and the supernatant was collected, and usually concentrated prior to SDS-PAGE analysis.

144

Large-scale protein purification from yeast

Large-scale purification of dynein from yeast was performed in a similar way as the small-scale purification, except after TEV cleavage, anion exchange chromatography and size exclusion chromatography steps were added. The first part of the purification was performed using the same buffers, and mostly the same protocol except for scaling each step proportionally to the volume of yeast powder being used. When purifying dynein from yeast grown in the 120L fermentor, up to 2L of yeast powder could be feasibly purified at one time.

One difference is that the TEV digest was performed overnight at 4°C rather than for 1hr at 16°C. After the TEV digest step, the degree of cleavage was determined by SDS-PAGE analysis. If cleavage was sufficient, the sample was diluted into buffer A (50mM Tris (pH = 8.0), 20mM KCl and 1mM DTT) and loaded onto a MonoQ column. The column was then washed with 10 column volumes of buffer A to remove unbound protein. Then, a salt gradient was applied using Buffer A as the low-salt component and Buffer B (50mM Tris (pH = 8.0), 500mM KCl and 1mM DTT) as the high salt component, until the protein eluted. The eluted protein's identity was confirmed using SDS-PAGE and then it was concentrated and loaded on a Superose6 column for size exclusion chromatography. This run exchanged the protein sample back down into TEV buffer, which the sample was kept in for further experiments, or for long-term storage at -80°C.

5.2.4   Expression in Sf9 cells

P2 virus amplification

50mL of Sf9 cells at 1-2x10^6 cells/mL was placed in a 1L disposable Erlenmeyer flask. The cells were infected with 500μL of P1 virus and then incubated at 27°C with 124RPM shaking for 72hr. With virus that originated from EMBacY cells, it was possible to check fluorescence of the cells visually as a qualitative indicator of infection and expression levels. Cells were harvested by centrifuging at 3750RPM for 10min at RT. The supernatant was preserved in sterile conditions and stored at 4°C in the dark as the P2 virus stock. For new constructs, the pellet from the P2 generation was often used for a small-scale test purification. At minimum, the crude lysate was analysed by SDS-PAGE to determine whether the construct had expressed and was soluble or not before scaling up protein expression.

Large-scale protein expression

500mL aliquots of Sf9 cells were grown to 1-2x10^6 cells/mL in 3L roller bottles with the caps left slightly loose for aeration. 5mL of P2 virus was added to each 500mL aliquot of cells and the cells were left to grow at 27°C with 124RPM shaking for 72hr. Cells were harvested by centrifuging at 3750RPM for 10min at 4°C. The supernatant was discarded (unless making a P3 virus stock) and cells were resuspended in $\sim$ 40mL of ice-cold PBS. The cells were centrifuged again at 3750RPM for 10min at 4°C. The supernatant was discarded and pellets were flash frozen in LN$_2$ before storing at -80°C.

5.2.5 Purification from Sf9 cells

Dynein was purified from Sf9 cells using the same method as for purifying dynein from yeast. The only difference was that cell lysis was achieved by resuspending the frozen pellet in 1X lysis buffer and then using 50 strokes in a glass tissue homogeniser. From this point on the purification was identical to a yeast one, although the yield was much higher, so more IgG Sepharose beads had to be use for the same mass of a cell pellet.

5.2.6 Expression in *E. coli*

Expression of DHC-E in *E. coli*

The plasmid designed for the expression of DHC-E in *E. coli* (described above) was transformed into SoluBL21 cells (Genlantis), which were grown in 2XTY media, supplemented with 7mg/L of ampicillin and 1mM $MgSO_4$ at 37°C with 210RPM shaking. The cultures were grown to an OD_{600} between 0.3 and 0.4, cooled to 16°C and induced overnight with 1mM IPTG. The cells were pelleted by centrifuging at $8,000g$ for 20min and then flash frozen in liquid nitrogen and stored at -80°C.

5.2.7 Purification from *E. coli*

Purification of DHC-E from *E. coli*

Cell pellets from 6L of bacterial culture from the expression of DHC-E were thawed on ice in 150mL of bacterial lysis buffer (30mM Hepes-NaOH (pH = 7.4), 200mM NaCl, 2mM $MgCl_2$, 1mM PMSF, 10mM imidazole, 10mM BME and 0.1mM Mg-ATP) supplemented with protease inhibitors (Complete-EDTA Free, Roche Applied Science). This suspension was sonicated on ice at maximum amplitude for 3min with 3s on and 7s off pulses. The lysate was cleared by centrifuging in a Ti45 rotor (Beckman Coulter) at 40,000RPM and 4°C for 40min, syringe filtered through a 0.22μm filter (Elkay) and loaded onto two tandem 5mL HisTrap columns (GE Healthcare) that were pre-equilibrated in buffer NiA (30mM Hepes-KOH (pH = 7.4), 200mM NaCl, 2mM $MgCl_2$, 10mM imidazole, 10mM BME and 0.1mM Mg-ATP). Unbound proteins were washed off the columns with 50mL of buffer NiA. The protein of interest was eluted in a single 15mL fraction with buffer NiB (30mM Hepes-KOH (pH = 7.4), 200mM NaCl, 2mM $MgCl_2$, 500mM imidazole, 10mM BME and 0.1mM Mg-ATP). This fraction was exchanged back into buffer NiA by concentrating (30kDa molecular weight cut-off concentrator, Amicon Ultracel, Merck-Millipore) to $\sim$ 1mL and diluting with 15mL of buffer NiA twice. Subsequently, the sample was incubated with 1mL of TEV protease (3.7mg/mL) and incubated on a roller at 4°C overnight. After TEV cleavage (confirmed by SDS-PAGE), the sample was run back through two tandem 5mL HisTrap columns (GE Healthcare), pre-equilibrated in buffer NiA. The flow-through was collected, concentrated to approximately 5mL and run through a HiLoad 26/600 Superdex200 column (GE Healthcare) that had been pre-equilibrated in GF buffer (30mM Hepes-NaOH (pH = 7.4), 200mM NaCl, 2mM $MgCl_2$, 10mM BME and 0.1mM Mg-ATP). This run resulted in two distinct peaks (Fig. 4.3), one near the void volume and the other at the expected elution volume for a dimer of DHC-E. The dimer peak was pooled, concentrated to $\sim$ 5mL

and run back through the same column, which was now equilibrated in crystallisation buffer (20mM Tris-NaOH (pH = 8.0, 150mM KCl, 10% glycerol (v/v), 1mM DTT and 0.1mM Mg-ATP). This run resulted in a single peak of dimerised DHC, which was pooled, concentrated to ~ 5mg/mL, aliquoted and flash frozen in liquid nitrogen before storing at -80°C.

5.3 Crystallography

5.3.1 Crystallisation of DHC-E

The DHC-E purified as described above was crystallised using the hanging drop method in EasyXtal 15-well plates (QIAGEN). Frozen protein was thawed on ice and 1μL was applied to the screw cap followed by either 0.75μL or 1μL of reservoir solution (0.1M Tris-NaOH (pH = 8.5), 0.2M Li_2SO_4 and $10 - 12.5\%$ PEG (MW = $8,000$, v/v)). Before screwing the cap back on the well, 500μL of Al's oil (Hampton Research) was applied on top of the 500μL of reservoir solution already in the well. Crystals grew overnight at 18°C to a maximum size of 500μm x 80μm x 80μm. When harvesting, individual crystals were briefly soaked in a cryo-protectant solution (0.1M Tris-NaOH (pH = 8.5), 0.2M Li_2SO_4, $10-12.5\%$ PEG (MW = $8,000$, v/v) and 30% ethylene glycol (v/v)) before being flash frozen in liquid nitrogen.

5.3.2 Selenomethionine derivative of DHC-E

Since no homology models were available for phasing via molecular replacement, we prepared a selenometionine (SeMet) derivative of DHC-E for experimental phasing. The same plasmid described above was transformed into B834 (DE3) cells (Novagen), which were grown in M9 media supplemented with 2mM $MgSO_4$,0.4% glucose (w/v), 25mg/L $FeSO_4 \cdot 7H_2O$, 7mg/L ampicillin, 1mg/L each of riboflavin, niacinamide, pyridoxine monohydrochloride and thiamine, 40mg/L each of L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-cysteine, L-glutamic acid, L-glutamine, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-tyrosine, L-valine and seleno-L-methionine. Cells were incubated at 30°C with shaking at 210RPM until OD_{600} reached between 0.8 and 1.0. They were induced with 0.3M IPTG and grown at 16°C for another 12hr and harvested as above. SeMet protein was purified in the same way as native protein except to minimise oxidation, all buffers were made with double the amount of BME or DTT and were thoroughly degassed before use. SeMet crystals were obtained and harvested in the same was as native crystals.

5.3.3 Data collection, phasing and model building

All data were collected at beamline i03 of Diamond Light Source (Didcot, UK). The data were integrated using IMOSFLM (Battye *et al.* 2011) and scaled using AIMLESS (Evans & Murshudov 2013). Experimental phase information was obtained employing the MIRAS approach (multiple isomorphous replacement with anomalous scattering) in autoSHARP (Vonrhein *et al.* 2007) with a native data set and three SeMet data sets as derivatives (Table 4.8). All data sets were cut at 5Å because including higher resolution data did not improve map quality. The resulting 5Å map

revealed density for two protein molecules in the asymmetric unit. The map was of sufficient quality to resolve α-helices, a β-sheet and several loops, which allowed an initial model to be constructed in COOT (Emsley *et al.* 2010). Model building was aided by an anomalous difference map calculated in the CCP4 program FFT (K. Cowtan 1994) using experimental phases and the amplitudes from the SeMet peak data set.

