Investigating the role of Deleted in Primary Ciliary Dyskinesia (DPCD) protein in *Trypanosoma brucei*

Emily Ablett

This thesis is submitted for the degree of Msc (by research) Biomedical Science

Department of Biomedical and Life Sciences

September 2022

I declare that this thesis is my own work and has not been submitted in part, or as a whole, for the award of a higher degree elsewhere.

Table of Contents

List of Tables

<u>Acknowledgements</u>

I would like to thank my supervisors Paul McKean and Sarah Allinson for all their support and guidance throughout this project. I would also like to thank the members of the McKean lab.

Abstract

Primary Ciliary Dyskinesia (PCD) is the most common ciliary syndrome in humans, associated with the reduction or loss of function of motile cilia due to inherited mutations. This results in life-altering symptoms spanning multiple organ systems, including airway infections and infertility. Since its discovery in 2004, Deleted in Primary Ciliary Dyskinesia (DPCD) protein has become a novel gene candidate for PCD, with a predicted role in the formation of cilia by ciliogenesis. However, the precise role of this protein in ciliogenesis remains obscure. To provide further insight into the role of this enigmatic protein, the *Trypanosoma brucei* homologue of DPCD was investigated. Bioinformatics interrogation revealed that DPCD is strongly conserved across ciliated eukaryotes but that a homologue of DPCD is only found in eukaryotes with motile cilia supporting a proposed role in ciliogenesis. However, endogenous tagging of the *Tb*DPCD protein in procyclic *T. brucei* cells localised the protein to the cell cytoplasm, while RNAi induced ablation of *Tb*DPCD produced no observable effect on cell growth, morphology, motility, or cell cycle progression. This suggests that *Tb*DPCD is not essential for ciliogenesis in *T. brucei* or potentially has a redundant function. The in-silico interrogation of DPCD protein structure/function using Phyre2 and Alphafold predicted that DPCD has structural similarity to proteins that have chaperone or co-chaperone activity, including small heat-shock proteins. Consequently, the *Tb*DPCD protein was overexpressed and purified from *Escherichia coli* to determine whether *Tb*DPCD demonstrated chaperone activity under heat stress conditions using a thermal aggregation assay. These experiments could not demonstrate that *Tb*DPCD possesses chaperone activity under heat stress, however, the results indicated that *Tb*DPCD is not stable at elevated temperatures. Overall, the in-silico analysis gives a strong indication of the role of DPCD as a chaperone, however in vivo and biochemical studies did not provide evidence to support this.

Section 1: Introduction

Cilia and flagella are highly conserved organelles implicated in a wide range of key biological processes in eukaryotes (Wan and Jekely, 2020). Due to the importance of cilia in maintaining human health, mutations in genes encoding proteins important for ciliary function and formation result in a diversity of diseases, collectively known as ciliopathies (Shoemark et al., 2020). This literature review will discuss in detail the structure and function of cilia, referring to cilia and flagella interchangeably, and how ciliary dysfunction can result in ciliopathies, with a particular focus on Primary Ciliary Dyskinesia (PCD). PCD is associated with reduction or loss of function of motile cilia, most commonly as a result of the loss of, or defects in, axonemal dynein arms. While a number of mutations resulting in PCD have been identified, with 65% patients having a known genetic basis (Tilley et al., 2015), the remaining cases have no identifiable genetic cause, suggesting other genes remain to be discovered. A potential gene candidate for PCD is Deleted in Primary Ciliary Dyskinesia (DPCD), however the biochemical function of this protein and its connection to ciliogenesis remains unresolved (Zariwala et al., 2004). The DPCD protein is the focus of the experimental work described in this thesis, where its function is interrogated using the protozoan parasite *T. brucei* as a model organism.

Section 1.1: The structure, function, and formation of cilia

1.1.1: The evolution of cilia

Cilia and flagella are microtubule-based organelles that protrude from the surface of many eukaryotic cells and have important motility and sensory functions (Wan and Jekely, 2020). These organelles contain a ciliary axoneme which is surrounded by the ciliary membrane (Garcia et al., 2018). The axoneme develops from and is anchored to the basal body, which has a role in microtubule organisation (Waters and Beales, 2011). Both the axoneme and basal body are connected by the transition zone (TZ) which helps maintain compartmentalisation (Nevers et al., 2017). The axoneme consists of nine doublet microtubules and characterises the cilia as motile

or nonmotile (primary), depending on the presence of a central pair of singlet microtubules (Waters and Beales, 2011). With motile cilia referred to as 9+2 and primary cilia as 9+0.

The organelle is thought to have ancient evolutionary origins in unicellular microorganisms, before the divergence of the Last Eukaryotic Common Ancestor (LECA) (Wan and Jekely, 2020). It is thought that eight gene duplications prior to the LECA led to the evolution of the complexity of the cilium and expansion of Heavy Chain Dynein motor proteins, which make up the ciliary axoneme (Hartman and Smith, 2009). Ciliary evolution is also thought to be linked to the evolution of mitosis due to the relationship between mitosis and the basal body (Vaughan and Gull, 2016). The LECA likely possessed a typical 9+2 cilia which is indistinguishable from cilia today in both structure and function (Mitchell, 2017). With sensory function arising first followed by motility alongside the evolution of the axoneme (Satir et al., 2008). Most current ciliary proteins were also present in the LECA suggesting ciliary evolution occurred before the divergence of extant eukaryotic clades (Mitchell, 2017).

Since the LECA, the organelle has diversified in modern eukaryotes, including changes in architecture and regulation of beating patterns (Mitchell, 2017). There has also been reduction or loss of cilia, with many lineages including most seed plants, Fungi and Amoebozoa being non-ciliated (Nevers et al., 2017).

1.1.2: Structure and function of nonmotile (primary) cilia

As mentioned above, cilia possess either a 9+2 or 9+0 axonemal structure. 9+0 cilia are known as primary cilia and carry out chemosensory, osmosensory and phototransduction functions. These are found on the apical surface of nearly all cell types, most commonly epithelial cells (Wheway et al., 2018). As with all cilia, the microtubule axoneme is made up of alpha and beta tubulin, which forms a radial array of nine microtubule doublets, encased in the ciliary membrane. Unlike the motile cilia, there is an absence of a central pair of microtubules and the related dynein arm structures (Satir and Christensen, 2007). The structure of nonmotile primary cilia can be seen in Figure 1.1. In addition, the formation of primary cilia occurs during G0 or G2 phase of the cell cycle, reliant on the process of Intraflagellar transport (IFT) (Wheway et al., 2018), this process is described in greater detail in section 1.1.4.
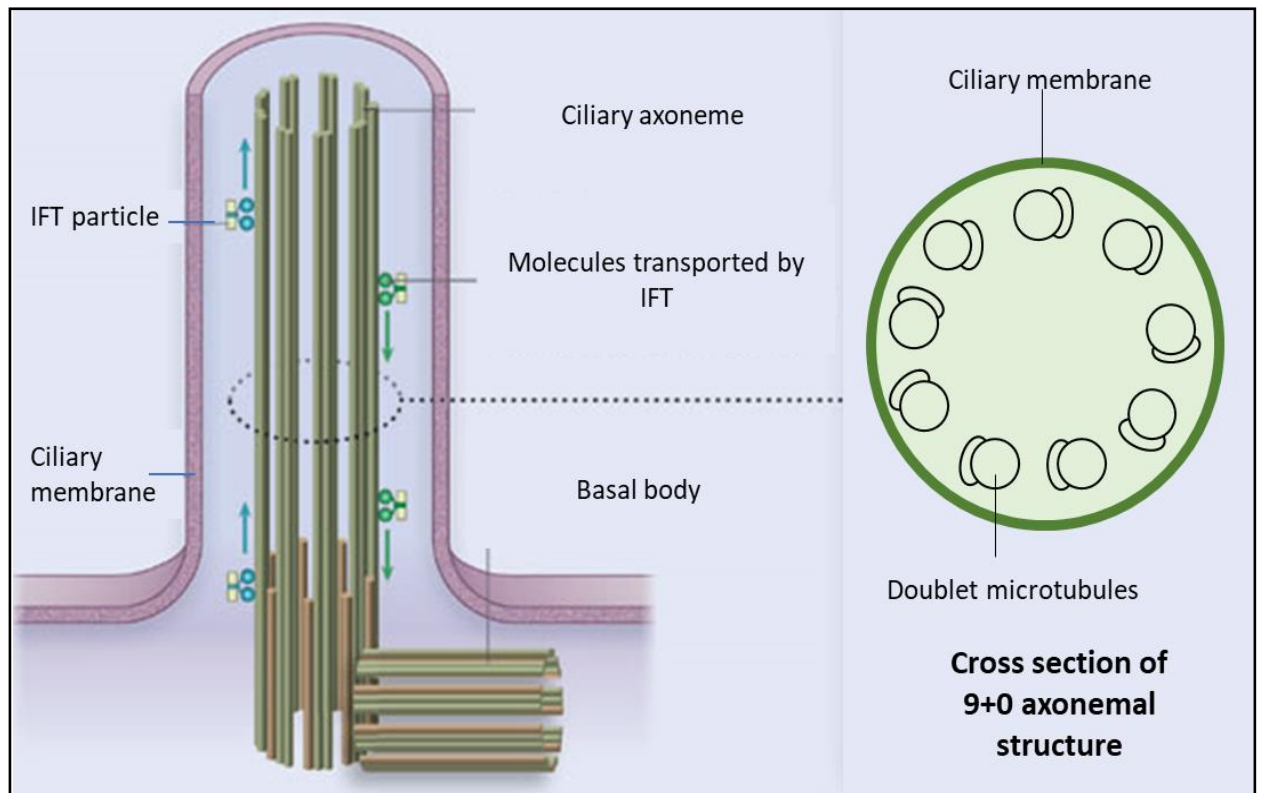
Figure 1.1: Diagram of primary cilia structure. Primary cilia are membrane bound organelles, containing the ciliary axoneme made up of 9 doublet microtubules. They are anchored to the cell by the basal body and assembled by IFT. Figure edited from: Ainsworth et al, 2007.

Primary cilia have an important role in cell signalling in the human body. This includes during the normal function and cell differentiation of kidney epithelium, with primary cilia acting as mechanosensors which respond to changes in the osmolarity and flow of urine by modulating signalling pathways (Mitchinson and Valente, 2016). The loss of cilia has been linked to abnormal kidney function and cell division in the form of polycystic kidney disease (PKD), first described in mice with mutated IFT particle IFT88 (Tg737) (Pazour et al., 2000).

In addition, primary cilia found on the ductal cells and islets of Langerhans of the pancreas, are essential for the modulation of Sonic hedgehog (Shh), Wnt and Notch signalling during development of the pancreas (Mitchinson and Valente, 2016). The link between primary cilia and Shh signalling was made by Huangfu et al., 2003, where mice with mutations in genes encoding IFT proteins, including *wimple* (a mouse homologue of IFT172) and *flexo* (a mouse homologue of IFT88) had defects in embryonic development due to Shh signalling defects. In addition, the role of primary cilium in Wnt signalling was described in Corbit et al., 2007, where

mouse embryos, primary fibroblasts, and embryonic stem cells with mutations in ciliogenesis factors KIF3A, IFT88 and OFD1, had disruptions in the canonical Wnt pathway.

Furthermore, the retina of the eye is dependent on a ciliary organelle, known as the outer segment, for detection of light by the cone and rod photoreceptors. The outer segment is connected to the cell body through 9+0 cilium which regulate the movement of proteins, including rhodopsin, in and out of the outer segment through IFT. This is demonstrated in the ciliopathy, retinitis pigmentosa, where CC2D2A which regulates 9+0 cilia and outer segment extension, is mutated (Mitchinson and Valente, 2016).

1.1.3: Motile cilia

1.1.3.1: Structure of motile cilia

As well as 9+0 primary cilia, cilia can exist as 9+2 structures that possess motility functions. As mentioned in section 1.1, motile cilia consist of an axoneme made up of nine microtubule doublets and a central pair of singlet microtubules. Radial spoke proteins protrude from each microtubule doublet towards the central pair, connecting them (Ishikawa, 2017). Moreover, unlike in the primary cilia, microtubule doublets are linked by dynein, a family of ATPase motor proteins, and the nexin- dynein regulatory complex (N-DRC). The dynein proteins form inner and outer rows of arms, consisting of polypeptides of different sizes, including heavy chains of 400-500kDa, intermediate chains of 45-110kDa and light chains of 8-55kDa (Ibanez-Tallon et al., 2003). Each axoneme possesses four identical outer dynein arms (ODAs) with 24nm spacing, which consist of three heavy chains (α, β and γ), packaged with two intermediate (IC69 and IC78) and eight light chains, and one set of inner dynein arms (IDAs) (Ishikawa, 2017; King, 2016). The IDAs are more complex with multiple isoforms being identified, including two-headed and single-headed forms. Each isoform consists of at least seven heavy chain isoforms and different combinations of intermediate and light chains (Satir and Christensen, 2008). An example of this is the two-headed isoform I1 (isoform f) which contains two heavy chains (1α and 1β), three intermediate chains (IC97, IC138 and IC140) and three light chains (Ibanez-Tallon et al., 2003). A diagram of the structure of 9+2 motile cilia is seen in Figure 1.2.
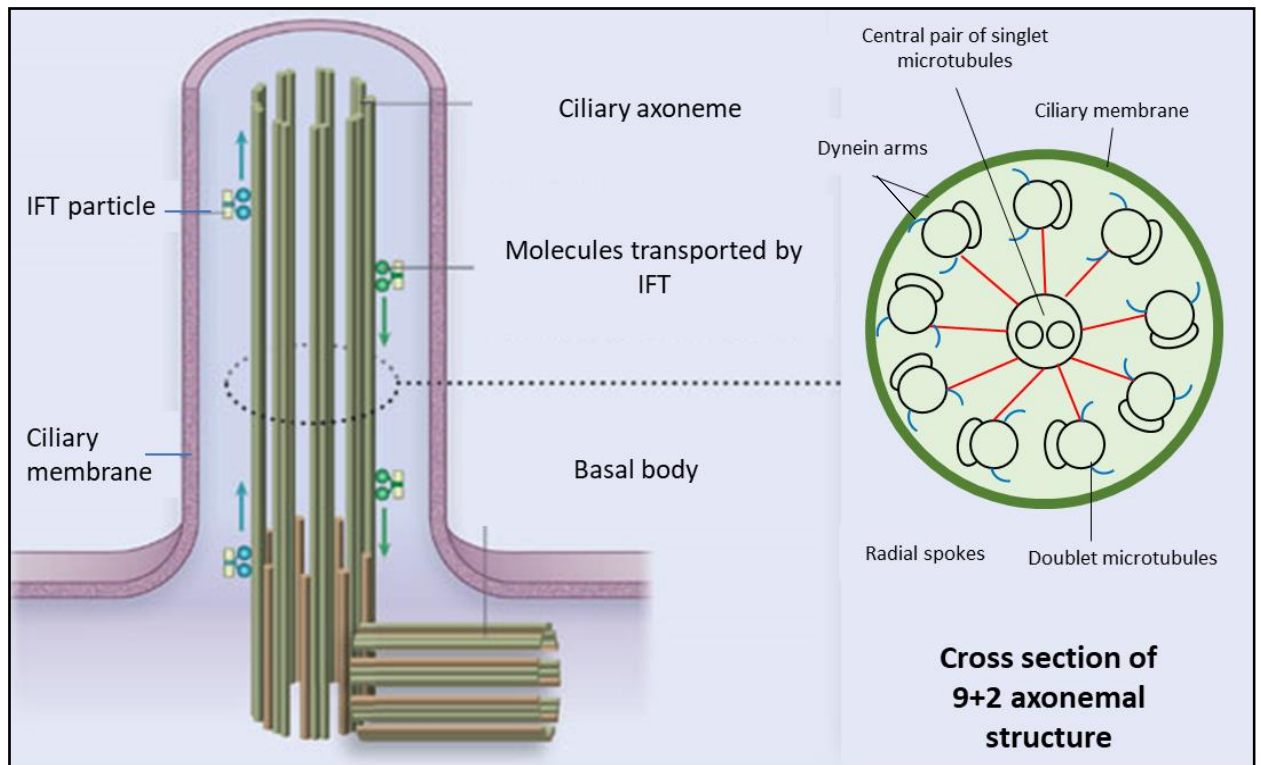
Figure 1.2: Diagram of motile cilia structure. Motile cilia are membrane bound organelles, containing the ciliary axoneme made up of 9 doublet microtubules and a central pair of single microtubules, connected by radial spoke proteins. Inner and outer dynein arms are present on the doublet microtubules. They are anchored to the cell by the basal body and assembled by IFT. Figure edited from: Ainsworth et al: 2007.

The axonemal dynein motor proteins of the inner and outer dynein arms are responsible for ciliary beating via ATPase activity. IDA and ODA heavy chains contain a catalytic domain made up of six AAA+ domains, AAA1 and AAA3 are sites of ATP hydrolysis while AAA4 acts as a microtubule binding domain (Ishikawa, 2017). The chemical energy created by ATP hydrolysis on the catalytic domain is converted into mechanical work by causing microtubule doublets to slide, through the attachment of the α-tubule of one doublet to the β-tubule on the adjacent doublet (King, 2016). After this, the sliding force is converted to bending, controlled by the phosphorylation of radial spoke proteins by IDAs, leading to interactions between radial spoke heads and central-pair microtubule projections (Satir et al., 2014). This causes progression of activity around the axoneme, leading to the helical beat (Satir and Christensen, 2007).

The full back-and-forth motion of the cilium, known as the ciliary stroke, is caused by the transfer from the resting state to the recovery stroke by swinging 180° backwards and extending. The

cilium then enters the effective stroke at 110°, before the cycle is repeated (Salathe, 2007). The ciliary beat frequency (CBF) increases when all three of these phases are shortened. Evidence suggests that the ODAs are primarily responsible for the control of CBF through cAMP-dependant ODA light chain phosphorylation, which increases doublet sliding velocity (Hamasaki, 1999; Christensen et al., 2001). In addition, several different factors have been found to influence CBF in mammals, including calcium concentration [Ca^{2+}]. CBF was found to be controlled by low ranges of Ca^{2+} acting on the axonemal site at the ciliary base, by Lansley and Sanderson, 1999, with maximum beat frequency induced by increasing [Ca^{2+}]. The source of calcium was determined to come from intracellular stores or influx through the plasma membrane. Also, the regulation of CBF by calcium is rapid, occurring within one beat cycle (Salathe, 2007). Furthermore, changes in pH have also been found to effect CBF. These changes affect the axoneme directly, with the increase of intracellular pH leading to faster CBF, while decreases attenuate CBF (Sutto et al., 2004).

Moreover, the orientation of the central-pair complex has been linked to difference in planar beating of cilia between eukaryotes. In members of the SAR and Plantae superclades, the central pair rotates during ciliary bending causing the central pair plane to be parallel to the bend plane. While in members of the excavata superclade the orientation of the central pair is fixed causing the central pair plane to be perpendicular to the bend plane (Mitchell, 2017).

1.1.3.2: Function of motile cilia

In comparison to primary cilia, there is a greater understanding of the role of motile cilia in the human body. The beating of the motile cilia leads to the movement of extracellular fluid. This has an important role in a wide range of bodily functions including, mucociliary clearance (MCC) in the lungs, fertility, and embryonic development. MCC is the primary innate immune mechanism of the lungs through mediating the propulsion of the mucus gel layer, along with pathogens and inhaled particles trapped within it, out of the airways (Bustamante-Marin and Ostrowski, 2017). The mucus gel layer is propelled in a cephalad direction, from the lower respiratory tract towards the head, where it can be swallowed (Tilley et al., 2015). This is mediated by highly coordinated in-plane beating of approximately 10^{12} cilia present on multiciliated cells, aided by lubrication provided by the periciliary layer (Bustamante-Marin and

Ostrowski, 2017). This generates a wave-like movement across the epithelial surface which propels the mucus gel layer (Mitchison and Valente, 2016).

In addition, motile cilia play an important role in both male and female fertility. In males, spermatozoa cells possess a flagellum with a 9+2 axoneme. This propels the male gametes towards the uterus for fertilization (Inaba and Mizuno, 2016). In female reproduction, the role of motile cilia is more complex. The first evidence of motile cilia in female fertility came from Eddy et al., 1978, where ovum transport and ciliary activity was examined in 35 female rabbits after the microsurgical reversal or double transection of segments of the distal, middle, or proximal ampulla and distal and proximal isthmus of the oviduct. It was determined that reversal of segments of all parts of the ampulla prevented pregnancy in the female rabbits while reversal in the isthmus had no effect. Analysis of ciliary activity confirmed that cilia within the reversed segments of the ampulla beat in the direction of the ovaries, opposite to the rest of the oviduct, suggesting the critical role of tubal cilia in transport of the ovum in the ampulla of the oviduct. More recently Yuan et al., 2021, determined that the motile cilia in the infundibulum of the oviduct are essential for oocyte pickup while other parts of the oviduct facilitate gamete and embryo transport. This was determined by analysing mice oviducts with infundibulum that lack cilia, through knockout of two microRNA clusters. This oviduct was found to not pick up ovulated oocytes and had a reduced efficiency of sperm migration to the ampulla and embryos to the uterus.

Additionally, motile cilia are also known to have a role in embryonic development, specifically in the determination of left-right asymmetry and the development of unpaired organs. This is carried out by nodal cilia present on the embryonic node. Unlike other motile cilia, nodal cilia possess a 9+0 axoneme with the absence of the central pair of singlet microtubules (Ostrowski et al., 2011). The lack of the central pair creates a clockwise rotating motion, generating a leftward fluid flow of extraembryonic fluid (Mitchison and Valente, 2016). Exactly how nodal cilia are responsible for left-right asymmetry is still debated, with multiple models hypothesised. This includes the 'two-cilia model' where a nodal flow is generated by nodal cilia at the centre of the node and is sensed by non-motile sensory cilia at the periphery of the node (Fliegauf et al., 2007). These sensory cilia then transmit signals to the lateral plate mesoderm to activate asymmetrical gene cascades to the downstream effector Nodal to induce asymmetrical transcription of downstream genes (Mitchison and Valente, 2016). Moreover, the 'morphogen gradient model' hypothesises that directional flow causes a greater concentration of a particular morphogen to the left side of the left-right organiser, triggering downstream asymmetrical signalling (Dasgupta and Amack, 2016).

1.1.4: The formation of cilia by Intraflagellar transport (IFT)

As mentioned earlier, the formation of both motile and primary cilia, known as ciliogenesis, is reliant on the process of IFT. This is the bidirectional movement of structural components of the axoneme and the ciliary membrane along the doublet microtubules, from the cell body to the ciliary tip (Nevers et al., 2017). This involves components of the axoneme, and membrane being loaded onto IFT particles, these are composed of two IFT protein complexes, IFT-A with 6 protein subunits and IFT-B with 13 protein subunits (Waters and Beales, 2011). The IFT complexes are transported using kinesin and dynein molecular motors. Specifically, kinesin-II, KIF3A, KAP3 and KIF3B or KIF3C, transports cargo in an anterograde manner towards the proximal ciliary tip and dynein-II (1b) facilitates the retrograde transport back towards the distal basal body (Basten and Giles, 2013). Figure 1.3 demonstrates the action of kinesin and dynein motors in IFT.
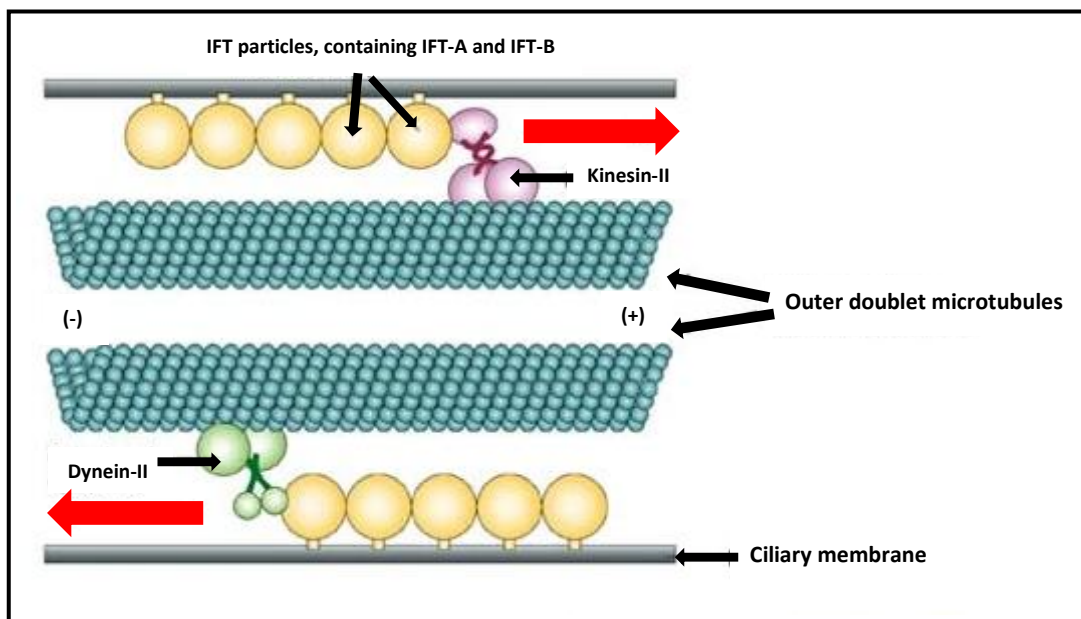
Figure 1.3: Intraflagellar transport machinery. Diagram shows the process of IFT. IFT particles (yellow) are transported towards the proximal ciliary tip (+) by kinesin-II (pink) and the distal basal body (-) by dynein-II (green), along the outer microtubule doublets of the axoneme. IFT particles are connected to the ciliary membrane (grey). Edited from: Rosenbaum and Witman, 2002.

Ciliogenesis by IFT occurs in several stages. Firstly, the cells must exit the mitotic cell cycle to free the centrioles from axonemal nucleation (Breslow and Holland, 2019). The centrioles then become basal bodies after the addition of distal appendages and docking to the ciliary vesicle. This ciliary vesicle comprises ciliary cargo arranged into ciliary-specific vesicles at the *trans* Golgi network (Smith et al., 2020). This then fuses with the plasma membrane, along with the basal body (Knodler et al., 2010). After this, microtubules nucleate at the basal body to initiate axoneme formation by IFT, where cilium elongation occurs after docking of the nascent cilium at the cell surface, to form the mature axoneme (Smith et al., 2020).

The IFT machinery is well conserved across eukaryotes and is required for ciliogenesis in nearly all these organisms (Satir et al., 2008). It is thought that the IFT machinery evolved from components of coat protein I (COPI) and clathrin-coated vesicles (Jekely and Arendt, 2006). Many proteins have been identified to be involved in IFT, this began with the discovery of the role of IFT88 in IFT. Work done by, Pazour et al, 2008, determined that IFT88 (Tg737), a subunit of the IFT particle is required for flagella assembly by IFT in mice. This research was also the first step in identifying the link between ciliary dysfunction and human health. More recently, Delaval et al., 2011, determined that IFT88 is needed for proper spindle formation and orientation through IFT88 depletion in human cells. Additional IFT proteins have been identified, including IFT27. The depletion of IFT27 by RNA interference reduces the levels of IFT-A and IFT-B complex proteins, suggesting it has a role in the stability of these complexes. While knockdown of the protein leads to cell cycle elongation and impaired cytokinesis (Qin et al., 2007).

Furthermore, each of the stages involved in ciliogenesis requires several different proteins. This includes cilium-related proteins that regulate the cell cycle, including NdeI which interacts with the cytoplasmic dynein subunit LC8 leading to increased cilia length and delayed cell cycle entry and Tctex-1, which when phosphorylated induces ciliary resorption and promotes S phase entry (Smith et al., 2020). Moreover, cell cycle modulators CDC14b and CDC42 have been found to also influence ciliogenesis. CDC14 is a family of phosphatases which counteracts Cdk1 phosphorylation allowing exit from mitosis and cytokinesis. Insertional mutation of the *cdc14b* gene in zebrafish by Clement et al., 2011, caused phenotypes reminiscent of zebrafish mutants with defective cilia, including hydrocephaly and left-right asymmetry defects, suggesting a role in the regulation of cilia length. CDC42 interacts with the exocyst complex and is required for its localization to the primary cilium. Research by Zuo et al., 2011, found that CDC42 colocalizes

with SEC10 at the primary cilium, and shRNA-mediated knockdown of CDC42 inhibits ciliogenesis in canine kidney cells.

Additionally, the protein CEP164 has been determined to be involved in the interaction of the basal body with the ciliary vesicle (Basten and Giles, 2013). Also, work by Schroder et al., 2011, determined that depletion of the microtubule tip tracking protein EB1 and related protein EB3 causes significant depletion of cilia in mammalian cell lines. Microscopy of these EB1 and EB3 defective cell lines determined that they lack microtubule minus-end anchoring at the centrosome and/or basal body and possess abnormally short cilia stumps surrounded by vesicles. While GST pull-down assays found that EB1 and EB3 interact with proteins implicated in minus-end anchoring or vesicular trafficking to the cilia base. Similarly, Rab small GTPases are found to be key regulators of membrane trafficking and essential for ciliogenesis, including Rab8, 11, 23 and 34 (Knodler et al., 2010; Oguchi et al., 2020; Yoshimura et al., 2007). Docking of the ciliary vesicles at the cell membrane is known to require proteins OFD1, OFD2, Ninein, MKS1, MKS3, CEP164, POC5 and CEP123 aswell as Rab8a which is modified by Rabin8 (Basten and Giles, 2013; Nachury et al., 2007). The CEP290 protein is also thought to be involved in membrane attachment to the transition zone of the cilium (Tsang et al., 2008).

Unlike the rest of the ciliary axoneme, dynein arms present on motile cilia are preassembled in the cytoplasm before being transported to the cilia by IFT. Evidence suggests that the R2TP complex, an Hsp-90 co-chaperone, in cooperation with dynein axonemal assembly factors (DNAAFs), plays a crucial role in this. The R2TP complex comprises of RuvB-like protein 1 (RUVBL1), RuvB-like protein 2 (RUVBL2), RNA polymerase-associated protein 3 (RPAP3), PIH1 domain-containing protein 1 (PIH1D1)) and R2TP-like complexes (Fabczak and Osinka, 2019).

1.1.5: The diversity of ciliogenesis across eukaryotes

All cilia are formed by microtubule extension from basal bodies, but the process of ciliogenesis is variable across eukaryotes. In many eukaryotic species, the basal body is formed from a triplet centriole embedded in the centrosome, with ciliogenesis occurring under control of the cell cycle, as mentioned in the previous section. In the trypanosomes motile flagella are always present. Here, basal bodies are never associated with spindle poles, instead existing as a basal body pair of a mature basal body BB1 and an immature pro-basal body BB2 (Vaughan and Gull, 2016). The pro-basal body matures to form the new flagellum during the cell cycle, alongside

the old flagellum present on the mature basal body (Vaughan and Gull, 2016). Additionally, in the slime mould *Physarum flavicomum*, the formation of cilia by ciliogenesis is dependent on the presence water. When the amoeboid phase encounters water, the centriole pair docks at the cell membrane and extends to form one long and one short cilium from the mature centriole and immature pro-centriole (Aldrich, 1968).

Moreover, as mentioned earlier, in most cases mature basal bodies dock to the membrane and elongation occurs through IFT. However, in some organisms such as in *Plasmodium spp.* elongation occurs in the cytoplasm in an IFT-independent manner (Hodges et al., 2012). The subsequent axonemes extend out from a protrusion of the plasma membrane, known as the perikinetosomal basket (Sinden et al., 1976). As expected, *Plasmodium spp* have a complete absence of IFT genes. (Briggs et al., 2004).

Many eukaryotes do not produce cilia at all, including most seed plants such as *Oryza sativa* and *Arabidopsis thaliana*. Interestingly, some conserved proteins with an inferred ciliary function have been identified in these non-ciliated plants. Work done by Hodges et al., 2011, identified a core set of proteins which are conserved in ciliary species. These included those with ciliary-specific roles such as IFT proteins, outer- and inner-dynein arms, and radial spoke proteins. Fluorescence microscopy of transformed *T. brucei* cells demonstrated that all proteins have a localisation pattern consistent with a role in ciliary function. A number of these proteins, were found to be expressed in non-ciliated land plants, as seen in Figure 1.4. This suggests that evolutionary scenarios of either sub- or neo-functionalisation occurred with these proteins likely co-opted to perform novel functions before the loss of cilia, some of which appear to be related to the formation of the male gametes.

Figure 1.4: Expression of ciliary proteins in ciliated organisms (blue), non-ciliated organisms (black) and non-ciliated land plants (green). Plot shows the expression of ciliary proteins in non-ciliated land plants being unusually high, compared to other non-ciliated organisms. E-values indicated by size of circle; larger circle indicates more significant. Edited from: Hodges et al., 2011.

Section 1.2: What happens when cilia are defective?

When certain genes encoding ciliary proteins are mutated, including those involved in ciliogenesis mentioned in the previous section 1.1.4, this leads to ciliary dysfunction. Diseases that are linked to defects in ciliary assembly or function are known as ciliopathies, which are phenotypically heterogeneous with different disease phenotypes manifesting from variations at a single gene locus or similar phenotypes resulting from mutations at several different loci (Waters and Beales, 2011). Ciliopathies are often complex multisystem disorders, involving all major organs, characterised by hydrocephalus, infertility and polycystic diseases of the kidney and pancreas (Shapiro et al., 2016). Defects in both primary and motile cilia have been linked to ciliopathic diseases.

1.2.1: Ciliopathies related to defects in primary cilia

Due to the role of primary cilia in cell signalling, defects in theses organelles leads to a wide range of disease phenotypes in the human body, including disruption during embryonic development. This is demonstrated by the autosomal recessive lethal malformation disorder Meckel–Gruber syndrome associated with occipital encephalocele, polycystic kidneys, and polydactyly (Turkyilmaz et al., 2021). Here, mutations in *MKS1-MKS*13 encoding ciliary proteins MSK1 and Meckelin, disrupt ciliary signalling, specifically involving Wnt, during development (Hartill et al., 2017). Also, mutations in genes encoding IFT motors including *Kif3a* and *Dync2h1* have been linked to developmental defects due to disruption in Shh signalling (Fliegauf et al., 2007).

Furthermore, as mentioned previously primary cilia are involved in cell signalling of the kidney epithelium for the maintenance of normal cell function and differentiation. Disruption of this ciliary function can lead to Polycystic Kidney Disease (PKD). This is a group of monogenic disorders that is characterised by cysts in the kidney and liver, which can result in kidney enlargement and eventually renal failure. (Brown and Witman, 2014; Waters and Beales, 2011). PKD is inherited in either an autosomal dominant (ADPKD) or recessive (ARPKD) way, with ADPKD being the most common form (1 in 400-1000) (Waters and Beales, 2011). Two genes are known to be mutated in PKD, *PKD1* and *PKD2,* encoding proteins polycystin-1 and polycystin-2 respectively (Hildebrandt et al., 2011). Disruption to these proteins was first linked to PKD by Pazour et al., 2000, when sequencing of IFT-particle proteins in *Chlamydomonas* found that IFT88 protein, one of the polycystin-1 and polycystin-2 proteins, was a homologue of a mouse protein of then-unknown function, *Tg737^orpk^,* was defective in the murine model for ARPKD. Examination of mouse kidneys with IFT88 mutations showed that the protein was necessary for the assembly of the primary cilia in the kidney collecting ducts and tubules, linking ARPKD to the malfunction of IFT and the failure to form kidney primary cilia. Following this, Pazour et al., 2002 determined that the polycystin-2 protein was specifically localised to the primary cilium of both mouse and human kidney cells. Polycystin-1 and polycystin-2 were found to interact to from a receptor-channel complex involved in epithelial cell differentiation and proliferation and that the primary cilium acts as a sensory antenna to relay signals to the cell body. Furthermore, polycystin-1 and polycystin-2 proteins have been found to be essential for Ca^{2+} influx, in response to primary cilium bending (Nauli et al., 2003). Overall, mutation of genes encoding

either of the two ciliary proteins, or disruption of the formation of the cilium by IFT, results in defective ciliary signalling leading to PKD.

Moreover, defects in primary cilium are linked to ciliopathies that affect sight and hearing. This includes retinal degeneration diseases including Retinitis pigmentosa (RP), which is implicated in 25% of vision loss in adults, and Leber's congenital amaurosis (LCA), which is associated with vision loss in infancy or early childhood (Bujakowska et al., 2017). As mentioned earlier, retinal rod and cone cells contain 9+0 cilia which connect the inner segment of photoreceptors with the outer segment, where phototransduction takes place through stacked membranous discs containing visual pigments and other phototransduction proteins (Bujakowska et al., 2017; Fliegauf et al., 2007). Mutations in both photoreceptor-specific and ciliary genes are associated with retinal degenerations, including *RPGR*, which interacts with IFT88 and other microtubule motor proteins, likely having a role in IFT (Fliegauf et al., 2007). Genes including *GUCY2D*, *RPE65*, *SPATA7* and *CEP290*, which is a transition zone and connecting cilium protein, have been linked specifically To LCA (Brown and Witman, 2014; Waters and Beales, 2011). Also, thirteen axonemal and basal body proteins have been linked to retinal degeneration, including *C2ORF71, FAM161A* and *LCA5* (Bujakowska et al., 2017). In addition, Ushers syndrome, an autosomal recessive dual impairment of vision and sensorineural hearing, specifically type II (USH2) has been linked to mutations in genes encoding the periciliary membrane complex adjacent to the transition zone of the photoreceptor cilia, including *USH2A*, *GPR98*, and *DFNB31* (Bujakowska et al., 2017; Tilley et al., 2015).

1.2.2: Ciliopathies related to defects in motile cilia

Unlike the above-mentioned primary ciliopathies, ciliopathies associated with motile cilia are caused mainly by the disruption of ciliary beating. Many of these are classified as syndromes consisting of a range of symptoms, involving all ciliated organs in the body. This includes Joubert syndrome (JBTS), which is characterised by hypotonia, ataxia, psychomotor delay, and cerebellar and brain stem malformations (Waters and Beales, 2011). Several genes are known to be mutated in JBTS, this includes the ciliary genes *INPP5E* and *CC2D2A*, which encode proteins which localise to the transition zone of cilia (Bielas et al., 2009; Noor et al., 2008). Also, mutations in the *CEP290* gene, which encodes a centrosomal protein involved in ciliary assembly and trafficking, are found in around 50% of JBTS cases (Brancati et al., 2007; Valente et al., 2006).

The phenotype of JBTS overlaps with the ciliopathy Meckel-Gruber syndrome (MKS). Mutations in *MKS1* and *MKS3*, encoding cilium basal body proteins are most commonly associated with MKS (Kyttala et al., 2006; Smith et al., 2006). Similar to JBTS, mutations in *CEP290* and *CC2D2A* genes are also associated with MKS (Tallila et al., 2008).

Furthermore, Oral-facial-digital syndrome type 1 (OFD1) is an X-linked dominant disorder which is associated with malformations of the oral cavity, face, and digits, as well as central nervous system (CNS) abnormalities and cystic kidney disease. This results in the death of affected males in utero (Waters and Beales, 2011). This syndrome is caused by mutations in the *OFD1* gene, which encodes a centrosomal protein which localises to the basal body of cilia (Franco and Thauvin-Robinet, 2016). Work done by Singla et al., 2010, determined that *OFD1* is a component of the distal centriole that controls centriole length and in *OFD1* absence, distal regions of the centriole elongate abnormally. *OFD1* was also found to be involved in the recruitment of the IFT protein IFT88, showing that the OFD1 protein is associated with ciliogenesis. In addition, Bardet-Biedl syndrome (BBS) is characterised by rod-cone dystrophy, polydactyly, and renal anomalies leading to end-stage renal disease (Shoemark et al., 2015). Heterogeneous mutations in at least 16 loci are associated with the syndrome, *BBS1-16.* This includes mutations in genes encoding MKS1, MKS3, CEP290 and SDCCAG8, all of which are involved in ciliary assembly and function (Niederlova et al., 2019).

1.2.2.1: The motile ciliopathy, Primary Ciliary Dyskinesia (PCD)

The first ciliary syndrome to be discovered and by far the most common motile ciliopathy, is Primary Ciliary Dyskinesia (PCD). The first case of PCD was reported in the early 1900s and was initially known as Kartagener syndrome characterised by chronic sinusitis, bronchiectasis, and situs inversus (Leigh et al., 2009). The term 'immotile cilia syndrome' was used to describe the disorder but studies showing that motile cilia exhibited defective beating meant the name was changed (Brown and Witman, 2014). As a motile ciliopathy, PCD is associated with reduction or loss of function of motile cilia, most commonly this is in the form of loss of or defects in dynein arms. This includes the absence of the whole or a large part of the outer dynein arm structure and or the inner dynein arm structure (Shoemark et al., 2020). These defects are usually associated with complete loss of ciliary motility (Bustamante-Marin and Ostrowski, 2017). Additionally, defects in the central pair complex, including the absence of one or both central

microtubules and or the translocation of the outer microtubule to the central region are also associate with the disease (Shoemark et al., 2020). Defects in radial spoke components and the mis localisation of basal bodies have also been identified in PCD (Shoemark et al., 2020). As well as the loss of ciliary motility, PCD is also associated with a reduction in ciliary beat frequency, usually caused by defects in dynein arms and cytoplasmic assembly factors (King, 2016).

All these defects are caused by mutations in genes encoding axonemal or non-axonemal proteins. The genetic basis can be identified in only approximately 65% of PCD patients, with the remaining cases having no identifiable cause (Tilley et al., 2015). This suggests that there are many undiscovered genes that are mutated in PCD. One of the first signs of PCD occurs 24hours after birth, with 80% of patients experiencing neonatal respiratory distress (Bustamante-Marin and Ostrowski, 2017). This is due to defective motile cilia in the upper and lower airways that mediate MCC. The impaired MCC means that mucus and pathogens cannot be effectively cleared from the airways, leading to coughing, congestion, and recurrent infections. 80% of children with PCD have recurrent lower respiratory tract infections, *Staphylococcus aureus* and *Streptococcus pneumonia* being the most common (Leigh et al., 2019). This can lead to chronic infection and inflammation of the upper and lower airways resulting in bronchiectasis and chronic lung failure in adulthood. Also, PCD children often experience sinus problems, including recurrent otitis media (glue ear) in 80% of patients (Leigh et al., 2019).

Furthermore, PCD causes infertility in both male and female patients. Virtually all males with PCD are infertile due to impaired sperm flagella motility and a reduced sperm count (Brown and Witman, 2014). Female infertility is less common, with subfertility likely due to dysfunctional motile cilia in the fallopian tubes and uterine lining, resulting in improper oocyte transport through the oviducts (Vanaken et al., 2017). Female PCD patients also have an increased likelihood of ectopic pregnancy (Leigh et al., 2019). Evidence suggests that the phenotypical severity of this is dependent on ciliary ultrastructure and the involved gene mutations, with fertility affected by the number of multiciliated cells, the number of cilia on each ciliated cell, and the motility of the cilia (Vanaken et al., 2017; Yuan et al., 2021). Approximately 50% of PCD patients possess situs inversus totalis, which is the mirror-image reversal of all visceral organs with no apparent physiologic consequences (Shapiro et al., 2016). This is caused by impaired nodal cilia function during embryonic development and the disruption to left-right patterning (Satir and Christensen, 2008). Figure 1.5 shows how situs inversus totalis is presented in PCD patients. Additionally, a less common symptom of PCD is hydrocephalus, which is the abnormal accumulation of cerebrospinal fluid in the ventricles of the brain (Brown and Witman, 2014).

This occurs in PCD due to the loss of motile cilia function in the brain ventricles which reduced the flow of cerebrospinal fluid (Satir and Christensen, 2007).
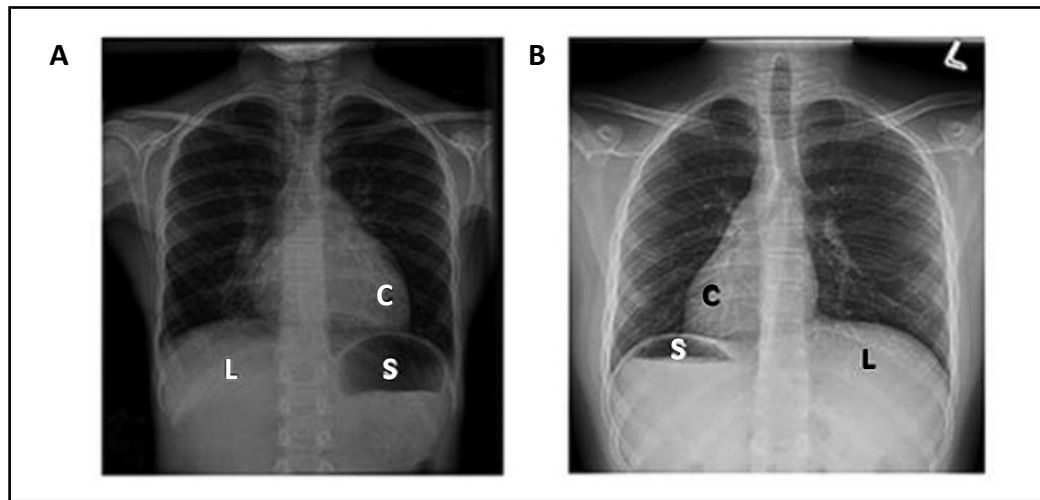
Figure 1.5: Chest X-ray showing patients with normal organ arrangement (situs solitus) and a PCD patient with situs inversus totalis. (A) shows a patient with situs solitus, or normal organ arrangement, with left cardiac apex (C), left-sided stomach (S), and right-sided liver (L). (B) shows a PCD patient with situs inversus totalis, or mirror-image organ arrangement, with right cardiac apex (C), right-sided stomach (S), and left-sided liver (L). Edited from: Shapiro et al., 2016.

1.2.2.2: Mutations in axonemal proteins that cause PCD

As mentioned above, PCD is caused by mutations in genes encoding proteins that have a role in motile ciliary formation and function. These can be split into axonemal and non-axonemal proteins, axonemal proteins being the most common. Mutations in axonemal proteins usually cause one of the ultrastructural axonemal defects as mentioned earlier, including the absence of the outer dynein arm and or the inner dynein arms associated with microtubular disorganization and/or central pair defects (Leigh et al., 2019). Also, certain PCD symptoms are associated with specific gene mutations. This was demonstrated by research carried out by Kennedy et al., 2007, where 337 PCD patients with situs inversus totalis were analysed for ciliary defects. Here, it was determined that patients with situs totalis had more ciliary outer dynein arm defects but fewer inner dynein arm and central apparatus defects, as well as more mutations in ciliary outer dynein arm genes including *DNAI1* and *DNAH5*, compared to patients

without situs defects. Similarly, Vanaken et al., 2017 found that PCD patients with mutations in one of four genes, *CCDC39, CCDC40, DNAAF1* or *LRRC6,* are very likely to be infertile, due to high expression levels in fallopian tubes and testicular cells. A full list of the axonemal proteins known to be mutated in PCD, along with their locations, the percentage of PCD cases they are found in, and the associated ciliary ultrastructure defects and PCD symptoms, can be seen in Table 1.1.

Table 1.1: Axonemal protein genes mutated in PCD. Table shows the genes names, protein locations, % occurrence in PCD cases, ciliary ultrastructure defects and associated PCD symptoms. Edited from Leigh et al., 2019; Lobo et al., 2015; Zariwala et al., 2007.

Gene name	Protein location	% PCD cases	Ultrastructure phenotype	PCD symptoms
DNAH5	ODA heavy chain	15-29%	ODA and IDA defects (Shoemark et al., 2020).	Upper and lower airway symptoms (Ibanez-Tallon et al., 2003). Situs inversus (Olbrich et al., 2002).
DNAI1	ODA intermediate chain	2-10%	ODA and IDA defects including microtubular disorganisation (Shoemark et al., 2020).	Situs abnormalities.
DNAI2	ODA intermediate chain	<2%	Complete or partial absence of the ODAs. (Praveen et al., 2015).	Situs abnormalities.
DNAL1	ODA light chain	<2%	ODA absent or truncated (Zariwala et al., 2001).	Situs abnormalities.
TXNDC3 (*NME8*)	ODA intermediate/light chain	<2%	Partial ODA defect (66% cilia defective) (Leigh et al., 2019).	Situs abnormalities.
CCDC114 (*ODAD1*)	ODA docking complex (Horani et al., 2016).	<2%	ODA absent or truncated.	Situs abnormalities.
CCDC151 (*ODAD3*)	ODA docking complex (Shoemark et al., 2020).	3%	ODA absent or truncated. Complete or partial absence of the ODAs. (Praveen et al., 2015).	Situs abnormalities.
ARMC4 (*ODAD4*)	ODA transport component/ docking complex (Shoemark et al., 2020).	3%	ODA absent or truncated. ODA defect attachment to microtubules (Praveen et al., 2015).	Situs abnormalities.
DNAAF1 (*LRRC50*)	Cytoplasmic dynein arm preassembly factor (Mitchison et al., 2012).	<2%	ODA+IDA absent or truncated. Defect involving assembly of the ODAs and IDAs (Mitchison et al., 2012).	Situs abnormalities. Infertility (Vanaken et al., 2017).

DNAAF2 (*KTU*)	Cytoplasmic dynein arm preassembly factor (Reiter and Leroux, 2017).	<2%	ODA + IDA absent or truncated (Omran et al., 2008).	Situs abnormalities Hydrocephalus (Mitchison et al., 2012).
DNAAF3 (*C19ORF51*)	Cytoplasmic dynein arm preassembly factor (Reiter and Leroux, 2017).	<2%	ODA+IDA absent or truncated.	Situs abnormalities Hydrocephalus (Mitchison et al., 2012).
CCDC103	Cytoplasmic dynein arm attachment factor, involved in dynein arm attachment.	4%	ODA + IDA absent or defective.	Situs abnormalities.
DNAAF4 (*DYX1C1*)	Cytoplasmic assembly of dynein arms (Bustamante-Marin and Ostrowski, 2017).	<2%	ODA + IDA defects.	Situs abnormalities.
DNAAF11 (*LRRC6*)	Cytoplasmic dynein arm preassembly and transport.	<2%	ODA and IDA absent or truncated.	Situs abnormalities. Infertility (Vanaken et al., 2017).
DNAAF5 (*HEATR2*)	Cytoplasmic dynein arm preassembly and transport (Horani et al., 2016). Localised to the cytoplasm (Horani et al., 2012).	<2%	ODA and IDA absent or truncated.	Situs abnormalities.
SPAG1	Cytoplasmic dynein arm preassembly and transport (Knowles et al., 2013).	4%	Absence of ODA and IDA proteins (Knowles et al., 2013).	Situs abnormalities.
CCDC39	Dynein regulatory complex (N-DRC).	4-9%	Microtubular disorganisation and inner dynein arm defect absent inner dynein arm, misplaced radial spokes, and microtubular disorganization (Davis et al., 2015; Merveille et al., 2011).	Situs abnormalities, worse lung function, bronchiectasis, poor weight gain. Infertility (Vanaken et al., 2017).
CCDC40	Dynein regulatory complex (N-DRC).	3-4%	Microtubular disorganisation and inner dynein arm defect (Shoemark et al., 2020).	Infertility (Vanaken et al., 2017). Severe lung disease and situs inversus (Davis et al., 2015).
RSPH1	Radial spoke component. (Werner et al., 2015).	<2%	Central pair defects (Bustamante-Marin and Ostrowski, 2017; Shoemark et al., 2020). Head-containing distal regions of radial spokes missing (Lin et al., 2014).	Mild disease phenotype, pulmonary symptoms (Lin et al., 2014).
RSPH9	Radial spoke component. (Werner et al., 2015).	<2%	Central pair defects (Bustamante-Marin and Ostrowski, 2017; Shoemark et al., 2020).	Unknown.

HYDIN	Radial spoke component. (Praveen et al., 2015).	<2%	Normal ciliary ultrastructure (Praveen et al., 2015).	Hydrocephalus (Olbrich et al., 2012).
DNAH11	ODA heavy chain.	6-9%	Normal ciliary ultrastructure abnormal motility without a clear structural defect (Horani et al., 2016; Knowles et al., 2012).	Situs abnormalities.
CCNO	Required for cilia biogenesis.	<2%	Mislocalisation of basal bodies with few or no cilia (Shoemark et al., 2020).	Unknown.
MCIDAS	Required for cilia biogenesis.	<2%	Mislocalisation of basal bodies with few or no cilia (Shoemark et al., 2020).	Unknown.

As seen in Table 1.1, PCD occurs due to mutations in genes encoding ODA proteins, including *DNAH5* encoding a component of the ODA heavy chain and *CCDC114* encoding a component of the ODA docking complex (Horani et al., 2016; Leigh et al., 2019). Also, mutations encoding radial spoke components and the N-DRC, including *RSPH1* and *CCDC39,* can result in PCD (Leigh et al., 2019; Werner et al., 2015). Moreover, PCD can also result from the aberrant assembly and or processing of ciliary proteins in the cytoplasm, including defects in cytoplasmic preassembly of dynein arms due to mutations in *DNAAF1* and *SPAG1* (Knowles et al., 2013; Mitchison et al., 2012). This demonstrates that disruption of ciliary function in PCD can result from mutations in cytoplasmic proteins as well as axonemal proteins.

1.2.2.3: Mutations in non-axonemal proteins that cause PCD

In addition, mutation in genes encoding non-axonemal proteins can cause PCD. These are mutations in *RPGR* and *OFD1. RPGR* encodes the outer segment of rod and cone photoreceptors and is essential for photoreceptor maintenance and viability (Zariwala et al., 2007). A mentioned in previous section 1.2.1, *RPGR* mutation is usually associated with the ciliopathy retinitis pigmentosa (RP). It has been determined that PCD can be transmitted through X-linked inheritance from patients with RP, leading to partial dynein arm defects (Moore et al., 2006). Similarly, mutations in *OFD1*, which are usually associated with oral-facial-digital type 1 syndrome (OFD1S), have also been found in PCD patients, suggesting that PCD is part of a clinical spectrum of OFD1-related disorders (Hannah et al., 2019). *OFD1* encodes a centriole component which is required for cilia biogenesis (Lobo et al., 2015). Mutations in the gene are associated

with a typical PCD phenotype in men without severe symptoms due to defective cilia motility patterns (Bukowy-Bieryllo et al., 2019).

Overall, the genetic basis of PCD is known in 65% of patients, including mutations in axonemal or non-axonemal proteins. This leads to several different ciliary ultrastructure defects including to ciliary axonemal components and to the cytoplasmic preassembly of ciliary proteins. However, 20-30% of PCD patients have no identifiable pathogenic variants and gene mutations. This suggests that a number of PCD gene mutations remain undiscovered.

Section 1.3: DPCD and its potential role in PCD

In 2004, an uncharacterized gene, named Deleted in Primary Ciliary Dyskinesia (DPCD) was discovered and was predicted to be a novel candidate gene for PCD. Research carried out by Kobayashi et al., 2002, determined that knockdown of DNA polymerase lambda (Pol λ) in mice resulted in PCD characteristics. Here, Pol λ-/- mice were generated by homologous recombination and examination of the mice found that they had a high mortality rate after birth, with approximately half of the Pol λ-/- mice dying by 3 weeks of age. The Pol λ-/- mice also began exhibiting enlarged and dome shaped heads at 2-4 weeks, which is characteristic of hydrocephalus, and 5 of the 26 total Pol λ-/- mice showed situs inversus totalis, Figure 1.6. In addition, male Pol λ-/- mice were infertile, determined by analysis of spermatozoa mobility which showed that Pol λ-/- mice had spermatozoa with truncated tails and abnormally shaped heads compared to Pol λ+/+ and Pol λ+/- mice. All these characteristics are often associated with PCD.
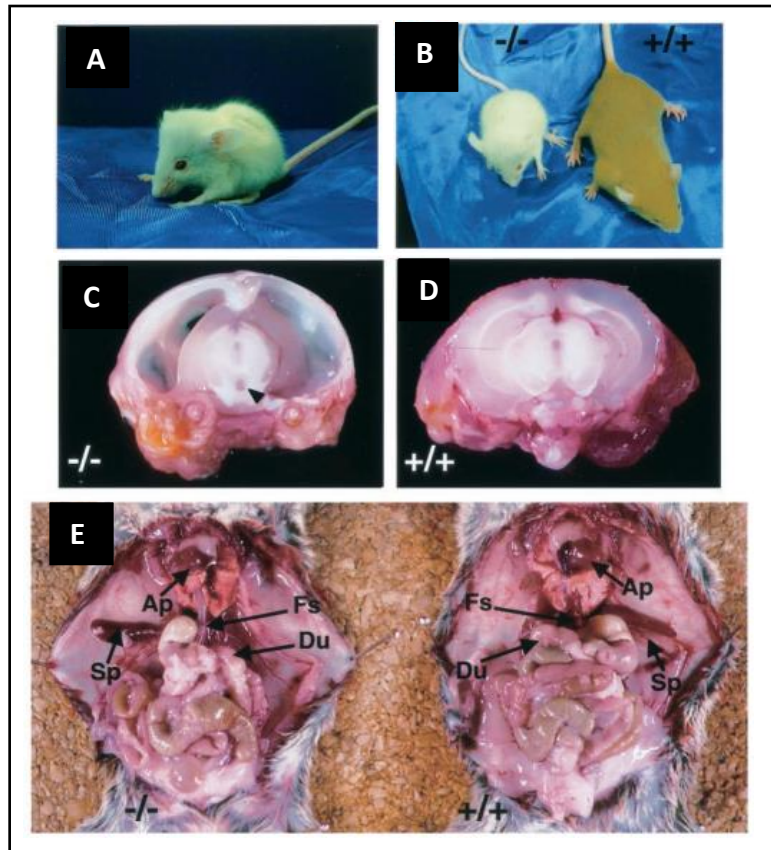
Figure 1.6: Growth defects, hydrocephalus, and situs inversus totalis in Pol λ-/- mice. (A) A Pol λ-/- mouse at 4 weeks of age seen with a dome shaped head. **(B)** A 4-week-old Pol λ-/- mouse on the left compared to a 4-week-old wild-type mouse on the right, showing growth defects in the Pol λ-/- mouse. **(C** and **D)** Show sagittal sections of brains from Pol λ-/- mice **(C)** and wild-type mice **(D)** at 5 weeks of age. Enlarge lateral ventricles seen in Pol λ-/- mouse brain section. **(E)** Dissection of Pol λ-/- (left) and wild-type (right) mice, showing situs inversus totalis in the Pol λ-/- mouse with the reversal of abdominal organs including heart (AP), stomach (FS), spleen (SP) and duodenum (Du). Edited from: Kobayashi et al., 2002.

To investigate whether the Pol λ-/- mice had any defects to motile cilia, ultrastructural analysis of the ependymal and respiratory epithelium motile cilia was carried out. This determined that the IDA of both ependymal and respiratory epithelium cilia were defective or absent, Figure 1.7. This suggested that the PCD phenotype seen in the Pol λ-/- mice was caused by defective ciliary beating due to IDA defects.
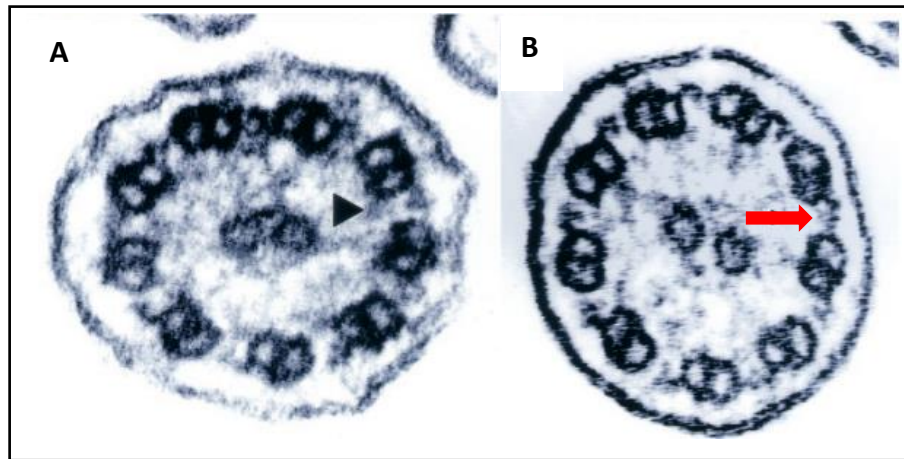
Figure 1.7: Ultrastructure analysis of cilia from wild-type and Pol λ-/- mice. Electron micrographs of respiratory epithelia motile cilia from wild-type (**A**) and Pol λ-/- (**B**) mice. Wild-type mice cilia possess an expected 9+2 microtubule arrangement with ODA and IDA present, shown by black arrowhead. While the IDA are absent from the Pol λ-/- mouse cilia, indicated by red arrow. Edited from: Kobayashi et al., 2002.

However, it remained unclear as to how a mutation in a DNA polymerase would result in a specific axonemal defect. This led to further research by Zariwala et al., 2004, who analysed the genomic region surrounding the Pol λ gene (*Poll*) to understand the knockdown targeting construct used previously. This study identified that an uncharacterised gene, now known as *DPCD* was predicted to be transcribed from the opposite strand relative to *Poll,* and that deletion of the *Poll* gene caused the loss of exon 1 of the *DPCD* gene, including the first ATG. Exon 1 of *DPCD* lies 75 bases from exon 1 of *Poll*, within the same region of knocked down by Kobayashi et al., 2002. This suggests that the knockdown of *Poll* carried out in 2002 would have disrupted the function of both DPCD and Pol λ. This hypothesis was supported by research carried out by Bertocci et al., 2002, which demonstrated that when Pol λ was deleted in mice by homozygous knockout using the catalytic domain of the *Poll* gene, which did not disturb the *DPCD* gene, did not show any PCD characteristics. Figure 1.8 shows the chromosomal location of *DPCD* and *Poll* in mice and the site of *Poll* deletions carried out previously (Bertocci et al., 2002; Kobayashi et al., 2002).
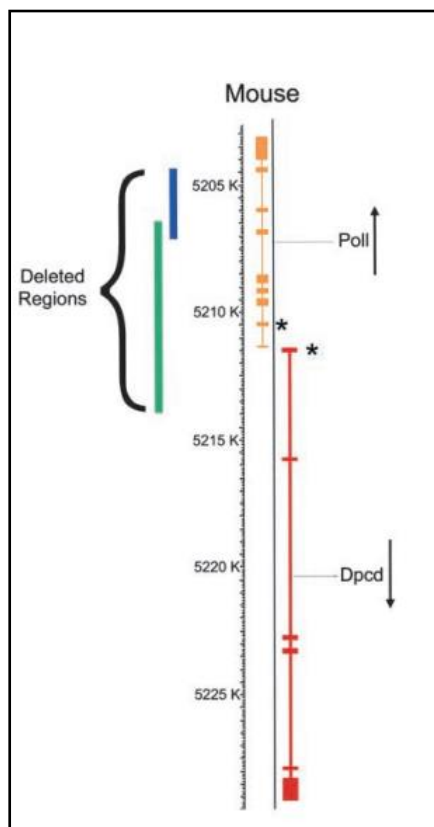
Figure 1.8: Diagram of chromosomal location of DNA polymerase lambda gene (*Poll*) and *DPCD* gene on mouse chromosome 19. Diagram shows the *DPCD* gene located on the opposite strand to *Poll* gene on mouse chromosome 19. Start codon indicated by *. Targeted deletion of region in green is predicted to disrupt both *Poll* and *DPCD* genes in mouse model of PCD, carried out by Kobayashi et al., 2002. Deletion within the blue region does not disrupt the *DPCD* gene and leads to mice with normal phenotypes, carried out by Bertocci et al., 2002. PCD phenotype seen in 2002 caused by disruption of DPCD. Edited from: Zariwala et al., 2004.

DPCD was investigated further by Zariwala and colleagues (Zariwala et al., 2004), who reported that the gene encodes a 23kDa protein with no homology to any conserved protein domains. The pattern of DPCD expression during ciliogenesis was also observed to understand its potential role in PCD. Here, the expression of *DPCD* during ciliogenesis of airway epithelial cells was examined and northern blot analysis revealed that DPCD expression increases during ciliated cell differentiation while the expression of *Poll* decreases. This suggests that DPCD has a role in the formation and or function of ciliated cells. However, the sequencing of the coding sequence of DPCD from 51 unrelated PCD patients failed to identify any disease-causing

33

mutations. As PCD is a heterogeneous disease it is not unexpected that no disease-causing mutations were identified in this study. Also, it is possible that mutations in *DPCD* may not be compatible with life, meaning they would not be present in PCD patient populations. Therefore, DPCD remains a novel candidate gene for PCD.

Section 1.4: Investigation of DPCD function in other animal models

Since the initial discovery of DPCD, little further research has been carried out to elucidate the biochemical role of DPCD and its potential link to PCD. To date, work by Funfak and colleagues (Funfak et al., 2015) investigating the role of DPCD in *Paramecium* is the only available research. In this study, expression of *DPCD* along with three other genes associated with cilia function or ciliopathies in humans (*MKS1, St6* and *T16*), was ablated using RNAi to study the effect on cilium function in *Paramecium*. The swimming velocity of each RNAi strain, as well as control cells, was measured using high-speed video at high magnification. As can be seen in Figure 1.9, the mean swimming velocity of *DPCD*, *T16, MKS1* and *St6* RNAi cells were significantly lower than that of wildtype control cells. However, the swimming velocity of DPCD depleted cells decreased less than cells depleted for *MKS1, St6* and *T16* RNAi strains.
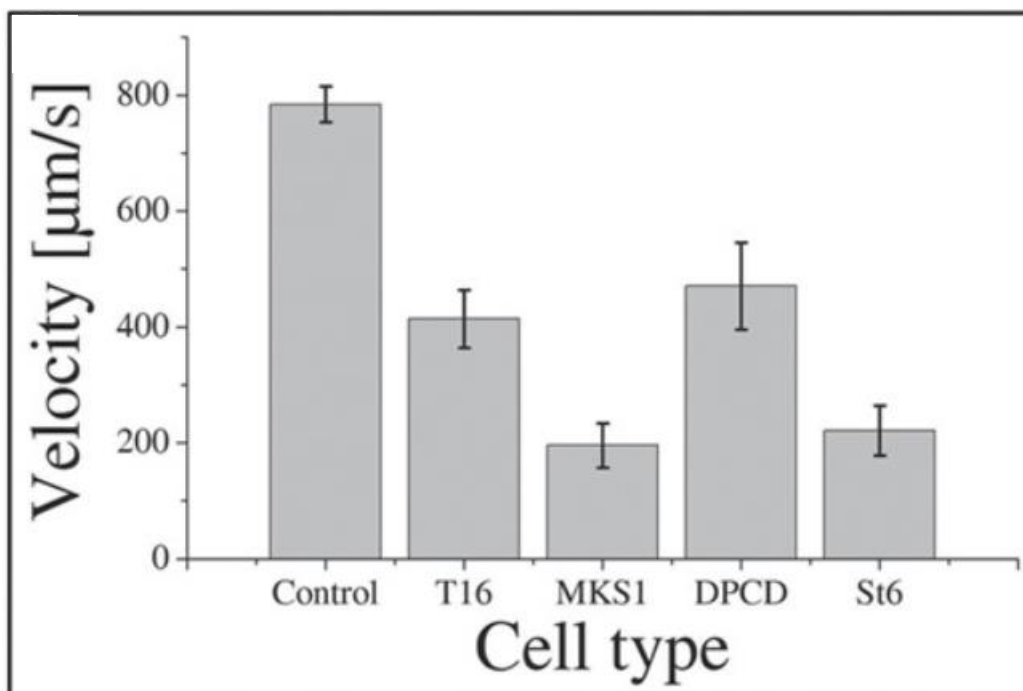
Figure 1.9: Mean swimming velocity of RNAi *Paramecium* strains compared to wildtype cells. Mean swimming velocities (µm/s) of each RNAi *Paramecium* strain compared to wildtype control cells. Standard Errors seen in error bars. *T16, MKS1*, *DPCD* and *St6* RNAi cells all had significantly lower swimming velocity compared to control cells. Edited from: Funfak et al., 2015.

In addition, the cell shape and size of each RNAi induced cell line was measured and compared to wildtype control cells, figure 1.10. This showed that *DPCD* RNAi cells appeared to have a 'fatter' cell body, with cells being longer and wider than control cells, a phenotype that was not observed in the other RNAi induced strains. However, this difference was not statistically significant.

Cell type	Length [µm]	Width [µm]
Control	124 ± 20	46 ± 5
St6	112 ± 10	54 ± 8
T16	114 ± 13	45 ± 6
MKS1	115 ± 7	53 ± 6
DPCD	125 ± 13	59 ± 9
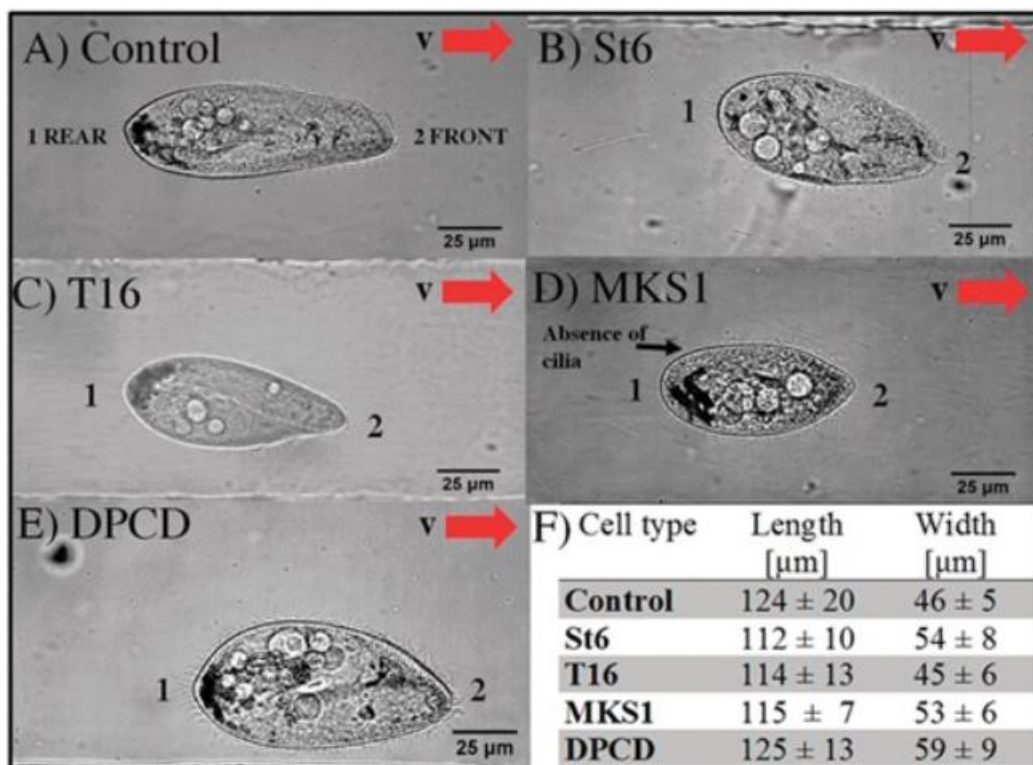
Figure 1.10: Cell shape, length and width of RNAi *Paramecium* strains and control wildtype cells. The shape of control (A), *St6* RNAi (B), *T16* RNAi (C), *MKS1* RNAi (D) and DPCD RNAi (E) *Paramecium* cells, red arrows indicated swimming direction. (F) Shows the respective cell dimensions, as mean length and width (µm), of RNAi strain and control cells. Only *DPCD* RNAi cells have larger cell length and width compared to control cells, however this is not a significant difference. Edited from: Funfak et al., 2015.

Furthermore, the frequency of cilium beating (CBF) was measured for *DPCD*, *T16, MKS1* and *St6* RNAi strains and compared to control cells, showing that there was no significant impact on CBF in cells in which *DPCD* and *T16* had been ablated by RNAi, but a significant reduction in CBF in *St6* and *MKS1* RNAi cells. Overall, this research shows that knockdown of DPCD in *Paramecium* causes a decrease in swimming velocity and an increase in cell length and width. However, *DPCD* RNAi has no effect on CBF. This suggests that DPCD is not essential for ciliogenesis in *Paramecium* as RNAi mediated knockdown of DPCD expression does not have the same effect of that of the other ciliary proteins.

Section 1.5: Aims of project

In summary, DPCD is a novel gene candidate for the motile ciliopathy PCD, with a potential but as yet undetermined function in ciliogenesis. This project aims to further our understanding of the role of DPCD using the protozoan parasite *Trypanosoma brucei* as a model organism. *T. brucei* is an excellent model system to study ciliary proteins due to its possession of a single flagellum, which is essential for motility, morphogenesis, and cell division (Reviewed in Benmerah et al., 2015), and for its experimental tractability (Dean et al., 2016). In addition to DPCD, the *T. brucei* proteome contains orthologs of many proteins that are important for the assembly of human flagella/cilia, (Broadhead et al., 2006; Dean et al., 2016).

The project had four aims:

1. Carry out a comprehensive phylogenetic survey of DPCD occurrence in eukaryotes to determine whether its conservation in ciliated eukaryotes has a biological basis.

2. Overexpress *Tb*DPCD protein in *T. brucei* to investigate whether over-expression of the protein leads to a phenotype.

3. Carry out RNAi mediated knockdown of *Tb*DPCD expression in *T. brucei* to investigate whether loss of the protein has a phenotypical consequence for the cell and in particular on flagellum formation/function.

4. Express and purify *Tb*DPCD protein in *E. coli* for biochemical studies.

Section 2: Materials and Methods

Section 2.1: Buffers and reagents

Table 2.1: Buffers used in this project

Buffer	Components
50X Tris-Acetate-EDTA (TAE) buffer	2M TRIS base, 5.7% glacial acetic acid, 50mM EDTA, H_2O.
1X Tris-Acetate-EDTA (TAE) buffer	10ml 50X TAE buffer, 490ml distilled water
LB broth	LB broth base (Invitrogen; 10g SELECT Peptone 140, 5g SELECT Yeast Extract, and 5g NaCl), distilled water.
LB agar plates	LB broth base (Invitrogen 10g SELECT Peptone 140, 5g SELECT Yeast Extract, and 5g NaCl), Select Agar powder (Invitrogen), 1L of distilled water. Additions: ITPG- 0.1mM X-gal - 20 mg/ml Carbenicillin- 60µg/ml Kanamycin- 50ng/µL Chloramphenicol- 50ng/µL
1X Phosphate-buffered saline (PBS)	PBS 100ml Tablets (Melford), distilled water
1X Phosphate-buffered saline (PBS) Tween 20	1XPBS, 0.05% Tween 20
10X SDS PAGE running buffer	30.3g TRIS, 144.2g Glycine, 50ml 20% SDS solution, distilled water to 1L
1X SDS PAGE running buffer	100ml 10X SDS PAGE running buffer, 900ml distilled water
1X TRIS-Glycine transfer buffer	3.03g Tris, 14.4g Glycine, 10% methanol per litre, distilled water to 1L
Bovine serum albumin (BSA)	PBS tween 20, BSA crystals (Sigma)
1X SDS Loading buffer	for 1ml- 100µL Dithiothreitol (DTT), 250µL 4Xlaemmli sample buffer (BioRad), 650µL 1XPBS
NI-NTA buffer A	50 mM NaH_2PO_4, 1 M NaCl, 10 mM imidazole, pH 8.0
NI-NTA buffer B	50 mM NaH_2PO_4, 1 M NaCl, 250 mM imidazole, pH 8.0
50mM Tris- 2M EDTA buffer	50mM Tris base, 2M EDTA, distilled water to 1L

40mM HEPES buffer	1M stock-238.3 g HEPES (MW 238.3 g/mol), distilled water to 1L. Diluted to 40mM with distilled water.

Table 2.2: Antibiotics used in this project

Antibiotic	Stock concentration	Working concentration	Use
Doxycycline	1 mg/mL	1 µg/ml	Inducible expression for *T. brucei*
Hygromycin	20 µg/µL	50 µg/mL	Selection for *T. brucei* cell lines
Phleomycin	7.5 mg/mL	3 µg/mL	Selection for *T. brucei* cell lines
Puromycin	7.5 mg/mL	2 µg/mL	Selection for *T. brucei* cell lines

Table 2.3: Antibodies used in this project

Antibody	Stock concentration	Working concentration	Incubation conditions
Primary antibodies			
Anti-Myc tag antibody [Myc.A7] (ab18185)	1mg/ml	1µg/ml	2 hours at 4°C.
Anti-Tubulin Antibody, beta, clone KMX-1 (Sigma-Aldrich)	1mg/ml	0.2µg/ml	40 minutes, room temperature
Anti-6X His tag® primary antibody [HIS.H8] (ab18184)	1mg/ml	0.125-0.25µg/ml	1 hour, room temperature
Anti-groEL antibody [9A1/2] (ab82592)	1mg/ml	5µg/ml	1 hour, room temperature
Secondary antibodies			
Rabbit Anti-Mouse IgG H&L (HRP) (ab6728)	2 mg/ml	0.2µg/ml-0.25µg/ml	1 hour, room temperature

A

pGEM®-T Easy Vector
(3015bp)

T7 ↓	1 start
ApaI	14
AatII	20
SphI	26
BstZI	31
NcoI	37
BstZI	43
NotI	43
SacII	49
EcoRI	52
SpeI	64
EcoRI	70
NotI	77
BstZI	77
PstI	88
SalI	90
NdeI	97
SacI	109
BstXI	118
NsiI	127
	141
↑ SP6	

B

pDex377
8239 bp

C

p2T7-177
6034 bp

39

D

Xho I(158)
Not I(166)
Eag I(166)
Hind III(173)
Sal I(179)
Sac I(190)
EcoR I(192)
BamH I(198)
Nhe I(231)
Nde I(238)
Nco I(296)

Bpu1102 I(80)

Xba I(335)
Bgl II(401)
SgrA I(442)
Sph I(598)

Dra III(5127)

f1 origin (4903-5358)

Kan (3995-4807)

Pvu I(4426)
Sgf I(4426)
Sma I(4300)
Cla I(4117)
Nru I(4083)
Eco57 I(3772)
AlwN I(3640)
BssS I(3397)
BspLU11 I(3224)
Sap I(3108)
Bst1107 I(2995)
Tth111 I(2969)

ori (3286)

pET-28a(+)
(5369bp)

lacI (773-1852)

Mlu I(1123)
Bcl I(1137)
BstE II(1304)
Apa I(1334)
BssH II(1534)
EcoR V(1573)
Hpa I(1629)

PshA I(1968)

Bgl I(2187)
Fsp I(2205)
Psp5 II(2230)

Figure 2.1: Vector maps. (A) pGEM®-T Easy vector (Dawson and House, 2010). **(B)** pDEX-377 vector used for constitutive expression of myc-tagged proteins in procyclic trypanosomes (Kelly et al., 2007). **(C)** p2T7-177 vector was used for a doxycycline-inducible expression of RNAi (Wickstead et al., 2002). **(D)** pET28a (+) used for IPTG-induced protein expression in *E. coli* cells (Novagen).

Section 2.2: Bioinformatics

2.2.1: BlastP

The *Homo sapiens* DPCD amino acid sequence was gained from protein Blast search (BlastP) of the predicted *T. brucei brucei* TREU927 DPCD amino acid sequence, found on the 'TriTrypDB' kinetoplastid database (Accession: Tb927.11.8900). BlastP was then used to find putative DPCD protein homologues in ciliated and non-ciliated eukaryotes, see full list in appendix table 1. Putative DPCD homologues in each organism were determined by E value above e^{-10} and reciprocal searches of putative homologue sequences back into *H. sapiens.* Sequence

alignments of putative homologue and *Hs*DPCD sequences were carried out using the 'ClustalOmega' multiple sequence alignment tool (Madeira et al., 2022). To demonstrate which eukaryotes possessed DPCD putative homologues, coulson plots were generated using the Coulson plot generator (CPG) (Field et al., 2013). These were colour coded by ciliated and non-ciliated eukaryotes, Figure 3.2, and by the presence of motile primary cilia, Figure 3.3.

Searches were carried out for putative homologues of ciliary proteins gained from Hodges et al., 2011, full list seen in appendix table 2. *Arabidopsis thaliana* protein accession numbers were used to gain amino acid sequences from *A. thaliana* 'ensemblePlant' database (Yates et al., 2022). The amino acid sequences of ciliary proteins were inputted into BlastP to find putative homologues in the same eukaryotes used earlier. Putative homologues in each organism were determined by E values above e^{-10} and reciprocal searches of putative homologue sequences back into *A. thaliana.* As above, coulson plots were generated using the CPG, to demonstrate whether putative homologues of the ciliary proteins are present in each eukaryote, comparing to the expression of DPCD. Again, these were colour coded by ciliated and non-ciliated eukaryotes, Figure 3.4, and by the presence of motile or primary cilia, Figure 3.5.

Available literature was used to investigate the similarities between ciliated eukaryotes that possess motile cilia but do not possess DPCD, including evolutionary origins and habitats.

2.2.2: Phyre2

The *T. brucei brucei* TREU927 DPCD amino acid sequence found on 'TriTrypDB' kinetoplastid database (Accession: Tb927.11.8900) was inputted into Phyre2 for protein folding prediction (Kelley et al., 2015).

2.2.3: Alphafold

The Alphafold protein structure prediction software was used to gain the predicted *T. brucei brucei* DPCD structure (Jumper et al., 2021; Varadi et al., 2022). The PDB file was exported and inputted into three different protein databases to compare predicted protein structures,

including DALI (Holm, 2022). Here, the Alphafold *T. brucei brucei* DPCD was searched against PDB50. This was chosen to keep redundancy in the results low, as initial search against the whole PDB database found only three different proteins. The top 50 hits were taken, restricting the Z score to above 2.