5.4 Electron microscopy

5.4.1 Negative stain EM of the NH dynein tail complex

Negative stain was performed on the NH dynein tail complex at $\sim$ 20nM in TEV buffer. The sample was applied to plasma-cleaned carbonfilm on 400-square-mesh copper grids (Electron Microscopy Sciences) and stained with 2% (w/v) uranyl acetate. $\sim$ 200 electron micrographs were recorded on a Gatan Ultrascan 1,000 XPCCD fitted to a FEI Tecnai G2 Spirit transmission electron micro-scope operating at 120 kV with a 26,000$\times$ nominal magnification (4 Å/pix, 30e/Å) at 1.5 μm underfocus. $\sim$ 10,000 particles were picked manually. Image analysis and 2D classification was performed using RELION (Scheres 2012). The initial 3D model was calculated using EMAN2 (G. Tang *et al.* 2007).

5.4.2 Fitting the DHC-E crystal structure into EM density

To fit the DHC-E crystal structure into the TDB EM density map, the full crystal structure was divided into three fragments: the dimerisation domain (residues 1 to 179 from both chains A and B) and the two elongated domains (chain A and chain B). All three fragments were fitted into the EM helices in the density independently using the "jiggle" command in COOT (Emsley *et al.* 2010). The positions of individual helices in the elongated domains were fitted by minor adjustments.

References

1. Akhmanova, A. & Steinmetz, M. O. Tracking the ends: a dynamic protein network controls the fate of microtubule tips. *Nature reviews Molecular cell biology* **9,** 309–322 (2008).

2. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25,** 3389–3402 (1997).

3. Asante, D., Stevenson, N. L. & Stephens, D. J. Subunit composition of the human cytoplasmic dynein-2 complex. *Journal of cell science* **127,** 4774–4787 (2014).

4. Asherie, N. Protein crystallization and phase diagrams. *Methods* **34,** 266–272 (2004).

5. Barbar, E. Dynein light chain LC8 is a dimerization hub essential in diverse protein networks. *Biochemistry* **47,** 503–508 (2008).

6. Barbar, E. Native disorder mediates binding of dynein to NudE and dynactin. *Biochemical Society Transactions* **40,** 1009 (2012).

7. Barber, C. F., Heuser, T., Carbajal-González, B. I., Botchkarev, V. V. & Nicastro, D. Three-dimensional structure of the radial spokes reveals heterogeneity and interactions with dyneins in Chlamydomonas flagella. *Molecular biology of the cell* **23,** 111–120 (2012).

8. Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D: Biological Crystallography* **67,** 271–281 (2011).

9. Beck, T., Krasauskas, A., Gruene, T. & Sheldrick, G. M. A magic triangle for experimental phasing of macromolecules. *Acta Crystallographica Section D* **64,** 1179–1182 (2008).

10. Benison, G., Karplus, P. A. & Barbar, E. Structure and dynamics of LC8 complexes with KXTQT-motif peptides: swallow and dynein intermediate chain compete for a common site. *Journal of molecular biology* **371,** 457–468 (2007).

11. Bergen, L. G., Kuriyama, R. & Borisy, G. G. Polarity of microtubules nucleated by centrosomes and chromosomes of Chinese hamster ovary cells in vitro. *The Journal of cell biology* **84,** 151–159 (1980).

12. Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: Multigene Baculovirus-Based Eukaryotic Protein Complex Production. *Current protocols in protein science,* 5–20 (2008).

13. Blasier, K. R., Humsi, M. K., Ha, J., Ross, M. W., Smiley, W. R., Inamdar, N. A., Mitchell, D. J., Lo, K. W.-H. & Pfister, K. K. Live cell imaging reveals differential modifications to cytoplasmic dynein properties by phospho- and dephosphomimic mutations of the intermediate chain 2C S84. *Journal of neuroscience research* **92,** 1143–1154 (2014).

14. Brown, S. S. Myosins in yeast. *Current opinion in cell biology* **9,** 44–48 (1997).

15. Bui, K. H., Pigino, G. & Ishikawa, T. Three-dimensional structural analysis of eukaryotic flagella/cilia by electron cryo-tomography. *Journal of synchrotron radiation* **18,** 2–5 (2010).

16. Bui, K. H., Sakakibara, H., Movassagh, T., Oiwa, K. & Ishikawa, T. Asymmetry of inner dynein arms and inter-doublet links in Chlamydomonas flagella. *The Journal of cell biology* **186,** 437–446 (2009).

17. Bullock, S. L. & Ish-Horowicz, D. Conserved signals and machinery for RNA transport in Drosophila oogenesis and embryogenesis. *Nature* **414,** 611–616 (2001).

18. Burgess, S. A., Walker, M. L., Sakakibara, H., Knight, P. J. & Oiwa, K. Dynein structure and power stroke. *Nature* **421,** 715–718 (2003).

19. Carter, A. P., Cho, C., Jin, L. & Vale, R. D. Crystal structure of the dynein motor domain. *Science* **331,** 1159–1165 (2011).

20. Carter, N. J. & Cross, R. Mechanics of the kinesin step. *Nature* **435,** 308–312 (2005).

21. Casey, D. M., Inaba, K., Pazour, G. J., Takada, S., Wakabayashi, K.-i., Wilkerson, C. G., Kamiya, R. & Witman, G. B. DC3, the 21-kDa subunit of the outer dynein arm-docking complex (ODA-DC), is a novel EF-hand protein important for assembly of both the outer arm and the ODA-DC. *Molecular biology of the cell* **14,** 3650–3663 (2003).

22. Caviston, J. P., Ross, J. L., Antony, S. M., Tokito, M. & Holzbaur, E. L. Huntingtin facilitates dynein/dynactin-mediated vesicle transport. *Proceedings of the National Academy of Sciences* **104,** 10045–10050 (2007).

23. Chaplin, P. J., Camon, E. B., Villarreal-Ramos, B., Flint, M., Ryan, M. D. & Collins, R. A. Production of interleukin-12 as a self-processing 2A polypeptide. *Journal of interferon & cytokine research* **19,** 235–241 (1999).

24. Chayen, N. E. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. *Journal of Applied Crystallography* **30,** 198–202 (1997).

25. Chayen, N. E., Stewart, P. D. S. & Blow, D. M. Microbatch crystallization under oil—a new technique allowing many small-volume crystallization trials. *Journal of Crystal Growth* **122,** 176–180 (1992).

26. Chowdhury, S., Ketcham, S. A., Schroer, T. A. & Lander, G. C. Structural organization of the dynein-dynactin complex bound to microtubules. *Nature structural & molecular biology* (2015).

27. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM newsletter on protein crystallography* **31,** 34–38 (1994).

28. Crépieux, P., Kwon, H., Leclerc, N., Spencer, W., Richard, S., Lin, R. & Hiscott, J. IκBα physically interacts with a cytoskeleton-associated protein through its signal response domain. *Molecular and cellular biology* **17,** 7375–7385 (1997).

29. De Felipe, P., Luke, G. A., Hughes, L. E., Gani, D., Halpin, C. & Ryan, M. D. E unum pluribus: multiple proteins from a self-processing polyprotein. *TRENDS in Biotechnology* **24,** 68–75 (2006).

30. Dell, K. R., Turck, C. W. & Vale, R. D. Mitotic phosphorylation of the dynein light intermediate chain is mediated by cdc2 kinase. *Traffic* **1,** 38–44 (2000).

31. Diamant, A. G. & Carter, A. P. in *Encyclopedia of Biophysics* (ed Roberts, G. C. K.) 552–558 (Springer Heidelberg, 2013).

32. Dienstbier, M., Boehl, F., Li, X. & Bullock, S. L. Egalitarian is a selective RNA-binding protein linking mRNA localization signals to the dynein motor. *Genes & development* **23,** 1546–1558 (2009).

33. Dillman, J. F. & Pfister, K. K. Differential phosphorylation in vivo of cytoplasmic dynein associated with anterogradely moving organelles. *The Journal of cell biology* **127,** 1671–1681 (1994).

34. Dodd, R. B., Allen, M. D., Brown, S. E., Sanderson, C. M., Duncan, L. M., Lehner, P. J., Bycroft, M. & Read, R. J. Solution structure of the Kaposi's sarcoma-associated herpesvirus K3 N-terminal domain reveals a Novel E2-binding C4HC3-type RING domain. *Journal of Biological Chemistry* **279,** 53840–53847 (2004).

35. Döhner, K., Wolfstein, A., Prank, U., Echeverri, C., Dujardin, D., Vallee, R. & Sodeik, B. Function of dynein and dynactin in herpes simplex virus capsid transport. *Molecular biology of the cell* **13,** 2795–2809 (2002).

36. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and Development of *Coot. Acta Crystallographica Section D - Biological Crystallography* **66** (2010).

37. Enrique, M., Wells, A. L., Rosenfeld, S. S., Ostap, E. M. & Sweeney, H. L. The kinetic mechanism of myosin V. *Proceedings of the National Academy of Sciences* **96,** 13726–13731 (1999).

38. Eshel, D., Urrestarazu, L., Vissers, S., Jauniaux, J., van Vliet-Reedijk, J., Planta, R. & Gibbons, I. Cytoplasmic dynein is required for normal nuclear segregation in yeast. *Proceedings of the National Academy of Sciences* **90,** 11172–11176 (1993).

39. Espindola, F. S., Suter, D. M., Partata, L. B., Cao, T., Wolenski, J. S., Cheney, R. E., King, S. M. & Mooseker, M. S. The light chain composition of chicken brain myosin-Va: calmodulin, myosin-II essential light chains, and 8-kDa dynein light chain/PIN. *Cell motility and the cytoskeleton* **47,** 269–281 (2000).

40. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallographica Section D: Biological Crystallography* **69,** 1204–1214 (2013).

41. Farkasovsky, M. & Küntzel, H. Cortical Num1p interacts with the dynein intermediate chain Pac11p and cytoplasmic microtubules in budding yeast. *The Journal of cell biology* **152,** 251–262 (2001).

42. Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., *et al.* Pfam: the protein families database. *Nucleic acids research,* gkt1223 (2013).

43. Follit, J. A., Xu, F., Keady, B. T. & Pazour, G. J. Characterization of mouse IFT complex B. *Cell motility and the cytoskeleton* **66,** 457–468 (2009).

44. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., Bairoch, A., *et al.* in *The proteomics protocols handbook* 571–607 (Springer, 2005).

45. Gee, M. A., Heuser, J. E. & Vallee, R. B. An extended microtubule-binding structure within the dynein motor domain. *Nature* **390,** 636–639 (1997).

46. Gibbons, I., Gibbons, B. H., Mocz, G. & Asai, D. J. Multiple nucleotide-binding sites in the sequence of dynein β heavy chain. *Nature* **352,** 640–643 (1991).

47. Gilbert, S. P. & Sloboda, R. D. A squid dynein isoform promotes axoplasmic vesicle translocation. *The Journal of cell biology* **109,** 2379–2394 (1989).

48. Gill, S., Cleveland, D. & Schroer, T. Characterization of DLC-A and DLC-B, two families of cytoplasmic dynein light chain subunits. *Molecular biology of the cell* **5,** 645–654 (1994).

49. Gill, S. R., Schroer, T. A., Szilak, I., Steuer, E. R., Sheetz, M. P. & Cleveland, D. W. Dynactin, a conserved, ubiquitously expressed component of an activator of vesicle motility mediated by cytoplasmic dynein. *The Journal of Cell Biology* **115,** 1639–1650 (1991).

50. Goldstein, L. S. & Yang, Z. Microtubule-based transport systems in neurons: the roles of kinesins and dyneins. *Annual review of neuroscience* **23,** 39–71 (2000).

51. Granger, B. L. & Lazarides, E. Desmin and vimentin coexist at the periphery of the myofibril Z disc. *Cell* **18,** 1053–1063 (1979).

52. Grava, S. & Philippsen, P. Dynamics of multiple nuclei in Ashbya gossypii hyphae depend on the control of cytoplasmic microtubules length by Bik1, Kip2, Kip3, and not on a capture/shrinkage mechanism. *Molecular biology of the cell* **21,** 3680–3692 (2010).

53. Ha, J., Lo, K. W.-H., Myers, K. R., Carr, T. M., Humsi, M. K., Rasoul, B. A., Segal, R. A. & Pfister, K. K. A neuron-specific cytoplasmic dynein isoform preferentially transports TrkB signaling endosomes. *The Journal of cell biology* **181,** 1027–1039 (2008).

54. Habura, A., Tikhonenko, I., Chisholm, R. L. & Koonce, M. P. Interaction Mapping of a Dynein Heavy Chain identification of dimerization and intermediate-chain binding domains. *Journal of Biological Chemistry* **274,** 15447–15453 (1999).

55. Hayashi, S. & Shingyoji, C. Mechanism of flagellar oscillation–bending-induced switching of dynein activity in elastase-treated axonemes of sea urchin sperm. *Journal of cell science* **121,** 2833–2843 (2008).

56. Hirokawa, N. Kinesin and dynein superfamily proteins and the mechanism of organelle transport. *Science* **279,** 519–526 (1998).

57. Hodge, T., Jamie, M. & Cope, T. A myosin family tree. *Journal of Cell Science* **113,** 3353–3354 (2000).

58. Hódi, Z., Németh, A. L., Radnai, L., Hetényi, C., Schlett, K., Bodor, A., Perczel, A. & Nyitray, L. Alternatively spliced exon B of myosin Va is essential for binding the tail-associated light chain shared by dynein. *Biochemistry* **45,** 12582–12595 (2006).

59. Holt, L. J., Tuch, B. B., Villén, J., Johnson, A. D., Gygi, S. P. & Morgan, D. O. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325,** 1682–1686 (2009).

60. Hom, E. F., Witman, G. B., Harris, E. H., Dutcher, S. K., Kamiya, R., Mitchell, D. R., Pazour, G. J., Porter, M. E., Sale, W. S., Wirschell, M., *et al.* A unified taxonomy for ciliary dyneins. *Cytoskeleton* **68,** 555–565 (2011).

61. Hoogenraad, C. C., Akhmanova, A., Howell, S. A., Dortland, B. R., De Zeeuw, C. I., Willemsen, R., Visser, P., Grosveld, F. & Galjart, N. Mammalian Golgi-associated Bicaudal-D2 functions in the dynein–dynactin pathway by interacting with these complexes. *The EMBO Journal* **20,** 4041–4054 (2001).

62. Howard, J. The movement of kinesin along microtubules. *Annual review of physiology* **58,** 703–729 (1996).

63. Howard, T. H. & Meyer, W. H. Chemotactic peptide modulation of actin assembly and locomotion in neutrophils. *The Journal of cell biology* **98,** 1265–1271 (1984).

64. Huang, J., Roberts, A. J., Leschziner, A. E. & Reck-Peterson, S. L. Lis1 acts as a "clutch" between the ATPase and microtubule-binding domains of the dynein motor. *Cell* **150,** 975–986 (2012).

65. Hunter, A. W., Caplow, M., Coy, D. L., Hancock, W. O., Diez, S., Wordeman, L. & Howard, J. The kinesin-related protein MCAK is a microtubule depolymerase that forms an ATP-hydrolyzing complex at microtubule ends. *Molecular cell* **11,** 445–457 (2003).

66. Huxley, H. & Hanson, J. Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature,* 973–6 (1954).

67. Jaffrey, S. R. & Snyder, S. H. PIN: an associated protein inhibitor of neuronal nitric oxide synthase. *Science* **274,** 774–777 (1996).

68. Jonassen, J. A., SanAgustin, J., Baker, S. P. & Pazour, G. J. Disruption of IFT complex A causes cystic kidneys without mitotic spindle misorientation. *Journal of the American Society of Nephrology* **23,** 641–651 (2012).

69. Kaan, H. Y. K., Hackney, D. D. & Kozielski, F. The structure of the kinesin-1 motor-tail complex reveals the mechanism of autoinhibition. *Science* **333,** 883–885 (2011).

70. Kagami, O. & Kamiya, R. Translocation and rotation of microtubules caused by multiple species of Chlamydomonas inner-arm dynein. *Journal of Cell Science* **103,** 653–664 (1992).

71. Kamiya, R. Functional diversity of axonemal dyneins as studied in Chlamydomonas mutants. *International review of cytology* **219,** 115–155 (2002).

72. Kardon, J. R., Reck-Peterson, S. L. & Vale, R. D. Regulation of the processivity and intracellular localization of Saccharomyces cerevisiae dynein by dynactin. *Proceedings of the National Academy of Sciences* **106,** 5669–5674 (2009).

73. Kardon, J. R. & Vale, R. D. Regulators of the cytoplasmic dynein motor. *Nature reviews Molecular cell biology* **10,** 854–865 (2009).

74. Karki, S. & Holzbaur, E. L. Cytoplasmic dynein and dynactin in cell division and intracellular transport. *Current opinion in cell biology* **11,** 45–53 (1999).

75. Karki, S., Tokito, M. K. & Holzbaur, E. L. Casein kinase II binds to and phosphorylates cytoplasmic dynein. *Journal of Biological Chemistry* **272,** 5887–5891 (1997).

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

77. King, S. J., Bonilla, M., Rodgers, M. E. & Schroer, T. A. Subunit organization in cytoplasmic dynein subcomplexes. *Protein Science* **11,** 1239–1250 (2002).

78. King, S. J. & Schroer, T. A. Dynactin increases the processivity of the cytoplasmic dynein motor. *Nature cell biology* **2,** 20–24 (2000).

79. King, S. M. *Dyneins: structure, biology and disease* (Academic Press, 2011).

80. Kishino, A. & Yanagida, T. Force measurements by micromanipulation of a single actin filament by glass needles. *Nature* **518,** 74–76 (1988).

81. Knäblein, J., Neufeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Löwe, J., Steipe, B. & Huber, R. Ta 6 Br 12 2+, a tool for phase determination of large biological assemblies by X-ray crystallography. *Journal of molecular biology* **270,** 1–7 (1997).

82. Kon, T., Mogami, T., Ohkura, R., Nishiura, M. & Sutoh, K. ATP hydrolysis cycle–dependent tail motions in cytoplasmic dynein. *Nature structural & molecular biology* **12,** 513–519 (2005).

83. Kon, T., Nishiura, M., Ohkura, R., Toyoshima, Y. Y. & Sutoh, K. Distinct functions of nucleotide-binding/hydrolysis sites in the four AAA modules of cytoplasmic dynein. *Biochemistry* **43,** 11266–11274 (2004).

84. Kon, T., Oyama, T., Shimo-Kon, R., Sutoh, K. & Kurisu, G. The 2.8-Å Crystal Structure of the Dynein Motor Domain. *Biophysical Journal* **102,** 367a–368a (2012).

85. Kon, T., Sutoh, K. & Kurisu, G. X-ray structure of a functional full-length dynein motor domain. *Nature structural & molecular biology* **18,** 638–642 (2011).

86. Korn, E. D. Actin polymerization and its regulation by proteins from non-muscle cells. *Physiological Reviews* **62,** 672–737 (1982).

87. Kozminski, K. G., Beech, P. L. & Rosenbaum, J. L. The Chlamydomonas kinesin-like protein FLA10 is involved in motility associated with the flagellar membrane. *The Journal of cell biology* **131,** 1517–1527 (1995).

88. Kull, F. J., Vale, R. D. & Fletterick, R. J. The case for a common ancestor: kinesin and myosin motor proteins and G proteins. *Journal of Muscle Research & Cell Motility* **19,** 877–886 (1998).

89. Larkin, M. A., Blackshields, G., Brown, N., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947–2948 (2007).

90. Lawrence, C. J., Dawe, R. K., Christie, K. R., Cleveland, D. W., Dawson, S. C., Endow, S. A., Goldstein, L. S., Goodson, H. V., Hirokawa, N., Howard, J., *et al.* A standardized kinesin nomenclature. *The Journal of cell biology* **167,** 19–22 (2004).

91. Lazarides, E. Intermediate filaments as mechanical integrators of cellular space. *Nature* **283,** 249–256 (1980).

92. Lee, W.-L., Kaiser, M. A. & Cooper, J. A. The offloading model for dynein function differential function of motor subunits. *The Journal of cell biology* **168,** 201–207 (2005).