The PDB file was also inputted into Foldseek and searched against PDB100 and 50% of the Alphafold database, taking the top 50 hits from each, restricting e value to above e^{-10} (Kempen et al., 2022). Finally, the Alphafold PDB file was inputted into FATCAT and searched against PDB40, restricting P value to below 0.05 (Ye and Godzik, 2003).

Section 2.3: Molecular cloning

2.3.1: PCR

PCR was carried out using 1:2 concentration Q5 polymerase master mix (New England Biolabs Q5® High-Fidelity DNA Polymerase). The master mix also contained *T. brucei* TREU427 procyclic form 2TI DNA template and *Tb*927.118900 forward and reverse primers. Primers were synthesised by Eurofins Genomics and were designed to amplify the *Tb*DPCD open reading frame along with cloning sites for specific vector ligation.

Table 2.4: Primers used in this project. Restriction sites used for cloning seen in yellow

Primer ID	Direction	Nucleotide sequence	Restriction site
*Tb*927.118 900_pPOT	Forward	GTCTGCGTGATGCATCAACGTAGCTGCCAGCGTTTTTGCGCTACTCCATA TTTTGAGAGCGAGTCCGGCTGGTGAAAAGAGTATAATGCAGCCTGCTGC	N/A
*Tb*927.118 900_pPOT	Reverse	CCATCGACAAACTTTGATGTTATGCGACGACGTCCGTTGACAATGACAGA TGACGTCCGTTTGACAATGACAGATGACTTAGGCTCTGAAAGGGTTACAC TCATACTACCCGATCCTGATCC	N/A
*Tb*927.118 900_pET28 *a*	Forward	GCG<mark>CATATG</mark>AGTGTAACCCTTTCAGAGCCTAAG	NdeI
*Tb*927.118 900_pET28 a	Reverse	GCG<mark>AAGCTT</mark>CTATTGCTGATTGCAACTGGTCCT	HindIII
*Tb*927.118 900_pDEX3 77	Forward	GCG<mark>AAGCTT</mark>ATGAGTGTAACCCTTTCAGAGCC	HindIII

*Tb*927.118 900_pDEX3 77	Reverse	GCG<mark>CTCGAG</mark>TTGCTTGATTGCAACTGGTCCTAGA	XhoI
*Tb*927.118 900_p2T7-177	Forward	GCG<mark>AGATCT</mark>ATGAGTGTAACCCTTTCAGAGCCT	BglII
*Tb*927.118 900_p2T7-177	Reverse	GCG<mark>AAGCTT</mark>CTATTGCTGATTGCAACTGGTCCT	HindIII

PCR was performed using Bio-Rad T100 Thermal Cycler machine with the following conditions:

Initial denaturation- (1 cycle) 94°C for 5 minutes.

Denaturation- 94°C for 30 seconds.

Annealing- 55°C for 30 seconds.

Elongation- 72°C for 2:30 minutes. (29 cycles).

Final elongation- (30 cycles)72°C for 5 minutes.

Hold- 4°C.

2.3.2: Purification of PCR products

PCR products were purified using GeneJET PCR Purification Kit (Thermo Fisher Scientific). A 1:1 volume of PCR product and binding buffer was made up and transferred to PCR purification spin columns and centrifuged at 13,000 rpm for 60 seconds. The flow-through was discarded and 700µL of wash buffer was added to the column and centrifuged at 13,000 rpm for 60 seconds. The flow-through was discarded and the empty column was centrifuged at 13,000 rpm for 60 seconds. Spin columns were transferred to Eppendorf tubes and 30µL of elution buffer was added before being centrifuged at 13,000 rpm for 2 minutes. DNA concentration was measured using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific).

2.3.3: A-tailing PCR products

25µL of each purified PCR product was A-tailed by mixing with 2X DreamTaq Green PCR Master Mix (Thermo Fisher Scientific). PCR was performed using Bio-Rad T100 Thermal Cycler at 72°C for 15 minutes. The PCR products were then purified using same the same method in section 2.3.2.

2.3.4: Agarose Gel Electrophoresis

0.8% Agarose gels were produced using 0.8g Agarose (Melford) dissolved in 100ml 1XTris-Acetate-EDTA (TAE) buffer. 5µL of SYBR Safe DNA Gel Stain (Thermo Fisher Scientific) was added and gently mixed. The mixture was heated in the microwave until boiling and cooled before pouring into a gel tray with combs in place. This was then allowed to set before being placed into a gel tank and submerged in 1XTAE buffer.

DNA samples were mixed with 6X DNA loading buffer and run alongside 1Kb Plus DNA Ladder (Invitrogen) at 90mV for 45 minutes using a power pack. Gels were imaged on a GelDoc EZ Imager (Bio-Rad) using Image Lab Software (Bio-Rad).

2.3.5: Vector ligations

DNA fragments were ligated into specific DNA vectors, table 2.5. The concentration in ng of each DNA fragment needed was calculated using the Insilico ligation calculator, using a 1:3 vector to insert ratio. DNA fragments were mixed with (1%) T4 DNA ligase (Thermo Fisher Scientific), (1%) 10X T4 DNA ligase buffer (Thermo Fisher Scientific) and 50ng of the specific vector. A vector only control was also made up without PCR products. Ligations were left at room temperature for 1 hour and then at -4°C overnight.

Table 2.5: Plasmids used in this project

Vector	Concentrations
pGEM-T Easy	50ng/ µL
p2T7-177	50ng/ µL

pDEX377	50ng/ μL
pET28a	50ng/ μL

2.3.6: Bacterial transformation

25μL DH5α *E.coli* competent cells (Invitrogen) were incubated with 1μL of prepared vector ligations on ice for 30 minutes, then heat shocked at 42°C for 45 seconds before cooling on ice for a further 2 minutes. 700μL LB broth media was added to the mixture and incubated at 37°C for 45 minutes whilst shaking. Mixtures were then plated onto LB agar plates with added antibiotics, see table 2.1 for specific antibiotics used. Plates were incubated at 37°C overnight and white colonies were streaked to single colony on Cb only LB plates and incubated at 37°C overnight. Single colonies were picked from each plate and incubated in LB broth with added Cb (60μg/ml), shaking at 37°C for 16 hours.

2.3.7: Plasmid purification using a Miniprep kit

The bacterial suspensions were centrifuged at 25,000 rpm for 10 minutes at 21°C. Bacterial pellets were lysed and plasmids were purified using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific).

2.3.8: Restriction enzyme digestion

Purified plasmid DNA was digested with restriction enzymes, at up to 10% of total mixture. Restriction enzymes corresponding to cloning sites and the associated digest buffer to DNA samples and incubated for the appropriate time at 37°C, following manufacturer instructions. Samples were run on 0.8% agarose gel, as described in section 2.3.4.

2.3.9: Analysis of cloned inserts

100ng/μL samples of each pGEM-T Easy plasmid DNA was sent to 'SourceBioscience' for Sanger nucleotide sequencing analysis, using MI3F primers. Results were inputted into 'Chromas' to gain nucleotide sequence and translated using 'Expasy' translate tool to gain the amino acid open reading frame. The ORF was inputted into 'ClustalOmega' and aligned against the *T. brucei*427 DPCD sequence (Tb427_110095600), to confirm the presence of the DPCD insert.

2.3.10: DNA gel extraction

Plasmid DNA digested with specific restriction enzymes were loaded into large wells on 0.8% agarose gel, without Sybr Safe. These were run at 90mV for 45 minutes, in the same way as mentioned in section 2.3.4. The gel was then soaked in 1XTAE plus 10μL Sybr Safe for 30 minutes on a rocker. The gel was then put on a UV light box for visualisation of DNA. DNA bands corresponding to the excised DNA inserts were cut out using a scalpel blade and put into pre-weighed Eppendorf tubes. The Eppendorf tubes were then weighed gain to determine the weight on the DNA insert in mg.

2.3.11: Purification of DNA inserts

The GeneJET PCR Purification Kit (Thermo Fisher Scientific) was used to purify the DNA inserts. A 1:1 volume of binding buffer was added to DNA inserts and was melted in a 50°C water bath. Mixtures were then added to PCR purification spin columns and centrifuged at 13,000 rpm for 60 seconds. The flow-through was discarded and 700μL of wash buffer was added to the column and centrifuged at 13,000 rpm for 60 seconds. The flow-through was discarded and the empty

column was centrifuged at 13,000 rpm for 60 seconds before transferring the column to an Eppendorf tube and adding 30µL of elution buffer and centrifuging at 13,000 rpm for 2 minutes.

2.4.1: *T. brucei* cell culture

SP9 procyclic form *T. brucei* cells were grown in SMD-79 medium, supplemented with 10% foetal calf serum and Hemin (0.008mM), at 28°C. All cell culture work was carried out in a Class II Microbiology safety cabinet.

2.4.2: Preparation of plasmid DNA for transfection into *T. brucei*

Purified plasmids (p2T7-177 and pDEX377-based constructs) were digested with NotI and the linearised plasmids ethanol precipitated by adding 1/10th volume of 3M sodium acetate (pH5.2) and 2.5X volume 100% ethanol. DNA was precipitated at -80°C overnight. Precipitated DNA was centrifuged at 13,000 rpm for 15 minutes, the supernatant removed, and the pellet washed in 250µl of 70% ethanol before the centrifugation step was repeated. The 70% ethanol was removed from the pellet DNA inside a Class II Microbiology safety cabinet and residual ethanol allowed to evaporate. The DNA pellet was re-suspended in 50µl of sterile filtered Elution Buffer and stored at -20°C until required. For pPOT tagging, PCR reactions were purified and directly transfected into *T. brucei* (Dean et al., 2015).

2.4.3: *T. brucei* DNA transfections

DNA transfections in SP9 procyclic form *T. brucei* cells were performed using $3X10^7$ cells per transfection. Cells were grown to mid-log phase ($4X10^6$-$9X10^6$) on no drug media beforehand. 9.5ml of no drug media was incubated at 27°C in two culture flasks, one for transfection and one for control. $3X10^7$ cells/ml of SP9 cells were added to a 50ml falcon tube and centrifuged at 2000 rpm for 10 minutes at 21°C, the supernatant removed, and the cell pellet resuspended in 1ml of Zimmerman's post fusion medium with glucose (ZMG) buffer. The resuspended pellet was

added to a Gene Pulser electroporation cuvette (Bio-Rad) and 50μL of linearised DNA added. Cells were electroporated using a BTX Electro Square Porator ECM830 (3 x 100 μs pulses at 1700 V with 200 ms intervals) and recovered into warmed no drug media and incubated at 27°C for 24hours. Following overnight recovery cells were diluted to 5x105 cells ml in SMD-79 media and selected using appropriate antibiotics, table 2.2

2.4.4: Preparation of slides for fluorescence microscopy

1ml of healthy mid-log phase *T. brucei* cells were transferred to an Eppendorf tube and centrifuged at 6000 rpm for 3 minutes. The supernatant was removed, and the cell pellet was resuspended in 1ml of 1XPBS buffer and centrifuged again at 6000 rpm for 3 minutes. The supernatant was removed, and pellet was resuspended in 100μL 1XPBS buffer and pipetted into a square drawn on a frosted edge microscope slide (Deltalab) with a hydrophobic barrier pen (Vector laboratories).

The cells were left to settle before being fixed with 4% Paraformaldehyde for 5 minutes and then in 100% methanol at -20°C for at least 20 minutes. Slides were then rehydrated in PBS Tween 20 for 15 minutes, before mounting with VECTASHIELD Antifade Mounting Medium with DAPI (Vector laboratories) and a coverslip. Prepared slides were then imaged using an Applied Precision DeltaVision Deconvolution microscope and images were processed using SoftWoRx software. All images were subsequently processed using Adobe Photoshop.

Kinetoplast (K) and Nucleus (N) counts were carried out by counting the number of cells with 1K1N, 2K1N and 2K2N from approximately 100 cells per cell type.

2.4.5: Preparation of *T. brucei* protein samples

Cells engineered for RNAi and protein overexpression were induced using doxycycline. In brief, $5X10^5$ cells/ml of transfected cells were incubated in 200ml no drug media and split into two T500 culture flasks. Doxycycline was added into one flask. Both flasks were incubated at 27°C, taking cell counts, samples for fluorescent imaging and protein samples every 24hours for 166hours in total.

Protein samples were made by taking $5X10^7$ cells/ml in a falcon tube and centrifuging at 2000 rpm for 10 minutes. The supernatant was removed, and pellet was resuspended in 10ml 1XPBS buffer. This was centrifuged again at 2000 rpm for 10 minutes and supernatant was removed, and pellet was resuspended in 1ml 1XPBS buffer and transferred to an Eppendorf tube. This was centrifuged at 6000 rpm for 5 minutes and the supernatant was removed, and the cell pellet was resuspended in 100μL 1X SDS loading buffer. This was heated at 95°C for 5 minutes before storing at -20°C, ready to run on SDS-PAGE gel.

2.4.6: SDS-PAGE electrophoresis

Proteins were separated by size using sodium-dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

1.5mm 12% and 10% polyacrylamide gels were produced using the following:

Table 2.6: Production of 12% and 10% polyacrylamide gels

	dH$_2$0	30% Acrylamide	1.5M Tris (pH8.8)	10% SDS	10% ammonium persulphate (APS)	TEMED
Resolving gel (12%)	3.2ml	4ml	2.6ml	0.1ml	100μL	10μL
Resolving gel (10%)	3.8ml	3.4ml	2.6ml	0.1ml	100μL	10μL
	dH$_2$0	30% Acrylamide	1.5M Tris (pH6.8)	10% SDS	10% ammonium persulphate (APS)	TEMED
Stacking gel	3.6ml	6.25ml	0.05ml	0.67ml	50μL	5μL

Prepared protein samples were heated at 95°C for 5 minutes before loading onto SDS-PAGE gels. SDS-PAGE gels were run in a Mini-Protean Tetra Cell tank (BioRad) containing 1XSDS-running buffer (10XSDS-running buffer diluted with distilled water) at 120mV for 1.5-2hours or until the dye reached the bottom of the gel. Samples were run alongside 5μL Precision Plus Protein™ All Blue Prestained Protein Standards (BioRad). SDS-GELS were stained with InstantBlue™ protein stain (C.B.S Scientific).

2.4.7: Immunoblotting

SDS-PAGE gels were transferred onto a polyvinylidene difluoride (PVDF) membrane which was activated with 100% methanol before use. Transfer occurred in a wet transfer pack, Mini Trans-Blot® Cell (BioRad), for 60 minutes at 100mV. Membranes were blocked with 2% Bovine Serum Albumin (BSA) (Sigma) overnight at 4°C. After blocking, the membrane was incubated with primary antibodies diluted with 1% BSA to concentrations specified in table 2.3.

Blots were washed for 2X 10 minutes with PBS Tween 20 then incubate with the appropriate secondary antibody. Probed blots were then developed with ECL SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) and imaged with the Chemidoc (BioRad).

2.4.8: *T. brucei* heat shock assay

T. brucei procyclic cells were subject to heat shock by 10ml of cells being transferred to a 50ml falcon tube and heated in a 41°C water bath for 1 hour. At 0-6hours after heating, 500μL cells were prepared for fluorescent imaging as described in section 2.4.4. These slides were imaged using the DeltaVision fluorescent microscope and K and N counts were carried out as described in section 2.4.4.

Section 2.5: Biochemical protein analysis

2.5.1: Protein test expression

Rosetta(DE3)pLYsS *E. coli* competent cells were incubated with DNA pET28a ligations using the same method as described in section 2.3.6. Rosetta *E. coli* was used here as they can be used to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli.* This was plated onto LB agar plates with added Kanamycin (Kn) and Chloromphenicol (Chl). Plates

were incubated at 37°C overnight and single colonies were picked from each plate and incubated in 5ml LB broth with added Kn and Chl, shaking at 37°C for 8 hours.

To test for protein expression, the bacterial suspensions were used to inoculate 50ml of LB with added Kn and Chl, shaking at 37°C for 3 hours in a conical flask. After 3 hours, 1ml of this culture was removed and centrifuged at 13000 rpm for 1 minute before supernatant was removed and cell pellet was resuspended in 100µL 1XSDS loading buffer. This was heated at 95°C for 5 minutes. IPTG (1mM) was added to the 50ml culture before incubating for a further 3 hours at 37°C, taking 1ml samples after 30 minutes, 1 hour, 2 hours and 3 hours preparing in the same way as above. These 1ml samples were run on 12% SDS-PAGE gel, see section 2.4.6, and SDS-PAGE gels were transferred to PVDF membrane and incubated with appropriate primary and secondary antibodies.

To prevent co-purification of GroEL proteins, test expression was carried out at 30°C. After the initial incubation at 37°C for 3 hours, the culture was cooled in an ice bath for the addition of IPTG (1mM) and incubated at 30°C for 5 hours, taking 1ml protein samples every hour.

Cells were harvested from the 50ml culture by centrifuging in a 50ml falcon tube at 4000g for 15 minutes at 4°C. The supernatant was removed, and cell pellets were taken forward for protein purification.

2.5.2: Protein purification

Protein purification was carried out using NI-NTS affinity chromatography. Cell pellets prepared by test expression, section 2.5.1, homogenised with 10ml NI-NTA Buffer A with the addition of 1 protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail, Roche). This was sonicated for 7X 15 second intervals on ice, before being centrifuged in at 20,000 rpm for 30 minutes at 4°C. The supernatant was removed, keeping pellet as the protein lysate, and added to a NI-NTA column in a 15ml falcon tube, made up from 200µL NI-NTA resin (Thermo Fisher Scientific), and left for 10 minutes before centrifuging at 700g for 2 minutes at 4°C. The supernatant was removed and kept as the flow through, the pellet was resuspended in 800µL NI-NTA Buffer A and run through a column made of glass wool (Sigma-Aldrich) in a 5ml stripette (Corning), wash 1. A mixture of 10% NI-NTA Buffer B in NI-NTA Buffer A was made up and run through the column twice, wash 2 and 3. Finally, 400µL Ni-NTA Buffer B was run through the

column, protein elution. The lysate and flow through were diluted with NI-NTA Buffer A and 4Xlaemmli sample buffer (BioRad), while the wash 1, wash 2, wash 3 and elution samples were mixed with 4Xlaemmli sample buffer only. All were run on 12% SDS-PAGE gel and transferred membranes were incubated with appropriate primary and secondary antibodies.

To prevent co-purification of GroEL proteins, NI-NTA chromatography was repeated with the addition of Adenosine triphosphate (ATP) and Glycerol to improve protein stability and catalyse protein refolding. This was carried out using the same method as above, instead resuspending the flow through pellet in 800µL 35% Glycerol in NI-NTA Buffer A with the addition of 1M ATP. This was left for 30 minutes at room temperature before being run through the column and carrying out wash steps as above.

2.5.3: Protein dialysis

Dialysis of protein elution sample and commercially available citrate synthase into stable buffers, table 2.7. This was carried out using 15ml Slide-A-Lyzer™ MINI Dialysis Devices (Thermo Fisher Scientific), filling each tube with the specific protein and buffers. These were incubated at 4°C shaking overnight.

Table 2.7: Buffers used for protein dialysis

Protein	Buffer
DPCD	50mM TRIS, 2mM EDTA
Citrate synthase	40mM HEPES

2.5.4: Thermal aggregation assay

The thermal aggregation assay was carried out following the protocol from Haslbeck and Buchner, 2015. The DPCD protein and citrate synthase samples from dialysis tubes were centrifuged at 14,000 rpm for 10 minutes at 4°C. The concentration of each was measured using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific) at absorbance 280nm. This

absorbance reading and the extinction coefficient of each portion was used to determine the volume of each protein needed. 4.8µM DPCD was added to well A1 of a 96-well plate. 1:2 dilutions using 40mM HEPES buffer were carried out leaving 2.4, 1.2, 0.6, 0.3 and 0.15µM in wells A2-A6. 40mM HEPES was added to wells A7-A8 alone. This was heated to 43°C in a 1420 Victor2 Microplate Reader (Wallac). 0.6µM of citrate synthase was then added to wells A1-A7, and absorbance was measured at 390nm at 43°C for 45 repeats in the plate reader. The data was exported and the mean value from the negative control, well A8, was subtracted from all other data points.

Section 3: Results

Section 3.1: Investigating the conservation of DPCD across ciliated eukaryotes

The DPCD protein is strongly conserved across ciliated eukaryotes. This is demonstrated by the alignment of the *Homo sapiens* and *Trypanosoma brucei* DPCD amino acid sequences, Figure 3.1.

```
TbDPCD      ------MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGG      54
HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA      60
            :.    :   *:::: :***::   * ** ** ****  *.:**:**:* :::**.

TbDPCD      FSAWEVEVGNEAS--TRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAV      112
HsDPCD      MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV      120
            :. *::***: *   : **. **: *:...* .:::*:   : :*******  *:*:**:*

TbDPCD      EHSKPTDMGEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVV      172
HsDPCD      DQKE----RCIIVRTTNKKYYKKFSIPDLDRHQLPLDDALLSFAHANCTLIISYQKPKEV      176
            ::.:     *:***:****:*::****::*::* ** * ***   : **** *:**  *

TbDPCD      LAAESAAKKERASLPAKRIDDADSRTSCNQQ      203
HsDPCD      VVAESELQKELKKVKTAHSNDGDCKTQ----      203
            :.***  :**  .: : : :*.*.:*.
```

Figure 3.1: Alignment of *H. sapiens* and *T. brucei* DPCD sequences. Alignment of the *H. sapiens* and *T. brucei* DPCD amino acid sequences. The *T. brucei* DPCD amino acid sequence was gained from the 'TriTrypDB' database, accession number Tb927.11.8900. The *H. sapiens* DPCD amino acid sequence was gained from a 'BlastP' search against *T. brucei* DPCD sequence. The sequences were then aligned in 'ClustalOmega'. This shows strong conservation of the amino acid sequence across these two species.

To investigate this conservation, a protein database search using 'BlastP' of the *H. sapiens* DPCD sequence was carried out in a range of ciliated and non-ciliated eukaryotes, some of which possess motile cilia, to find putative DPCD homologues. The sequence hits that had E value above e^{-10} were taken and reciprocal searches of these sequences back into *H. sapiens* were carried out, to ensure that the top hits from this are the initial *H. sapiens* DPCD sequence. The results from these searches can be seen in Figure 3.2. For the full list of sequence hits see appendix table 1.

DPCD

Homo sapiens
Xenopus laevis
Danio rerio
Drosophila melanogaster
Caenorhabditis elegans
Nematostella vectensis
Hydra magnipapillata
Amphimedon queenslandica
Trichoplax adhaerens
Monosiga brevicollis
Rozella allomycis
Thecamonas trahens
Phytophthora infestans
Aphanomyces invadans
Aphanomyces cochlioides
Saprolegnia parasitica
Pythium ultimum
Albugo candida
Paramecium tetraurelia
Plasmodium falciparum 3D7
Toxoplasma gondii
Reticulomyxa filosa
Plasmodiophora brassica
Micromonas pusilla
Chlamydomonas reinhardtii
Trypanosoma brucei
Leishmania major
Perkinsela sp.
Naegleria gruberi
Trichomonas vaginalis
Tritrichomonas foetus
Giardia intestinalis
Neurospora crassa
Saccharomyces cerevisae
Schizosaccharomyces pombe
Ustilago maydis
Cryptococcus neoformans
Mucor circinelliodes
Encephalitozoon cuniculo
Acanthamoeba castellani
Entamoeba histolytica
Dictyostelium discoideum
Cryptosporidium hominis
Theileria parva
Cyanidioschyzon merolae
Galdieria sulphuria
Chondrus crispus
Klebsormidium nitens
Arabidopsis thaliana
Oryza sativa

Figure 3.2: Coulson plot of putative DPCD homologues present in ciliated and non-ciliated eukaryotes. Coulson plot shows the eukaryotes that possess DPCD homologues, represented by filled circles. Organisms in blue are ciliated eukaryotes and organisms in red are non-ciliated eukaryotes. All homologues have an E value above e^{-10} and were confirmed by reciprocal 'BlastP' search. This shows that only ciliated organisms possess DPCD homologues, but this is not absolute.

As seen in Figure 3.2, DPCD homologues are only found in ciliated eukaryotes. However, this is not absolute as not all ciliated eukaryotes possess the protein. To investigate this further, these ciliated eukaryotes were grouped by ones that have motile cilia or primary cilia only, Figure 3.3. This showed that within these ciliated eukaryotes, DPCD homologues are only present in organisms with motile cilia. However, this is also not absolute, as putative homologues were not found in all the organisms with motile cilia. This suggests that DPCD may have a role in ciliogenesis but is not essential.
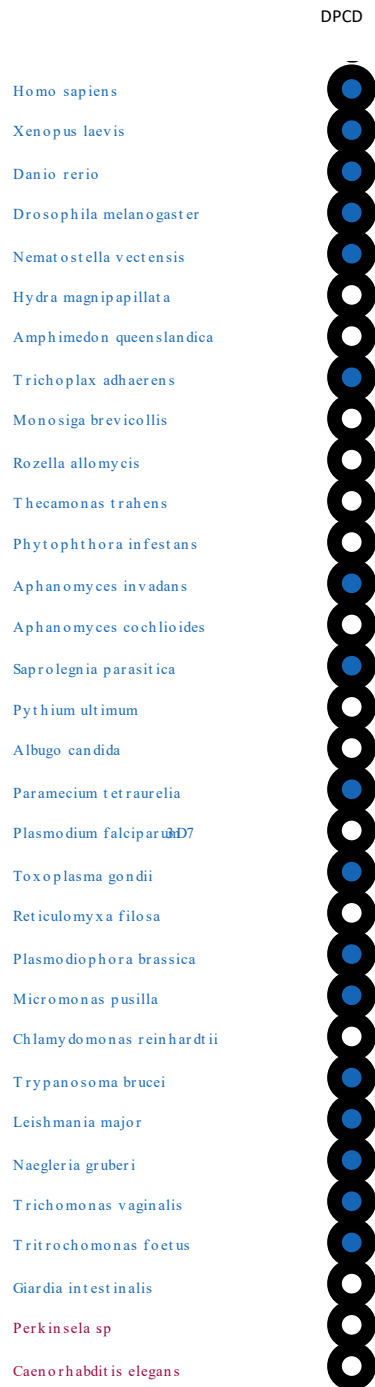
.

DPCD

Homo sapiens
Xenopus laevis
Danio rerio
Drosophila melanogaster
Nematostella vectensis
Hydra magnipapillata
Amphimedon queenslandica
Trichoplax adhaerens
Monosiga brevicollis
Rozella allomycis
Thecamonas trahens
Phytophthora infestans
Aphanomyces invadans
Aphanomyces cochlioides
Saprolegnia parasitica
Pythium ultimum
Albugo candida
Paramecium tetraurelia
Plasmodium falciparum 3D7
Toxoplasma gondii
Reticulomyxa filosa
Plasmodiophora brassica
Micromonas pusilla
Chlamydomonas reinhardtii
Trypanosoma brucei
Leishmania major
Naegleria gruberi
Trichomonas vaginalis
Tritrochomonas foetus
Giardia intestinalis
Perkinsela sp
Caenorhabditis elegans

Figure 3.3: Coulson plot of putative DPCD homologues present in ciliated eukaryotes with either motile or primary cilia. Coulson plot shows the ciliated eukaryotes that possess DPCD homologues, represented by filled circles. Organisms in blue are ciliated eukaryotes with motile cilia and organisms in red are ciliated eukaryotes with only primary cilia. All homologues have an E value above e^{-10} and were confirmed by reciprocal 'BlastP' search. This shows that only ciliated organisms with motile cilia possess DPCD homologues, but this is not absolute.

To prove that the 'BlastP' hits gained are putative DPCD homologues, amino acid sequences of all hits were aligned against the *H. sapiens* DPCD sequence using 'ClustalOmega', all of which show strong conservation of amino acid sequence, suggesting that these are putative homologues of the DPCD protein, see appendix figure 1.

Furthermore, to further investigate the conservation of DPCD, protein database searches using 'BlastP' of proteins known to be involved in ciliogenesis were carried out in the ciliated and non-ciliated eukaryotes used in above searches, to find putative homologues. These ciliary proteins were identified by Hodges et al., 2011, as a core set of proteins with an inferred ciliary function and flagellar localisation in trypanosomes. Several of these proteins are maintained in non-ciliated land plants, suggesting an additional role potentially co-opted before the loss of the cilia. This includes in the non-ciliated land plant *Arabidopsis thaliana*. It will be useful to compare the expression of these proteins to that of DPCD, especially in the organisms that possess motile cilia but no DPCD homolog. To do this, the *A. thaliana* 'ensemblePlant' database accession numbers of each of these proteins were inputted into 'BlastP' and searched in the ciliated and non-ciliated eukaryotes. The *A. thaliana* accession numbers were chosen as it is an organism that is not expected to possess these proteins, allowing a useful comparison to DPCD presence in non-ciliated organisms. The full list of ciliary proteins and 'ensemblePlants' database accession numbers can be seen in appendix, table 2. As above, only sequence hits with E value above e^{-10} and confirmed by reciprocal search back into *A. thaliana* were considered as putative homologues. The results of these searches can be seen in Figure 3.4. This shows that unlike DPCD, a large amount of the ciliary proteins are present in a number of non-ciliated organisms. However, there is a correlation between the expression of DPCD and the ciliary proteins, with being present mainly in ciliated eukaryotes.

Figure 3.4: Coulson plot of putative homologues of DPCD and known ciliary proteins present in ciliated and non-ciliated eukaryotes. Coulson plot shows the ciliated and non-ciliated eukaryotes that possess DPCD and ciliary protein homologues, represented by filled circles. Organisms in blue are ciliated eukaryotes and organisms in red are non-ciliated ciliated eukaryotes. All homologues have an E value above e^{-10} and were confirmed by reciprocal 'BlastP' search. This shows that unlike DPCD, some ciliary proteins are present non-ciliated organisms.

To investigate this further, the ciliated eukaryotes were grouped by ones that have motile cilia or primary cilia only, Figure 3.5.
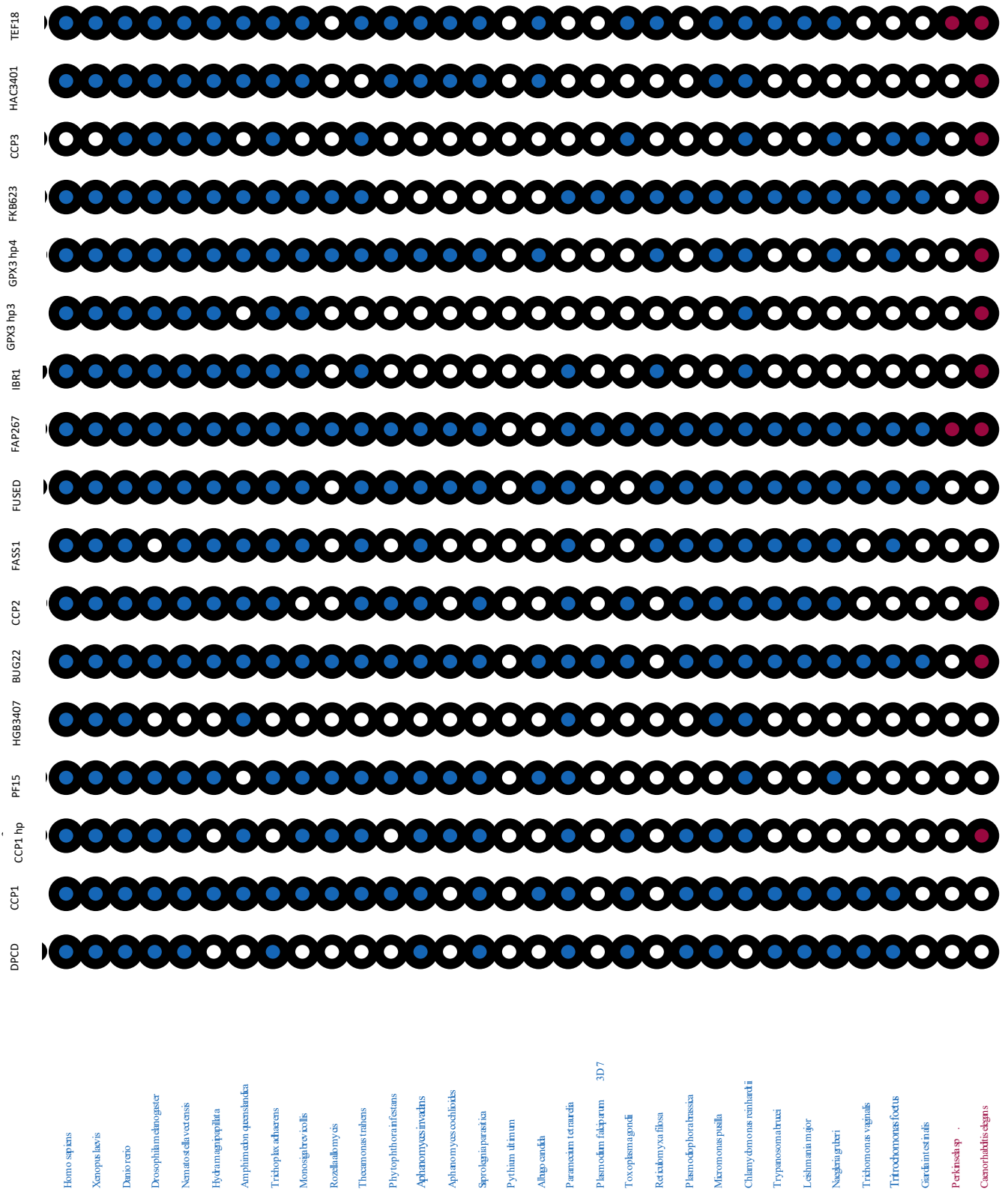
Figure 3.5: Coulson plot of putative homologues of DPCD and ciliary proteins present in ciliated eukaryotes with either motile or primary cilia. Coulson plot shows the ciliated eukaryotes that possess DPCD and ciliary protein homologues, represented by filled circles. Organisms in blue are ciliated eukaryotes with motile cilia and organisms in red are ciliated eukaryotes with only primary cilia. All homologues have an E value above e^{-10} and were confirmed by reciprocal 'BlastP' search. This shows a correlation between the DPCD and ciliary protein expression, with ciliary protein homologues found mainly in ciliated organisms with motile cilia. However, the organisms with motile cilia that do not possess DPCD still possess many the ciliary proteins.

As seen in Figure 3.5, there is a correlation between DPCD and ciliary protein expression, being mainly in ciliated organisms with motile cilia. However, this analysis does not explain the lack of DPCD in some organisms with motile cilia, as these organisms have homologues of many these ciliary proteins, ruling out the prediction that all these organisms possess different machinery for ciliogenesis.

To try to understand why these organisms with motile cilia do not contain DPCD homologues, similarities between these organisms were determined. To do this, the literature on different factors including the evolution and habitat of these organisms was assessed. From this analysis, two main factors were found to be common amongst these organisms. This includes that a number belong to early diverging metazoan groups. Five of these organisms, *Hydra magnipallata, Amphimedon queenslandica, Monosiga brevicollis, Rozella allomycis* and *Thecamonas trahens,* belong to groups that diverged early on in eukaryotic evolution. *Hydra magnipallata* is known to have diverged 570 million years ago, while *Amphimedon queenslandica* belongs to the Porifera phylum which is a type of sponge recognized as the oldest surviving metazoan phyletic lineage (Srivastava et al., 2010). In addition, *Monosiga brevicollis* belongs to the choanoflagellate group, which are the closest known relatives of metazoans (King et al., 2008). While *Rozella allomycis* and *Thecamonas trahens* both belong to early diverging phyla, Cryptomycota and Apusomonads, respectively (Letcher and Powell, 2018; Torruella et al., 2017). This suggests that these organisms may have evolved before the DPCD protein, which may explain the absence in these organisms. In addition, the second factor which links these organisms is that they belong to the SAR eukaryotic 'supergroup', containing the stramenopiles, alveolates and rhizaria (Derelle et al., 2016). This *includes Reticulomyxa filosa, Phytopthora infestans, Aphanomyces cochloides, Pythium ultimum, Albugo candida* and *Plasmodium falciparum.* It is known that *Plasmodium falciparum* possesses different machinery for

ciliogenesis compared to other eukaroytes, by carrying out cilia assembly in the cytoplasm rather than the cell membrane in an IFT-independent manner (Hodges et al., 2012). This is likely why this organism does not possess DPCD. It is also possible that the other members of this 'supergroup' also contain different ciliogenesis machinery compared to other eukaryotes. However, it is important to note that not all members of the SAR 'supergroup' have an absence of DPCD, including *Paramecium tetraurelia, Toxoplasma gondii* and *Plasmodiophora brassica*.

In conclusion, from the bioinformatic analysis of DPCD across eukaryotes, DPCD may have role in ciliogenesis, being present in only ciliated eukaryotes with motile cilia. However, it may not be essential for the formation of cilia as DPCD homologues are not present in all ciliated organisms with motile cilia. These organisms with motile cilia may not possess DPCD due to differences in ciliogenesis machinery or potentially as they belong to early diverging group, suggesting DPCD evolved more recently.

Section 3.2: Generation of DPCD plasmid constructs for analysis of *T. brucei* DPCD protein function through knockdown

To gain an understanding of the role of DPCD in *T. brucei*, two DPCD constructs were generated. These include a pPOT amplicon for tagging an endogenous DPCD allele and a DPCD p2T7-177 construct to enable inducible ablation of the protein in *T. brucei*. Both constructs were generated by PCR using *T. brucei* DPCD (*Tb*927.118900) specific for each plasmid.

3.2.1: Generation of pPOTmSCARLET- based construct for endogenous tagging of DPCD in *T. brucei*

T. brucei procyclic cells were engineered to enable the inducible RNAi mediated knockdown of DPCD protein in cells previously engineered to endogenously express a *Tb*DPCD::myc::mSCARLET fusion protein. To generate this cell line the PCR only tagging (pPOT) approach was used for endogenous tagging of the DPCD gene. This also allows a DNA sequence encoding a tag (in this case mSCARLET) to be amplified for fluorescent visualisation of the DPCD

protein. As mentioned above, the DPCD pPOTmSCARLET construct was generated by PCR using *Tb*927.118900_ppot primers and mSCARLET DNA template. The product from this amplification was then purified, Figure 3.6.
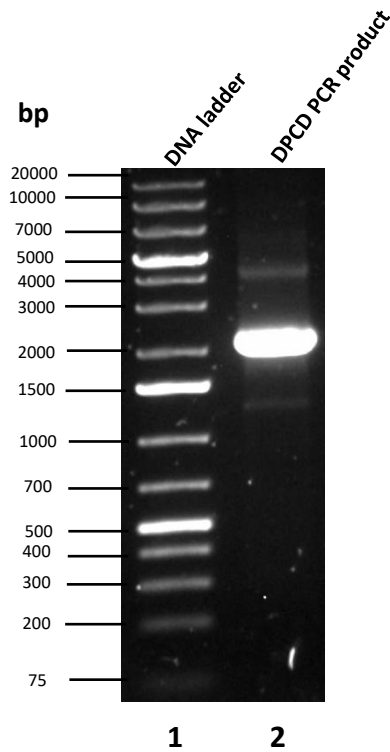
Figure 3.6: Agarose gel electrophoresis of DPCD pPOTmSCARLET PCR product. Agarose gel showing amplified DPCD pPOTmSCARLET product. Lane 1 contains 1kb plus DNA ladder, Lane 2 contains the purified DPCD pPOTmSCARLET PCR product, amplified with Q5 polymerase using Tb927.118900_ppot primers.

3.2.2: Generation of p2T7-177- based construct for RNAi of DPCD in *T. brucei*

Next, a DPCD p2T7-177 insert was generated by PCR. To create a p2T7-based vector to allow for the inducible ablation of the DPCD protein, primers were designed to allow the PCR amplification of the entire DPCD open reading frame. This amplified product was then A-tailed with DreamTaq polymerase. This adds a single deoxyadenosine (A) to the 3′end of the PCR product, amplified initially with high-fidelity polymerase, to allow ligation into the pGEM-T Easy

vector. This product was then purified, Figure 3.7, which showed the major PCR product at 650bp, which corresponds to the expected size of DPCD.
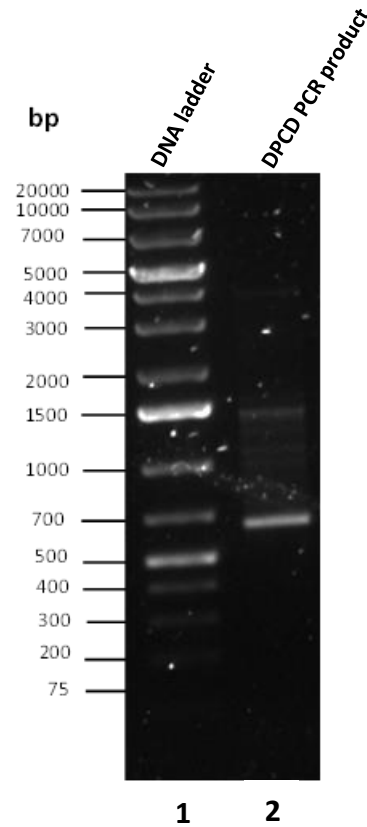
Figure 3.7: Agarose gel electrophoresis of DPCD PCR product for p2T7-177 vector ligation. Agarose gel showing amplified DPCD PCR product. Lane 1 contains 1kb plus DNA ladder, Lane 2 contains the purified DPCD PCR product, amplified with Q5 polymerase and A- tailed with DreamTaq polymerase, using Tb927.118900_p2T7 primers for ligation into p2T7-177 plasmid. The major PCR product is seen at 650bp, corresponding to the expected size of the DPCD insert.