93. Liang, J., Jaffrey, S. R., Guo, W., Snyder, S. H. & Clardy, J. Structure of the PIN/LC8 dimer with a bound peptide. *Nature Structural & Molecular Biology* **6,** 735–740 (1999).

94. Liu, C., Fallen, M. K., Miller, H., Upadhyaya, A. & Song, W. The actin cytoskeleton coordinates the signal transduction and antigen processing functions of the B cell antigen receptor. *Frontiers in biology* **8,** 475–485 (2013).

95. Lo, K. W.-H., Kan, H.-M., Chan, L.-N., Xu, W.-G., Wang, K.-P., Wu, Z., Sheng, M. & Zhang, M. The 8-kDa dynein light chain binds to p53-binding protein 1 and mediates DNA damage-induced p53 nuclear accumulation. *Journal of Biological Chemistry* **280,** 8172–8179 (2005).

96. Lodish, H., Berk, A., Matsudaira, P., Kaiser, C. A., Krieger, M., Scott, M. P., Zipursky, L. & Darnell, J. *Molecular Cell Biology (5th, Fifth Edition)* 5th (WHFreeman, 2004).

97. Lorens, J., Pearsall, D., Swift, S., Peelle, B., Armstrong, R., Demo, S., Ferrick, D., Hitoshi, Y., Payan, D. & Anderson, D. Stable, stoichiometric delivery of diverse protein functions. *Journal of biochemical and biophysical methods* **58,** 101–110 (2004).

98. Luft, J. R., Wolfley, J. R., Said, M. I., Nagel, R. M., Lauricella, A. M., Smith, J. L., Thayer, M. H., Veatch, C. K., Snell, E. H., Malkowski, M. G., *et al.* Efficient optimization of crystallization conditions by manipulation of drop volume ratio and temperature. *Protein science* **16,** 715–722 (2007).

99. Lupas, A., Van Dyke, M., Stock, J., *et al.* Predicting coiled coils from protein sequences. *Science* **252,** 1162–1164 (1991).

100. Markus, S. M., Punch, J. J. & Lee, W.-L. Motor-and tail-dependent targeting of dynein to microtubule plus ends and the cell cortex. *Current Biology* **19,** 196–205 (2009).

101. Matanis, T., Akhmanova, A., Wulf, P., Del Nery, E., Weide, T., Stepanova, T., Galjart, N., Grosveld, F., Goud, B., De Zeeuw, C. I., *et al.* Bicaudal-D regulates COPI-independent Golgi–ER transport by recruiting the dynein–dynactin motor complex. *Nature Cell Biology* **4,** 986–992 (2002).

102. Mazumdar, M., Mikami, A., Gee, M. A. & Vallee, R. B. In vitro motility from recombinant dynein heavy chain. *Proceedings of the National Academy of Sciences* **93,** 6552–6556 (1996).

103. McDonald, H. B., Stewart, R. J. & Goldstein, L. S. The kinesin-like ncd protein of Drosophila is a minus end-directed microtubule motor. *Cell* **63,** 1159–1165 (1990).

104. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16,** 404–405 (2000).

105. McKenney, R. J., Huynh, W., Tanenbaum, M. E., Bhabha, G. & Vale, R. D. Activation of cytoplasmic dynein motility by dynactin-cargo adapter complexes. *Science* **345,** 337–341 (2014).

106. McKenney, R. J., Weil, S. J., Scherer, J. & Vallee, R. B. Mutually exclusive cytoplasmic dynein regulation by NudE-Lis1 and dynactin. *Journal of Biological Chemistry* **286,** 39615–39622 (2011).

107. Mehta, A. D., Rock, R. S., Rief, M., Spudich, J. A., Mooseker, M. S. & Cheney, R. E. Myosin-V is a processive actin-based motor. *Nature* **400,** 590–593 (1999).

108. Mikami, A., Tynan, S. H., Hama, T., Luby-Phelps, K., Saito, T., Crandall, J. E., Besharse, J. C. & Vallee, R. B. Molecular structure of cytoplasmic dynein 2 and its distribution in neuronal and ciliated cells. *Journal of cell science* **115,** 4801–4808 (2002).

109. Miki, H., Setou, M., Kaneshiro, K. & Hirokawa, N. All kinesin superfamily protein, KIF, genes in mouse and human. *Proceedings of the National Academy of Sciences* **98,** 7004–7011 (2001).

110. Milisav, I. Dynein and dynein-related genes. *Cell motility and the cytoskeleton* **39,** 261–272 (1998).

111. Mitchell, D. J., Blasier, K. R., Jeffery, E. D., Ross, M. W., Pullikuth, A. K., Suo, D., Park, J., Smiley, W. R., Lo, K. W.-H., Shabanowitz, J., *et al.* Trk activation of the ERK1/2 kinase pathway stimulates intermediate chain phosphorylation and recruits cytoplasmic dynein to signaling endosomes for retrograde axonal transport. *The Journal of Neuroscience* **32,** 15495–15510 (2012).

112. Mizianty, M. J. & Kurgan, L. Sequence-based prediction of protein crystallization, purification and production propensity. *Bioinformatics* **27,** i24–i33 (2011).

113. Moritz, M., Braunfeld, M. B., Guénebaut, V., Heuser, J. & Agard, D. A. Structure of the γ-tubulin ring complex: a template for microtubule nucleation. *Nature cell biology* **2,** 365–370 (2000).

114. Movassagh, T., Bui, K. H., Sakakibara, H., Oiwa, K. & Ishikawa, T. Nucleotide-induced global conformational changes of flagellar dynein arms revealed by in situ analysis. *Nature structural & molecular biology* **17,** 761–767 (2010).

115. Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography* **67,** 355–367 (2011).

116. Neuwald, A. F., Aravind, L., Spouge, J. L. & Koonin, E. V. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome research* **9,** 27–43 (1999).

117. Nicastro, D., McIntosh, J. R. & Baumeister, W. 3D structure of eukaryotic flagella in a quiescent state revealed by cryo-electron tomography. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 15889–15894 (2005).

118. Nicastro, D., Schwartz, C., Pierson, J., Gaudette, R., Porter, M. E. & McIntosh, J. R. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science* **313,** 944–948 (2006).

119. Noguchi, T. & Mabuchi, I. Reorganization of actin cytoskeleton at the growing end of the cleavage furrow of Xenopus egg during cytokinesis. *Journal of cell science* **114,** 401–412 (2001).

120. Nyarko, A. & Barbar, E. Light chain-dependent self-association of dynein intermediate chain. *Journal of Biological Chemistry* **286,** 1556–1566 (2011).

121. Ogawa, K. Four ATP-binding sites in the midregion of the β heavy chain of dynein. *Nature* **352,** 643–645 (1991).

122. Okada, Y., Yamazaki, H., Sekine-Aizawa, Y. & Hirokawa, N. The neuron-specific kinesin superfamily protein KIF1A is a uniqye monomeric motor for anterograde axonal transport of synaptic vesicle precursors. *Cell* **81,** 769–780 (1995).

123. Olmsted, J. B. & Borisy, G. G. Microtubules. *Annual review of biochemistry* **42,** 507–540 (1973).

124. Ori-McKenney, K. M. & Vallee, R. B. Neuronal migration defects in the Loa dynein mutant mouse. *Neural Dev* **6,** 26 (2011).

125. Ottenheijm, C., Heunks, L. & Dekhuijzen, R. Diaphragm adaptations in patients with COPD. *Respir Res* **9,** 12 (2008).

126. Palmer, K. J., Hughes, H. & Stephens, D. J. Specificity of cytoplasmic dynein subunits in discrete membrane-trafficking steps. *Molecular biology of the cell* **20,** 2885–2899 (2009).

127. Paschal, B. M., Shpetner, H. S. & Vallee, R. B. MAP 1C is a microtubule-activated ATPase which translocates microtubules in vitro and has dynein-like properties. *The Journal of cell biology* **105,** 1273–1282 (1987).

128. Pazour, G. J., Dickert, B. L. & Witman, G. B. The DHC1b (DHC2) isoform of cytoplasmic dynein is required for flagellar assembly. *The Journal of cell biology* **144,** 473–481 (1999).

129. Pernigo, S., Lamprecht, A., Steiner, R. A. & Dodding, M. P. Structural basis for kinesin-1: cargo recognition. *Science* **340,** 356–359 (2013).

130. Perrone, C. A., Tritschler, D., Taulman, P., Bower, R., Yoder, B. K. & Porter, M. E. A novel dynein light intermediate chain colocalizes with the retrograde motor for intraflagellar transport at sites of axoneme assembly in Chlamydomonas and mammalian cells. *Molecular biology of the cell* **14,** 2041–2056 (2003).

131. Pfister, K. K., Shah, P. R., Hummerich, H., Russ, A., Cotton, J., Annuar, A. A., King, S. M. & Fisher, E. M. Genetic analysis of the cytoplasmic dynein subunit families. *PLoS genetics* **2,** e1 (2006).

132. Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **21,** 3435–3438 (2005).

133. Pullikuth, A. K., Ozdemir, A., Cardenas, D., Bailey, E., Sherman, N. E., Pfister, K. K. & Catling, A. D. Epidermal growth factor stimulates extracellular-signal regulated kinase phosphorylation of a novel site on cytoplasmic dynein intermediate chain 2. *International journal of molecular sciences* **14,** 3595–3620 (2013).

134. Qualmann, B., Kessels, M. M. & Kelly, R. B. Molecular links between endocytosis and the actin cytoskeleton. *The Journal of cell biology* **150,** F111–F116 (2000).

135. Ramakrishnan, V., Finch, J., Graziano, V., Lee, P. & Sweet, R. Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *nature* **362,** 219–223 (1993).

136. Rao, L., Romes, E. M., Nicholas, M. P., Brenner, S., Tripathy, A., Gennerich, A. & Slep, K. C. The yeast dynein Dyn2-Pac11 complex is a dynein dimerization/processivity factor: structural and single-molecule characterization. *Molecular biology of the cell* **24,** 2362–2377 (2013).