The DNA fragment encoding the DPCD ORF was then ligated into pGEM-T Easy vector and transformed into *E. coli* competent cells onto Carbenicillin IPTG/XGAL plates for blue-white screening. The plasmid DNA in five white colonies, which are putative recombinants containing the DPCD insert, was then purified, and digested with EcoRI enzyme, to determine whether the DPCD insert is present, Figure 3.8.

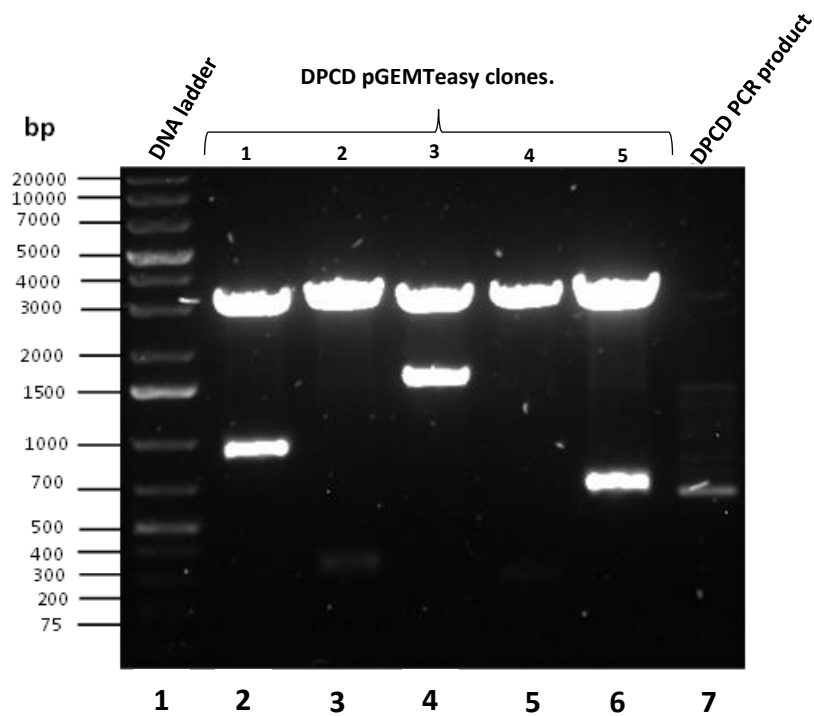**Figure 3.8: Agarose gel electrophoresis of DPCD pGEM-T Easy clones digested with EcoRI.** Agarose gel showing DNA constructs digested with EcoRI enzyme. Lane 1 contains 1kb plus DNA ladder. Lanes 2-6 contain the digested DPCD pGEM-T Easy clones. The DPCD insert is present in DPCD pGEM-T Easy clone 5 (lane 6), at the expected size of 650bp. Lane 7 contains 2µL of the DPCD PCR product. The DPCD insert is present in DPCD pGEM-T Easy clone 5 (lane 6), at the expected size of 650bp.

As seen in Figure 3.8, out of the five DPCD pGEM-T Easy clones that were selected, only clone 5 (lane 6) contained the pGEM-T Easy fragment at 3000bp and a fragment at the expected size of 650bp, lining up with the fragment present in the DPCD PCR product (lane 7). Clone 1 (lane 2) contains the pGEM-T Easy fragment and another fragment at 1000bp, which is not the expected size of the DPCD insert. Clone 2 (lane 3) also contains the pGEM-T Easy fragment and a fragment at 350bp, which also does not fit the expected size of the DPCD insert. Clone 3 (lane 4) contains the pGEM-T Easy fragment and another fragment at 1500bp, which is not the expected size of the DPCD insert while clone 4 (lane 5) contains only the pGEM-T Easy fragment.

To confirm that DPCD pGEM-T Easy clone 5 contained the DPCD open reading frame, it was sent off for nucleotide sequencing to confirm the DPCD insert sequence as wildtype. The resulting

nucleotide sequence provided by 'SourceBioscience' was then translated into the amino acid sequence using the 'Expasy' translate tool. The ORF was then inputted into 'ClustalOmega' and aligned against the *T. brucei*427 DPCD sequence (Tb427_110095600), confirming that the DPCD insert has 100% identity to *Tb*427 DPCD, Figure 3.9. The full nucleotide sequence of the DCPD insert can be seen in appendix figure 2.

```
Clone       MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV 60
Tb427DPCD   MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV 60
            ***********************************************************

Clone       EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM 120
Tb427DPCD   EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM 120
            ***********************************************************

Clone       GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK 180
Tb427DPCD   GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK 180
            ***********************************************************

Clone       KERASLPAKRIDDADSRTSCNQQ 203
Tb427DPCD   KERASLPAKRIDDADSRTSCNQQ 203
            ***********************
```

Figure 3.9: Alignment of DPCD insert and *Tb*427 DPCD sequence. Alignment of the ORF of DPCD insert from pGEM-T Easy clone against the *T. brucei*427 DPCD sequence, using 'ClustalOmega'. This shows 100% identity between the two sequences, indicating the DPCD insert as wildtype.

This DPCD pGEM-T Easy clone was digested with restriction enzymes BglII and HindIII, to allow for cloning into p2T7-177. The digested clone was then run on 0.8% agarose gel and the insert was excised, purified, and then ligated into p2T7-177 vector for transformation into *E. coli*. Five clones were selected, and plasmid DNA was purified and digested with SpeI and HindIII restriction enzymes, Figure 3.10. As BamHI and HindIII were chosen as the cloning sites on the p2T7-177 vector, and BglII was used to cut the DPCD insert, neither BamHI nor BglII could be used for the diagnostic digest. Instead, SpeI was used for this as it is upstream of the BglII site on the vector.

Figure 3.10: Agarose gel electrophoresis of DPCD p2T7-177 clones digested with SpeI and HindIII. Agarose gel showing DPCD p27-177 clones digested with SpeI and HindIII enzymes. Lane 1 contains 1kb plus DNA ladder. Lanes 2-6 contain the digested DPCD p2T7-177 clones. The DPCD insert is present in all clones, at the expected size of 650bp.

As seen in Figure 3.10, all the DPCD p2T7-177 clones (lanes 2-6) contain the p2T7-177 fragment at 7000bp and the DPCD insert at the expected size of 650bp. Clone 1 was taken forward and linearised using NotI enzyme, Figure 3.11.
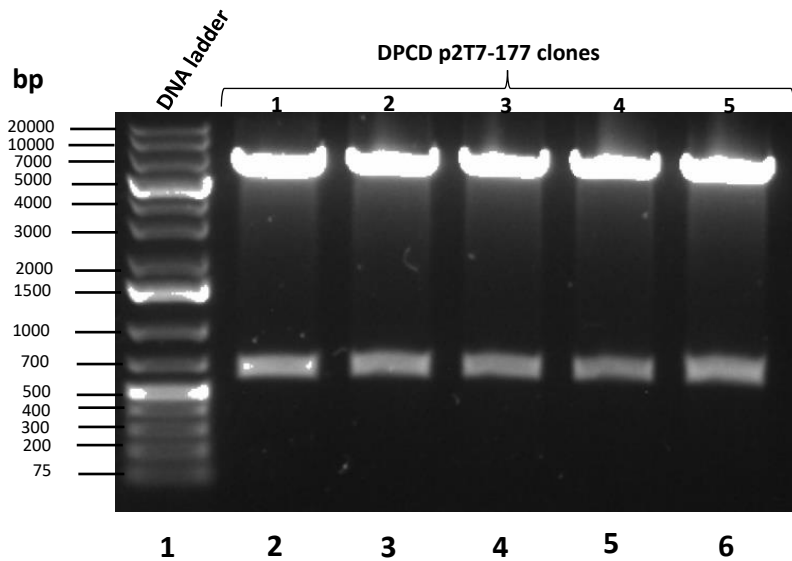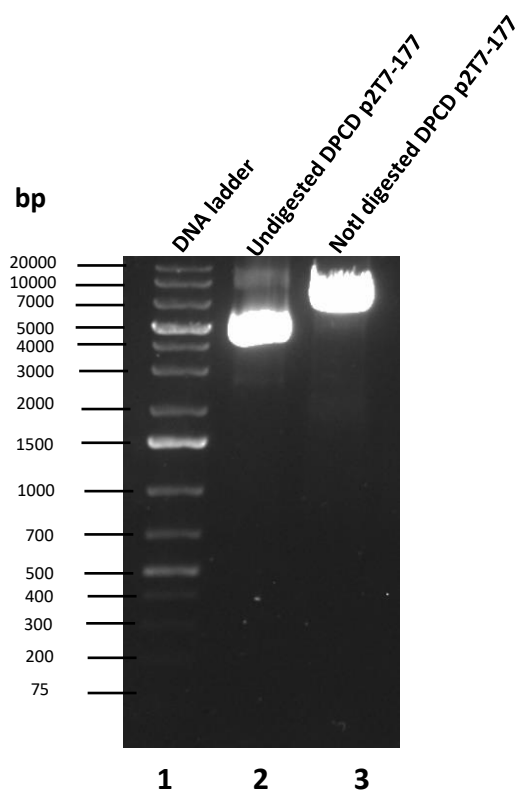
Figure 3.11: Agarose gel electrophoresis of DPCD p2T7-177 clone digested with NotI. Agarose gel showing DPCD p2T7-177 clone linearised with NotI enzyme. Lane 1 contains 1kb plus DNA ladder. Lane 2 contains the undigested DPCD p2T7-177 clone and lane 3 contains the DPCD p2T7-177 clone digested with NotI. The linearisation of the construct indicated by the linearised fragment running at a larger size of 7000bp compared to the undigested clone at 5000bp.

Section 3.3: Prediction of T. brucei DPCD protein function

3.3.1: RNAi *T. brucei* DPCD to predict protein function

To investigate the function of DPCD in *T. brucei*, induced ablation of the protein was carried out. To do this, the precipitated DPCD pPOT:mSCARLET and DPCD p2T7-177 DNA, prepared as described in previous section 3.2, was transfected into SP9 procyclic *T. brucei* cells. The DPCD pPOT:mSCARLET DNA was transfected into the cells first. This cell line was checked for the presence of the DPCD pPOT:mSCARLET by fluorescent imaging, against the mSCARLET signal. These were compared to the SP9 background cell line, Figure 3.12.
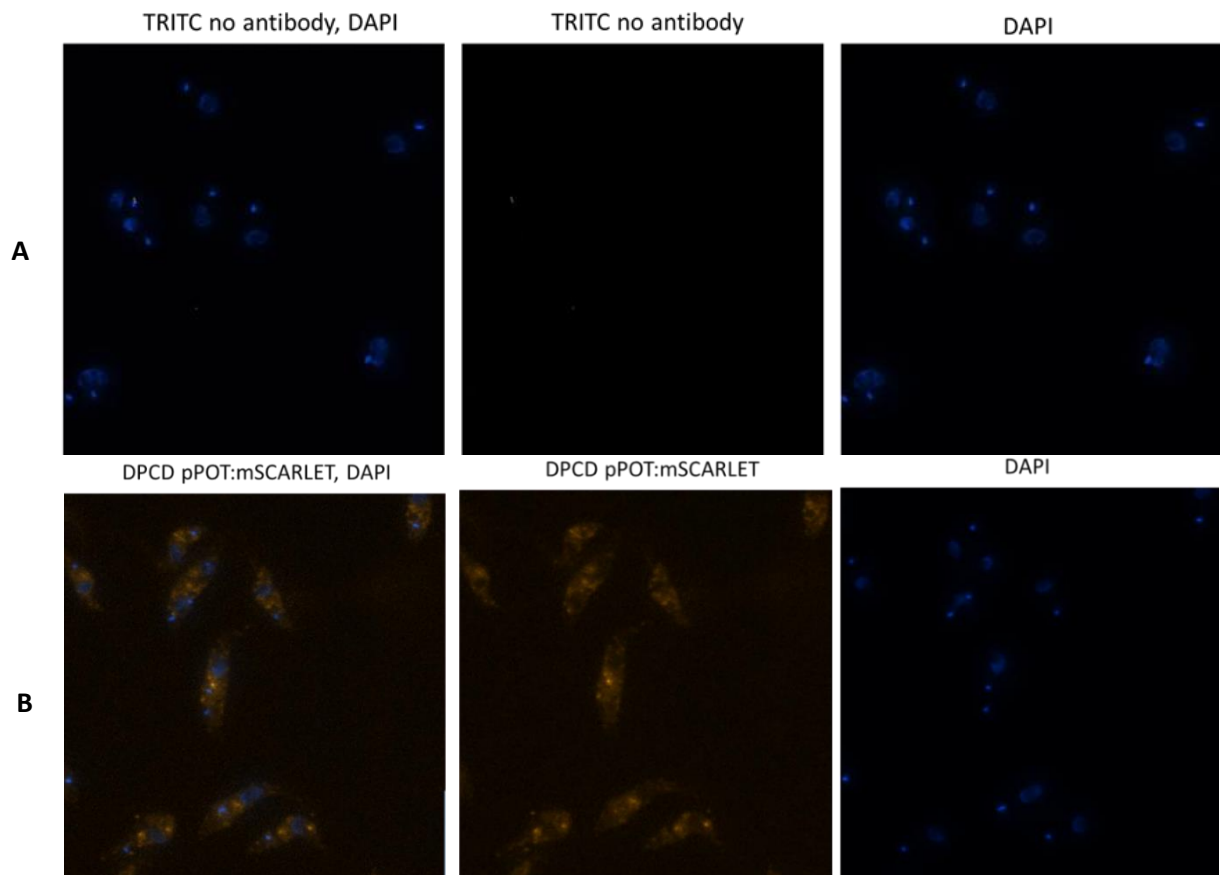
Figure 3.12: Fluorescent images of *T. brucei* SP9 background cell line and DPCD pPOT:mSCARLET transfected cells. Fluorescent images showing signal through the TRITC channel. (**A**) *T. brucei* SP9 background cell line showing TRITC, DAPI and merged signals. This shows that there is no background fluorescence through the TRITC channel. (**B**) DPCD pPOT:mSCARLET transfected cells showing DPCD pPOT:mSCARLET, DAPI and merged signals. This shows that the DPCD pPOT:mSCARLET has been successfully transfected, indicating DPCD localisation in the cell cytoplasm.

As seen in Figure 3.12, the DPCD pPOT:mSCARLET DNA was successfully transfected into the *T. brucei* cells, with a strong signal through the TRITC channel, which is not present in the SP9 background cell line. The DPCD pPOT:mSCARLET images also show that DPCD localises to the cell cytoplasm.

The DPCD p2T7-177 DNA was then transfected on top of the DPCD pPOT:mSCARLET cells. Induced ablation of the DPCD protein was then carried out on these cells. To do this, doxycycline was added to cells at a concentration of 5×10^5 cells/ml to induce DPCD depletion. Cell counts were taken every 24hours up to 166hours, for induced and non-induced cells, to determine if DPCD knockdown has an effect on cell growth, Figure 3.13.
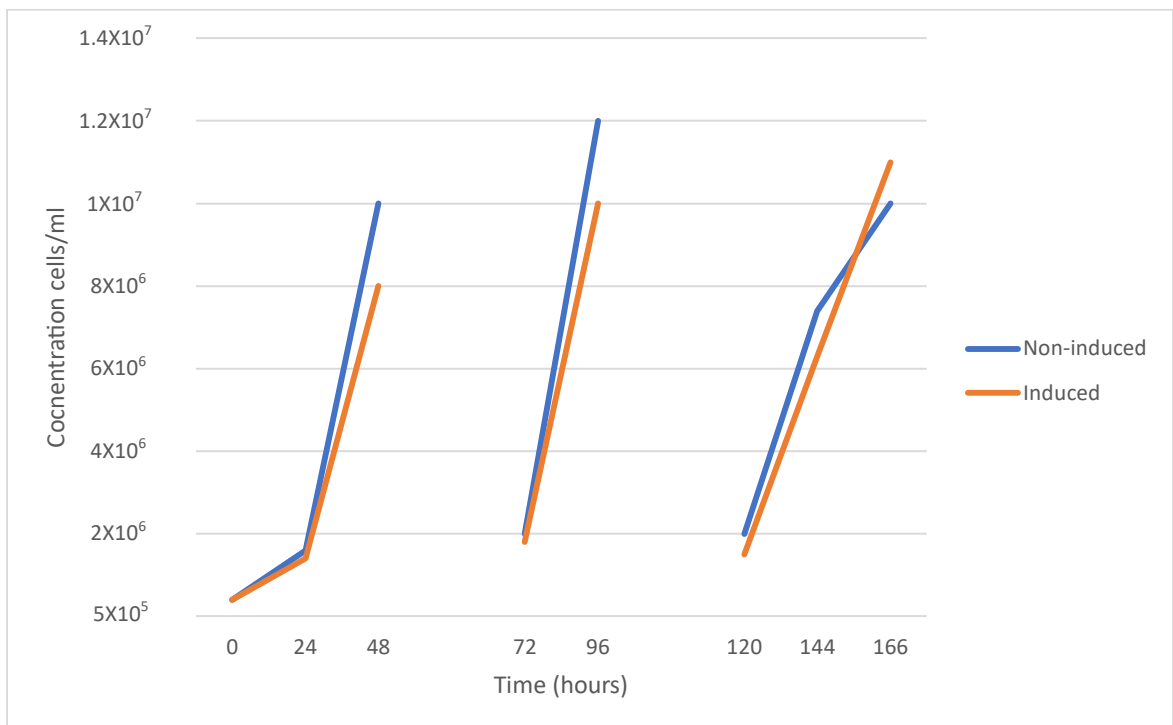
Figure 3.13: Growth curve of non-induced and induced DPCD p2T7-177::pPOT:mSCARLET cells. Growth curve shows the concentration (cells/ml) of DPCD p2T7-177::pPOT:mSCARLET cells with time. Orange lines represent cells induced for DPCD ablation. Blue lines represent non-induced cells. Cells subbed back to $5X10^5$ cells/ml after 48 and 96 hours. This shows that there is no difference in the growth of induced and non-induced cells.

As seen in Figure 3.13, there was no significant difference between the growth of the DPCD induced knockdown cells and the non-induced cells, suggesting that DPCD knockdown does not have any effect on cell growth. In addition, to determine if DPCD knockdown affects cell morphology, samples of induced and non-induced DPCD p2T7-177::pPOT:mSCARLET cells were taken every 24hours to investigate morphology and the depletion of the DPCD pPOT:mSCARLET signal through fluorescent imaging, Figure 3.14. These images show that the DPCD protein is being depleted in the induced population, as the DPCD pPOT:mSCARLET signal is weaker in the induced cell images compared to the non-induced cells, at all timepoints. Furthermore, the phase images show that there are no differences in cell morphology between the non-induced and induced cells, suggesting that DPCD knockdown does not have an effect on cell morphology. In addition, no effects on cell motility were observed in the DPCD depleted population.
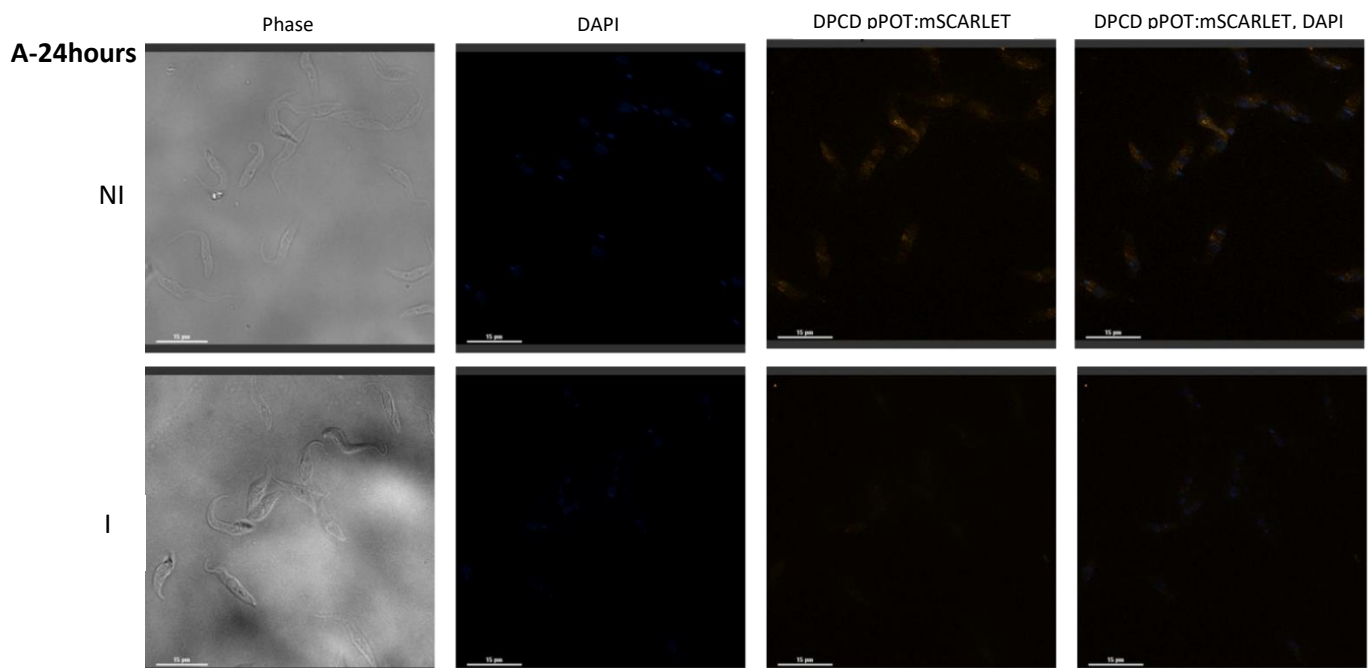
B-48hours

NI
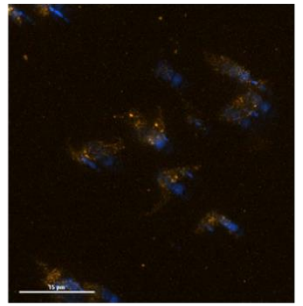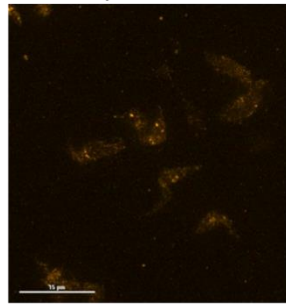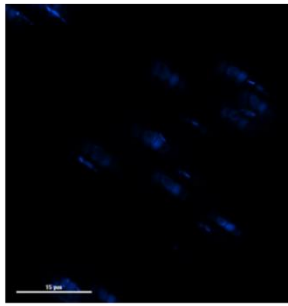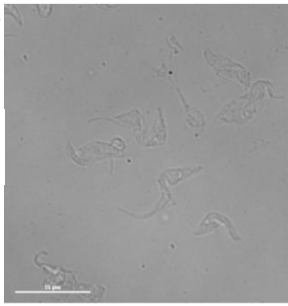
I
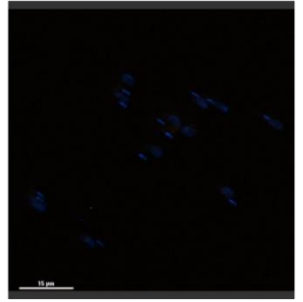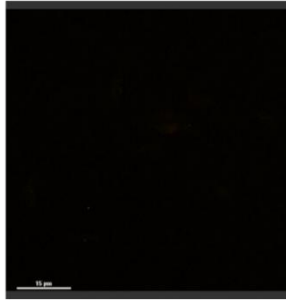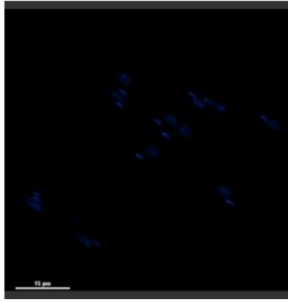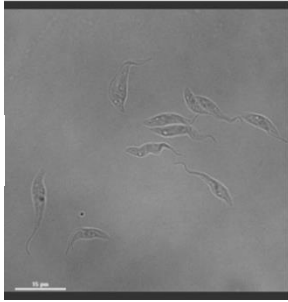
C-72hours

NI
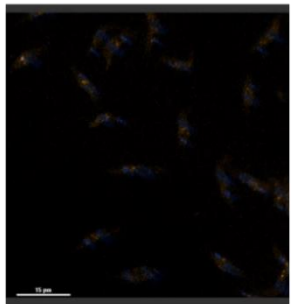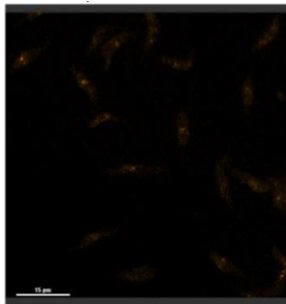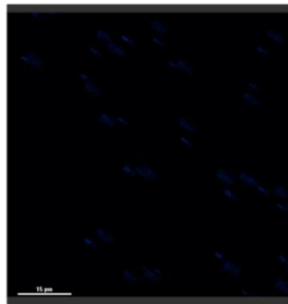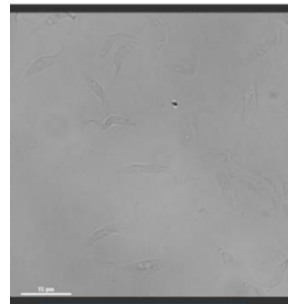
I
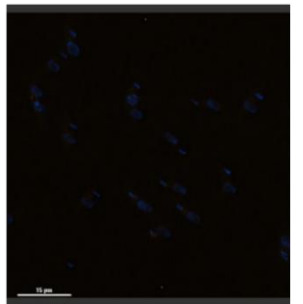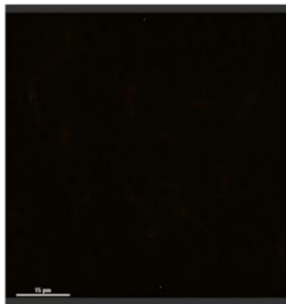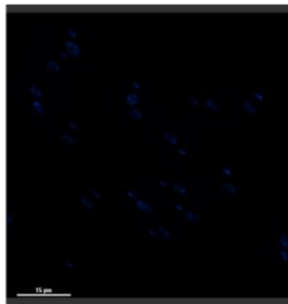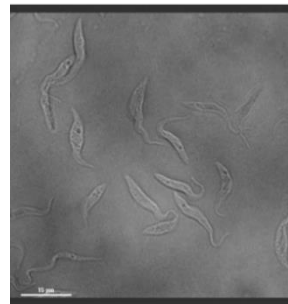
D-96hours

NI
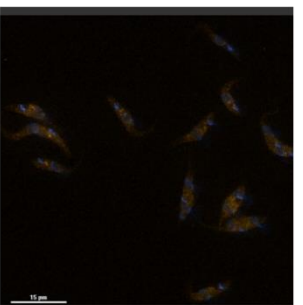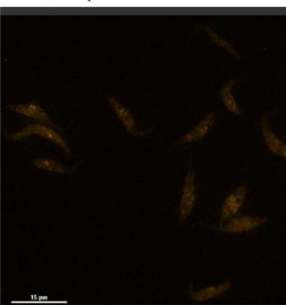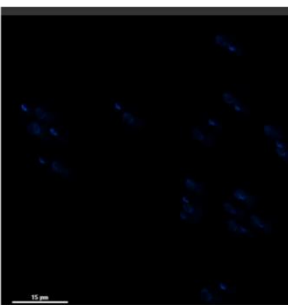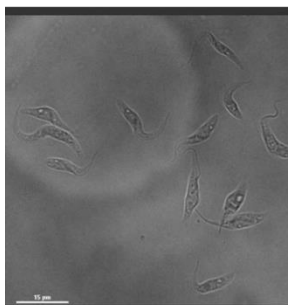
I
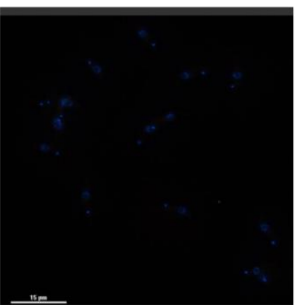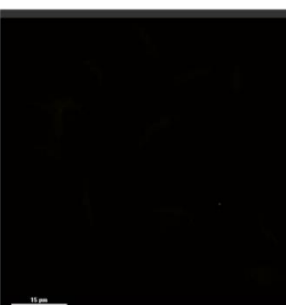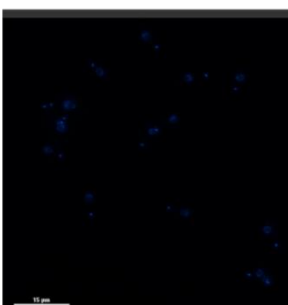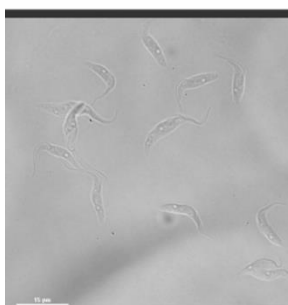
E-120hours

NI
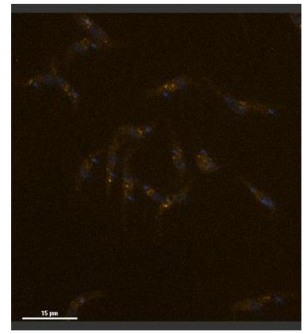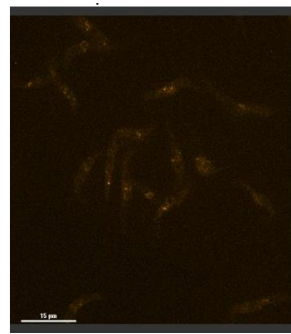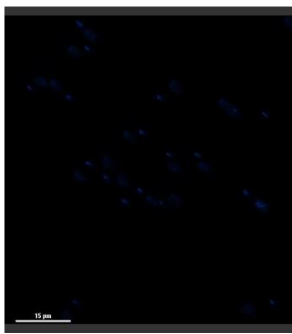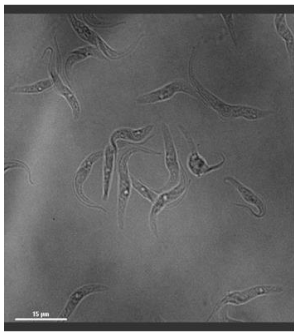
I
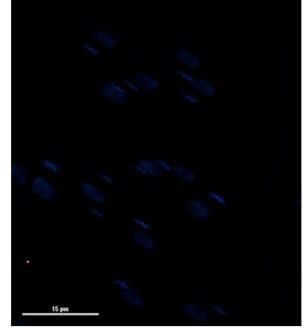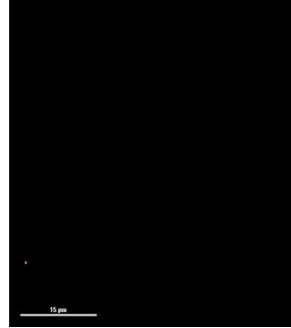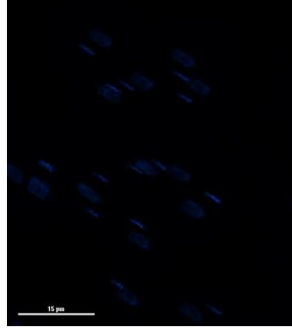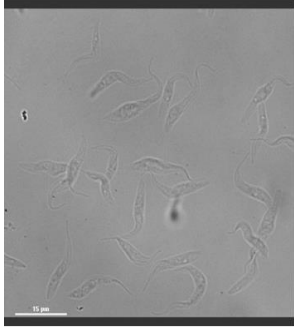
F-144hours

NI
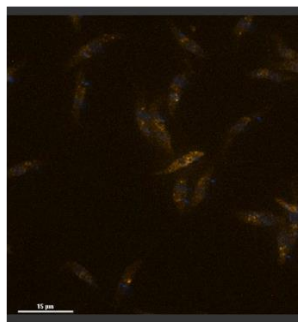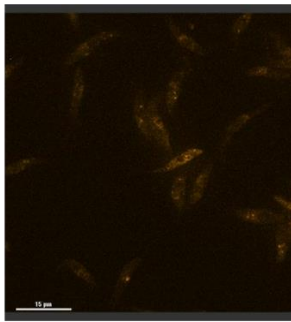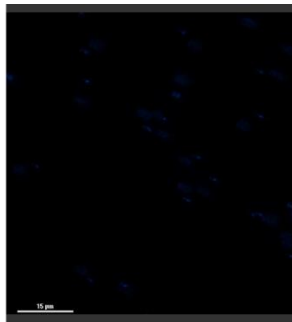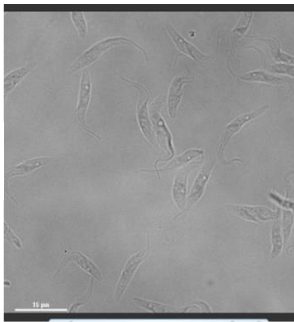
I
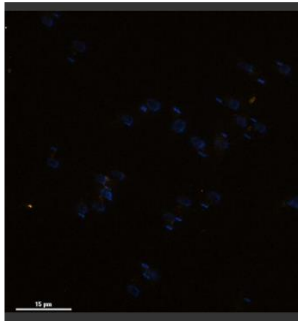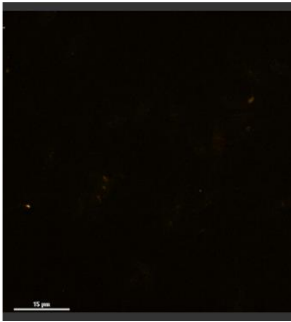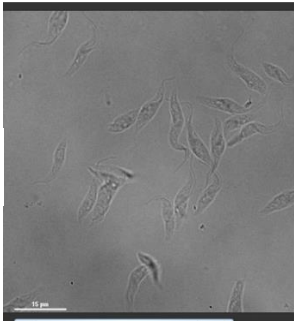
G-166hours

NI
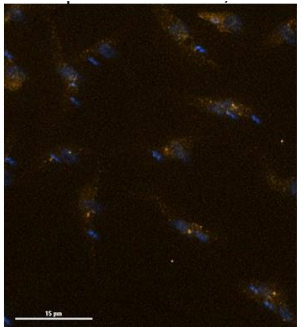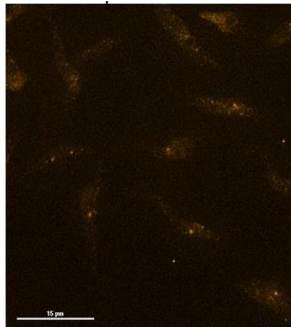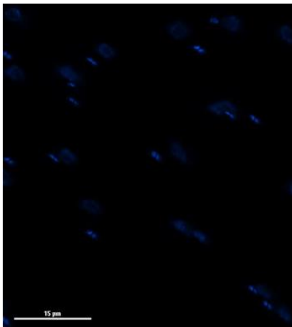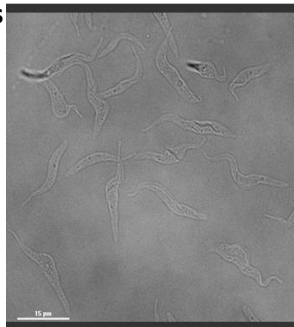
I
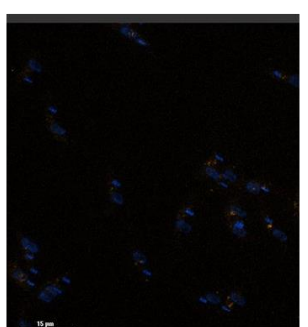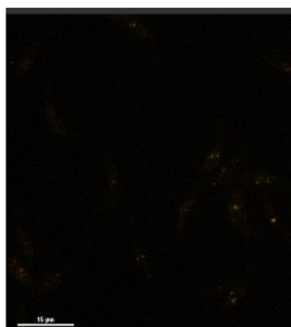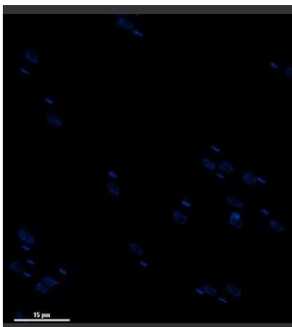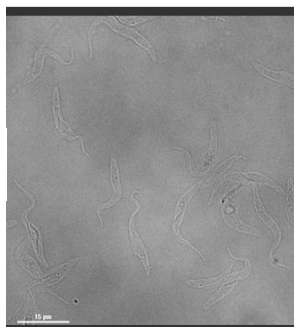
Figure 3.14: Fluorescent images of non-induced (NI) and induced (I) DPCD p2T7-177::pPOT:mSCARLET cells. Fluorescent images comparing morphology and fluorescent signal of DPCD p2T7-177::pPOT:mSCARLET cells with induced ablation of the DPCD protein (induced) and non-induced cells. (**A**) shows phase, DAPI, DPCD pPOT:mSCARLET and merge images of NI and I cells 24hours after induction, (**B**) shows NI and I cells 48hours after induction, (**C**) shows NI and I cells 72hours after induction, (**D**) shows NI and I cells 96hours after induction, (**E**) shows NI and I cells 120hours after induction, (**F**) shows NI and I cells 144hours after induction and (**G**) shows NI and I cells 166hours after induction. These images show that the DPCD protein is being depleted in the induced cells, as at all time points, the DPCD pPOT:mSCARLET signal is reduced in the induced population. The phase images show that there are not any morphological differences between the non-induced and induced cells.

To determine if induced ablation of the DPCD protein has an effect on cell cycle progression, kinetoplast and nucleus counts were carried out on approximately 100 non-induced and induced cells at each timepoint, Figure 3.15.
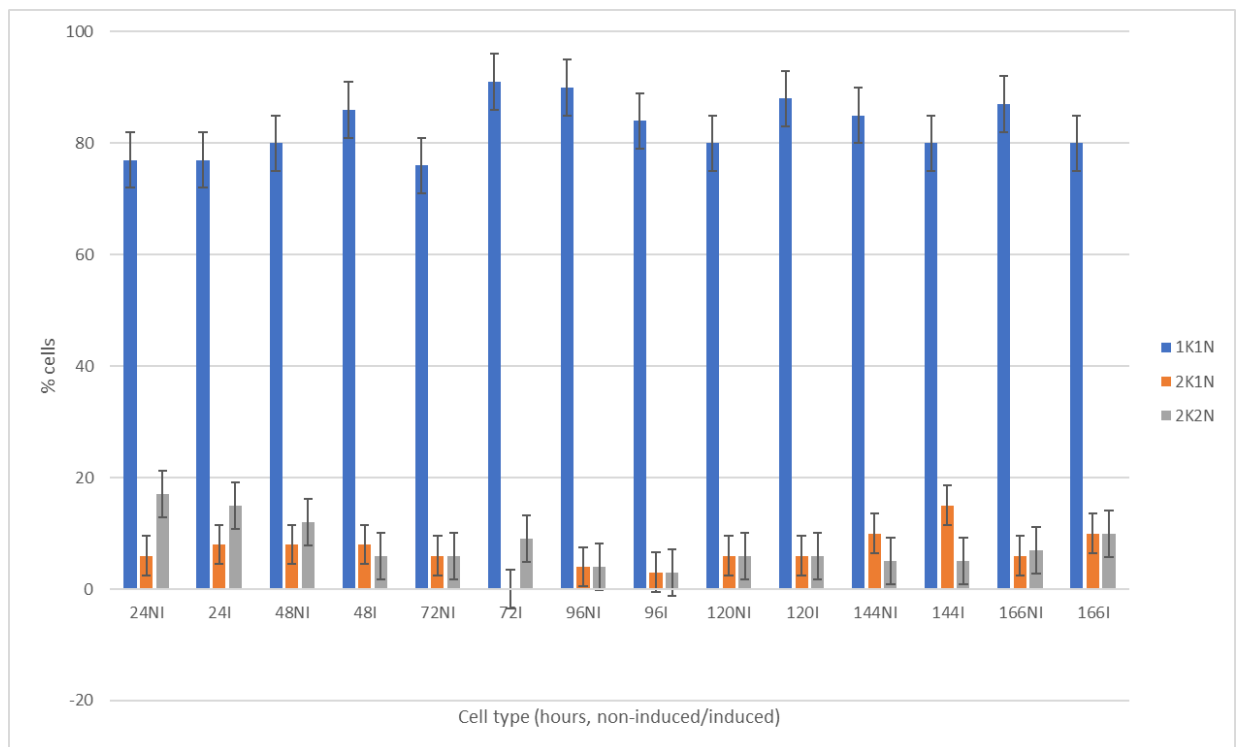
Figure 3.15: Bar chart of the percentage of cells at each cell cycle stage for non-induced (NI) and induced (I) DPCD p2T7-177::pPOT:mSCARLET cells at 24-166hours. Bar chart showing the %cells at 1K1N (blue), 2K1N (orange) and 2K2N (grey) in the NI and I population at each 24hour timepoint. Error bars show the standard deviation between NI and I cells of each cell cycle stage. A repeated measures ANOVA statistical test was carried out. The P value was >0.05 for all cell cycle stages (1K1N p=0.327, 2K1N p=0.693 and 2K2N p=0.693). This determined there was no significant difference between the % of cells at each cell cycle stage between NI and I populations at each timepoint.