137. Rapali, P., Radnai, L., Süveges, D., Harmat, V., Tölgyesi, F., Wahlgren, W. Y., Katona, G., Nyitray, L. & Pál, G. Directed evolution reveals the binding motif preference of the LC8/DYNLL hub protein and predicts large numbers of novel binders in the human proteome. *PLoS One* **6,** e18818 (2011).

138. Ravikumar, B., Acevedo-Arozena, A., Imarisio, S., Berger, Z., Vacher, C., O'Kane, C. J., Brown, S. D. & Rubinsztein, D. C. Dynein mutations impair autophagic clearance of aggregate-prone proteins. *Nature genetics* **37,** 771–776 (2005).

139. Rayala, S. K., den Hollander, P., Balasenthil, S., Yang, Z., Broaddus, R. R. & Kumar, R. Functional regulation of oestrogen receptor pathway by the dynein light chain 1. *EMBO reports* **6,** 538–544 (2005).

140. Read, R. & Schierbeek, A. A phased translation function. *Journal of Applied Crystallography* **21,** 490–495 (1988).

141. Reck-Peterson, S. L., Provance, D. W., Mooseker, M. S. & Mercer, J. A. Class V myosins. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1496,** 36–51 (2000).

142. Reck-Peterson, S. L. & Vale, R. D. Molecular dissection of the roles of nucleotide binding and hydrolysis in dynein's AAA domains in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 1491–1495 (2004).

143. Reck-Peterson, S. L., Yildiz, A., Carter, A. P., Gennerich, A., Zhang, N. & Vale, R. D. Single-molecule analysis of dynein processivity and stepping behavior. *Cell* **126,** 335–348 (2006).

144. Reggiani, C. & Bottinelli, R. in *Myosins* 125–169 (Springer, 2008).

145. Rodríguez-Crespo, I., Yélamos, B., Roncal, F., Albar, J. P., Ortiz de Montellano, P. R. & Gavilanes, F. Identification of novel cellular proteins that bind to the LC8 dynein light chain using a pepscan technique. *FEBS Letters* **503,** 125–141 (2001).

146. Rogers, K. R., Weiss, S., Crevel, I., Brophy, P. J., Geeves, M. & Cross, R. KIF1D is a fast non-processive kinesin that demonstrates novel K-loop-dependent mechanochemistry. *The EMBO journal* **20,** 5101–5113 (2001).

147. Romes, E. M., Tripathy, A. & Slep, K. C. Structure of a yeast Dyn2-Nup159 complex and molecular basis for dynein light chain-nuclear pore interaction. *Journal of Biological Chemistry* **287,** 15862–15873 (2012).

148. Rosenbaum, J. L. & Witman, G. B. Intraflagellar transport. *Nature Reviews Molecular Cell Biology* **3,** 813–825 (2002).

149. Sakakibara, H., Kojima, H., Sakai, Y., Katayama, E. & Oiwa, K. Inner-arm dynein c of Chlamydomonas flagella is a single-headed processive motor. *Nature* **400,** 586–590 (1999).

150. Samsó, M., Radermacher, M., Frank, J. & Koonce, M. P. Structural characterization of a dynein motor domain. *Journal of molecular biology* **276,** 927–937 (1998).

151. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **180,** 519–530 (2012).

152. Schlager, M. A., Hoang, H. T., Urnavicius, L., Bullock, S. L. & Carter, A. P. In vitro reconstitution of a highly processive recombinant human dynein complex. *The EMBO journal* **33,** 1855–1868 (2014).

153. Schmidt, H., Gleave, E. S. & Carter, A. P. Insights into dynein motor domain function from a 3.3-Å crystal structure. *Nature structural & molecular biology* **19,** 492–497 (2012).

154. Schmidt, H., Zalyte, R., Urnavicius, L. & Carter, A. P. Structure of human cytoplasmic dynein-2 primed for its power stroke. *Nature* (2014).

155. Scholey, J. M. & Anderson, K. V. Intraflagellar transport and cilium-based signaling. *Cell* **125,** 439–442 (2006).

156. Schroer, T. A. *Structure and function of dynactin* in *Seminars in Cell & Developmental Biology* **7** (1996), 321–328.

157. Schroer, T. A. Dynactin. *Annu. Rev. Cell Dev. Biol.* **20,** 759–779 (2004).

158. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallographica Section D: Biological Crystallography* **66,** 479–485 (2010).

159. Siglin, A. E., Sun, S., Moore, J. K., Tan, S., Poenie, M., Lear, J. D., Polenova, T., Cooper, J. A. & Williams, J. C. Dynein and dynactin leverage their bivalent character to form a high-affinity interaction. *PloS one* **8,** e59453 (2013).

160. Silvanovich, A., Li, M.-g., Serr, M., Mische, S. & Hays, T. S. The third P-loop domain in cytoplasmic dynein heavy chain is essential for dynein motor function and ATP-sensitive microtubule binding. *Molecular biology of the cell* **14,** 1355–1365 (2003).

161. Smith, E. F. & Sale, W. S. Microtubule binding and translocation by inner dynein arm subtype I1. *Cell motility and the cytoskeleton* **18,** 258–268 (1991).

162. Stephens, R. & Edds, K. Microtubules: structure, chemistry, and function. *Physiological reviews* **56,** 709–777 (1976).

163. Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D. & Eisenberg, D. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences* **103,** 8060–8065 (2006).

164. Stuchell-Brereton, M. D., Siglin, A., Li, J., Moore, J. K., Ahmed, S., Williams, J. C. & Cooper, J. A. Functional interaction between dynein light chain and intermediate chain is required for mitotic spindle positioning. *Molecular biology of the cell* **22,** 2690–2701 (2011).

165. Takazaki, H., Liu, Z., Jin, M., Kamiya, R. & Yasunaga, T. Three outer arm dynein heavy chains of Chlamydomonas reinhardtii operate in a coordinated fashion both in vitro and in vivo. *Cytoskeleton* **67,** 466–476 (2010).

166. Tan, S. C., Scherer, J. & Vallee, R. B. Recruitment of dynein to late endosomes and lysosomes through light intermediate chains. *Molecular biology of the cell* **22,** 467–477 (2011).

167. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. EMAN2: an extensible image processing suite for electron microscopy. *Journal of structural biology* **157,** 38–46 (2007).

168. Tang, X., Germain, B. S. & Lee, W.-L. A novel patch assembly domain in Num1 mediates dynein anchoring at the cortex during spindle positioning. *The Journal of cell biology* **196,** 743–756 (2012).

169. Toropova, K., Zou, S., Roberts, A. J., Redwine, W. B., Goodman, B. S., Reck-Peterson, S. L. & Leschziner, A. E. Lis1 regulates dynein by sterically blocking its mechanochemical cycle. *eLife* **3,** e03372 (2014).

170. Tynan, S. H., Gee, M. A. & Vallee, R. B. Distinct but overlapping sites within the cytoplasmic dynein heavy chain for dimerization and for intermediate chain and light intermediate chain binding. *Journal of Biological Chemistry* **275,** 32769–32774 (2000).

171. Urnavicius, L., Zhang, K., Diamant, A. G., Motz, C., Schlager, M. A., Yu, M., Patel, N. A., Robinson, C. V. & Carter, A. P. The structure of the dynactin complex and its interaction with dynein. *Science* **347,** 1441–1446 (2015).

172. Vale, R. D. Intracellular transport using microtubule-based motors. *Annual review of cell biology* **3,** 347–378 (1987).

173. Vale, R. D. The molecular motor toolbox for intracellular transport. *Cell* **112,** 467–480 (2003).

174. Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **288,** 88–95 (2000).

175. Vale, R. D., Reese, T. S. & Sheetz, M. P. Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* **42,** 39–50 (1985).

176. Vale, R. D., Schnapp, B. J., Mitchison, T., Steuer, E., Reese, T. S. & Sheetz, M. P. Different axoplasmic proteins generate movement in opposite directions along microtubules in vitro. *Cell* **43,** 623–632 (1985).

177. Varadi, A., Johnson-Cadwell, L. I., Cirulli, V., Yoon, Y., Allan, V. J. & Rutter, G. A. Cytoplasmic dynein regulates the subcellular distribution of mitochondria by controlling the recruitment of the fission factor dynamin-related protein-1. *Journal of cell science* **117,** 4389–4400 (2004).

178. Varga, V., Leduc, C., Bormuth, V., Diez, S. & Howard, J. Kinesin-8 motors act cooperatively to mediate length-dependent microtubule depolymerization. *Cell* **138,** 1174–1183 (2009).

179. Vijayachandran, L. S., Viola, C., Garzoni, F., Trowitzsch, S., Bieniossek, C., Chaillet, M., Schaffitzel, C., Busso, D., Romier, C., Poterszman, A., *et al.* Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. *Journal of structural biology* **175,** 198–208 (2011).

180. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. in *Macromolecular crystallography protocols* 215–230 (Springer, 2007).

181. Wang, Z., Khan, S. & Sheetz, M. P. Single cytoplasmic dynein molecule movements: characterization and comparison with kinesin. *Biophysical journal* **69,** 2011 (1995).

182. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 (2009).

183. Wells, A. L., Lin, A. W., Chen, L.-Q., Safer, D., Cain, S. M., Hasson, T., Carragher, B. O., Milligan, R. A. & Sweeney, H. L. Myosin VI is an actin-based motor that moves backwards. *Nature* **401,** 505–508 (1999).

184. Westphal, M., Jungbluth, A., Heidecker, M., Mühlbauer, B., Heizer, C., Schwartz, J.-M., Marriott, G. & Gerisch, G. Microfilament dynamics during cell movement and chemotaxis monitored using a GFP–actin fusion protein. *Current Biology* **7,** 176–183 (1997).

185. Williams, J. C., Roulhac, P. L., Roy, A. G., Vallee, R. B., Fitzgerald, M. C. & Hendrickson, W. A. Structural and thermodynamic characterization of a cytoplasmic dynein light chain–intermediate chain complex. *Proceedings of the National Academy of Sciences* **104,** 10028–10033 (2007).

186. Williams, J. C., Xie, H. & Hendrickson, W. A. Crystal structure of dynein light chain TcTex-1. *Journal of Biological Chemistry* **280,** 21981–21986 (2005).

187. 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., McCoy, A., *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67,** 235–242 (2011).

188. Wittmann, T., Hyman, A. & Desai, A. The spindle: a dynamic assembly of microtubules and motors. *Nature cell biology* **3,** E28–E34 (2001).

189. Xiang, X., Beckwith, S. M. & Morris, N. R. Cytoplasmic dynein is involved in nuclear migration in Aspergillus nidulans. *Proceedings of the National Academy of Sciences* **91,** 2100–2104 (1994).

190. Yagi, T., Uematsu, K., Liu, Z. & Kamiya, R. Identification of dyneins that localize exclusively to the proximal portion of Chlamydomonas flagella. *Journal of cell science* **122,** 1306–1314 (2009).

191. Yang, Z. Y., Jing, G., Ning, L., Min, Q., Wang, S. N. & Zhu, X. L. Mitosin/CENP-F is a conserved kinetochore protein subjected to cytoplasmic dynein-mediated poleward transport. *Cell research* **13,** 275–283 (2003).

192. Yildiz, A., Forkey, J. N., McKinney, S. A., Ha, T., Goldman, Y. E. & Selvin, P. R. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *science* **300,** 2061–2065 (2003).

193. Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. Kinesin walks hand-over-hand. *Science* **303,** 676–678 (2004).
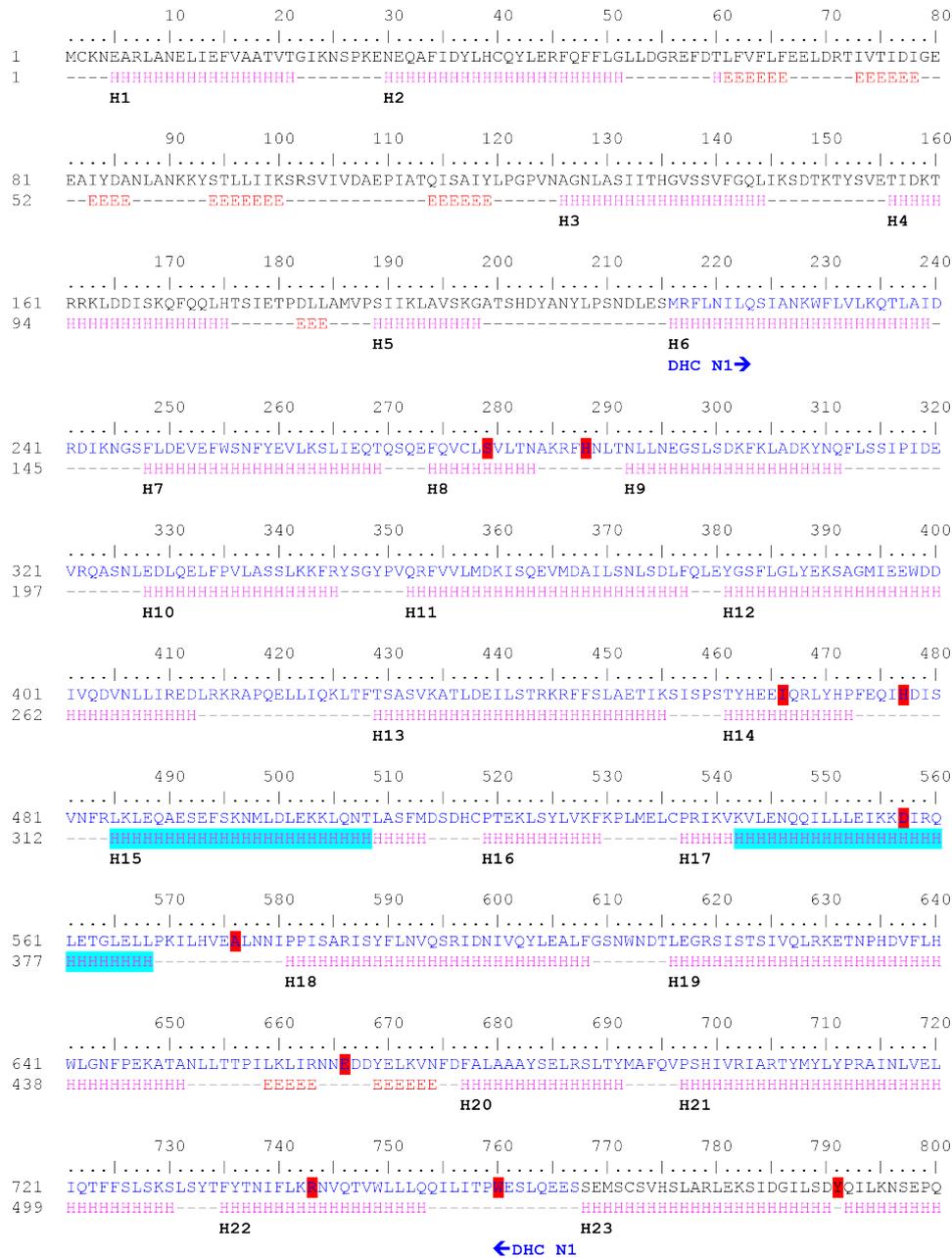
Appendix A

Supplemental figures and tables

```
              10        20        30        40        50        60        70        80
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1    MCKNEARLANELIEFVAATVTGIKNSPKENEQAFIDYLHCQYLERFQFFLGLLDGREFDTLFVFLFEELDRTIVTIDIGE
1    ----HHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHHH--------HEEEEEE------EEEEEE--
         H1                      H2

              90       100       110       120       130       140       150       160
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
81   EAIYDANLANKKYSTLLIIKSRSVIVDAEPIATQISAIYLPGPVNAGNLASIITHGVSSVFGQLIKSDTKTYSVETIDKT
52   --EEEE-------EEEEEEE-------------EEEEEE------HHHHHHHHHHHHHHHHHH-----------HHHHH
                                                  H3                            H4

             170       180       190       200       210       220       230       240
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
161  RRKLDDISKQFQQLHTSIETPDLLAMVPSIIKLAVSKGATSHDYANYLPSNDLESMRFLNILQSIANKWFLVLKQTLAID
94   HHHHHHHHHHHHHH------EEE----HHHHHHHHHH-----------------HHHHHHHHHHHHHHHHHHHHHHHH-
                              H5                        H6
                                                        DHC N1→

             250       260       270       280       290       300       310       320
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
241  RDIKNGSFLDEVEFWSNFYEVLKSLIEQTQSQEFQVCLEVLTNAKRFSNLTNLLNEGSLSDKFKLADKYNQFLSSIPIDE
145  -------HHHHHHHHHHHHHHHHHHHH----HHHHHHHHH--------HHHHHHHHHHHHHHHHHHHH---------
             H7                     H8                H9

             330       340       350       360       370       380       390       400
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
321  VRQASNLEDLQELFPVLASSLKKFRYSGYPVQRFVVLMDKISQEVMDAILSNLSDLFQLEYGSFLGLYEKSAGMIEEWDD
197  -------HHHHHHHHHHHHHHHH------HHHHHHHHHHHHHHHHHHHHHHHH---HHHHHHHHHHHHHHHHHHHHHHH
             H10                    H11                          H12

             410       420       430       440       450       460       470       480
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
401  IVQDVNLLIREDLRKRAPQELLIQKLTFTSASVKATLDEILSTRKRFFSLAETIKSISPSTYHEEIQRLYHPFEQIHDIS
262  HHHHHHHHHHH---------------HHHHHHHHHHHHHHHHHHHHHHHHHH-----HHHHHHHHHHH--------
                                    H13                        H14

             490       500       510       520       530       540       550       560
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
481  VNFRLKLEQAESEFSKNMLDLEKKLQNTLASFMDSDHCPTEKLSYLVKFKPLMELCPRIKVKVLENQQILLLEIKKDIRQ
312  ----HHHHHHHHHHHHHHHHHHHHHHH-HHHHH-----HHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHHHHH
         H15                            H16               H17

             570       580       590       600       610       620       630       640
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
561  LETGLELLPKILHVEHLNNIPPISARISYFLNVQSRIDNIVQYLEALFGSNWNDTLEGRSISTSIVQLRKETNPHDVFLH
377  HHHHHHHHH-----------HHHHHHHHHHHHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHHHHHHHHHH
                        H18                                    H19

             650       660       670       680       690       700       710       720
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
641  WLGNFPEKATANLLTTPILKLIRNNIDDYELKVNFDFALAAAYSELRSLTYMAFQVPSHIVRIARTYMYLYPRAINLVEL
438  HHHHHHHHHHH------EEEEE----EEEEEE--HHHHHHHHHHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHH
                        H20                                    H21

             730       740       750       760       770       780       790       800
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
721  IQTFFSLSKSLSYTFYTNIFLKRNVQTVWLLLQQILITPYESLQEESSEMSCSVHSLARLEKSIDGILSDYQILKNSEPQ
499  HHHHHHHHH----HHHHHHHHHHHHHHHHHH-------------HHHHHHHHHHHHHHHHHHHHHHHHHHH-HHHHHHH
             H22                          H23
                          ←DHC N1
```