As seen in Figure 3.15, there is no significant difference between the % of cells at each cell cycle stage between NI and I populations at each timepoint, with cells at 1K1N being the highest percentage in all samples. This suggests that DPCD knockdown does not have any effect on cell cycle progression in *T. brucei* cells.

To prove that the DPCD protein is being depleted in the induced population, $5X10^7$ cells/ml protein samples were taken from the non-induced and induced population at 0-166hours after doxycycline induction. These samples were run on 12% SDS-PAGE gel and probed by western blotting with an anti-Myc tag primary antibody against the myc tag present on the DPCD p2T7-177 construct, Figure 3.16 (A). The samples were also probed with an anti-tubulin antibody, KMX1, as a loading control, Figure 3.16 (B).
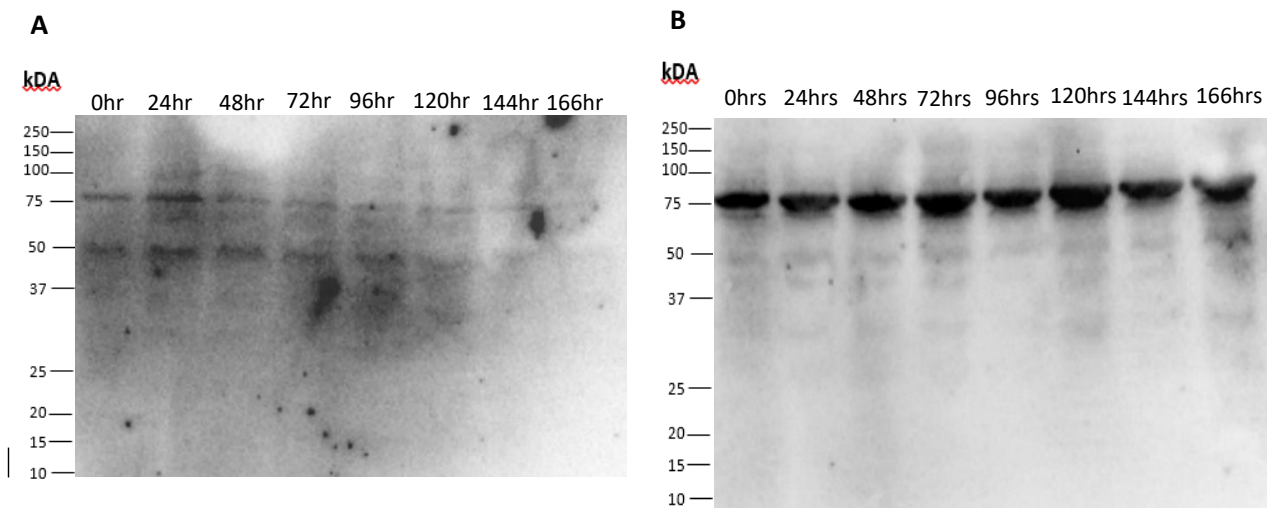
Figure 3.16: Western blots showing DPCD depletion with time compared to loading control.
(**A**) Western blot of DPCD p2T7-177::pPOT:mSCARLET protein samples 0, 24, 48, 72, 96, 120, 144 and 166 hours after induced ablation of the DPCD protein with doxycycline. The membrane was probed with mouse anti-MYC tag primary antibody (1:1000) and Rabbit-anti-mouse secondary antibody conjugated to HRP (1:10000). Bands seen at 50kDA, which is the expected size of the DPCD p2T7-177::pPOT:mSCARLET construct, decreasing in intensity with time. Bands are also seen at 75kDA, which is likely due to the DPCD protein being modified, as these are also being depleted. (**B**) Western blot of the same DPCD p2T7-177::pPOT:mSCARLET protein samples. The membrane was probed with loading control in the form of an anti-tubulin antibody, KMX1, (1:5000) and Rabbit-anti-mouse secondary antibody conjugated to HRP (1:10000). Bands seen at 75kDA, which is the expected size of tubulin, with the same intensity at each timepoint, proving equal loading and concentration of cells in each sample.

As seen in Figure 3.16 (A), the DPCD protein is being depleted in the induced cells, shown by the bands of decreasing intensity at 50 and 75kDA. Figure 3.16 (B) shows that there was equal loading of the samples and an equal concentration of cells.

3.3.2: Heat shock analysis of *T. brucei* DPCD to investigate function

To further investigate what the function of DPCD is in *T. brucei,* heat shock analysis on DPCD p2T7-177::pPOT:mSCARLET transfected cells was carried out, to determine if the DPCD protein has any effect on the heat shock response. To do this, mid-log phase cells were heat treated in a 41°C water bath for 1hour, taking samples for fluorescent imaging at 0-6hours post heat shock (induced). These samples were compared to DPCD p2T7-177::pPOT:mSCARLET cells that were not heat shocked (non-induced), Figure 3.17.

A-0hr after heat | Phase | DAPI | DPCD pPOT:mSCARLET | DPCD pPOT:mSCARLET, DAPI

N
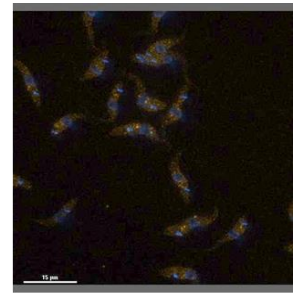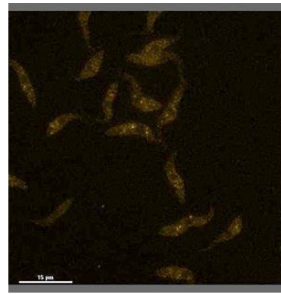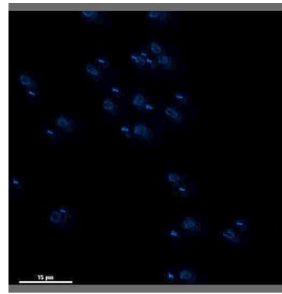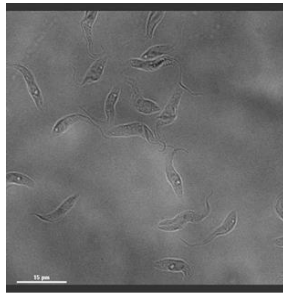
I
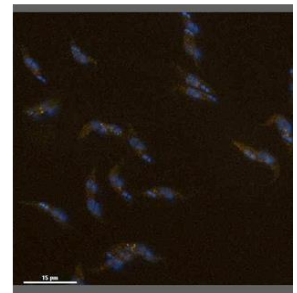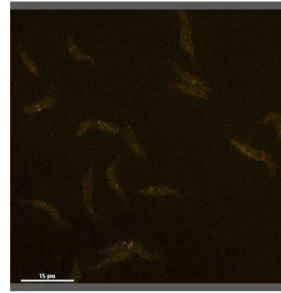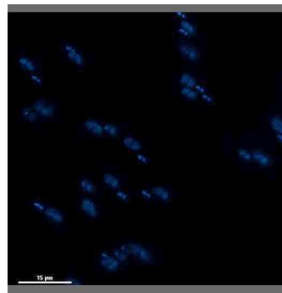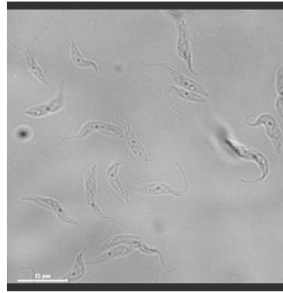
B-1hr after heat shock

NI
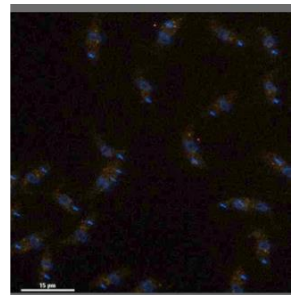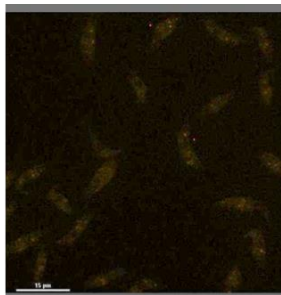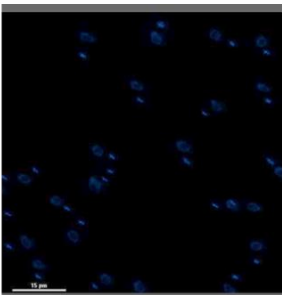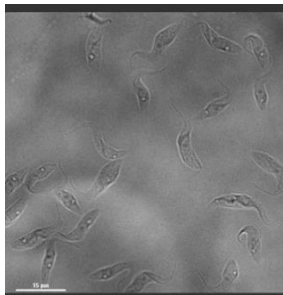
I
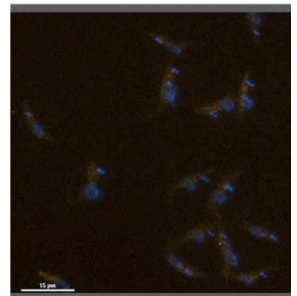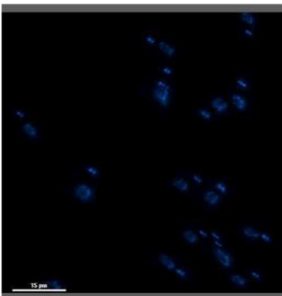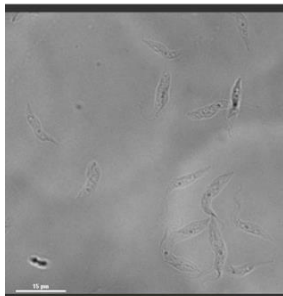
C-2hr after heat shock

NI
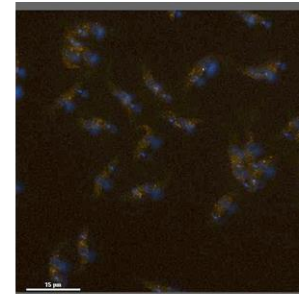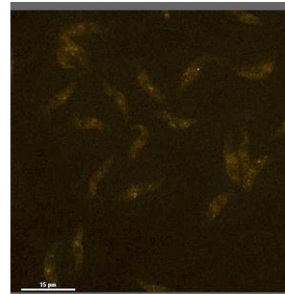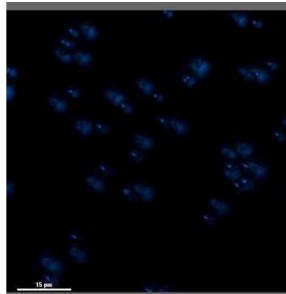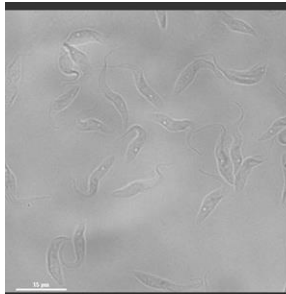
I
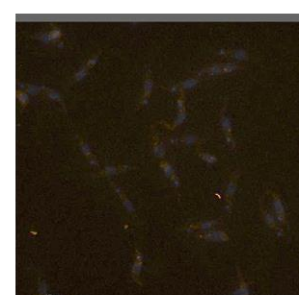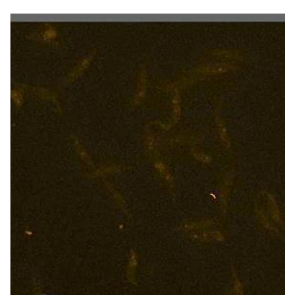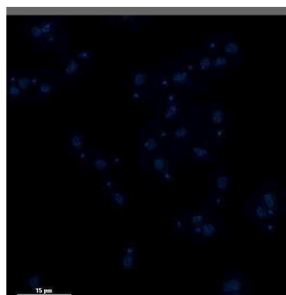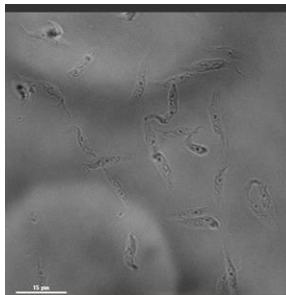
D-3hr after heat shock

NI

I
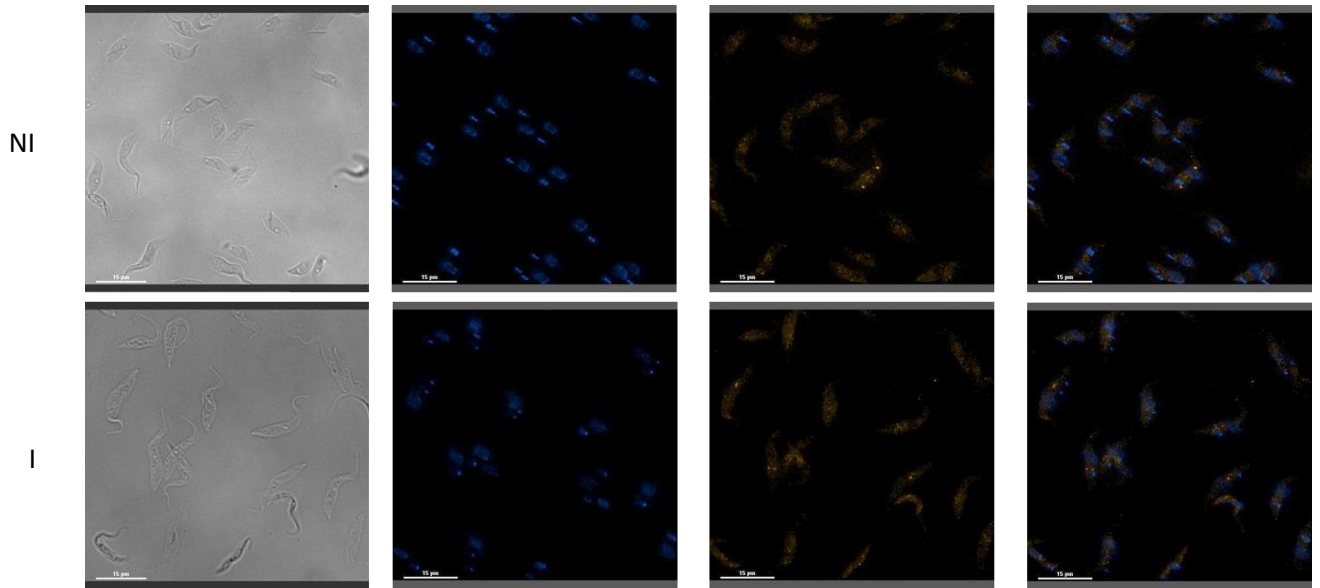
E-4hr after heat shock

NI

I
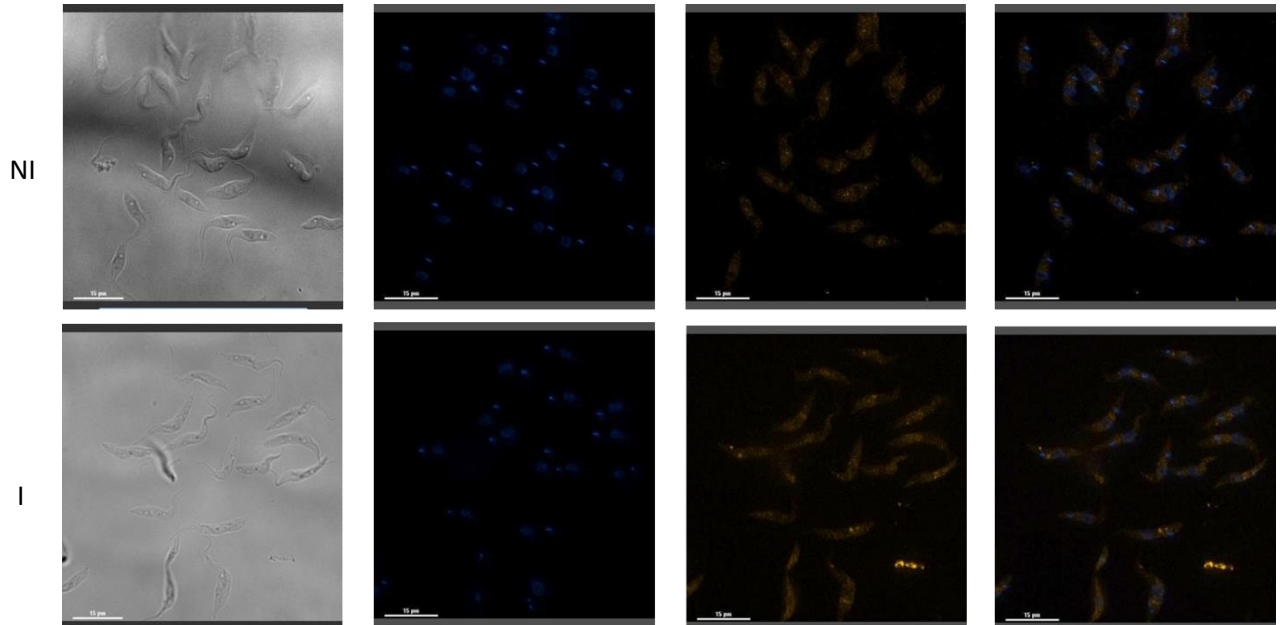
F-5hr after heat shock

NI

I
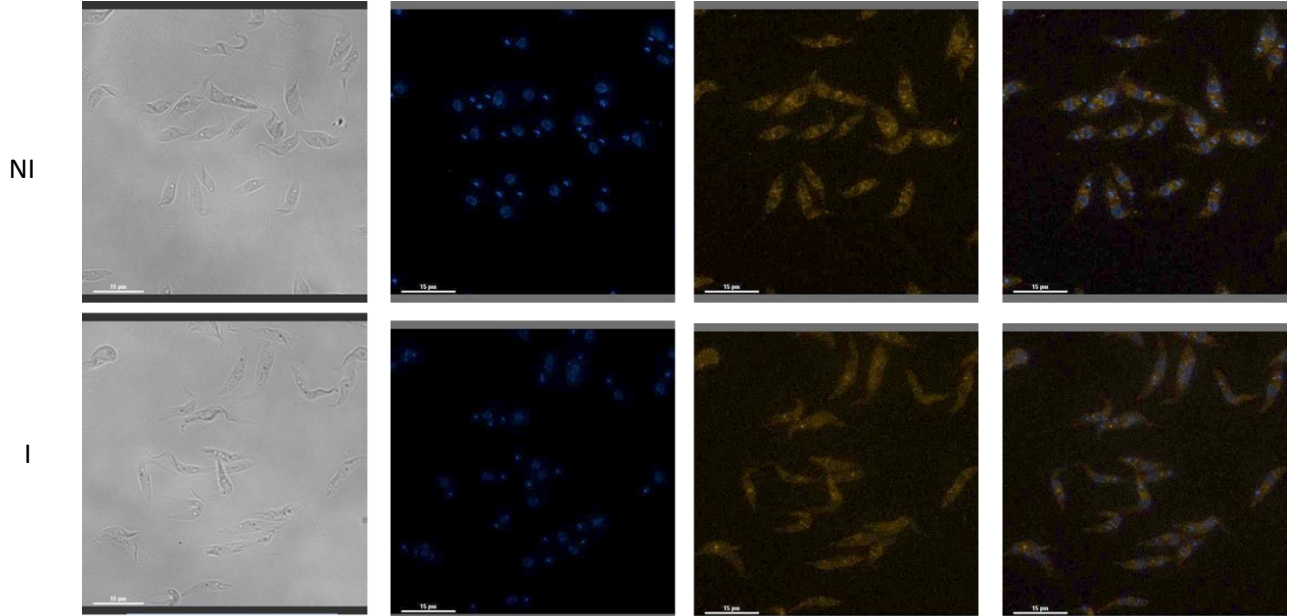
G-6hr after heat shock

NI

I
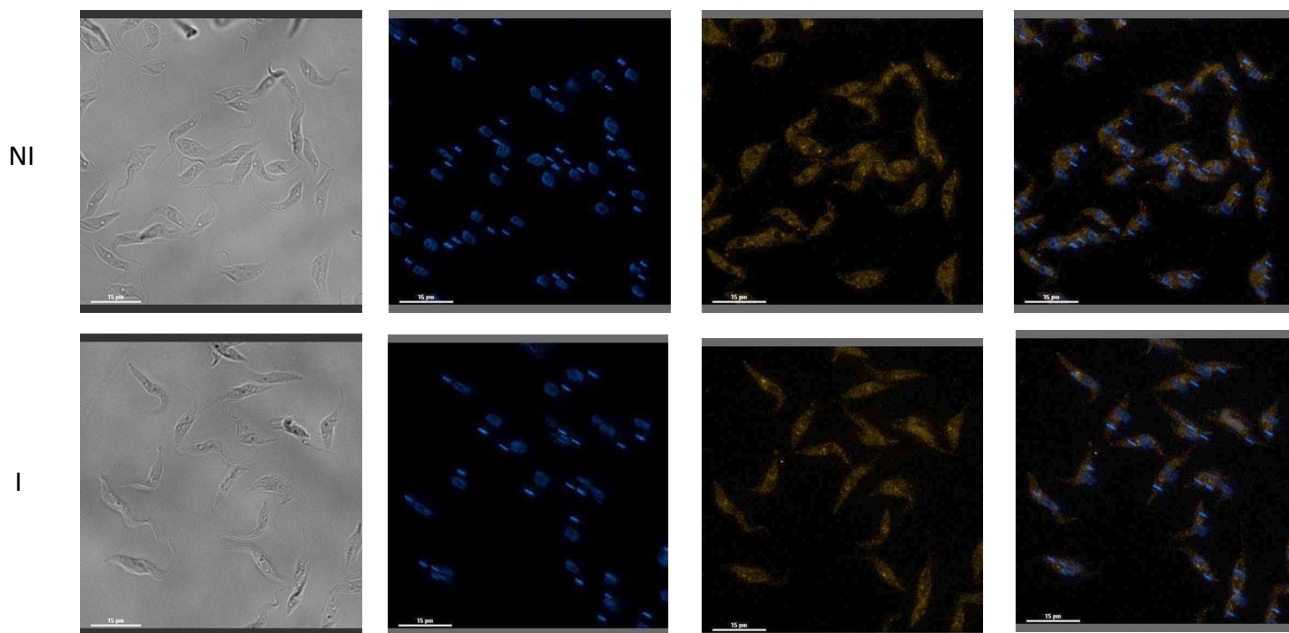
Figure 3.17: Fluorescent images of heat shocked (I) and non-heat shocked (NI) DPCD p2T7-177::pPOT:mSCARLET cells. Fluorescent images comparing morphology and fluorescent signal of DPCD p2T7-177::pPOT:mSCARLET cells that were heat shocked (I), compared to the non-heat shocked cells (NI). (**A**) shows phase, DAPI, DPCD pPOT:mSCARLET and merge images of NI and I cells 0hours after heat shock, (**B**) shows NI and I cells 1hours after heat shock, (**C**) shows NI and I cells 2hours after heat shock, (**D**) shows NI and I cells 3hours after heat shock, (**E**) shows NI and I cells 4hours after heat shock, (**F**) shows NI and I cells 5hours after heat shock and (**G**) shows NI and I cells 6hours after heat shock. These images show that the DPCD protein does not have any effect on the heat shock response in the *T. brucei* cells, with no differences between the DPCD pPOT:mSCARLET signal or the cell morphology seen in phase images, between the non-induced and induced populations.

As seen in figure 3.17, there are no differences in DPCD pPOT:mSCARLET signal and cell morphology between the heat shocked and non-heat shocked cells at all timepoints. This suggests that DPCD does not have a role in the heat shock response of *T. brucei* cells.

In addition, kinetoplast and nucleus counts were also carried out on approximately 100 cells heat shocked and non-heat shocked cells at each timepoint to see if DPCD affects cell cycle progression in response to heat shock, Figure 3.18.
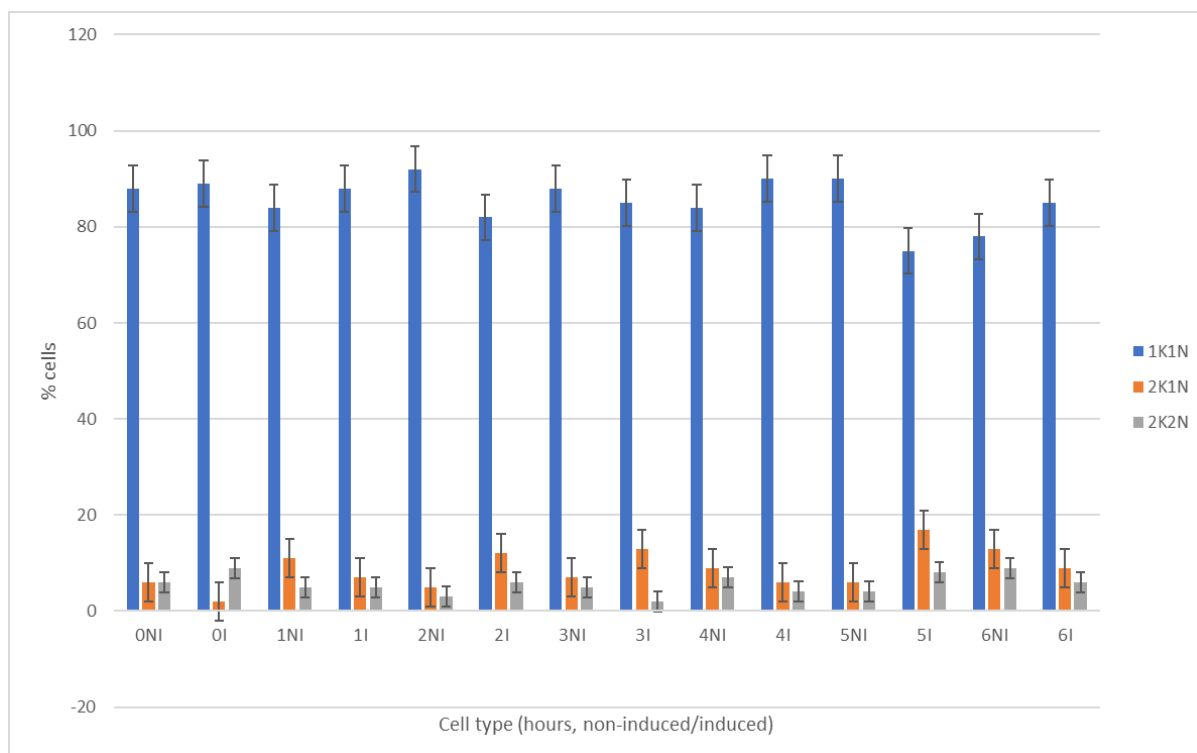
Figure 3.18: Bar chart of the percentage of cells at each cell cycle stage for non-heat shocked (NI) and heat shocked (I) DPCD p2T7-177::pPOT:mSCARLET cells at 0-6hours post heat shock. Bar chart showing the %cells at 1K1N (blue), 2K1N (orange) and 2K2N (grey) in the NI and I population at each timepoint post heat shock. Error bars show the standard deviation between NI and I cells of each cell cycle stage. A repeated measures ANOVA statistical test was carried out. The P value was >0.05 for all cell cycle stages (1K1N p=0.668, 2K1N p=0.618 and 2K2N p=0.909). This determined there was no significant difference between the % of cells at each cell cycle stage between NI and I populations at each timepoint.

As seen in Figure 3.18, there is no significant difference between the % of cells at each cell cycle stage between heat shocked and non-heat shocked DPCD p2T7-177::pPOT:mSCARLET cells at each timepoint, with cells at 1K1N being the highest percentage in all samples. This suggests that the DPCD does not have any effect on cell cycle progression in *T. brucei* cells after heat shock.

In conclusion, from the data presented in section 3.3, it is shown that induced ablation of DPCD in *T. brucei* cells does not have any affect on cell growth or morphology. This leaves the question

of what the role of the protein is in *T. brucei* largely unknown. However, the DPCD protein is unlikely to be involved in control of cell cycle progression. In addition, from the heat shock analysis carried out on DPCD transfected cells, the protein is unlikely to be involved in the heat shock response of *T. brucei* cells.

Section 3.4: Generation of pDEX377- based construct for overexpression of DPCD in *T. brucei*

To further investigate the role of DPCD in *T. brucei*, *T. brucei* procyclic cells were engineered to overexpress the DPCD protein. To do this a DCPD pDEX377 insert was generated by amplifying the DPCD ORF, except the stop codon, by PCR. It important that this PCR product did not contain the stop codon as it needs to read through in frame to allow for cloning into the pDEX377 vector which contains a triple myc tag. As described in earlier section 3.2, the amplified product was A-tailed and purified, Figure 3.19, which showed the major PCR product at 650bp, which corresponds to the expected size of the DPCD insert.
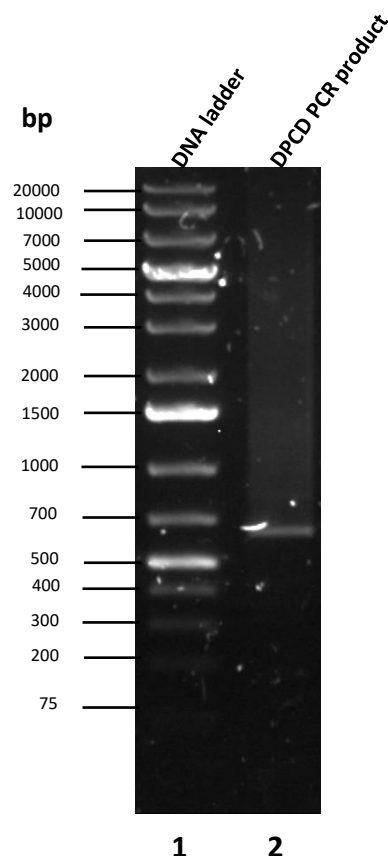
Figure 3.19: Agarose gel electrophoresis of DPCD PCR product for pDEX377 vector ligation. Agarose gel showing amplified DPCD PCR product. Lane 1 contains 1kb plus DNA ladder, Lane 2 contains the purified DPCD PCR product, amplified with Q5 polymerase and A- tailed with DreamTaq polymerase, using Tb927.118900_pDEX377 primers for ligation into pDEX377 plasmid. The major PCR product is seen at 650bp, corresponding to the expected size of the DPCD insert.

As described in earlier section 3.2.2, the DNA fragment encoding the DPCD ORF was ligated into pGEM-T Easy vector and transformed into *E. coli* competent cells. Four recombinant colonies were then streaked to single colony, before plasmid DNA was purified and digested with EcoRI enzyme, to determine whether the DPCD insert is present, Figure 3.20.
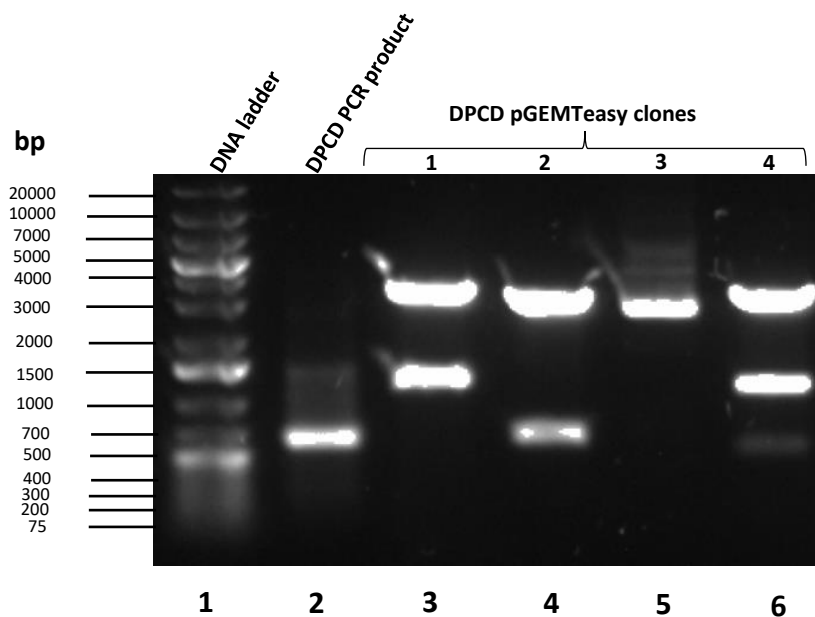
Figure 3.20: Agarose gel electrophoresis of DPCD pGEM-T Easy clones digested with EcoRI. Agarose gel showing DNA constructs digested with EcoRI enzyme. Lane 1 contains 1kb plus DNA ladder. Lanes 2 contains the DPCD PCR product. Lanes 3-6 contain the digested DPCD pGEM-T Easy clones. The DPCD insert is present in DPCD pGEM-T Easy clone 2 (lane 4), at the expected size of 650bp.

As seen in Figure 3.20, out of the four DPCD pGEM-T Easy clones that were selected, only clone 2 (lane 4) contained the pGEM-T Easy fragment at 3000bp and a fragment at the expected size of 650bp, lining up with the fragment present in the DPCD PCR product (lane 2). Clones 1 and 4 (lanes 3 and 6) contain the pGEM-T Easy fragment and another fragment at 1500bp, which is not the expected size of the DPCD insert, while clone 3 (lane 5) contains only the pGEM-T Easy fragment.

To confirm that DPCD pGEM-T Easy clone 2 contained the DPCD ORF without the stop codon, it was sent off for nucleotide sequencing to confirm the DPCD insert sequence as wildtype. The resulting nucleotide sequence provided by 'SourceBioscience' was analysed in the same way as described in section 3.2.2, before being aligned against the *T. brucei*427 DPCD sequence (Tb427_110095600), confirming that the DPCD insert has 100% identity to *Tb*427 DPCD, Figure 3.21. The full nucleotide sequence of the DCPD insert can be seen in appendix figure 3.

```
Clone       MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV  60
Tb427DPCD   MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV  60
            ***********************************************************

Clone       EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM 120
Tb427DPCD   EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM 120
            ***********************************************************

Clone       GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK 180
Tb427DPCD   GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK 180
            ***********************************************************

Clone       KERASLPAKRIDDADSRTSCNQQLERNH 208
Tb427DPCD   KERASLPAKRIDDADSRTSCNQQ----- 203
            **********************
```

-stop codon.

Figure 3.21: Alignment of DPCD insert and Tb427 DPCD sequence. Alignment of the ORF of DPCD insert from pGEM-T Easy clone against the *T. brucei*427 DPCD sequence, using 'ClustalOmega'. This shows 100% identity between the two sequences, indicating the DPCD insert as wildtype. – represents a stop codon, this indicates that the DPCD insert from pGEM-T Easy clone does not contain the stop codon. Instead reading through into the HindIII sequence, as represented by LERHN.

This alignment confirms that the DPCD insert has 100% identity to *Tb*427 DPCD. It also confirms that the DPCD insert does not contain the stop codon, instead reading through into the HindIII site. This DPCD pGEM-T Easy clone was digested with restriction enzymes XhoI and HindIII, to allow for cloning into pDEX377. The digested clone was then run on 0.8% agarose gel and the insert was excised, purified, and then ligated into pDEX377 vector for transformation into *E. coli*. Six clones were selected and streaked to single colony, before plasmid DNA was purified and digested with XhoI and HindIII restriction enzymes, Figure 3.22.
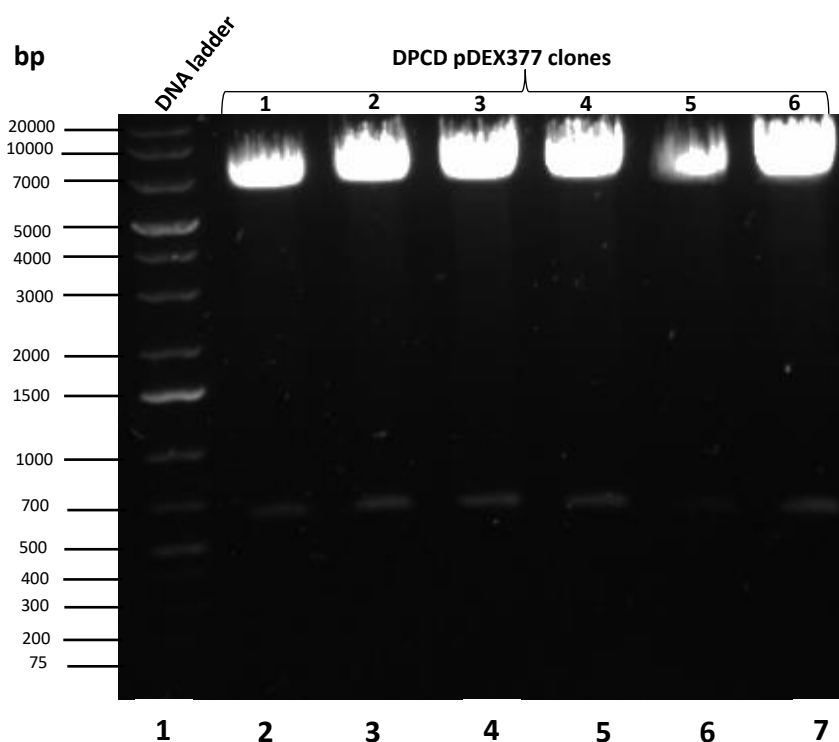
Figure 3.22: Agarose gel electrophoresis of DPCD pDEX377 clones digested with XhoI and HindIII. Agarose gel showing DPCD pDEX377 clones digested with XhoI and HindIII enzymes. Lane 1 contains 1kb plus DNA ladder. Lanes 2-7 contain the digested DPCD pDEX377 clones. The DPCD insert is present in all clones, at the expected size of 650bp.

As seen in Figure 3.22, all DCPD pDEX377 clones possess the pDEX377 vector fragment at 7000bp and the and the DPCD insert at the expected size of 650bp. Two of these clones, 1 and 2, were sent off for nucleotide sequencing to confirm the DPCD insert encodes the DPCD ORF and the

triple c-myc tag epitope, Figure 3.23. The full nucleotide sequence of the DCPD insert can be seen in appendix figure 4. This shows that DPCD pDEX377 clone 2 reads through in frame to the triple c-myc epitope, EQKLISEEDL.

```
Clone       MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV 60
Tb427DPCD   MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV 60
            ***********************************************************

Clone       EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM 120
Tb427DPCD   EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM 120
            ***********************************************************

Clone       GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK 180
Tb427DPCD   GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK 180
            ***********************************************************

Clone       KERASLPAKRIDDADSRTSCNQQLEQKLISEEDLLGSEEQKLISEEDLLGSEEQKLISEE 240
Tb427DPCD   KERASLPAKRIDDADSRTSCNQQ*------------------------------------ 203
            ***********************

Clone       DL*    242
Tb427DPCD   --     203
```

cmyc epitope- EQKLISEEDL
-stop codon

Figure 3.23: Alignment of DPCD insert and Tb427 DPCD sequence. Alignment of the ORF of DPCD insert from DPCD pDEX377 clone against the *T. brucei*427 DPCD sequence, using 'ClustalOmega'. This shows 100% identity between the two sequences, indicating the DPCD insert as wildtype, a swell as that the DPCD ORF in the pDEX377 clones does not possess the stop codon (-) and instead reads through to the triple c-myc epitope (EQKLISEEDL).

DPCD pDEX377 clone 2 was taken forward and linearised using NotI enzyme, Figure 3.24.
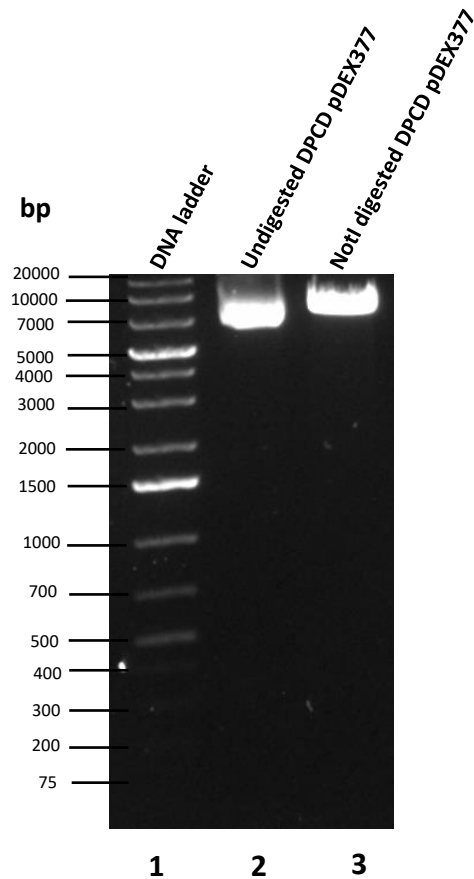
Figure 3.24: Agarose gel electrophoresis of DPCD pDEX377 clone digested with NotI. Agarose gel showing DPCD pDEX377 clone linearised with NotI enzyme. Lane 1 contains 1kb plus DNA ladder. Lane 2 contains the undigested DPCD pDEX377 clone and lane 3 contains the DPCD pDEX377 clone digested with NotI. The linearisation of the construct indicated by the linearised fragment running at a larger size of 8000bp compared to the undigested clone at 7000bp.

The linearised DNA was precipitated and transfected into SP9 procyclic *T. brucei* cells. To aim to understand the role of DPCD in T. brucei, these transfected cells were engineered to overexpress the DPCD protein, by induction with doxycycline. However, this was unsuccessful with no immunofluorescent or western blotting evidence to suggest that the DPCD protein was being overexpressed. It is unlikely that this was caused by fault in the DPCD insert present in the DPCD pDEX377 construct as nucleotide sequencing confirmed that this had 100% identity to Treu427 DPCD and that it read through in frame into the triple c-myc epitope.

Section 3.5: Biochemical analysis of the DPCD protein

3.5.1: Prediction of DPCD protein function using Phyre2

To predict the potential function of the DPCD protein, the *T. brucei* DPCD sequence was inputted into Phyre2 (Kelley et al, 2015). The top three predictions from this search are seen in Figure 3.25.
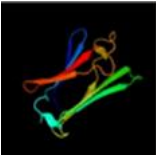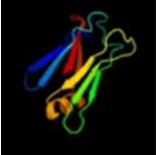
3D model	Confidence	% Identity	Description
1	93.2	18	**PDB header:**chaperone **Chain: A: PDB Molecule:**heat shock protein-related protein; **PDBTitle:** hsp17.7 from deinococcus radiodurans **PDB Entry:** PDBe RCSB PDBj
2	84.4	15	**PDB header:**chaperone **Chain: D: PDB Molecule:**pih1 domain-containing protein 1; **PDBTitle:** crystal structure of the human rpap3(tpr2)-pih1d1(cs) complex **PDB Entry:** PDBe RCSB PDBj
3	83.8	23	**PDB header:**chaperone **Chain: L: PDB Molecule:**25.3 kda heat shock protein, chloroplastic; **PDBTitle:** hsp21 dodecamer, structural model based on cryo-em and homology2 modelling **PDB Entry:** PDBe RCSB PDBj

Figure 3.25: Phyre2 prediction of DPCD protein folding. Top three predictions of protein folding of *Tb*DPCD, including (1) heat shock protein-related protein, (2) pih1 domain-containing protein 1 and (3) 25.3kDA heat shock protein. The confidence, % identity and a predicted 3D model of protein folding seen. Figure edited from (Kelley et al., 2015).

As seen in Figure 3.25, all three predicted structures are associated with proteins that have chaperone activity. These include heat-shock related protein 17.7 and a 25.3kDA heat shock protein HSP21. Heat shock proteins are generated when cells are exposed to high temperatures and include HSP100 (100 kDa), HSP90 (90 kDa), HSP70 (70 kDa), HSP60 (60 kDa), and small HSP (20–30 kDa) families. These are often combined into multi-subunit complexes which act as chaperones to aid in protein folding and stability under non-physiological temperatures (Waters et al., 1996). The third structural prediction from the Phyre2 analysis was the protein PIH1D1, which is a part of the R2TP complex involved in the cytoplasmic pre-assembly of dynein arms in

ciliated cells. It is thought that the R2TP complex also acts as a HSP90 co-chaperone (Fabczak and Osinka, 2019).

In summary, the Phyre2 protein folding predictions indicate that DPCD has a similar structure to proteins that have chaperone activity under heat stress conditions. This raised the possibility that suggests that DPCD functions as a chaperone or co-chaperone involved in the heat shock response.

3.5.2: Biochemical analysis of *Tb*DPCD protein

To investigate whether the *Tb*DPCD protein acts as a chaperone, as predicted by Phyre2, the protein was expressed in Rosetta(DE3)pLYsS *E. coli* competent cells and purified for functional analysis.

The *Tb*DPCD open reading frame was amplified by PCR using the hi-fidelity polymerase, purified and then A-tailed using DreamTaq polymerase, for the same reason mentioned in previous section 3.2. The product was then purified and analysed by agarose gel electrophoresis, which showed the PCR product was of the expected size of 650bp, Figure 3.26.
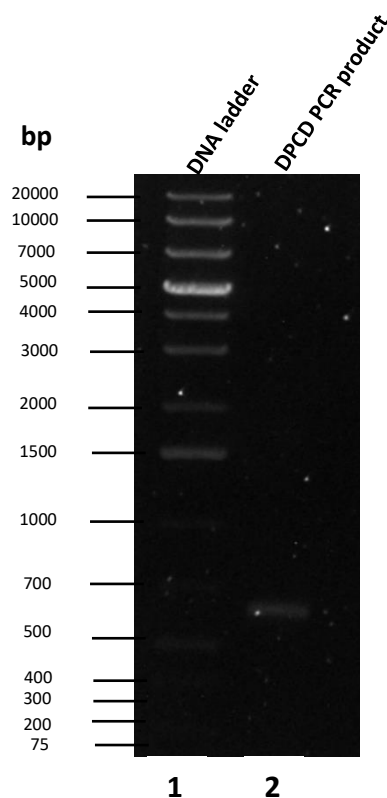
Figure 3.26: Agarose gel electrophoresis of *Tb*DPCD PCR product generated to enable protein expression using the pET28a vector. Agarose gel showing amplified DPCD PCR product. Lane 1 contains 1kb plus DNA ladder, Lane 2 contains the purified DPCD PCR product, amplified with Q5 polymerase and A- tailed with DreamTaq polymerase, using Tb927.118900_pET28a primers for ligation into pET28a plasmid. The major PCR product is seen at 650bp, corresponding to the expected size of the DPCD insert.

In the same way as described in section 3.2.2, the DNA fragment encoding the DPCD ORF was then ligated into pGEM-T Easy vector and transformed into competent cells for blue-white screening. Purified plasmid DNA from three clones was digested with EcoRI, to determine presence of the DPCD insert, Figure 3.27.
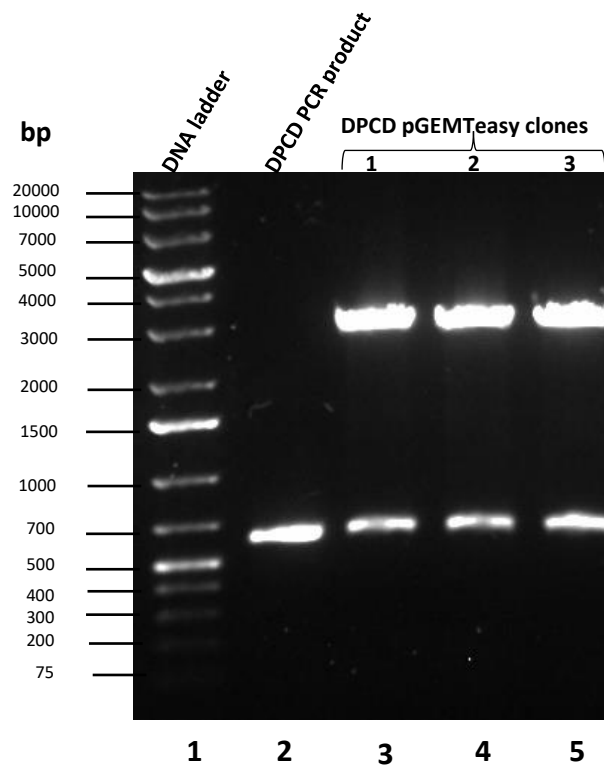
Figure 3.27: Agarose gel electrophoresis of DPCD pGEM-T Easy clones digested with EcoRI. Agarose gel showing DNA constructs digested with EcoRI enzyme. Lane 1 contains 1kb plus DNA ladder. Lane 2 contains the DPCD PCR product. Lanes 3-5 contain the digested DPCD pGEM-T Easy clones. The DPCD insert is present in all three DPCD pGEM-T Easy clones, at the expected size of 650bp.

As seen in Figure 3.27, all three of the putative *Tb*DPCD pGEM-T Easy clones, contained a DNA fragment of 3kb corresponding to the expected size of the pGEM-T Easy vector and a DNA fragment of 650bp, which aligns with the size of the purified DPCD PCR product (lane 2).

However, to confirm that these clones contained the *Tb*DPCD ORF, all three were sent for nucleotide sequencing, to identify the insert as DPCD and confirm that they were wildtype with respect to sequence. The resulting nucleotide sequences provided by 'Source Bioscience' were conceptually translated into amino acid sequence and for alignment against the *T. brucei* 427 DPCD sequence (Tb427_110095600). This alignment confirmed that the DPCD insert in only one of the pGEM-T Easy clones (clone 2) had 100% identity to *Tb*427 DPCD, Figure 3.28. The full nucleotide sequence of the DCPD insert can be seen in appendix figure 5.

```
Clone       MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV    60
Tb427DPCD   MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGGFSTWEV    60
            ************************************************************

Clone       EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM   120
Tb427DPCD   EVGNEASTRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAVEHSKPTDM   120
            ************************************************************

Clone       GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK   180
Tb427DPCD   GEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVVLAAESAAK   180
            ************************************************************

Clone       KERASLPAKRIDDADSRTSCNQQ 203
Tb427DPCD   KERASLPAKRIDDADSRTSCNQQ 203
            ***********************
```

Figure 3.28: Alignment of DPCD insert and *Tb*427 DPCD sequence. Alignment of the ORF of DPCD insert from pGEM-T Easy clone against the *T. brucei*427 DPCD sequence, using 'ClustalOmega'. This shows 100% identity between the two sequences, indicating the DPCD insert as wildtype.

This DPCD pGEM-T Easy clone was digested with restriction enzymes NdeI and HindIII, to allow for cloning into pET28a. The digested clone was resolved on a 0.8% agarose gel and the insert excised, purified, and then ligated into pET28a vector for transformation into *E. coli*. Following

successful transformation, four colonies were selected, and plasmid DNA prepared and digested with NdeI and HindIII restriction enzymes, Figure 3.29.
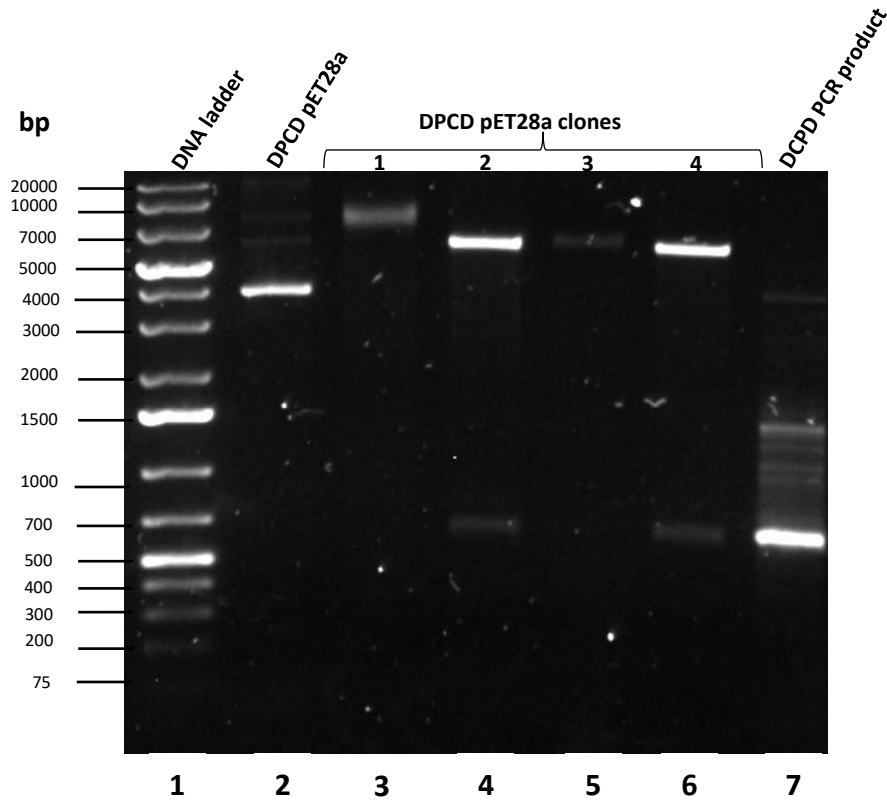
Figure 3.29: Agarose gel electrophoresis of DPCD pET28a clones digested with NdeI and HindIII. Agarose gel showing DPCD pET28A clones digested with NdeI and HindIII enzymes. Lane 1 contains 1kb plus DNA ladder. Lanes 2 contains undigested DPCD pET28a, Lanes 3-6 contain the digested DPCD pEt28a clones. Lane 7 contains 3μL of the DPCD PCR product. The DPCD insert is present in clones 2 and 4 (lane 4 and 6), at the expected size of 650bp.

As seen in Figure 3.29, only DPCD pET28a clones 2 and 4 contain the pET28a fragment at 7000bp and the DPCD insert at the expected size of 650bp, which aligns with the size of DPCD PCR product. DPCD pET28a clone 1, does not contain the DPCD insert fragment at, and the pET28a fragment is larger than expected at around 10,000bp, suggesting that this clone was not digested by restriction enzymes due to absence of one or both cloning sites. Clone 3 also does not contain the DPCD insert fragment but does contain the pET28a fragment at 7000bp.

3.5.3: Expression and purification of *Tb*DPCD protein for functional analysis

In order to purify the *Tb*DPCD protein, it was first expressed in Rosetta(DE3)pLYsS *E. coli*. To do this, the two DPCD pET28a clones containing the DPCD insert, clones 2 and 4, were transformed into the Rosetta(DE3)pLYsS *E. coli* competent cells and grown on kanamycin/chloramphenicol plates.

To determine whether the transformed Rosetta(DE3)pLYsS *E. coli* cells were expressing *Tb*DPCD protein, a test expression was carried out on clone 2 at 37°C, with protein samples being prepared at 0, 0.5, 1 and 2 hours after induction with IPTG. These samples were run on 12% SDS-PAGE gel, showing total protein content Figure 3.30 (A), and probed with an anti-HIS primary antibody, Figure 3.30 (B).
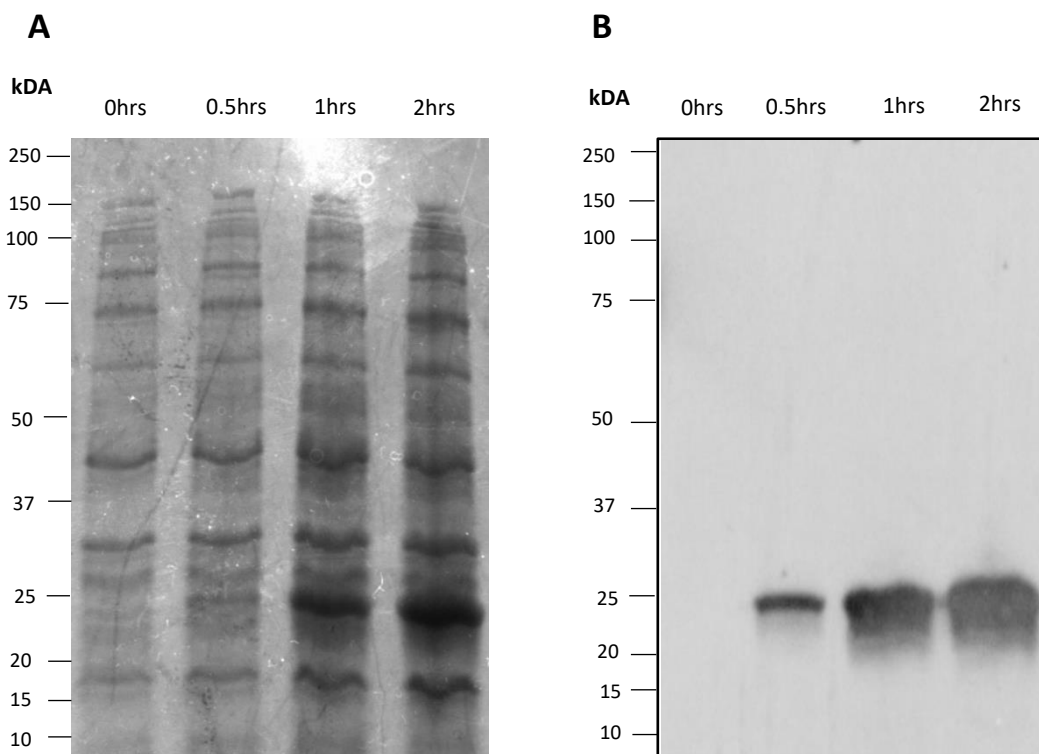
Figure 3.30: SDS-PAGE gel and western blot showing *Tb*DPCD expression in Rosetta(DE3)pLYsS *E. coli.* (A) SDS-PAGE gel of DPCD pET28a Rosetta *E. coli* clones 0, 0.5, 1 and 2 after induction with IPTG. Gel stained with Instant Blue Coomassie stain. Strongest bands seen in 1hr and 2hr samples at approx. 23kDA, which is the expected size of the *Tb*DPCD protein. **(B)** Western blot

of the same DPCD pET28a Rosetta *E. coli* clones 0, 0.5, 1 and 2 after induction with IPTG. The membrane was probed with mouse anti-HIS primary antibody (1:4000) and Rabbit-anti-mouse secondary antibody conjugated to HRP (1:8000). Bands seen in 0.5-, 1- and 2-hour samples at expected size of 23kDA, showing increasing *Tb*DPCD expression with time.

As seen in Figure 3.30, *Tb*DPCD is being expressed in the *E. coli* cells after induction with IPTG, with bands seen in both the SDS-PAGE gel and western blot at the expected size of 23kDA.

The remainder of this 50ml culture inoculated with DPCD pET28a *E. coli* cells was pelleted by centrifugation and taken forward for protein purification. To do this Ni-NTA affinity chromatography was used to purify the recombinant DPCD protein carrying a His tag. Samples of the protein lysate and flow through were taken before washing with increased concentration of imidazole. Here, wash 1, wash 2, wash 3 and elution samples were taken, with the eluted sample being the purest. All these protein samples were run on SDS-PAGE gel and stained with Instant blue Coomassie stain, Figure 3.31.
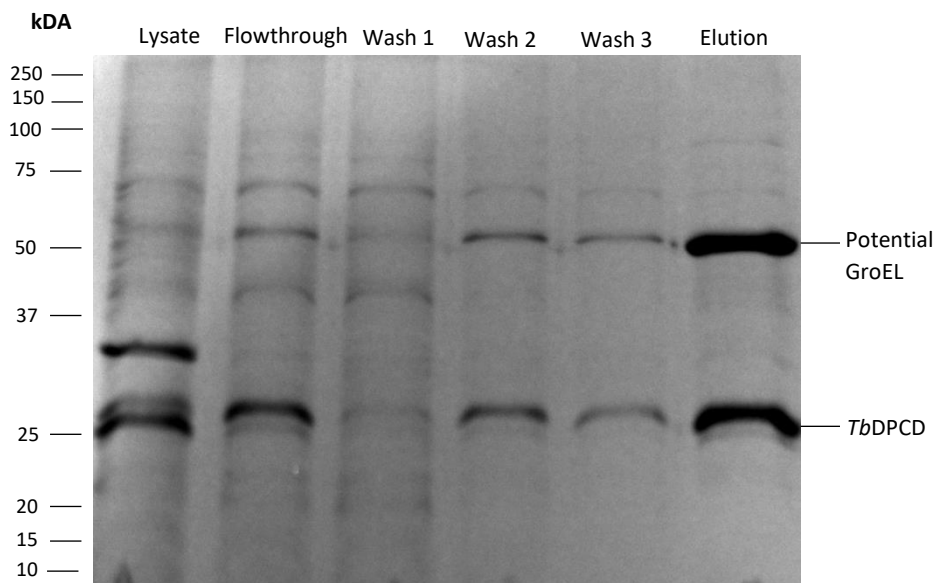
Figure 3.31: SDS-PAGE gel showing *Tb*DPCD protein purification samples. SDS-PAGE gel of *Tb*DPCD protein purification lysate, flowthrough, wash 1, wash 2, wash 3 and elution samples. Gel stained with Instant Blue Coomassie stain. Strongest bands seen in the elution sample at approx. 25kDA, which is the expected size of the *Tb*DPCD protein plus the HIS tag, and unexpectedly at approx. 55kDA. This fits the expected size of an *E.coli* GroEL chaperone protein.

As seen in Figure 3.31, two bands of protein are seen in the elution sample at 25kDA, which is likely the *Tb*DPCD protein, and at 50kDA, which may be an *E. coli* GroEL chaperone protein. Chaperone proteins, like GroEL, often co-purify with other proteins if that protein is partially folded, in this case DPCD, by binding to the misfolded construct (Katayama et al., 2008).

To investigate this and determine whether the band at 55kDA was a GroEL protein, the purification samples were blotted with an anti-GroEL primary antibody. The membrane was also probed the anti-HIS primary antibody, to detect the His tagged *Tb*DPCD protein, Figure 3.32. This experiment confirmed the eluted fraction contained both *Tb*DPCD and the *E.coli* GroEL chaperone protein.
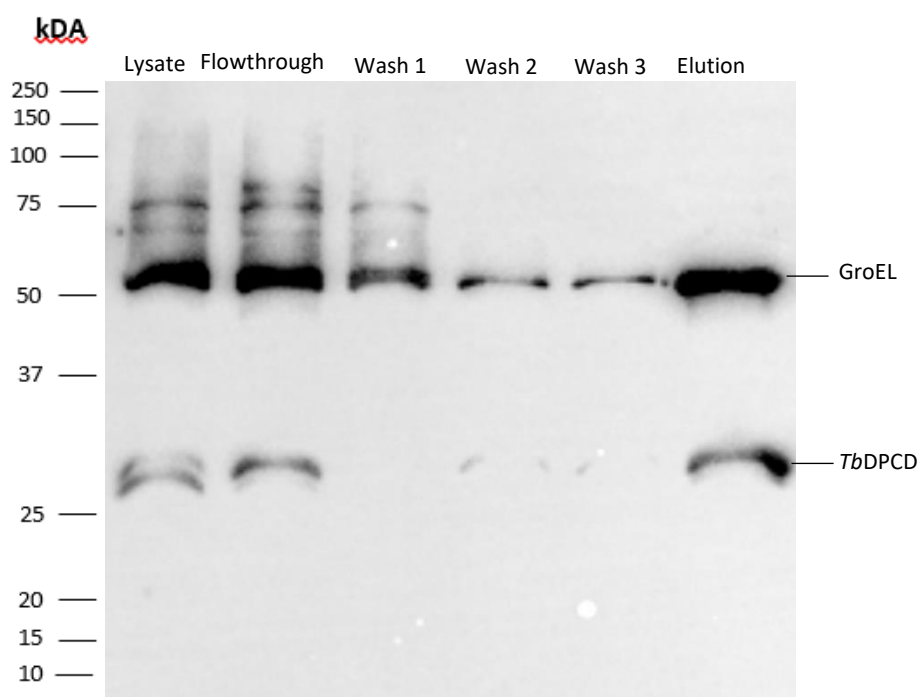
Figure 3.32: Western blot showing co-purification of the *Tb*DPCD protein and *E.coli* GroEL chaperone. Western blot showing the *Tb*DPCD protein purification lysate, flowthrough, wash 1, wash 2, wash 3 and elution samples. The membrane was probed with mouse anti-GroEL primary antibody (1:2000) and anti- mouse 6X His tag antibody (1:4000) followed by a rabbit-anti-mouse secondary antibody conjugated to HRP (1:8000). Bands seen in elution sample at 25kDA, which is the expected size of the *Tb*DPCD protein plus the HIS tag, and at approx. 55kDA, proving that the *Tb*DPCD protein is being copurified with an *E.coli* GroEL chaperone protein.

As the potential function of DPCD predicted by Phyre2 was a chaperone, section 3.5.1, the copurification of GroEL with *Tb*DPCD was problematic for the planned biochemical interrogation of the *Tb*DPCD protein function. Consequently, the *Tb*DPCD protein induction experiment was repeated at a lower temperature of 30°C, to prevent GroEL activation. The DPCD protein in these cells was then purified using the same method as mentioned above, but instead washing in 35% glycerol in NI-NTA buffer plus ATP. This was used as the ATP acts as a cofactor, enabling the GroEL to refold the partially unfolded target protein (DPCD). Additionally, the glycerol acts as an osmolyte to aid the GroEL in protein refolding in the absence of its partner, GroES (Katayama et al., 2008). The lysate, flow through, wash 1, wash 2, wash 3 and elution samples from this purification were resolved on an SDS-PAGE gel, Figure 3.33 (A) and probed with anti-GroEL primary antibody to determine whether the GroEL protein continued to be co-purified with the *Tb*DPCD protein, Figure 3.33 (B).
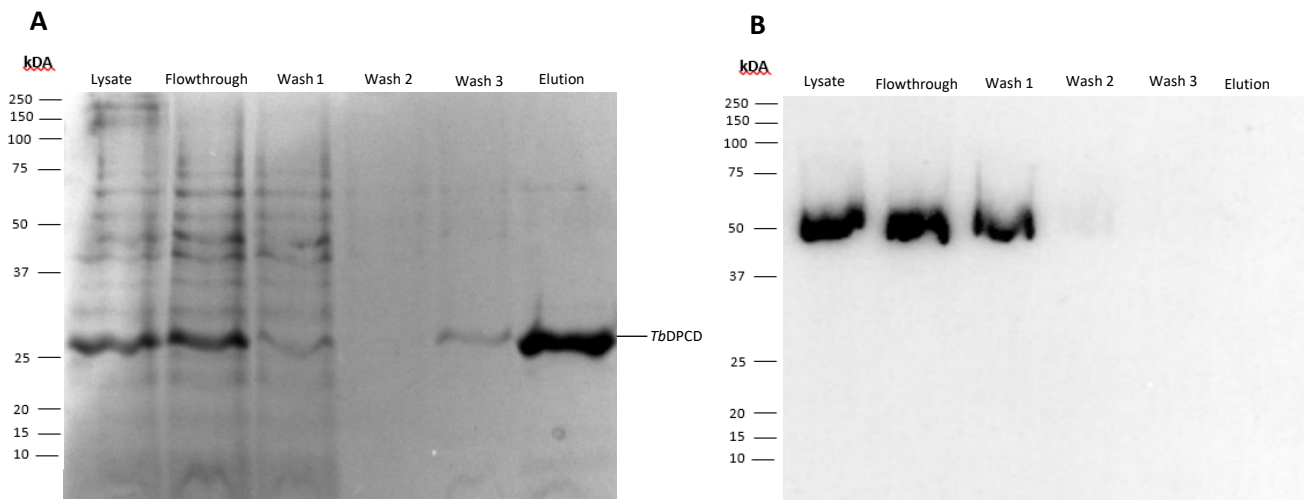
Figure 3.33: SDS-PAGE gel and western blot showing *Tb*DPCD purification and absence of GroEL copurification. (**A**) SDS-PAGE gel of *Tb*DPCD protein purification lysate, flowthrough, wash 1, wash 2, wash 3 and elution samples. Gel stained with Instant Blue Coomassie stain. Strongest bands seen in the elution sample at approx. 25kDA, which is the expected size of the *Tb*DPCD protein plus the HIS tag. No band is seen at 55kDA, indicating that the *E. coli* GroEL chaperone protein is no longer being purified. (**B**) Western blot of the same *Tb*DPCD protein purification samples. The membrane was probed with mouse anti-GroEL primary antibody (1:2000) and Rabbit-anti-mouse secondary antibody conjugated to HRP (1:8000). Bands present in the lysate,

flowthrough and wash 1 samples at 55kDA, indicate that the GroEL protein was present in the samples. However, these bands are no longer seen in the elution sample, showing that the GroEL protein in not present in this sample.

As seen in Figure 3.33 (A), only one band at 25kDA is present in the elution sample, indicating that the *Tb*DPCD protein was now being purified without contamination by *E. coli* GroEL chaperone protein. Furthermore, to confirm this an immunoblot was probe with the anti-GroEL antibody which demonstrated that the GroEL protein was still present in the lysate, flowthrough and wash 1 samples. However, the GroEL was not present in the wash 2, wash 3 and elution samples, indicating that the addition of the glycerol and ATP prevented the co-purification of the GroEL with the *Tb*DPCD protein, Figure 3.33 (B).

3.5.4: Thermal aggregation assay to investigate potential *Tb*DPCD chaperone function

To investigate whether *Tb*DPCD had the potential to function as a chaperone involved in the heat shock response, as predicted by Phyre2, section 3.5.1, a thermal aggregation assay using citrate synthase as a substrate was carried out (Haslbeck and Buchner, 2015). This assay measured the ability of DPCD to prevent the thermal aggregation of citrate synthase. To do this, the sample containing the purified *Tb*DPCD protein was dialysed into 40mM HEPES buffer and the citrate synthase protein, gained from commercially available citrate synthase, dialysed into 50mM TRIS + 0.5M EDTA. The concentration of each protein was determined after dialysis by measuring absorbance at 290nm and using the predicted extinction coefficients. The thermal aggregation assay was then carried out in a plate reader at 43°C, measuring the absorbance at 390nm of 4.8µM- 0.15µM DPCD plus 0.6µM citrate synthase, over 45 minutes. Before the addition of the citrate synthase, the solutions of DPCD were incubated at 43°C for 5 minutes. The absorbance of 0.6µM citric synthase alone was also measured, Figure 3.34. If the *Tb*DPCD protein possessed chaperone activity it would be expected that there would be less of an increase in absorbance in solutions containing the *Tb*DPCD protein, with a negative correlation between the increase in absorbance and concentration of *Tb*DPCD.

However, as seen in Figure 3.34, it is likely that *Tb*DPCD does not have chaperone activity, at least not in the context of heat response. As the addition of the *Tb*DPCD protein does not supress the thermal aggregation of the citrate synthase, with absorbance increasing with time at all *Tb*DPCD concentrations. This is demonstrated by the positive correlation between the amount of *Tb*DPCD in the reaction and the absorbance at 0 minutes, this may be explained by the incubation of the *Tb*DPCD solutions prior to addition of the citrate synthase. Also, when *Tb*DPCD was at 4-fold higher the amount of the citrate synthase, 2400nM, the point at which the absorbance reaches its maximum is delayed to around 30 minutes and starts to decrease after some time. The size of this change in absorbance is too high to be due to citrate synthase aggregation alone. A similar effect is seen when *Tb*DPCD was at 8-fold higher than the amount of the citrate synthase, 4800nM. However, the maximum absorbance is reached earlier at around 17 minutes and returns to a low level. This may indicate foldase chaperone activity on the *Tb*DPCD protein itself, but additional experimentation is needed to confirm this. Overall, from these results it appears that *Tb*DPCD does not protect citrate synthase from aggregation suggesting it does not function as a holdase. The confidence of the function of this assay is increased as in absence of *Tb*DPCD, 0nM, the time at which the absorbance reaches its maximum is the same, around 20 minutes, as in the published protocol (Haslbeck and Buchner, 2015).
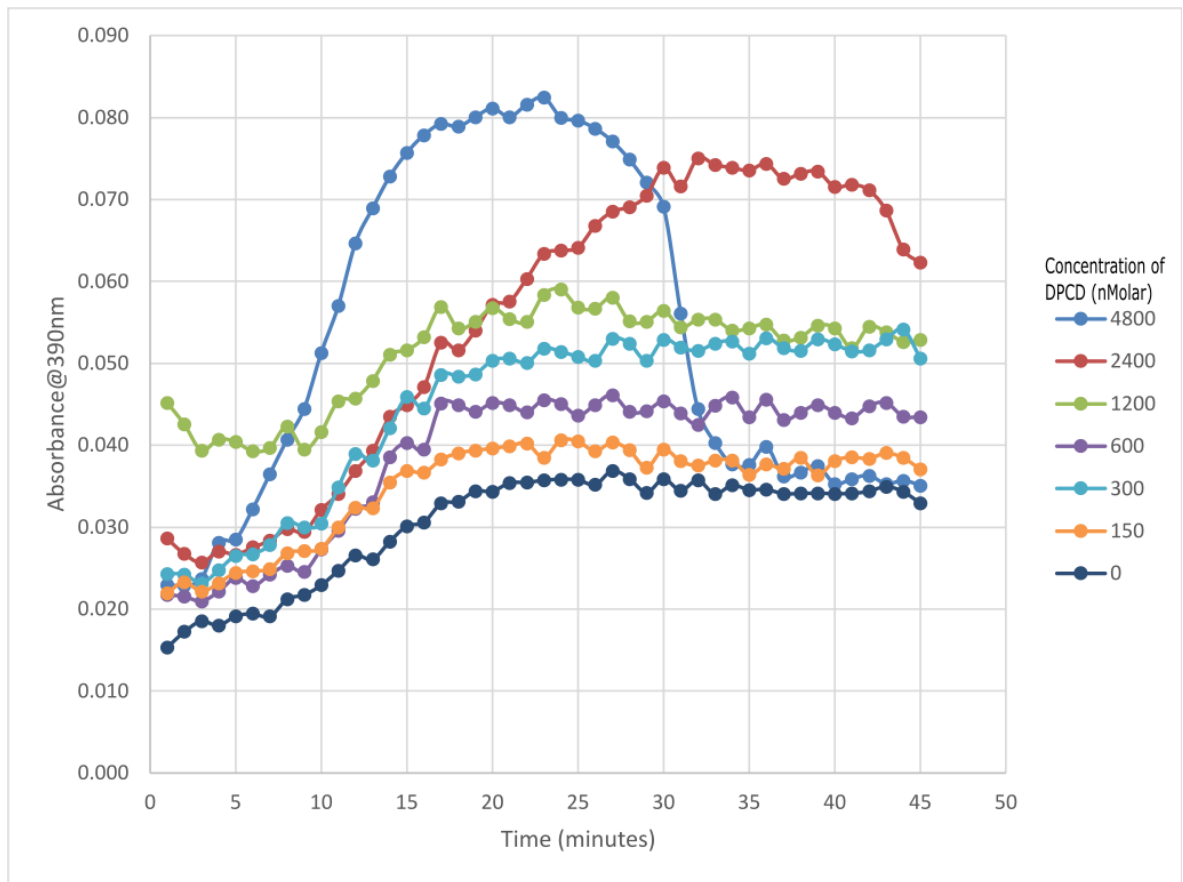
Figure 3.34: Thermal aggregation assay measuring suppression of citric synthase aggregation by *Tb*DPCD. Scatter plot showing the thermal aggregation of citric synthase at 43°C across 45 minutes, at increasing concentration of DPCD protein. Absorbance measured at 390nm, for 0-4.8μM DPCD and 0.6μM citric synthase. 0μM DPCD demonstrates spontaneous aggregation of citric synthase at 43°C. This shows that *Tb*DPCD does not possess chaperone activity, in the context of heat response, as the aggregation of citric synthase is not supressed, with absorbance increasing at all concentration of DPCD.

To investigate the effect of *Tb*DPCD on the thermal aggregation of citrate synthase in more detail, the absorbance across 45 minutes at 43°C, at only 0nM DPCD and the conditions where DPCD is in molar excess, 1200, 2400 and 4800nM were plotted. This was done by subtracting the absorbance measured at 1 minute from all subsequent timepoints for each condition, see Figure 3.35. Relative to citrate synthase alone, DPCD at 8-fold and 4-fold higher than the amount of the citrate synthase, the absorbance increased at a greater level. This suggests that DPCD

does not suppress the thermal aggregation of citrate synthase. However, DPCD at 2-fold the amount of the citrate synthase, 1200nM, the absorbance increased at a lower level compared to citrate synthase alone. This suggests that the DPCD protein has holdase activity at this concentration.
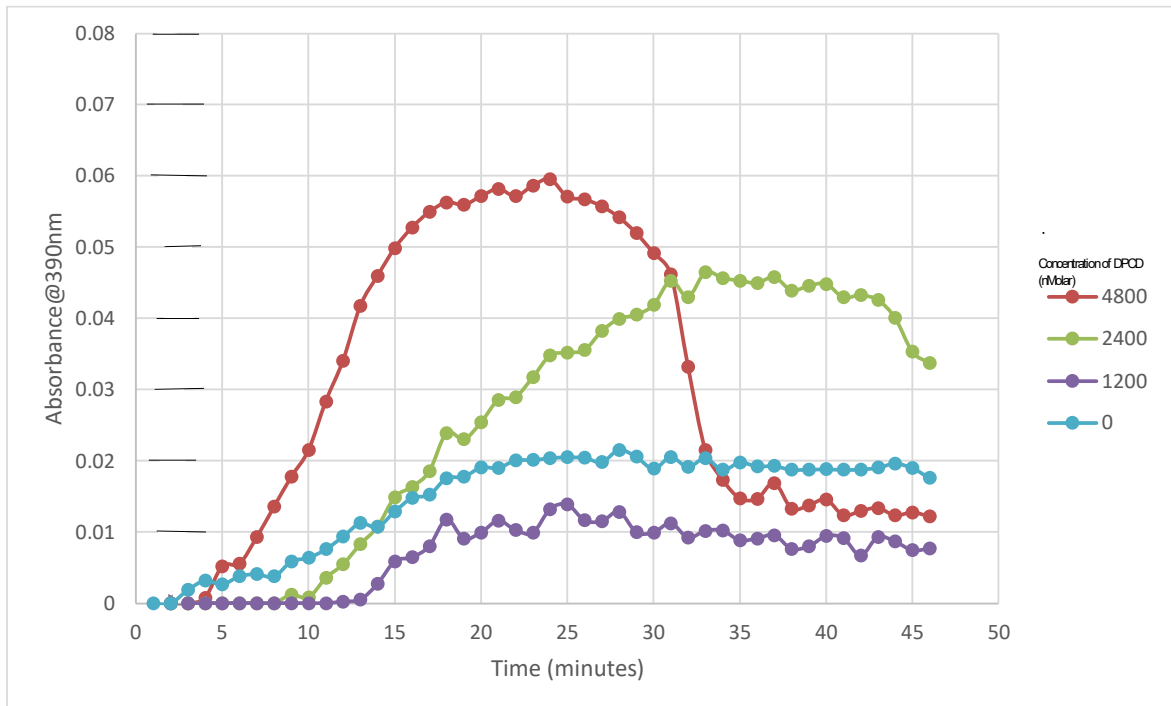
Figure 3.35: Thermal aggregation assay measuring suppression of citric synthase aggregation by *Tb*DPCD in molar excess. Scatter plot showing the thermal aggregation of citric synthase at 43°C across 45 minutes, at 4.8, 2.4, 1.2 and 0µM DPCD, demonstrating spontaneous aggregation of citric synthase at 43°C. This shows that *Tb*DPCD at 8-fold and 4-fold the concentration of citrate synthase does not possess chaperone activity. However, *Tb*DPCD at 2-fold the concentration of citrate synthase does show some chaperone activity, with absorbance increasing less compared to citrate synthase alone.

In conclusion, the biochemical analysis of the *Tb*DPCD protein carried out in this section, indicates *Tb*DPCD is unlikely to possess chaperone activity. Although holdase activity is suggested in Figure 3.35, this only occurs at a specific concentration of *Tb*DPCD and at a low level. This contradicts the prediction from the bioinformatics interrogation of the *Tb*DPCD

protein by Phyre2, however, the biochemical analysis carried out in this study suggests the *Tb*DPCD protein itself may not be stable at high temperatures, which may have affected the results of the thermal aggregation assay.

Section 3.6: Prediction of *Tb*DPCD protein function using Alphafold

As the results gained from the biochemical analysis of *Tb*DPCD did not support the functional prediction provide by Phyre2, the function of DPCD still remains unknown. Throughout the course of this project, many protein structures became available on the AlphaFold protein database (Jumper et al., 2021; Varadi et al., 2022). This included the *T. brucei brucei* DPCD structure. This has allowed the potential function of DPCD to be further investigated. To do this, the Protein Data Bank (PDB) file of the *T. brucei brucei* DPCD AlphaFold predicted structure was inputted into three different databases, to predict protein function. These included DALI, which takes 3D coordinates of protein structures and returns similar structures based on distance matrix comparison (Holm, 2022) and FATCAT, which carries out flexible structure alignment by minimizing the number of twists in the reference protein (Ye and Godzik, 2003). For both databases the structure was searched against the PBD database, PDB50 for DALI and PDB40 for FATCAT. These were chosen to reduce the redundancy in the output, while maintaining a high sequence identity. A newer database, Foldseek was also used, which compares protein structures based on tertiary interactions (Kempen et al., 2022). The only search option against the PDB database here was PDB100, which increases the redundancy in the output. The top hit from each of these databases was taken and compared, Table 3.1.

Table 3.1: Top hits against Alphafold DPCD *T. brucei brucei* structure in three protein databases. Table of top hits against Alphafold DPCD *T. brucei brucei* structure in DALI, Foldseek and FATCAT databases.