Figure A.1: **Secondary structure prediction of DHC N-terminus: Part 1**

```
              810       820       830       840       850       860       870       880
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
801  FAKEFSGLKSFDGTADDLHEVEEIISNIQAIFENLFTKGLTNVSDHISTFNNLIISIILEKVRLNLKKMSFPKHVLKLSF
560  HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH------HHHHHHHHHHHHHHHHHHHHHH--------EEEEEEEE
                                                        H24

              890       900       910       920       930       940       950       960
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
881  NEGRITSSPSLAAMKRSLLKDIEALLNKVVLINFLHDPDHPLSTTLTFNSLVIKLKDDIQNCIEQVQNLHCKINSYVKEW
626  E--EEEEE--HHHHHHHHHHHHHHHHHHHH------------------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
          H25                                       H26

              970       980       990       1000      1010      1020      1030      1040
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
961  QKMEFLWQITEEAFLEVVDNSTQRCFGILKGLLDSQSKFDLIISRNNFSKNLVLHTEDAQRHIRSKMDSWILYVSKHLLT
682  H-HHHHH---HHHHHHHH----HHHHHHHHHHHHHHHHHH-----E---EEEEE-HHHHHHHHHHHHHHHHHHHHHHHHH
         H27           H28                                        H29

              1050      1060      1070      1080      1090      1100      1110      1120
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1041 IYERDARKLHEDMNRDREAVEDMDINFTSLKNITVIIEAVNVNKRHLTERDIQIKLLGSVMRALTKLKVRFPSHFVYIDQ
745  HHHHHHHHHHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH------HHHHHHH
                                 H30                                       H31

              1130      1140      1150      1160      1170      1180      1190      1200
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1121 LDNDFSSLRQSLSYVEQELQKHRVVIAKSLEEGVENINNLSQSLNESWSVRKPISPTLTPPEALKILEFFNESITKLKKK
812  HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---------HHHHHHHHHHHHHHHHHHHHHHHH
                                                                H32

              1210      1220      1230      1240      1250      1260      1270      1280
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1201 MHSVAAAAKMLLIPVVLNDQLTHVVEEVKTYDLVWRSIKNLWEDVQRTFETPWCRVDVLLLQSDLANFLRRADELPRAVK
883  HHHHHHHHHH--------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHH----
                     H33                                        H34
                     DHC N2➜

              1290      1300      1310      1320      1330      1340      1350      1360
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1281 QFEMYKSLFSQVNMLTSVNKILVELKDGALKPRHWNMIFRDIGKRQIQKNLLDKLEFSLKDVMVLNLTLNEILLTKIIER
943  --HHHHHHHHHHHHHHHHHHHHH-----HHHHHHHHHH--------------HHHHH---HHHHHHHHHHHHHH
       H35                       H36                       H37      H38

              1370      1380      1390      1400      1410      1420      1430      1440
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1361 AQLEFVIEKSLNRIKKFWKEAQYEVIEHSSGLKLVREWDVLEQACKEDLEELVSMKASNYYKIFEQDCLDLESKLTKLSE
999  HHHHHHHHHHHHHHHH---EEEEEE-----EEE--HHHHHHHHHHHHHHHH-----HHHHHHHHHHHHHHHHHHH
       H1 (Linker)            H2                        H3

              1450      1460      1470      1480      1490      1500      1510      1520
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1441 IQVNWVEVQFYWLDLYGIYGENLDIQNFLPLETSKFRSLTSEYKMITTRAFQLDTTIEVIHIPNFDTTLKLTIDSLKMIK
1064 HHHHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHHHHHHHHHHHHHHH---HHHHH---HHHHHHHHHHHHHH
                     H4                                         H5

              1530      1540      1550      1560      1570      1580      1590      1600
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
1521 SSLSTFLERQRRQFPRFYFLGNDDLLKIIGSGKHHDQVSKFMKKMFGSIESIIFLEDFITGVRSVEGEVLNLNEKIELKD
1131 HHHHHHHHHHHHH---EEE---HHHHHHH-----HHHHHHHHHHHH---EEE----EEEEEE----EEEE----EE---
                     H6          H7
```

Figure A.2: **Secondary structure prediction of DHC N-terminus: Part 2**

Figure A.3: **An alignment of the DHC "N region" from thirteen different species.** Full-length DHCs from thirteen different species were aligned using ClustalX. The alignment was visualised with Jalview 2 using the "percentage identity" colour scheme (darker purple indicates higher conservation). Only the region preceding and including the DHC-NH truncation are included. The red arrow indicates residues S1004 of *S. cerevisiae*, which was used to make a human equivalent of the tail complex truncation. Ag=*Ashbya gossypii*, Ce=*Caenorhabditis elegans*, Ct=*Chaetomium thermophilum*, Dd=*Dictyostelium discoideum*, Dm=*Drosophila melanogaster*, Dr=*Danio rerio*, Gg=*Gallus gallus*, Hs=*Homo sapiens*, Kl=*Kluyveromyces lactis*, Mm=*Mus musculus*, Nc=*Neurospora crassa*, Sc=*Saccharomyces cerevisiae*, Tt=*Tetrahymena thermophila*.

Strain	Description	Genotype	Reference	Thesis section
CY176	S288C wildtype	MAT+_ SUC2 gal2 mal mel flo1 flo8-1 hap1 ho bio1 bio6	Commercial ATCC	2.2.1
CY223	DSY-5	MATalpha leu2 trp1 ura3-52 his3::GAL1-GAL4 pep4 prb1-1122	Commercial MoBiTec	2.2.1
CY282	BCY123	Pep4HIS3 prbLEU2 bar1:HISG lys2GAL1/10-GAL4 can1 ade2 ura3 leu2-3, 112	Nagai lab, MRC-LMB	2.2.1
CY286	Overexpresses full-length dynein complex	MATalpha leu2 trp1 ura3-52 his3::GAL1-GAL4 pep4 prb1-1122 HIS3-pGAL1::DYN2 LEU2-pGAL1::PAC11 URA3-pGAL1::DYN3 TRP1-pGAL1-ZZ-LTL-DYN1	New	2.2.2
CY296	CY286 but ura-	MATalpha leu2 trp1 ura3-52 his3::GAL1-GAL4 pep4 prb1-1122 HIS3-pGAL1::DYN2 LEU2-pGAL1::PAC11 URA3-pGAL1::DYN3 TRP1-pGAL1-ZZ-LTL-DYN1 ura3Δ(using flanking repeats)	New	2.2.3
CY298	DHC-A	CY296 with a stop codon and URA3 inserted after aa 279 of Dyn1	New	2.2.3
CY299	DHC-B	CY296 with a stop codon and URA3 inserted after aa 288 of Dyn1	New	2.2.3
CY300	DHC-D	CY296 with a stop codon and URA3 inserted after aa 477 of Dyn1	New	2.2.3
CY301	DHC-E	CY296 with a stop codon and URA3 inserted after aa 557 of Dyn1	New	2.2.3
CY302	DHC-F	CY296 with a stop codon and URA3 inserted after aa 576 of Dyn1	New	2.2.3
CY303	DHC-G	CY296 with a stop codon and URA3 inserted after aa 666 of Dyn1	New	2.2.3
CY304	DHC-H	CY296 with a stop codon and URA3 inserted after aa 743 of Dyn1	New	2.2.3
CY305	DHC-I	CY296 with a stop codon and URA3 inserted after aa 760 of Dyn1	New	2.2.3
CY306	DHC-J	CY296 with a stop codon and URA3 inserted after aa 791 of Dyn1	New	2.2.3
CY307	DHC-K	CY296 with a stop codon and URA3 inserted after aa 856 of Dyn1	New	2.2.3
CY308	DHC-L	CY296 with a stop codon and URA3 inserted after aa 870 of Dyn1	New	2.2.3
CY309	DHC-M	CY296 with a stop codon and URA3 inserted after aa 954 of Dyn1	New	2.2.3
CY310	DHC-N	CY296 with a stop codon and URA3 inserted after aa 962 of Dyn1	New	2.2.3
CY311	DHC-O	CY296 with a stop codon and URA3 inserted after aa 1048 of Dyn1	New	2.2.3
CY312	DHC-P	CY296 with a stop codon and URA3 inserted after aa 1065 of Dyn1	New	2.2.3
CY313	DHC-Q	CY296 with a stop codon and URA3 inserted after aa 1148 of Dyn1	New	2.2.3
CY314	DHC-R	CY296 with a stop codon and URA3 inserted after aa 1175 of Dyn1	New	2.2.3
CY315	DHC-S	CY296 with a stop codon and URA3 inserted after aa 1239 of Dyn1	New	2.2.3
CY316	DHC-T	CY296 with a stop codon and URA3 inserted after aa 1254 of Dyn1	New	2.2.3
CY317	DHC-U	CY296 with a stop codon i and URA3 nserted after aa 1300 of Dyn1	New	2.2.3
CY318	DHC-V	CY296 with a stop codon and URA3 inserted after aa 1310 of Dyn1	New	2.2.3
CY319	DHC-W	CY296 with a stop codon and URA3 inserted after aa 1340 of Dyn1	New	2.2.3
CY320	DHC-X	CY296 with a stop codon and URA3 inserted after aa 1346 of Dyn1	New	2.2.3
CY321	DHC-Y	CY296 with a stop codon and URA3 inserted after aa 1363 of Dyn1	New	2.2.3
CY322	DHC-Z	CY296 with a stop codon and URA3 inserted after aa 1459 of Dyn1	New	2.2.3
CY323	DHC-Ω	CY296 with a stop codon and URA3 inserted after aa 1477 of Dyn1	New	2.2.3
CY332	DHC-NA	CY296 with a stop codon and URA3 inserted after aa 958 of Dyn1	New	2.2.4
CY333	DHC-NB	CY296 with a stop codon and URA3 inserted after aa 962 of Dyn1	New	2.2.4
CY334	DHC-NC	CY296 with a stop codon and URA3 inserted after aa 968 of Dyn1	New	2.2.4
CY335	DHC-ND	CY296 with a stop codon and URA3 inserted after aa 979 of Dyn1	New	2.2.4
CY336	DHC-NE	CY296 with a stop codon and URA3 inserted after aa 981 of Dyn1	New	2.2.4
CY337	DHC-NF	CY296 with a stop codon and URA3 inserted after aa 990 of Dyn1	New	2.2.4
CY338	DHC-NG	CY296 with a stop codon and URA3 inserted after aa 995 of Dyn1	New	2.2.4
CY339	DHC-NH	CY296 with a stop codon and URA3 inserted after aa 1004 of Dyn1	New	2.2.4
CY340	DHC-NI	CY296 with a stop codon and URA3 inserted after aa 1016 of Dyn1	New	2.2.4
CY341	DHC-NJ	CY296 with a stop codon and URA3 inserted after aa 1025 of Dyn1	New	2.2.4
CY342	DHC-NK	CY296 with a stop codon and URA3 inserted after aa 1035 of Dyn1	New	2.2.4
CY343	DHC-NL	CY296 with a stop codon and URA3 inserted after aa 1063 of Dyn1	New	2.2.4
CY344	DHC-NA-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 958 of Dyn1	New	2.2.4
CY345	DHC-NB-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 962 of Dyn1	New	2.2.4
CY346	DHC-NC-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 968 of Dyn1	New	2.2.4
CY347	DHC-ND-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 979 of Dyn1	New	2.2.4
CY348	DHC-NE-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 981 of Dyn1	New	2.2.4
CY349	DHC-NF-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 990 of Dyn1	New	2.2.4
CY350	DHC-NG-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 995 of Dyn1	New	2.2.4
CY351	DHC-NH-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 1004 of Dyn1	New	2.2.4
CY352	DHC-NI-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 1016 of Dyn1	New	2.2.4
CY353	DHC-NJ-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 1025 of Dyn1	New	2.2.4
CY354	DHC-NK-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 1035 of Dyn1	New	2.2.4
CY355	DHC-NL-GFP	CY296 with a GFP, stop codon and URA3 inserted after aa 1063 of Dyn1	New	2.2.4

Figure A.4: **Yeast strains used in and created for the work described in the thesis.**

Plasmid	Description	Vector	Resistance	Thesis section
P61	ZZ-tag template, pDyn:pGal1:ZZ-LTEVLTEV-GST-D6	n/a	Amp	2.2.1
P107	pTBSN, TEV protease gene, S219N, from Jennifer Doudna's lab	n/a	Amp	2.2.1
P230	Polyprotein with Dyn1 1-1363	pYES2	Amp	2.2.1
P371	Polyprotein NH Dyn1 1-1004	pYES2	Amp	2.2.6
P474	Multiple cassettes expression of NH truncation in yeast	pYES2	AMP	2.2.6
P526	Nip100 fragment A 249-323	HLTV	Amp	3.2.5
P527	Nip100 fragment B 249-336	HLTV	Amp	3.2.5
P528	Nip100 fragment C 249-343	HLTV	Amp	3.2.5
P529	Nip100 fragment D 249-352	HLTV	Amp	3.2.5
P530	Nip100 fragment E 249-355	HLTV	Amp	3.2.5
P636	P474 with DIC-A	pYES2	Amp	3.2.4
P637	P474 with DIC-B	pYES2	Amp	3.2.4
P638	P474 with DIC-C	pYES2	Amp	3.2.4
P639	P474 with DIC-D	pYES2	Amp	3.2.4
P640	P474 with DIC-E	pYES2	Amp	3.2.4
P641	P474 with DIC-F	pYES2	Amp	3.2.4
P642	P474 with DIC-G	pYES2	Amp	3.2.4
P643	P474 with DIC-H	pYES2	Amp	3.2.4
P646	P474 with DIC-I	pYES2	Amp	3.2.4
P645	P474 with DIC-J	pYES2	Amp	3.2.4
P646	P474 with DIC-K	pYES2	Amp	3.2.4
P647	P474 with DIC-L	pYES2	Amp	3.2.4
P648	P474 with DIC-M	pYES2	Amp	3.2.4
P649	P474 with DIC-N	pYES2	Amp	3.2.4
P650	P474 with DIC-O	pYES2	Amp	3.2.4
P651	P474 with DIC-P	pYES2	Amp	3.2.4
P652	P474 with DIC-Q	pYES2	Amp	3.2.4
P653	P474 with DIC-R	pYES2	Amp	3.2.4
P654	P474 with DIC-S	pYES2	Amp	3.2.4
P655	P474 with DIC-T	pYES2	Amp	3.2.4
P656	P474 with DIC-U	pYES2	Amp	3.2.4
P660	P474 with DLIC-A	pYES2	Amp	3.2.3
P661	P474 with DLIC-B	pYES2	Amp	3.2.3
P662	P474 with DLIC-C	pYES2	Amp	3.2.3
P663	P474 with DLIC-D	pYES2	Amp	3.2.3
P664	P474 with DLIC-E	pYES2	Amp	3.2.3
P665	P474 with DLIC-F	pYES2	Amp	3.2.3
P666	P474 with DLIC-G	pYES2	Amp	3.2.3
P667	P474 with DLIC-H	pYES2	Amp	3.2.3
P668	P474 with DLIC-I	pYES2	Amp	3.2.3
P669	P474 with DLIC-J	pYES2	Amp	3.2.3

Figure A.5: **Plasmids used in and created for the work described in the thesis: Part 1.**

Plasmid	Description	Vector	Resistance	Thesis section
P670	P474 with DLIC-K	pYES2	Amp	3.2.3
P671	P474 with DLIC-L	pYES2	Amp	3.2.3
P672	P474 with DLIC-M	pYES2	Amp	3.2.3
P673	P474 with DLIC-N	pYES2	Amp	3.2.3
P674	P474 with DLIC-O	pYES2	Amp	3.2.3
P675	P474 with DLIC-P	pYES2	Amp	3.2.3
P676	P474 with DLIC-Q	pYES2	Amp	3.2.3
P677	P474 with DLIC-R	pYES2	Amp	3.2.3
P678	P474 with DLIC-S	pYES2	Amp	3.2.3
P679	P474 with DLIC-T	pYES2	Amp	3.2.3
P563	NH truncation of yeast tail for expression in Sf9 cells	pACEBAC1	Gen	2.2.7
P704	P474 plus TDA2 cassette	pYES2	Amp	3.2.2
P792	DHC-E for expression in Sf9 cells	pACEBAC1	Gen	4.2.1
P999	DHC-EA	HLTV	Amp	4.2.7
P1000	DHC-EB	HLTV	Amp	4.2.7
P1001	DHC-EC	HLTV	Amp	4.2.7
P1002	DHC-ED	HLTV	Amp	4.2.7
P1003	DHC-EE	HLTV	Amp	4.2.7
P1004	DHC-EF	HLTV	Amp	4.2.7
P1005	DHC-EG	HLTV	Amp	4.2.7
P1006	DHC-EH	HLTV	Amp	4.2.7
P1007	DHC-EI	HLTV	Amp	4.2.7
P1008	DHC-EJ	HLTV	Amp	4.2.7
P1009	DHC-EK	HLTV	Amp	4.2.7
P1010	DHC-EL	HLTV	Amp	4.2.7
P1382	DHC DimerA	HLTV	Amp	4.2.7
P1383	DHC ElongatedA	HLTV	Amp	4.2.7
P1384	DHC DimerB	HLTV	Amp	4.2.7
P1385	DHC ElongatedB	HLTV	Amp	4.2.7
P1386	DHC C-terminalA	HLTV	Amp	4.2.7
P1387	DHC C-terminalB	HLTV	Amp	4.2.7
P1388	DHC-E from *A. gossypii*	HLTV	Amp	4.2.7
P1389	DHC-E from *C. elegans*	HLTV	Amp	4.2.7
P1390	DHC-E from *C. thermophilum*	HLTV	Amp	4.2.7
P1391	DHC-E from *D. discoideum*	HLTV	Amp	4.2.7
P1392	DHC-E from *D. melanogaster*	HLTV	Amp	4.2.7
P1393	DHC-E from *D. rerio*	HLTV	Amp	4.2.7
P1394	DHC-E from *G. gallus*	HLTV	Amp	4.2.7
P1395	DHC-E from *H. sapiens*	HLTV	Amp	4.2.7
P1396	DHC-E from *K. lactis*	HLTV	Amp	4.2.7
P1397	DHC-E from *M. musculus*	HLTV	Amp	4.2.7
P1398	DHC-E from *N. crassa*	HLTV	Amp	4.2.7
P1399	DHC-E from *T. thermophila*	HLTV	Amp	4.2.7

Figure A.6: **Plasmids used in and created for the work described in the thesis: Part 2.**

	Native$_{comb}$	**Peak$_A$**	**Inflection$_A$**	**Remote$_A$**
Diffraction data				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)	$a = 71.9$	$a = 72.0$	$a = 72.1$	$a = 72.1$
	$b = 148.9$	$b = 149.0$	$b = 148.9$	$b = 149.2$
	$c = 179.6$	$c = 179.4$	$c = 179.5$	$c = 178.8$
Unit cell (°)	$\alpha = 90.0$	$\alpha = 90.0$	$\alpha = 90.0$	$\alpha = 90.0$
	$\beta = 90.0$	$\beta = 90.0$	$\beta = 90.0$	$\beta = 90.0$
	$\gamma = 90.0$	$\gamma = 90.0$	$\gamma = 90.0$	$\gamma = 90.0$
Resolution (Å)	52.6 - 5.0	64.9 - 5.0	61.0 - 5.0	64.9 - 5.0
R$_{Sym}$[1]	7.7 (28.7)*	9.9 (58.2)	7.5 (41.5)	7.5 (37.1)
$<I>/<\sigma I>$	15.8 (9.7)	12.9 (4.3)	14.7 (4.9)	14.5 (5.5)
Completeness (%)	99.9 (100.0)	99.8 (99.3)	99.8 (99.3)	99.8 (99.3)
Redundancy	11.4 (11.8)	9.9 (9.9)	9.8 (9.9)	9.6 (9.6)
No. reflections	101,399 (29,169)	87,406 (24,417)	87,249 (24,362)	84,886 (23,558)
No. unique reflections	8,866 (2,463)	8,870 (2,455)	8,870 (2,455)	8,871 (2,455)
Wavelength (Å)	0.9763	0.9795	0.9797	0.9393
CC$_{anom}$=0.5 cutoff (Å)	-	6.7	-	-
Phasing statistics				
FOM[2]	0.36	-	-	-

Table A.1: **Data collection and phasing statistics for the four datasets used to solve the SeMet DHC-E structure cut to 5.0Å.** *Values in parentheses refer to the highest resolution shell. [1]R$_{Sym}$ (I) = ($\Sigma_{hkl}\Sigma_i |I_{hkl,i} - <I_{hkl}>|$) / $\Sigma_{hkl}\Sigma_i I_{hkl,i}$ for n independent reflections and i observations of a given reflection. [2]FOM = mean figure of merit.