Database	Identity of top hit
DALI	Suppressor of G2 allele of SKP1 homolog
Foldseek	Suppressor of G2 allele of SKP1 homolog
FATCAT	Low molecular weight heat shock protein

As seen in Table 3.1, the top hit in DALI and Foldseek is a Suppressor of G2 allele of SKP1 homolog. This is also known as SGT1 and is known to be involved in protein stabilization by acting as a cochaperone of heat shock protein 90 (HSP90) chaperone. This is a key regulator of proteostasis under physiological and stress conditions in eukaryotic cells and has been linked to overexpression of oncoproteins in cancer cells and hypersensitive cell death in *Capsicum annuum* plant cells (Kim et al., 2015; Ogi et al., 2015). While the top hit from FATCAT is a low molecular weight heat shock protein, which as mentioned in section 3.5.1, are a type of molecular chaperone that bind to partially denatured proteins preventing protein inactivation and aggregation (Waters et al., 1996). These results support the bioinformatics-based prediction that DPCD may have a chaperone function. To look further into this, the top 50 this from all three of these databases were compared, determining which proteins are present in all three, Figure 3.36. For the DALI results, hits were restricted to those above a Z score of 2, while FATCAT hits were restricted to P value below 0.05. The Foldseek hits all had an e-value above e^{-10} which makes these hits less significant.
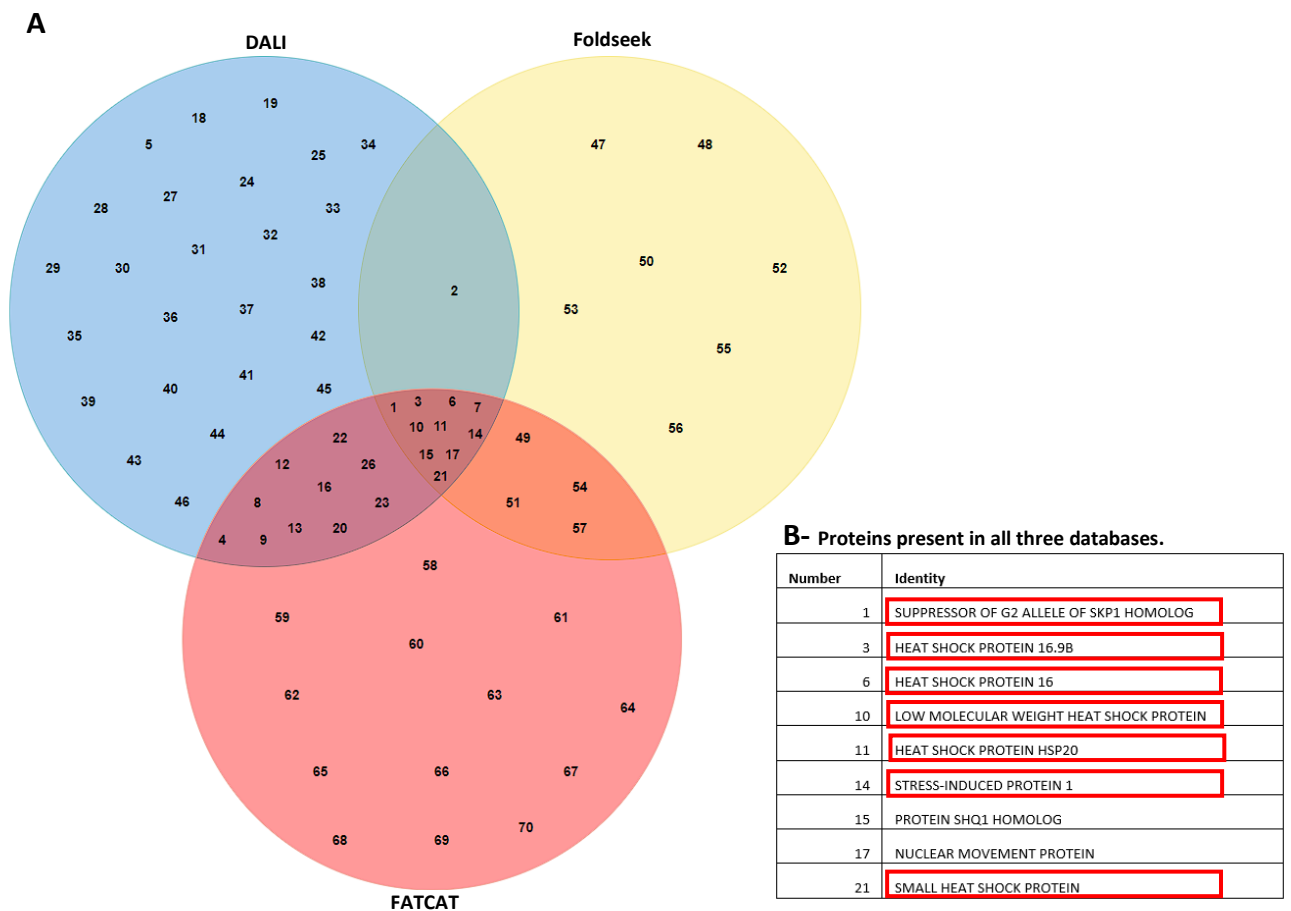
Figure 3.36: Venn diagram of top 50 protein hits against Alphafold DPCD *T. brucei brucei* structure present in DALI, FATCAT and Foldseek databases. (**A**) Shows Venn diagram of top 50 protein hits against the Alphafold DPCD *T. brucei brucei* structure in DALI (blue), FATCAT (red) and Foldseek (yellow). Numbers represent different proteins, refer to appendix table 3, for full list of protein identifies. (**B**) Table of proteins present in all three databases. Proteins with chaperone or co-chaperone activity outlined in red.

As seen in Figure 3.36 (B), nine different proteins are present in the top 50 this against the Alphafold DPCD *T. brucei brucei* structure of all three of the protein databases, seven of which possess chaperone or co-chaperone activity. This includes the Suppressor of G2 allele of SKP1 homolog (SGT1) and low molecular weight heat shock protein, both of which are described above. In addition, four of the other proteins are forms of heat shock proteins with chaperone activity, including heat shock protein 16.9B, heat shock protein 16, heat shock protein 20 and small heat shock protein. Stress induced protein 1 is also present in all three databases, this is a co-chaperone of heat shock proteins HSP70 and HSP90 (Tsai et al., 2018). Two other proteins,

protein SHQ1 homolog and nuclear movement protein, are also found in all three databases. These have different functions compared to the previously described proteins, with SHQ1 being involved of the biogenesis of H/ACA ribonucleoproteins, which process ribosomal RNAs and stabilizes telomerase and nuclear movement protein which is thought to have functions in mitosis and cytokinesis, in part by regulating microtubule organization (Aumais et al., 2003; Slieman et al., 2022). As a large majority of these proteins possess some form of chaperone activity, this further suggests that DPCD has a function as a chaperone or cochaperone, likely involved in heat shock.

In addition, the Foldseek database was also used to search for homologues against the Alphafold DPCD *T. brucei brucei* structure in the Alphafold database itself, see appendix table 4 for full list. This search was against 50% of the Alphafold database, with all hits having evalue below e^{-10}. All but two of the top 50 this, were DPCD homologues present in other organisms, including *Trypanosoma vivax* and *Bodo saltans.* This suggest that DPCD has a unique fold structure among the structure in the Alphafold database.

In summary, the results from this Alphafold analysis suggests that *T. brucei brucei* DPCD has a similar protein structure to chaperones, or co-chaperones, involved in heat shock. This is like what was predicted by Phyre2 but does not fit with the results gained from the thermal aggregation assay, section 3.5.4. It was also determined that DPCD has a unique fold structure.

Section 4: Discussion

DPCD is a novel gene candidate for the disease PCD and is hypothesised to have a role in ciliogenesis, however the precise function of this protein is unclear. The aim of this present study was to investigate whether DPCD had a role in the formation and function of motile flagella using *T. brucei* as a model organism.

To investigate the occurrence of DPCD in eukaryotic organisms, the human DPCD protein was used in a 'BlastP' search to determine the phylogenetic occurrence of this protein in eukaryotes. This confirmed that DPCD homologues are only found in ciliated eukaryotes with motile cilia, but this is not absolute. The lack of DPCD in these ciliated eukaryotes is unclear but may be due to their evolutionary history. This includes certain members of the SAR eukaryotic superclade. The early branching of excavates and the resulting evolution of the SAR superclade is supported

by the presence of a rotation of a helical central pair in the cilia of both SAR and Plantae. This may explain the lack of DPCD in some SAR superclade members, along with members of the Plantae superclade, due to the evolution of this synapomorphy in their most recent common ancestor (Mitchell, 2017). As it was determined that these organisms possess many of the other ciliary proteins, whether these ciliary proteins have a role in ciliogenesis in these organisms could be determined. In addition, as DPCD is conserved only in organisms with motile cilia, it could be hypothesised that the protein has a role in the function or formation of structures found only in motile cilia, including the axonemal dynein arms. During the writing of this project a publication was released which supports this theory. This work carried out by Morias and colleagues (Morias et al., 2022), investigated the role of DPCD in complex with RUVBL1/RUVBL2 ATPases in the co-chaperoning function of R2TP complexes. As mentioned previously, the R2TP complex is predicted to be involved in the cytoplasmic pre-assembly of dynein arms as a heat shock protein 70 and 90 co-chaperone (Fabczak and Osinka, 2019). DPCD was identified as a new R1R2 partner in vivo, interacting directly with R1 and R2 proteins to form a specific R1R2D complex. It was also found to associate with other R2TP components or cofactors including RPAP3 and PIH1D1 (Morias et al., 2022). The full list of identified DPCD-interacting proteins can be seen in Table 3.2. This corresponds to the results from the Phyre2 analysis, where *T. brucei brucei* DPCD protein folding was predicted be like that of the PIH1D1 protein. This suggests that DPCD has a role similar to this protein in the preassembly of the axonemal dynein arms of motile cilia. To confirm what was determined by Morias and colleagues, the role of DPCD as a component of the R2TP complex could be investigated in *T. brucei* using BioID approaches and co-immunoprecipitation techniques to confirm the protein interactions. However, this still does not explain why DPCD is not present in all ciliated organisms with motile cilia.

Table 3.2: DPCD-interacting proteins identified by Morias et al., 2022. Table of R2TP complex components that were found to interact with DPCD. Sourced from: Morias et al., 2022.

Protein	Function
RUVBL1 (R1)	ATPase involved in co-chaperoning function within R2TP or R2TP-like machineries
RUVBL2 (R2)	ATPase involved in co-chaperoning function within R2TP or R2TP-like machineries
RPAP3	R2TP complex core protein
PIHID1	R2TP complex core protein
WDR92	R2TP/R2TP-like/PAQosome component

POLR2E	R2TP/R2TP-like/PAQosome component
ZNHIT2	R2TP/R2TP-like/PAQosome component
EFTUD2	R2TP/R2TP-like/PAQosome component
HSP90/70	Chaperone acting alongside the R2TP complex

Cytoplasmic localisation of proteins does not preclude a role in ciliogenesis. This is observed in several PCD related genes, including *DNAAF1* and *SPAG1,* which have been linked to PCD cases, due to defects in cytoplasmic preassembly and or processing of ciliary proteins (Knowles et al., 2013; Mitchison et al., 2012). In this project, endogenous tagging of the *Tb*DPCD protein determined localisation of DPCD to the cell cytoplasm. This has also been observed through the N-terminus and C-terminus tagging of the *T. brucei brucei* DPCD protein on the 'TrypTag' protein localisation resource (Dean et al., 2017). From this and the findings by Marias and colleagues Morias et al., 2022, if DPCD had a role in non-redundant assembly or processing of motile cilia, RNAi mediated knock down is expected to result in a phenotype. However, the results gained from the RNAi mediated knockdown of DPCD in this project did not correspond with this. Induced ablation of *Tb*DPCD resulted in reduced protein expression in the *T. brucei* cell cytoplasm, however this did not have any observable effect on cell growth, morphology, motility or cell cycle progression. This result also contrasts to the observation by Funfak and colleagues, who reported that after RNAi ablation of DPCD in *Paramecium* (Funfak et al., 2015) caused a reduced swimming velocity and an increased cell length and width in DPCD ablated cells compared to wild type cells. The functional basis for this phenotype is unclear however, as the statement by Funak and colleagues that DPCD is a 'structural protein of the ciliary microtubule scaffold and therefore expected to influence a cilium's individual movement' is not supported by research presented by the authors work, nor in any other scientific literature on DPCD.

RNAi ablation of DPCD had no phenotypic consequences for *T. brucei*, the ablation of other *T. brucei* genes that are homologues of proteins implicated in PCD in humans have been shown to result in a deleterious effect on flagellum formation and function. This includes proteins associated with cytoplasmic preassembly of ciliary components, such as DNAAF1 and DNAAF11 (Duquesnoy et al., 2009; Morgan et al., 2005). Additionally, the deletion of *T. brucei* cytoplasmic proteins that are required for PFR formation, PFR assembly factors 1 and 2, caused reduced cell growth and disruption to cell cycle progression due to cytokinesis defects (Alves et al., 2020). The lack of observable phenotype in the *Tb*DPCD mutant cells may be due to DPCD redundancy leading to partial knockdown of the protein by RNAi having little effect. This would be the case

particularly if the protein has enzymatic or chaperone activity, where a 10-20% knockdown is not enough to lose protein function. The complete knockout of the gene could be attempted using the episome-based CRISPR/Cas9 system developed by Vasquez and colleagues for use in procyclic *T. brucei* cells (Vasquez et al., 2018). The effect of *Tb*DPCD knockout on cell growth, morphology and cell cycle progression would be observed. If knockout of the protein results in null cell lines, it would suggest that DPCD has an important function. The effect of DPCD knockout in mice was investigated in a newly published paper by Yoshida and colleagues, showing that knockout of DPCD results in characteristics of hydrocephalus, a known PCD symptom, including a significantly lower cilia-induced velocity and cilia motor function compared to wild-type mice (Yoshida et al., 2022). This provides further evidence for the role of DPCD in PCD and the deleterious phenotypical effects of DPCD knockout. The effect on flagella motility can also be investigated in more detail using direct observation of the flagella movement and cell swimming using video microscopy or by sedimentation assay. However, the potential redundancy of DPCD does not explain why the protein is so well conserved across ciliated eukaryotes with motile cilia and why it has been evolutionary retained.

As over expression analysis was inconclusive further study on the effects of overexpression of *Tb*DPCD is needed. Overexpression of proteins involved in ciliogenesis has been found to affect the formation of cilia, including DNAAF11 (*Tb*LRTP) where expression of excess protein suppressed new flagellum assembly (Morgan et al., 2005). This could be done by repeating the cloning of the DPCD open reading into a pDEX377 vector and transfecting this into *T. brucei* cells. A doxycycline induced over-expression assay could be carried out to determine the effect of excess expression of DPCD on cell morphology, growth and motility.

Analysis of the effect of heat shock on *Tb*DPCD transfected cells showed that heat treatment at 41°C for 1 hour had no effect on cell morphology, motility and cell cycle progression compared to non-heat shocked cells. Also, the expression of DPCD did not increase or decrease and an expected halt in the cell cycle was not observed. Contrastingly, other proteins involved in heat shock in procyclic form trypanosomes, including the zinc finger protein ZC3H11 chaperone are found to increase in expression during heat shock (Droll et al., 2013). However, previous work by Ooi and colleagues showed that heat treatment of bloodstream form *T. brucei* cells leads to a significant decrease in motile cells after 2 hours at 41°C (Ooi et al., 2020). This heat shock analysis could be repeated by heat treating *Tb*DPCD transfected cells for 2 hours at 41°C and comparing the effects on cell motility and cell cycle progression compared to that of the heat-treated background cell line. The effect of over-expression and RNAi ablation of *Tb*DPCD on the heat shock response could also be investigated and compared to wild-type cells.

As in vivo work failed to provide any insight into *Tb*DPCD function, in silico analysis was undertaken. As mentioned earlier, *T. brucei brucei* DPCD protein folding was predicted by Phyre2 to be like that of proteins with chaperone or co-chaperone activity, including small heat-shock proteins 'holdases' which are thought to be the first line of defence against protein aggregation under non-physiological temperatures (Haslbeck and Buchner, 2015). Chaperone and co-chaperone activity has been previously determined to be essential for the preassembly of cilia/flagella. This includes the above mentioned R2TP complex, which acts as a co-chaperone to HSP90 and HSP70 in the pre-assembly of dynein arms in cooperation with dynein axonemal assembly factors (DNAAFs) (Fabczak and Osinka, 2019). DNAAFs, including DNAAF2 (KTU) and DNAAF4 (DYX1C1) are predicted to assist HSP chaperones to promote subunit folding and cytoplasmic pre-assembly of dynein motors (Tarkar et al., 2013). Novel co-chaperones, including ZMYND10 have also been implicated in the cytoplasmic pre-assembly of axonemal dynein motors, working alongside HSP chaperones (Mali et al., 2018).

Therefore, the holdase activity of *Tb*DPCD and its ability to stabilize citrate synthase against thermal unfolding was investigated. This suggested that *Tb*DPCD does not act as a chaperone in the context of the heat stress response, contradicting what was predicted by Phyre2. However, after the data was normalised against the mean absorption at 1 minute, *Tb*DPCD at 2-fold molar excess of citrate synthase seemed to possess holdase activity. This should be repeated to determine if *Tb*DPCD at this specific concentration has chaperone activity. Also, the thermal aggregation assay should be repeated comparing the results to that of a known heat shock protein with holdase activity. This study also suggested that the *Tb*DPCD protein itself may not be stable at high temperatures, demonstrated by potential foldase chaperone activity, which may have affected the results of the thermal aggregation assay. This could be investigated by carrying out a chemically induced aggregation assay, to measure the ability of *Tb*DCPD to supress chemically induced aggregation of citrate synthase independent of temperature (Haslbeck and Buchner, 2015). This has been demonstrated previously by the ability of *Tb*Hsp70 to suppress the protein aggregation of chemically denatured rhodanese in a dose-dependent manner (Burger et al., 2014). It would be unlikely that *Tb*DPCD possesses foldase activity but no holdase activity, however it would be useful to determine if the protein has chaperone activity that is not related to heat stress.

Further analysis of *Tb*DPCD protein function using the AlphaFold protein database predicted that *Tb*DPCD has a similar protein structure to chaperones, or co-chaperones, involved in heat shock, similar to what predicted by Phyre2. It was also determined that DPCD has a unique fold structure. To predict whether *Tb*DPCD acts with similar function to these predicted chaperone

and co-chaperone proteins, how the proteins behave in a citrate synthase thermal aggregation assay was determined from available literature, comparing to that of *Tb*DPCD. A citrate synthase aggregation assay carried out on the SGT1 protein shows that it effectively attenuates aggregation of the enzyme induced by increased temperature (Zabka et al., 2008). Additionally, many publications have described the suppression of citrate synthase aggregation under high temperatures by a large number of heat shock proteins, including HSP21, HSP20, HSP18.1 and HSP17.7 (Ahrman et al., 2007; Lee et al., 1995; Lee et al., 1997; Li et al., 2012). No data on how stress induced protein 1 behaves in a citrate synthase thermal aggregation assay could be found. From this and the results from the thermal aggregation assay above, it is unlikely that *Tb*DPCD has the same chaperone or co-chaperone function as these proteins.

Section 5: Conclusion

The function of DPCD in *T. brucei* and its predicted role in ciliogenesis in relation to PCD remains undetermined. Bioinformatic analysis indicates that the DPCD protein is present only in the genomes of eukaryotes that build motile cilia and predicts that the protein has structural homology to proteins that function as chaperones or co-chaperones. However, heat shock experiment analysis on *T. brucei* procyclic cells did not demonstrate that this caused upregulation of the *Tb*DPCD protein and biochemical analysis of purified protein in a thermal aggregation assay did not provide evidence that *Tb*DPCD can act as a chaperone in the context of heat stress. As RNAi induced ablation of DPCD in *T. brucei* procyclic cells did not result in an observable phenotype it suggests that DPCD has a redundant function in this organism. While the role of DPCD in ciliogenesis was supported by the bioinformatics analysis, the in vivo studies and biochemical analysis were unable to provide insights into the protein's function.

Section 6: Appendix

Appendix 1: Bioinformatic analysis of DPCD conservation across eukaryotes

Table 1: Eukaryotic genomes analysed using 'BlastP' to determine the presence/absence of DPCD homologues. Table shows that full list of ciliated and non-ciliated eukaryotes used in *Hs*DPCD 'Blastp' searches to find putative homologues. Organisms in green are ciliated eukaryotes with motile cilia, organisms in blue are ciliated eukaryotes with only non-motile (primary) cilia and organisms in read are non-ciliated eukaryotes.

Phylum	Species
Ciliated with motile cilia	
Bilateria	*Homo sapiens*
	Xenopus laevis
	Danio rerio
	Drosophila melanogaster
Coelenterata	*Nematostella vectensis*
	Hydra magnipapillata
Porifera	*Amphimedon queenslandica*
Placozoa	*Trichoplax adhaerens*
	Monosiga brevicollis
Cryptomycota	*Rozella allomycis*
Apusomonads	*Thecamonas trahens*
Stramenopiles	*Phytophthora infestans*
	Aphanomyces invadans
	Aphanomyces cochlioides
	Saprolegnia parasitica
	Pythium ultimum
	Albugo candida
Alveolates	*Paramecium tetraurelia*
	Plasmodium falciparum
	Toxoplasma gondii
Rhizaria	*Reticulomyxa filosa*
	Plasmodiophota brassica
Chlorophyta	*Micromonas pusilla*
	Chlamydomonas reinhardtii
Euglenozoa	*Trypanosoma brucei*
	Leishmania major
Heterolobosea	*Naegleria gruberi*
Parabasalids	*Trichomonas vaginalis*
	Tritrichomonas foetus
Diplomonads	*Giardia intestinalis*
Ciliated with non-motile cilia	
Bilateria	*Caenorhabditis elegans*
Non-ciliated	
Ascomycota	*Neurospora crassa*
	Saccharomyces cerevisae
	Schizosaccharomyces pombe
Basidiomycota	*Ustilago maydis*
	Cryptococcus neoformans
Mucoromycota	*Mucor circinelloides*
Microsporidia	*Encephalitozoon cuniculi*

Discosea	*Acanthamoeba castellani*
Archamoebae	*Entamoeba histolytica*
Mycetozoa	*Dictyostelium discoideum*
Alveolates	*Cryptosporidium hominis*
	Theileria parva
Cyanidiophytes	*Cyanidioschyzon merolae*
Banglophytes	*Galdieria sulpharia*
Floridiophytes	*Chondrus crispus*
Charophyta	*Klebsormidium nitens*
Angiosperms	*Arabidopsis thaliana*
	Oryza sativa
Euglenozoa	*Perkinsela sp.*

1-protein DPCD [*Xenopus laevis*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA      60
XlDPCD      MALQGWLESLKAAQKTCILQDGRRKVHFLFTDGKEMAEEYDAKSHELIVRKWRQKSGLGA      60
            **: ******::****.:*********:** ********* *: **:***** **.***

HsDPCD      MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV      120
XlDPCD      YGQWQIEVGDPPLPGAGTIQPDFLKESSSNPTFTRKDTKSSFQWRIRNLPYPKEVYSVTV      120
             ****:*****   ***.: *:::***.:** * ***** *************:****.*

HsDPCD      DQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVAE      180
XlDPCD      DKKDRCCIIRTTNKKYYKKFSIPDLDRCHLDLNENAISFAHANNTLVVTYEKPKEILSIE      180
            *:*:** *:****************** :* *:: ::****** **::.*:****::  *

HsDPCD      SELQKELKKVKTAHSNDGDCKTQ      203
XlDPCD      EELQREVKSMKATTDGDVECKTQ      203
            .***:*:*.:*:: ..* :****
```

2-protein DPCD [*Danio rerio*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA      60
DrDPCD      MAVQQWADALKKAKKTALISDGKRKVHYQFEDGNEMAEEYDLKTDELVMRKWRHKTTFGG      60
            ***   * ::*:.*:****:.**:*****  *:***** **.******* **.**::.*:::*.

HsDPCD      MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV      120
DrDPCD      QGQWTLEVGETNPTGAT---SDLIKENSSNPLFMRRDTKTSFQWRIRNISYPLEVYSVTA      117
             *** ****:  * **      :****..:**.:**:***:*** *********: ** :****..

HsDPCD      DQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVAE      180
DrDPCD      EPMERCCVIRTSNKKYYKKFSIPDLERCQLPLESTALSFTHANNTLIISYKKPKEILTLE      177
            :  *** ::**:*************.:* ****:.::***:*** ******:****::. *

HsDPCD      SELQKELKKVKTAHSNDGDCKTQ      203
DrDPCD      QELLGELKKLKGTSEGDIDCKTQ      200
            .** ****:* : ..* *****
```

3- uncharacterized protein Dmel_CG13901, isoform B [*Drosophila melanogaster*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSA-LG      59
DmDPCD      MSYQNWLNYLQAAEKNSMISGRARKVLYKFPDGRQMAEEYNMDTGIVQRRAWKSKSNKIM      60
            *:  .**: *::*:*.::.. *** * ****::*****: .*. : * *: ** :

HsDPCD      AMGQWQLEVGDPA---------PLGAGN------LGPELIKESNANPIFMRKDTKMSFQW      104
DmDPCD      GESEWEIELGDEPRQLNWSGNKPPGSGEETDSLAGGDFTLRESNTAPLLTKRITKKNIEW      120
            . .:*::*:**           * *:*:         *    ::***: *:: :: ** .::*

HsDPCD      RIRNLPYPKDVYSVSVDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANC      164
DmDPCD      RIRNMPYSLDTYSVTADPEKRAIVVRTSNKKYYKVIPVPELDRCGVKPAQESLSVHHQFN      180
            ****:** *.***:.* ::*.*:***.****** : :*:*** :   : ***. *

HsDPCD      TLIISYQKPKEVVVAESELQKELKKVKTAHSNDGDCKTQ---  203
DmDPCD      TLIITYQKPDILCEMEAQVLLLLKNVDTETDMDDLLKGLMAK  222
            ****:****. :   *::: **:*.*  . *. *
```

4- predicted protein [*Nematostella vectensis*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      -----MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVK      55
NvDPCD      MAATGTREFPWLETLRNARKTSLVQDGKRKIHFTADGTQMVEEYSVANGELLVRKWKKK      60
                 ***:**.*:**:*:***.**:*: * **.:*.***.  ..*******: *

HsDPCD      SALGAMGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDV      115
NvDPCD      TTLGKECKWEYEIGEEV--QPVNIDSEHIMESRANPIFIRKDVNKAFQWRVRNLPYPLEV      118
            ::**    :*: *:*: .      *:. * * **.*****:***.: :****:****** :*

HsDPCD      YSVSVDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKE      175
NvDPCD      YSVTVDDDRKTITTRTNNKKYFKKFSIPDMVRAELPLDQSSISLAHANNTLIISYAKPPM      178
            ***:**:..: * .**.****:********: * ;****:*:*:**** ****** **

HsDPCD      VVVAESELQKELKKVKTAHSNDGDCKTQ          203
NvDPCD      ILEQEKLIKVELAKTQGSKEGDIECKPS          206
            ::  *. :: ** *.: ::..* :** .
```

5- hypothetical protein H310_13815 [*Aphanomyces invadans*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEK-TSELLVRKWRVKSALG      59
AiDPCD      MS-KAMFSVHPGGTSNCAIAGGRSRIHTTFSDGVEMVEEYTADGRPQLLVRRWKAPTALG      59
            *:  .. :.    . ... : .** ::*  * ** **.***   :****:*:. :***

HsDPCD      AMGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVS      119
AiDPCD      GDGKWEFEVGEPTLDTVVEKD-IGIVPSGTNPIFLCRESTATWEWRVRNLPYPKDTYVIT      118
            . *:*::***:*:   . :  . * *.:****: :::. :::**;********.* ::

HsDPCD      VDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVA      179
AiDPCD      IDQATQEIIIRTTNKKYFKRFRVPAMVRASLPLQQSSLSHTHSNNTLVVQYEKPDEVVEA      178
            :**   : **:*****:*:* :* : * .***::***.:*:* **::.*:**.*** *

HsDPCD      ESELQKELKKVKTAHSNDGDCKTQ          203
AiDPCD      EKNLLDERLS-ESKRAGDSQCRQQ          201
            *.:* .*  . :: ::.*.:*: *
```

6- hypothetical protein TRIADDRAFT_31259 [*Trichoplax adhaerens*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA      60
TaDPCD      --MSTWVEMLKSSQKTCLVQDGRKKIHYTFKNHTEMAEEYDMATFDLSVRKFRKKSTLGS      58
              ::  *:* *::.***.*:****.*:.** * :  .*******   * :* ***:* **:**:

HsDPCD      MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV     120
TaDPCD      SGQWEYEIGEATNI--TNIEELSIRESNTTPRFSAKYSKNAFQWRIRNLPYPLDNYSITV     116
             ***: *.*: : :   *:   *.***:.* *   * :* :*********** * **::*

HsDPCD      DQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVAE     180
TaDPCD      DDEKRQLVLRTKNKKYFKRISIPDMDRFHLPLDQNAVKLAHANNCLIISYQKPSQIIEHE     176
            *:::* :::**.****:*::****.**.:**** :::.:**** ********.:::  *

HsDPCD      SELQKELKKVKTAHSNDGDCKTQ     203
TaDPCD      KALHKELRKMKASKDGDVECNPS     199
            . *:***:*:*:::..* :*: .
```

7- hypothetical protein SPRG_05802 [*Saprolegnia parasitica* CBS 223.65]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEK-TSELLVRKWRVKSALG      59
SpDPCD      -----MYRSHSGGNSTVVVAGGRKRLHTTFADGVELVEEYTKDGPPELLVRRWKNTTALG      55
                 .*   ..*.::  .**:::*  * ** *:.*** :.   *****:*: .:***

HsDPCD      AMGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVS     119
SpDPCD      GDGRWEYEIGEPLPAEGRSRDIG-LAPSSTQPTFLCREAATHWEWRVRNLPYPKETYAIG     114
            . *:*: *:*:* *   . .   :  *..:* *: :::   ::**:*******:.*::

HsDPCD      VDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVA     179
SpDPCD      IDDDAQTIVIRTSNKKYFKRFQIPAMVRQSQRLSAAAVSFTHDNNTLVVHYKKPDAVMAV     174
            :*:. : *::**:****:*:.** : *:. *. *:::**:* * **:: *:**. *:..

HsDPCD      ESELQKELKKVKTAHSNDGDCKTQ     203
SpDPCD      EHQELR--DRLHGAGTSEPQCRQQ     196
            * : :  .::: * :.: :*: *
```

8- hypothetical protein [*Paramecium tetraurelia* strain d4-2]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      --MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSAL      58
PtDPCD      MQSSNNQYLQFVGKPHITAVIAGGRRRVNYTFEDKSELVEEYDINNHELISRKWKRVSNI      60
              : . :*: : .: **:: .***:*:* * * .*:.**** :. **: ***:  * :

HsDPCD      GAMGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSV     118
PtDPCD      K-ESQWIYEVGEAPIDQQGE-----LKLSNANPIFVRKDTPQHFQYRVRNLNYPEDVYKV     114
             .** ***:      *:      :* *******:****   **:*:*** **:***.*

HsDPCD      SVDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVV     178
PtDPCD      EVDENTQEIVIRTSNKKYYKRFAIPDMRRANLKLEQGKVTHVYKNNTLIVSYPKPDQILE     174
            .**:: : *::**:*****::*:*** * :* *::..:..: * ***:** **.:::

HsDPCD      AESELQKELKKVKTAHSNDGDCK--TQ 203
PtDPCD      REYQIRQEFDKLNKKKPKEGDLECVQQ 201
             * :::::*:.*::. : ::** :   *
```

9- DPCD family protein [*Toxoplasma gondii* TgCatPRC2]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      -----MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVK      55
TgDPCD      MSVFQKAENGPI-DTGKPQILSFIADGRRTVHTTYPDKSEKVEEYDLRDDTVLVRKFRRP      59
                 * .* :  .  . *   ::: ****.**  :** .* .**** : . :****:*


HsDPCD      S-ALGAMGQWQLEVGDPAPLGA-----GNLGPELIKESNANPIFMRKDTKMSFQWRIRNL     109
TgDPCD      PALFGQEGMWEYEIGAPRQGMDAVVASSDRNEIVMAPSTSSPLCVMRDTKGAFEWRIRNL     119
                :*  * *: *;* *         .: .  :: *.:.*: : :*** :*:******


HsDPCD      PYPKDVYSVSVDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIIS     169
TgDPCD      PYPRDTYIVQIDGDASEIVISTTNKKYYKRLRISDIHALGLQLNLESLSWTHQHNTLVVS     179
            ***:*.* *.:* .   *:: ********:: * *:.  * *:  ***::* : **::*


HsDPCD      YQKPKEVVVAESELQKELKKVKTAHSNDGDCKTQ    203
TgDPCD      YKKPQKVLDLERKRREEAKRTSLSVG--------    205
            *:**::*:  * : ::* *:.. : .
```

10- unnamed protein product [*Plasmodiophora brassicae*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA      60
PbDPCD      ----MGDGNDRAPRRTSIVSSGRRKVHTTFPDGREMVEEYDVRTEDLLVRKTRGKTILGA      56
                . *: ::*:::..****** ****:**.**** :*.:***** * *: ***


HsDPCD      MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV     120
PbDPCD      QQDWQWDVGKPASTAA--MPASGIRESPHNPVLVRKDTHDTIEFRIRNLPYPQDVYDVSV     114
             :** :**.**  .*  : . *:** **:::****: :::*:*********:*** ***


HsDPCD      DQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVAE     180
PbDPCD      DEVSQTIIIRTSNKKYFKRITVPDVQG----LRTRDLTWHWGNNTLVMYYRKPADLIANE     170
            *: .: **:**:****:*::::**::      *   .*:: .* **:: *:** :::. *


HsDPCD      SELQKELKKVKTAHSNDGD--------CKTQ    203
PbDPCD      QKTRAEIARLDSAEHADGNEGRPGPNECKQQ    201
            .: : *: ::..:*.  **:        ** *
```

11- predicted protein [*Micromonas pusilla* CCMP1545]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA      60
MpDPCD      -------MFRKGGESTSVNQGGRRKVHTTYPDGAEMIEEFDLKTDELLVRRRRGRTTLGK      53
                   : .:.*:: *.****** :*** ** **:* **.*****: * :::**


HsDPCD      MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMR-KDTKMSFQWRIRNLPYPKDVYSVS     119
MpDPCD      DTPWEYLVGEAPRTFAPDTG--TLIESGANPSFTRSQDAPGSFVWRVRNLPYPKETYDVR     111
               *: **:      * : *   : **.*** * * :**: ** **:********:.*.*


HsDPCD      VDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVA     179
MpDPCD      CDHEAQQVVIRTSNKKYFKRFEVPELKLLDKKLEDAALSWDYRNFTLIVQYKKPREAVVA     171
             *:: : :::**:****:*:.:.*.:*:.:*.  : * ***:.*:**:*.***.***


HsDPCD      ESELQKELK---KVKTAHSNDGDCKTQ 203
MpDPCD      EIEEKKDRMQLAAKKTPEEGDVDCKQQ 198
            * * :*:        ** ...* *** *
```

12- hypothetical protein, conserved [*Trypanosoma brucei brucei* TREU927]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD    MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA     60
TbDPCD    ------MSVTLSEPKSSVIVNGRRRITSKFVDGGEMIEEYDVITDDLLLRKYRSRTTLGG     54
              :.   :  *:::: :***::    * ** ** ****  *.:**;**:* :::**.

HsDPCD    MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV    120
TbDPCD    FSAWEVEVGNEAS--TRNLDKELVVETSGSPEVVKQDALEFYVFRIRNLPYAKEVFSVAV    112
          :. *::***: *   : **. **: *:...* .:::*:   : :******* *:*:**:*

HsDPCD    DQKE----RCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEV    176
TbDPCD    EHSKPTDMGEIVVRTSNKKYFKRLSIPDMNRRNLKLDPAQLSFDVQHNTLIIRYKKPLVV    172
          ::.:      *:***:****:*::****::*:;* ** *.***   : **** *:** *

HsDPCD    VVAESELQKELKKVKTAHSNDGDCKTQ----      203
TbDPCD    LAAESAAKKERASLPAKRIDDADSRTSCNQQ    203
          :.*** :** .: : : :*.*.:*.
```

13- conserved hypothetical protein [*Leishmania major* strain Friedlin]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD    MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA     60
LmDPCD    ---------MLAEPKSSVIRGDRKCITSKFTDGSELVDEYDVVTDSLLLRKRRTRNALGG     51
                 : :  *:::::..*: :   * **.*::.:*** *..**;** *.:.***.

HsDPCD    MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV    120
LmDPCD    FSEWSIEVGTEAPS--RNLDRALIAESSGSPVVVRQDTKESYVVRIRNLPYPKDVFSVAV    109
          :.:*.:*** **    **. ** **...*:.:*:*** *: ***********;**:*

HsDPCD    DQKE----RCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEV    176
LmDPCD    EREDGDSVGKIVVRTSNRKYFKRLAIPDLERAHIPLESAHLSYDVQHQTLIIQYKKPLSV    169
          ::::      *:***:*:*:*:::*****:* ::**:.* **:   : ****.*:** .*

HsDPCD    VVAESELQKELKKVKT--AHSNDGDCKTQ      203
LmDPCD    LAAEAAARKERASMPSKRADDSGPDCKQQ    198
          :.**: :** .: : *.... *** *
```

14- uncharacterized protein NAEGRDRAFT_72517 [*Naegleria gruberi*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD    ----------------MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKT     44
NgDPCD    MLSELLSKKTGDDQNTTSSSSEPSTTKYPYTTSVVADGRRKFHTTYENGVEMVEEFDLQT     60
              : :.  .: :   .*::: *****.*   : :* **.**:* :*

HsDPCD    SELLVRKWRVKSALGAMGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQW    104
NgDPCD    NQLLLRKRRTVSILG-EGTWTIEIGGEVK--SFNPIHDLIAPSTVNPIFMRKDTSTLFQW    117
          .:**;** *.* ** * * :*:*. .  : *  :** *..*********. ***

HsDPCD    RIRNLPYPKDVYSVSVDQKERCIIVRTTNKKYYKKFSIPDLDRH-QLPLDDASLSFAHAN    163
NgDPCD    RVRNCFYEPNVYKVNIDHEKQEIVISTTNKKYYKRFSIPDMKESPKTKLNDESLSFKLEN    177
          *:** * :**.*.:*:::: *:: ********:*****:.. : *:* **** *

HsDPCD    CTLIISYQKPKEVVVAESELQKELKKVKTA---HSNDGDCKTQ 203
NgDPCD    NTLIISYVKPDLILKKERAKEMELASMKSERPPRDGDVQCPQQ 220
          ****** **. :: *   : ** .:*:    :...* :* *
```

15- DPCD protein [*Trichomonas vaginalis* G3]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLFPDGKEMAEEYDEKTSELLVRKWRVKSALGA      60
TvDPCD      -------MTIAFPKAVAYNTGGYKTIVYDEKEGKQVVDICELDTGKLLARKERVRHTSGS      53
                  ::    : .*    .* :.: *    :**::.:    : .*.:**.** **: : *:

HsDPCD      MGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVSV     120
TvDPCD      DGPWEWTYGSPIEQK---QEVETISASSKNPVFLRLDTDNEWQWRVRNIPYPANFYNVTV     110
             * *:    *.*          * *. *. **:*:* **. ..:***:**:*** :.*.*:*

HsDPCD      DQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVAE     180
TvDPCD      DEPKNQIVIRTTNKKYFKRIDAPEGEKFV----PSKLQWAWSYNTLVIQHEKPPRVVAAD     166
            *: :. *::*******:*::. *: ::.        :.*.:* :  **:*.::** .**.*:

HsDPCD      SELQKELKKVKTAHSNDGDCKTQ           203
TvDPCD      KELCKWRRGLPLQEDKDPDCRV-           188
            .** *   : :    ..:* **:.
```

16- DPCD protein [*Tritrichomonas foetus*]

```
CLUSTAL O(1.2.4) multiple sequence alignment


HsDPCD      MAVTGWLESLRTAQKTALLQDGRRKVHYLF-PDGKEMAEEYDEKTSELLVRKWRVKSALG      59
TfDPCD      -------MTLKKPDATAYKVGGYKQIMYVDAKTGNDVMEVCELDTGILVARRERSKTIYG      53
                 :*:. : **   .* ::: *:    *::: *   : .*. *:.*: * *: *

HsDPCD      AMGQWQLEVGDPAPLGAGNLGPELIKESNANPIFMRKDTKMSFQWRIRNLPYPKDVYSVS     119
TfDPCD      ATSQWECTYGQIPS---EKKEIETIELSSKNPIFFRQDTPNQWSWRVRNIPYPPENYEVT     110
            * .**:  *:       :   * *: *. ****:*:** .:.**:**:*** : *.*:

HsDPCD      VDQKERCIIVRTTNKKYYKKFSIPDLDRHQLPLDDASLSFAHANCTLIISYQKPKEVVVA     179
TfDPCD      VDEAKNQIVVRTRNKKYFKRIDAPNGEKMVP----SEVNWNWNFNALVITHNKPPRVIAQ     166
            **: :. *:*** ****:*::. *: ::        :.:.:    :*:*:::** .*:.

HsDPCD      ESELQKELKKVKTAHSNDGDCKTQ          203
TfDPCD      DKDMQQWRAKLPCKEDNEPDCKTQ          190
            :.::*:   *:    ..*: *****
```

Figure 1: Alignments of DPCD putative homologue amino acid sequences against *Hs*DPCD sequence. Alignment of the amino acid sequences of DPCD putative homologues against the *Homo sapiens* DPCD amino acid sequence, using 'ClustalOmega'. 1-16- show alignments of putative DPCD homologues of all ciliated eukaryotes that produced 'Blastp' hits against the *Hs*DPCD sequence. All show strong sequence identity to the HsDPCD sequence, confirming each as a putative DPCD homologue.

Table 2: Ciliary proteins used for 'BlastP' searches to find putative homologues in ciliated and non-ciliated eukaryotes. The full list of ciliary proteins and their 'ensemblePlants' database accession numbers used for 'BlastP' searches. Proteins gained from **Hodges et al., 2011**.

Protein	Description	Accession number
CCP1	Putative katanin p60 catalytic subunit	AT2G34560.2
CCP1 hp	Zinc ion binding	AT5G42320.1
PF15	Putative katanin p80 subunit	AT5G23430.2
HMGB3407	ARID domain containing HMGB protein	AT1G76110.1
BUG22	Basal body protein, unknown function	AT3G12300.1
CCP2	Predicted ubiquitin-fold modifier 1 precursor	AT1G77710.1
FASS1	FASS1/ tonneau 2 protein phosphatase type 2B	AT5G18580.1
FUSED	Hypothetical protein	AT1G50240.2
FAP267	Flagella associated protein	AT1G77550.1
IBR1	Indole-3-butyric acid response 1	AT4G05530.1
GPX3 hp3	Hypothetical protein	AT2G47600.1
GPX3 hp4	Hypothetical protein	AT2G27900.2
FKB62	Peptidy-prolyl cis-trans isomerase	AT3G25230.1
MTP5	CDF transporter	AT1G51610.1
CCP3	Flagella associated protein	AT2G25240.1
HAC3401	CREB-binding protein	AT1G79000.1
TEF18	Glutathione S-transferase-related protein	AT5G42150.1

```
CLUSTAL O(1.2.4) multiple sequence alignment


Wildtype      ATGAGTGTAACCCTTTCAGAGCCTAAGTCATCTGTCATTGTCAACGGACGTCGTCGCATA      60
DPCDinsert    ATGAGTGTAACCCTTTCAGAGCCTAAGTCATCTGTCATTGTCAACGGACGTCGTCGCATA      60
              ************************************************************

Wildtype      ACATCAAAGTTTGTCGATGGCGGTGAGATGATTGAGGAATATGACGTCATTACTGATGAC     120
DPCDinsert    ACATCAAAGTTTGTCGATGGCGGTGAGATGATTGAGGAATATGACGTCATTACTGATGAC     120
              ************************************************************

Wildtype      CTTCTACTTCGCAAGTACCGCTCAAGGACCACACTTGGAGGGTTTAGCGCTTGGGAAGTT     180
DPCDinsert    CTTCTACTTCGCAAGTACCGCTCAAGGACCACACTTGGAGGGTTTAGCACTTGGGAAGTT     180
              *********************************************** ***********

Wildtype      GAGGTGGGAAATGAAGCCTCAACGCGTAACTTGGACAAAGAATTAGTTGTGGAGACAAGT     240
DPCDinsert    GAGGTGGGAAATGAAGCCTCAACGCGTAACTTGGACAAAGAATTAGTTGTGGAGACAAGT     240
              ************************************************************

Wildtype      GGCAGCCCAGAAGTGGTAAAACAGGATGCGTTGGAGTTTTATGTGTTCCGCATACGCAAT     300
DPCDinsert    GGCAGCCCAGAAGTGGTAAAGCAGGATGCGTTGGAGTTTTATGTGTTCCGCATACGCAAT     300
              ******************** ***************************************

Wildtype      CTTCCATACGCAAAGGAGGTGTTTAGTGTTGCCGTAGAGCATAGCAAACCAACCGATATG     360
DPCDinsert    CTTCCATACGCAAAGGAGGTGTTTAGTGTTGCCGTAGAGCATAGCAAACCAACCGATATG     360
              ************************************************************

Wildtype      GGGGAAATTGTGGTGCGTACGAGCAACAAAAAATACTTCAAGAGGCTCTCCATACCTGAT     420
DPCDinsert    GGGGAAATTGTGGTGCGTACGAGCAACAAAAAATACTTCAAGAGGCTCTCCATACCTGAT     420
              ************************************************************

Wildtype      ATGAACCGACGTAACCTTAAGTTGGATCCTGCTCAACTCTCCTTTGATGTACAGCACAAC     480
DPCDinsert    ATGAACCGACGTAACCTTAAGTTGGATCCTGCTCAACTCTCCTTTGATGTACAGCACAAC     480
              ************************************************************

Wildtype      ACGCTGATTATTCGGTACAAGAAGCCCCTAGTGGTTCTTGCCGCAGAAAGTGCGGCAAAG     540
DPCDinsert    ACGCTGATTATTCGGTACAAGAAGCCCCTAGTGGTTCTTGCCGCAGAAAGTGCGGCAAAG     540
              ************************************************************

Wildtype      AAGGAACGAGCCTCGCTCCCAGCGAAACGCATCGATGATGCCGACTCTAGGACCAGTTGC     600
DPCDinsert    AAGGAACGAGCCTCGCTCCCAGCGAAACGCATCGATGATGCCGACTCTAGGACCAGTTGC     600
              ************************************************************

Wildtype      AATCAGCAATAG      612
DPCDinsert    AATCAGCAA---      609
              *********
```

Figure 2: Alignment of DPCD insert nucleotide sequence of DPCD insert from pGEM-T Easy clone to be cloned into p2T7-177 against the wildtype *T. brucei* DPCD sequence. DPCD insert nucleotide sequence provided by 'SourceBioscience'. *T. brucei brucei* DPCD sequence gained from 'TriTrypDB' (Tb1125.11.8900). Sequences aligned using 'ClustalOmega'.

```
CLUSTAL O(1.2.4) multiple sequence alignment


Wildtype        ---------------ATGAGTGTAACCCTTTCAGAGCCTAAGTCATCTGTCATTGTCAAC    45
DPCDinsert      ATTCGATTGAAGCTTATGAGTGTAACCCTTTCAGAGCCTAAGTCATCTGTCATTGTCAAC    60
                               ********************************************

Wildtype        GGACGTCGTCGCATAACATCAAAGTTTGTCGATGGCGGTGAGATGATTGAGGAATATGAC   105
DPCDinsert      GGACGTCGTCGCATAACATCAAAGTTTGTCGATGGCGGTGAGATGATTGAGGAATATGAC   120
                ************************************************************

Wildtype        GTCATTACTGATGACCTTCTACTTCGCAAGTACCGCTCAAGGACCACACTTGGAGGGTTT   165
DPCDinsert      GTCATTACTGATGACCTTCTACTTCGCAAGTACCGCTCAAGGACCACACTTGGAGGGTTT   180
                ************************************************************

Wildtype        AGCGCTTGGGAAGTTGAGGTGGGAAATGAAGCCTCAACGCGTAACTTGGACAAAGAATTA   225
DPCDinsert      AGCACTTGGGAAGTTGAGGTGGGAAATGAAGCCTCAACGCGTAACTTGGACAAAGAATTA   240
                *** ********************************************************

Wildtype        GTTGTGGAGACAAGTGGCAGCCCAGAAGTGGTAAAACAGGATGCGTTGGAGTTTTATGTG   285
DPCDinsert      GTTGTGGAGACAAGTGGCAGCCCAGAAGTGGTAAAGCAGGATGCGTTGGAGTTTTATGTG   300
                *********************************** ************************

Wildtype        TTCCGCATACGCAATCTTCCATACGCAAAGGAGGTGTTTAGTGTTGCCGTAGAGCATAGC   345
DPCDinsert      TTCCGCATACGCAATCTTCCATACGCAAAGGAGGTGTTTAGTGTTGCCGTAGAGCATAGC   360
                ************************************************************

Wildtype        AAACCAACCGATATGGGGGAAATTGTGGTGCGTACGAGCAACAAAAAATACTTCAAGAGG   405
DPCDinsert      AAACCAACCGATATGGGGGAAATTGTGGTGCGTACGAGCAACAAAAAATACTTCAAGAGG   420
                ************************************************************

Wildtype        CTCTCCATACCTGATATGAACCGACGTAACCTTAAGTTGGATCCTGCTCAACTCTCCTTT   465
DPCDinsert      CTCTCCATACCTGATATGAACCGACGTAACCTTAAGTTGGATCCTGCTCAACTCTCCTTT   480
                ************************************************************

Wildtype        GATGTACAGCACAACACGCTGATTATTCGGTACAAGAAGCCCCTAGTGGTTCTTGCCGCA   525
DPCDinsert      GATGTACAGCACAACACGCTGATTATTCGGTACAAGAAGCCCCTAGTGGTTCTTGCCGCA   540
                ************************************************************

Wildtype        GAAAGTGCGGCAAAGAAGGAACGAGCCTCGCTCCCAGCGAAACGCATCGATGATGCCGAC   585
DPCDinsert      GAAAGTGCGGCAAAGAAGGAACGAGCCTCGCTCCCAGCGAAACGCATCGATGATGCCGAC   600
                ************************************************************

Wildtype        TCTAGGACCAGTTGCAATCAGCAATAG   612
DPCDinsert      TCTAGGACCAGTTGCAATCAGCAA---   624
                ***********************
```

Figure 3: Alignment of DPCD insert nucleotide sequence of DPCD insert from pGEM-T Easy clone to be cloned into pDEX377 against the wildtype *T. brucei* DPCD sequence. DPCD insert nucleotide sequence provided by 'SourceBioscience'. *T. brucei brucei* DPCD sequence gained from 'TriTrypDB' (Tb1125.11.8900). Sequences aligned using 'ClustalOmega'.

Figure 4: Alignment of DPCD insert nucleotide sequence of DPCD insert from pDEX377 clone against the wildtype *T. brucei* DPCD sequence. DPCD insert nucleotide sequence provided by 'SourceBioscience'. *T. brucei brucei* DPCD sequence gained from 'TriTrypDB' (Tb1125.11.8900). Sequences aligned using 'ClustalOmega'. The N-terminal extension containing the NdeI cloning site outlined in yellow. Triple c-myc epitope outlined in green.

```
CLUSTAL O(1.2.4) multiple sequence alignment


Wildtype      ATGAGTGTAACCCTTTCAGAGCCTAAGTCATCTGTCATTGTCAACGGACGTCGTCGCATA      60
DPCDinsert    ATGAGTGTAACCCTTTCAGAGCCTAAGTCATCTGTCATTGTCAACGGACGTCGTCGCATA      60
              ************************************************************

Wildtype      ACATCAAAGTTTGTCGATGGCGGTGAGATGATTGAGGAATATGACGTCATTACTGATGAC      120
DPCDinsert    ACATCAAAGTTTGTCGATGGCGGTGAGATGATTGAGGAATATGACGTCATTACTGATGAC      120
              ************************************************************

Wildtype      CTTCTACTTCGCAAGTACCGCTCAAGGACCACACTTGGAGGGTTTAGCGCTTGGGAAGTT      180
DPCDinsert    CTTCTACTTCGCAAGTACCGCTCAAGGACCACACTTGGAGGGTTTAGCACTTGGGAAGTT      180
              ***********************************************.**********

Wildtype      GAGGTGGGAAATGAAGCCTCAACGCGTAACTTGGACAAAGAATTAGTTGTGGAGACAAGT      240
DPCDinsert    GAGGTGGGAAATGAAGCCTCAACGCGTAACTTGGACAAAGAATTAGTTGTGGAGACAAGT      240
              ************************************************************

Wildtype      GGCAGCCCAGAAGTGGTAAAACAGGATGCGTTGGAGTTTTATGTGTTCCGCATACGCAAT      300
DPCDinsert    GGCAGCCCAGAAGTGGTAAAGCAGGATGCGTTGGAGTTTTATGTGTTCCGCATACGCAAT      300
              ********************.****************************************

Wildtype      CTTCCATACGCAAAGGAGGTGTTTAGTGTTGCCGTAGAGCATAGCAAACCAACCGATATG      360
DPCDinsert    CTTCCATACGCAAAGGAGGTGTTTAGTGTTGCCGTAGAGCATAGCAAACCAACCGATATG      360
              ************************************************************

Wildtype      GGGGAAATTGTGGTGCGTACGAGCAACAAAAAATACTTCAAGAGGCTCTCCATACCTGAT      420
DPCDinsert    GGGGAAATTGTGGTGCGTACGAGCAACAAAAAATACTTCAAGAGGCTCTCCATACCTGAT      420
              ************************************************************

Wildtype      ATGAACCGACGTAACCTTAAGTTGGATCCTGCTCAACTCTCCTTTGATGTACAGCACAAC      480
DPCDinsert    ATGAACCGACGTAACCTTAAGTTGGATCCTGCTCAACTCTCCTTTGATGTACAGCACAAC      480
              ************************************************************

Wildtype      ACGCTGATTATTCGGTACAAGAAGCCCCTAGTGGTTCTTGCCGCAGAAAGTGCGGCAAAG      540
DPCDinsert    ACGCTGATTATTCGGTACAAGAAGCCCCTAGTGGTTCTTGCCGCAGAAAGTGCGGCAAAG      540
              ************************************************************

Wildtype      AAGGAACGAGCCTCGCTCCCAGCGAAACGCATCGATGATGCCGACTCTAGGACCAGTTGC      600
DPCDinsert    AAGGAACGAGCCTCGCTCCCAGCGAAACGCATCGATGATGCCGACTCTAGGACCAGTTGC      600
              ************************************************************

Wildtype      AATCAGCAATAG      612
DPCDinsert    AATCAGCAA---      609
              *********
```

Figure 5: Alignment of DPCD insert nucleotide sequence of DPCD insert from pGEM-T Easy clone to be cloned into pET28a against the wildtype *T. brucei* DPCD sequence. DPCD insert nucleotide sequence provided by 'SourceBioscience'. *T. brucei brucei* DPCD sequence gained from 'TriTrypDB' (Tb1125.11.8900). Sequences aligned using 'ClustalOmega'.

Table 3: Protein hits from DALI, Foldseek and FATCAT databases against Alphafold *T. brucei brucei* structure. Table shows the identity of protein hits from DALI, FATCAT and Foldseek database, X represents the presence of this protein in each database. Numbers refer to numbers in Venn diagram, Figure 45.

Number	Protein identity	DALI	Foldseek	FATCAT
1	SUPPRESSOR OF G2 ALLELE OF SKP1 HOMOLOG	X	X	X
2	CYTOSOLIC HEAT SHOCK PROTEIN 90	X	X	
3	HEAT SHOCK PROTEIN 16.9B	X	X	X
4	SMALL HEAT SHOCK PROTEIN HSP16.5	X		X
5	RNA POLYMERASE II-ASSOCIATED PROTEIN 3	X		
6	HEAT SHOCK PROTEIN 16	X	X	X
7	SMALL HEAT SHOCK PROTEIN HSP20 FAMILY	X	X	X
8	HEAT SHOCK PROTEIN HSP20 FAMILY PUTATIVE	X		X
9	HEAT SHOCK PROTEIN-RELATED PROTEIN	X		X
10	LOW MOLECULAR WEIGHT HEAT SHOCK PROTEIN	X	X	X
11	HEAT SHOCK PROTEIN HSP20	X	X	X
12	MOLECULAR CHAPERONE (SMALL HEAT SHOCK PROTEIN	X		X
13	HSPA	X		X
14	STRESS-INDUCED PROTEIN 1	X	X	X
15	PROTEIN SHQ1 HOMOLOG	X	X	X
16	25.3 KDA HEAT SHOCK PROTEIN, CHLOROPLASTIC	X		X
17	NUCLEAR MOVEMENT PROTEIN	X	X	X
18	PROGESTERONE RECEPTOR P23	X		
19	TPR REPEAT-CONTAINING PROTEIN ASSOCIATED WITH HSP	X		
20	AGGREGATION SUPPRESSING PROTEIN	X		X
21	SMALL HEAT SHOCK PROTEIN	X	X	X
22	HSPB2,HEAT SHOCK PROTEIN BETA-2	X		X
23	HEAT SHOCK PROTEIN	X		X
24	14-3-3 PROTEIN SIGMA	X		
25	CYTOCHROME B5 REDUCTASE 4	X		
26	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE 19	X		X
27	ATP-DEPENDENT MOLECULAR CHAPERONE HSP82	X		
28	CALCYCLIN-BINDING PROTEIN	X		
29	TYPE IV PILUS BIOGENESIS AND COMPETENCE PROTEIN P	X		
30	KIAA1068 PROTEIN	X		
31	U2 SMALL NUCLEAR RIBONUCLEOPROTEIN A'	X		
32	SLR0280 PROTEIN;	X		
33	PRE-MRNA-SPLICING FACTOR PRP9	X		

34	N-ACETYLMURAMOYL-L-ALANINE AMIDASE AMIC	X		
35	NUCLEAR DISTRIBUTION GENE C HOMOLOG	X		
36	LIPOVITELLIN (LV-1N, LV-1C)	X		
37	U1 SNRNA	X		
38	K(+)/H(+) ANTIPORTER SUBUNIT KHTT	X		
39	UNCHARACTERIZED PROTEIN AT3G03773	X		
40	IG EPSILON CHAIN C REGION	X		
41	MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE S	X		
42	NUDC DOMAIN-CONTAINING PROTEIN 2	X		
43	CG1507-PB, ISOFORM B	X		
44	INNER KINETOCHORE SUBUNIT MIF2	X		
45	ALL4481 PROTEIN	X		
46	XRCC4-MYH7-1590-1657	X		
47	30S RIBOSOMAL PROTIEN S2		X	
48	NUP157/NUP170		X	
49	ALPHA-CRYSTALLIN A CHAIN		X	X
50	HEAT SHOCK PROTEIN 21		X	
51	ALPHA-CRYSTALLIN B CHAIN		X	X
52	STEROL REGULATORY ELEMENT-BINDING PROTEIN CLEAVAGE-ACTIVATING PROTEIN		X	
53	SMALL HEAT SHOCK PROTEIN HSP14.0		X	
54	18.1 kDa HEAT SHOCK PROTEIN		X	X
55	TETRATHIONATE HYDROLASE		X	
56	UL2		X	
57	PUTATIVE UNCHARACTERISED PROTEIN ST1653		X	X
58	17.1 kDa CLASS II HEAT SHCOK PROTEIN			X
59	SMALL HEAT SHOCK PROTEIN HSP26			X
60	HEAT SHOCK PROTEIN 26			X
61	CO-CHAPERONE PROTEIN SBA1			X
62	PROSTAGLANDIN E SYNTHASE 3			X
63	HEAT SHOCK PROTEIN BETA-1			X
64	CACTIN			X
65	HEAT SHOCK PROTEIN BETA-6			X
66	CYSTEIN-RICH PROTECTIVE ANTIGEN			X
67	PROTEIN INTERACTING WITH HSP90 1			X
68	NUCLEAR MIGRATION PROTEIN NUDC			X
69	UNCHARACTERISED PROTEIN YQZG			X
70	CPAP			X

Table 4: Alphafold database top 50 hits against Alphafold DPCD *T. brucei brucei* structure.

Top 50 protein hits from the Alphafold database against the Alphafold DPCD *T. brucei brucei* predicted structure, showing identity, species and E value.

Protein identity	Species	E value
DPCD	*Trypanosoma vivax*	1.337E-29
DPCD	*Bodo saltans*	6.657E-27
DPCD	*Dunaliella tertiolecta*	1.838E-21
DPCD	*Eutreptiella gymnastica*	1.838E-21
DPCD	*Chloropicon primus*	2.542E-21
DPCD	*Chiloscyllium punctatum*	2.54E-21
DPCD	*Panthera pardus*	4.60E-21
DPCD	*Ciona intestinalis*	5.41E-21
DPCD	*Tetraselmis chuii*	5.41E-21
DPCD	*Dissostichus mawsoni*	8.34E-21
DPCD	*Hemiselmis andersenii*	9.29E-21
DPCD	*Bugula neritina*	9.81E-21
DPCD	*Batrachochytrium dendrobatidis*	1.15E-20
DPCD	*Astyanax mexicanus*	1.36E-20
DPCD	*Seriola dumerili*	1.51E-20
DPCD	*Carlito syrichta*	2.09E-20
DPCD	*Echenis naucrates*	2.33E-20
DPCD	*Spongospora subterranea*	2.33E-20
DPCD	*Noctiluca scintillans*	2.33E-20
DPCD	*Streblomastix strix*	3.40E-20
DPCD	*Vitrella brassicaformis*	3.78E-20
DPCD	*Chromera velia*	4.70E-20
DPCD	*Fasciola hepatica*	6.15E-20
DPCD	*Taeniopygia guttata*	7.63E-20
DPCD	*Lotharella oceanica*	8.98E-20
DPCD	*Chlamydomonas eustigma*	1.06E-19
DPCD	*Schistocephalus solidus*	1.24E-19
DPCD	*Ophiophagus hannah*	1.24E-19
DPCD	*Temnothorax longispinosus*	1.38E-19
DPCD	*Brachionus calyciflours*	1.46E-19
DPCD	*Naegleria fowleri*	1.46E-19
DPCD	*Pyramimonas obovata*	2.02E-19
DPCD	*Tetraselmis sp.*	3.28E-19
DPCD	*Strombidium inclinatum*	3.28E-19
DPCD	*Callipepla squamata*	3.28E-19
DPCD	*Ursus maritimus*	3.46E-19
DPCD	*Colinus virginianus*	4.54E-19
DPCD	*Papilio xuthus*	4.54E-19
DPCD	*Acromyrmex charruanus*	4.79E-19
Hypothetical protein	*Polarella glacialis*	5.05E-19
DPCD	*Stentor coeruleus*	5.33E-19
DPCD	*Thraustotheca clavata*	5.33E-19

DPCD	*Strombidinopsis acuminata*	5.63E-19
Hypothetical protein	*Polarella glacialis*	7.37E-19
Hypothetical protein	*Rotaria sp.*	8.22E-19
DPCD	*Hermetia illucens*	9.66E-19
DPCD	*Giardia muris*	1.02E-18
DPCD	*Rhodnius neglectus*	1.02E-18
Elongator complex protein 6	*Timema tahoe*	1.02E-18
Hypothetical protein	*Rotaria socialis*	1.14E-18

Section 7: References

Ahrman E., Gustavsson N., Hultshig C., Boelens W. C. and Emanuelsson C. S. (2008) Small heat shock proteins prevent aggregation of citratesynthase and bind to the N-terminal region whichis absent in thermostable forms of citrate synthase. *Extremophiles*, **11:** 659-666.

Ainsworth C. (2007) Cilia, Tails of the unexpected. *Nature,* **448**: 638-641.

Aldrich H. C. (1968) The development of flagella in swarm cells of the myxomycete Physarum flavicomum. *J Gen Microbiol,* **50:** 217-222.

Alves A. A., Gabriel H. B., Bezerra M. J. R., Souza W., Vaughan S., Cunha-e-Silva N. L. and Sunter J. D. (2020) Control of assembly of extra-axonemal structures: the paraflagellar rod of trypanosomes. *Journal of Cell Science,* **133**: jcs242271.

Aumais J. P., Williams S. N., Luo W., Nishino M., Caldwell K. A., Caldwell G. A., Lin S. and Yu-Lee L. (2003) Role for NudC, a dynein-associated nuclear movement protein, in mitosis and cytokinesis. *J Cell Sci,* **116**: 1991-2003.

Basten S. G. and Giles R. H. (2013) Functional aspects of primary cilia in signaling, cell cycle and tumorigenesis. *Cilia*, **2:** 6.

Benmerah A., Durand B., Giles R. H., Harris T., Kohl L., Laclef C., Meilhac S. M., Mitchison H. M., Pederson L. B., Roepman R., Swoboda P., Ueffing M. and Bastin P. (2015) The more we know, the more we have to discover: an exciting future for understanding cilia and ciliopathies. *Cilia*, **4**: 5.

Bertocci B., Smet A. D., Flatter E., Dahan A. Bories J., Landreua C., Weill J. and Reynaud C. (2002) Cutting edge: DNA polymerases mu and lambda are dispensable for Ig gene hypermutation. *J Immunol,* **168:** 3702-3706.

Bielas S. L., Silhavy J. L., Brancati F., Kisseleva M. V., Al-Gazali L., Sztriha L., Bayoumi R. A., Zaki M. S., Abdel-Aleem A., Rosti O., Kayserili H., Swistun D., Scott L, C., Bertini E., Boltshauser E., Fazzi E., Travaglini L., Field S. J., Gayral S., Jacoby M., Schurmans S., Dallapiccola B., Majerus P. W., Valente E. M. and Gleeson J. G. (2009) Mutations in the *inositol polyphosphate-5-phosphatase E* gene link phosphatidyl inositol signaling to the ciliopathies. *Nat Genet,* **41**: 1032-1036.

Brancati F., Barrano G., Silhavy J. L., Marsh S. E., Travaglini L., Bielas S. L., Amorini M., Zablocka D., Kayserili H., Al-Gazali L., Bertini E., Boltshauser E., D'Hooghe M., Fazzi E., Fenerci E. Y., Hennekam R. C. M., Kiss A., Lees M. M., Marco E. and Swoboda K. J. (2007) CEP290 mutations are frequently identified in the oculo-renal form of Joubert syndrome-related disorders. *American Journal of Human Genetics*, **81,** 104-113.

Breslow D. K. and Holland A. J. (2019) Mechanism and Regulation of Centriole and Cilium Biogenesis. *Annual Review of Biochemistry,* **8:** 691-724.

Briggs L. J., Davidge J. A., Wickstead B., Ginger M. L. and Gull K. (2004) More than one way to build a flagellum: comparative genomics of parasitic protozoa. *Curr Biol,* **14:** 611-612.

Broadhead R., Dawe H. R., Farr H., Griffiths S., Hart S. R., Portman N., Shaw M. K., Ginger M. L., Gaskell S. J., McKean P. G. and Gull K. (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature,* **440**: 224-227.

Brown J. M. and Witman G. B. (2014) Cilia and diseases. *BioScience*, **64:** 1126-1137.

Bujakowska K. M., Liu Q. and Pierce E. A. (2017) Photoreceptor Cilia and Retinal Ciliopathies. *Cold Spring Harb Perspect Biol*, **9:** a028274.

Bukowy-Bieryllo Z., Rabiasz. A., Dabrowski M., Pogorzelski A., Wojda A., Dmenska H., Grzela K., Sroczynski J., Witt M. and Zietkiewicz E. (2019) Truncating mutations in exons 20 and 21 of OFD1 can cause primary ciliary dyskinesia without associated syndromic symptoms. *J Med Genet*, **56:** 769-777.

Burger A., Ludewig M H. and Boshoff A. (2014) Investigating the Chaperone Properties of a Novel Heat Shock Protein, Hsp70.c, from *Trypanosoma brucei. J Parasitol Res*, **2014:** 172582.

Bustamante-Marin X. M. and Ostrowski L. E. (2017) Cilia and Mucociliary Clearance. *Cold Spring Harb Perspect Biol*, **9**: a028241.

Clement A., Solnico-Krezel L. and Gould K. L. (2011) The Cdc14B phosphatase contributes to ciliogenesis in zebrafish. *Development*, **138:** 291-302.

Corbit K. C., Shyer A. E., Dowdle W. E. Gaulden J., Singla V. and Reiter J. F. (2007) Kif3a constrains β-catenin-dependent Wnt signalling through dual ciliary and non-ciliary mechanisms. *Nat Cell Biol,* **10:** 70-76.

Davis S. D., Ferkol T. W., Rosenfeld M., Lee H., Dell S. D., Sagel S. D., Milla C., Zariwala M. A., Pittman J. E., Shapiro A. J., Carson J. L., Krischer J. P., Hazucha M. J., Cooper M. L., Knowles M. R. and Leigh M. W. (2015) Clinical features of childhood primary ciliary dyskinesia by genotype and ultrastructural phenotype. *Am J Respir Crit Care Med,* **191**: 316-324.

Dasgupta A. and Amack J. D. (2016) Cilia in vertebrate left–right patterning. *Philos Trans R Soc Lond B Biol Sci,* **371:** 20150410

Dean S., Moreira-Leite F., Varga V. and Gull K. (2016) Cilium transition zone proteome reveals compartmentalization and differential dynamics of ciliopathy complexes. *PNAS*, **113**: 135-143.

Dean S., Sunter J. D. and Wheeler R. J. (2017) TrypTag.org: A Trypanosome Genome-wide Protein Localisation Resource. *Trends in Parasitology,* **33:** 80-82.

Delaval B., Bright A., Lawson N. D. and Doxsey S. (2011) The cilia protein IFT88 is required for spindle orientation in mitosis. *Nat Cell Biol,* **13:** 461-468.

Derelle R., Lopez-Garcia P., Timpana H. and Moreira D. (2016) A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Mol Biol Evol*, **33**: 2890-2898.

Droll D., Minia I., Fadda A., Singh A., Stewart M., Queiroz R. and Clayton C. (2013) Post-transcriptional regulation of the trypanosome heat shock response by a zinc finger protein. *PLoS Pathogens,* **9**: e1003286

Duquesnoy P., Escudier E., Vincensini L., Freshour J., Bridoux A. M., Coste A., Deschildre A., de Blic J., Legendre M., Montantin G., Tenreiro H., Vojtek A. M., Loussert C., Clément A., Escalier D., Bastin P., Mitchell D. R. and Amselem S. (2009) Loss-of-function mutations in the human ortholog of Chlamydomonas reinhardtii ODA7 disrupt dynein arm assembly and cause primary ciliary dyskinesia. *Am J Hum Genet.* **85**: 890-896.

Eddy C. A., Flores J. J., Archer D. R. and Pauerstein C. J. (1978) The role of cilia in fertility: an evaluation by selective microsurgical modification of the rabbit oviduct. *Am J Obstet Gynecol,* **132:** 814-821.

Fabczak H. and Osinka A. (2019) Role of the Novel Hsp90 Co-Chaperones in Dynein Arms' Preassembly. *International Journal of Molecular Sciences*, **20**: 6174.

Field H. I., Coulson R. M. R. and Field M. C. (2013) An automated graphics tool for comparative genomics: the Coulson plot generator. *BMC Bioinformatics,* **14:** 141.

Fliegauf M., Benzing T. and Orman H. (2007) When cilia go bad: cilia defects and ciliopathies. *Nature*, **8:** 880-893.

Franco B. and Thauvin-Robinet C. (2016) Update on oral-facial-digital syndromes (OFDS). *Cilia*, **5**: 12**.**

Funfak A., Fisch C., Motaal H. T. A., Diener J., Combettes L., Baroud C. N., Dupuis-Williams P. (2015) *Paramecium* swimming and ciliary beating patterns: a study on four RNA interference mutations. *Integrative Biology,* **7:** 90-100.

Garcia G., Raleigh D. R. and Reiter J. F. (2018) How the ciliary membrane is organized inside-out to communicate outside-in. *Current Biology*, **28**: 421-434.

Hamasaki T. (1999) Regulation of outer-arm-dynein activity by phosphorylation and control of ciliary beat frequency. *Protoplasma*, **206**:241-244.

Hannah W. B., DeBrosse S., Kinghorn B., Strausbaugh S., Aitken M. L., Rosenfeld M., Wolf W. E., Knowles M. R. and Zariwala M. A. (2019) The expanding phenotype of OFD1-related disorders: Hemizygous loss-of-function variants in three patients with primary ciliary dyskinesia. *Mol Genet Genomic Med*, **7:** e911.

Hartill V., Szymanska K., Sharif S. M., Wheway G. and Johnson C. A. (2017) Meckel–Gruber Syndrome: An Update on Diagnosis, Clinical Management, and Research Advances. *Front. Pediatr*, **5**: 244.

Hartman H. and Smith T. F. (2009) The evolution of the cilium and the eukaryotic cell. *Cell Motility,* **66:** 215-219.

Haslbeck M. and Buchner J. (2015) Assays to characterize molecular chaperone function in vitro. *Methods Mol Biol,* **1292**: 39-51.

Hildebrandt F., Benzing T. and Katsnasis N. (2011) Ciliopathies. *N Eng J Med,* **364:** 1533-1543.

Hodges M. E., Wickstead B., Gull K. and Langdale J. A. (2011) Conservation of ciliary proteins in plants with no cilia. *BMC Plant Biology*, **11:** 185.

Hodges M. E., Wiskstead B., Gull K. and Langdale J. A. (2012) The evolution of land plant cilia. *New Phytologist*, **195**: 526-540.

Holm L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Research*, **50**: 210-215.

Horani A., Ferkol T. W., Dutcher S. K. and Brody S. L. (2016) Genetics and Biology of Primary Ciliary Dyskinesia. *Paediatr Respir Rev,* **18:** 18-24.

Huangfu D., Liu A., Rakeman A. S., Murcia N. S., Niswander L. and Anderson K. V. (2003) Hedgehog signalling in the mouse requires intraflagellar transport proteins. *Nature*, **426:** 83-87.

Ibanez-Tallon I., Heintz N. and Omran H. (2003) To beat or not to beat: roles of cilia in development and disease. *Human Molecular Genetics*, **12**: 27-35.

Inaba K. and Mizuno K. (2016) Sperm dysfunction and ciliopathy. *Reprod Med Biol,* **15**: 77-94.

Ishikawa T. (2017) Axoneme structure from motile cilia. *Cold Spring Harb Perspect Biol,* **9:** a028076.

Jekely G. and Arendt D. (2006) Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. *Bioessays,* **28:**191-198.

Jumper, J., Evans, R., Pritzel, A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature,* **596**: 583–589.

Katayama H. M., McGill M., Kearns A., Brzozowski M., Degner N., Harnett B., Kornilayev B., Matkovic-Calogovic D., Holyoak T., Calvet J. P., Gogol E. P., Seed J. and Fisher M. T. (2008) Strategies for folding of affinity tagged proteins using GroEL and osmolytes. *Journal of Structural and Functional Genomics,* **10:** 57-66.

Kelley L. A., Mezulis S., Yates C. M., Wass M. N. and Sternberg M. J. E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **10:** 845-858.

Kelly S., Reed J., Kramer S., Ellis L., Webb H., Sunter J., Salje J., Marinsek N., Gull K., Wickstead B. and Carrington M. (2007) Functional genomics in *Trypanosoma brucei*: a collection of vectors for the expression of tagged proteins from endogenous and ectopic gene loci. *Molecular and Biochemical Parasitology*. **154**:103-109.

Kempen M., Kim S. S., Tumescheit C., Mirdita M., Gilchrist C. L. M., Soding J. and Steinegger M. (2022) Foldseek: fast and accurate protein structure search. *bioRxiv,* **479398.**

Kennedy M. P., Omran H., Leigh M. W., Dell S., Morgan L. Molina P. L., Robinson B. V., Minnix S. L., Olbrich H., Severin T., Ahrens P., Lange L., Morillas H. N., Noone P. G., Zariwala M. A. and

Knowles M. R. (2007) Congenital Heart Disease and Other Heterotaxic Defects in a Large Cohort of Patients With Primary Ciliary Dyskinesia. *Circulation,* **115**: 2814-2821.

Kim N. H., Kim D. S., Chung E. H. and Hwang B. K. (2014) Pepper suppressor of the G2 allele of skp1 interacts with the receptor-like cytoplasmic kinase1 and type III effector AvrBsT and promotes the hypersensitive cell death response in a phosphorylation-dependent manner. *Plant Physiol,* **165**: 76-91.

King N., Westbrook M. J. and Rokhsar D. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature,* **451**: 783-788.

King S. M. (2016) Axonemal dynein arms. *Cold Spring Harb Perspect Biol,* **8:** a028100.

Knodler A., Feng S., Zhang J., Zhang X., Das A., Peranen J. and Guo W. (2010) Coordination of Rab8 and Rab11 in primary ciliogenesis. *Proc Natl Acad Sci USA,* **107**: 6346-6351.

Knowles M. R., Leigh M. W., Carson J. L., Davis S. D., Dell S. D., Ferkol T. W., Olivier K. N., Sagel S. D., Rosenfeld M., Burns K. A., Minnix S. L., Armstrong M. C., Lori A., Hazucha M. J., Loges N. T., Olbrich H., Becker-Heck A., Schmidts M. A., Werner C., Omran H. and Zariwala M. A. (2012) Mutations of *DNAH11* in Primary Ciliary Dyskinesia Patients with Normal Ciliary Ultrastructure. *Thorax,* **67:** 433-441.

Knowles M. R., Leigh M. W., Ostrowski L. E., Huang L., Carson J. L., Hazucha M. J., Yin W., Berg J. S., Davis S. D., Dell S. D., Ferkol T. W., Rosenfeld M., Sagel S. D., Milla C. E., Olivier K. N., Turner E. H., Lewis A. P., Bamshad M. J., Nickerson D. A., Shendure J. and Zariwala M. A. (2013) Exome sequencing identifies mutations in CCDC114 as a cause of primary ciliary dyskinesia. *Am J Hum Genet,* **92:** 99-106.

Knowles M. R., Ostrowski L. E., Loges N. T., Hurd T., Leigh M. W., Huang L., Wolf W. E., Carson J. L., Hazucha M. J., Yin W., Davis S. D., Dell S. D., Ferkol T. W., Sagel S. D., Olivier K. N., Jahnke C., Olbrich H., Werner C., Raidt J., Wallmeier J., Pennekamp P., Dougherty G. W., Hjeij R., Gee H. Y., Otto E. A., Halbritter J., Chaki M., Diaz K. A., Braun D. A., Porath J. D., Schueler M., Baktai G., Griese M., Turner E. H., Lewis A. P., Bamshad M. J., Nickerson D. A., Hildebrandt F., Shendure J., Omran H. and Zariwala M. A. (2013) Mutations in SPAG1 cause primary ciliary dyskinesia associated with defective outer and inner dynein arms. *Am J Hum Genet,* **93**: 711-720.

Kobayashi Y., Watanabe M., Okada Y., Sawa H., Takai H., Nakanishi M., Kawase Y., Suzuki H., Nagashima K., Ikeda K. and Motoyama N. (2002) Hydrocephalus, Situs Inversus, Chronic

Sinusitis, and Male Infertility in DNA Polymerase λ-Deficient Mice: Possible Implication for the Pathogenesis of Immotile Cilia Syndrome. *Molecular and Cellular Biology,* **22**: 2769-2776.

Kyttala M., Tallila J., Salonen R., Kopra O., Kohlschmidt N., Paavole-Sakki P., Peltonen L. and Kestila M. (2006) MKS1, encoding a component of the flagellar apparatus basal body proteome, is mutated in Meckel syndrome. *Nat Genet,* **38**: 155-157.

Lansley A. B. and Sanderson M. J. (1999) Regulation of airway ciliary activity by Ca2+: simultaneous measurement of beat frequency and intracellular Ca2+. *Biophys J,* **77**: 629-638.

Lee G. J., Pokala N. and Vierling E. (1995) Structure and *in Vitro* Molecular Chaperone Activity of Cytosolic Small Heat Shock Proteins from Pea. *Journal of Biological Chemistry*, **270:** 10432-10438.

Lee G. J., Roseman A. M., Saibil H. R. and Vierling E. (1997) A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *The Embo Journal,* **16:** 659-671.

Leigh M. W., Pittman J. E., Carson J. L., Ferkol T. W., Dell S. D., Davis S. D., Knowles M. R. and Zariwala M. A. (2009) Clinical and genetic aspects of primary ciliary dyskinesia/Kartagener syndrome. *Genetics in Medicine,* **11**: 473-487.

Leigh M. W., Horani A., Kinghorn B., OConnor M. G., Zariwala M. A. and Knowles M. R. (2019) Primary ciliary dyskinesia (PCD): A genetic disorder of motile cilia. *Translational Science of Rare Diseases,* **4:** 51-75.

Letcher P. M. and Powell M. J. (2018) A taxonomic summary and revision of *Rozella* (*Cryptomycota*). *IMA Fungus,* **9**: 383-399.

Li D., Yang F., Lu B., Chen D. and Yang W. (2012) Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, *Sulfolobus solfataricus P2. Cell Stress and Chaperones*, **17:** 103-108.

Lin J., Yin W., Smith M. C., Song K., Leigh M. W., Zariwala M. A., Knowles M. R., Ostrowski L. E. and Nicastro D. (2014) Cryo-electron tomography reveals ciliary defects underlying human *RSPH1* primary ciliary dyskinesia. *Nat Commun,* **5:** 5727.

Lobo J., Zariwala M. A. and Noone P. G. (2015) Primary Ciliary Dyskinesia. *Semin Respir Crit Care Med*, **36:** 169-179.

Madeira F., Pearce M., Tivey A., Basutakar P., Lee J., Edbali O., Madhusoodanan N., Kolesnikov A. and Lopez R. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research,* **2022.**

Mali G. R., Yeyati P. L., Mizuno S., Dodd D. O., Tennant P. A., Keighren M. A., Lage P., Shoemark A., Garica-Munoz A., Shimada A., Takeda H., Edlich F., Takahashi S., Kreigsheim A., Jarman A. P. and Mill P. (2018) ZMYND10 functions in a chaperone relay during axonemal dynein assembly. *eLife,* **7**: e34389.

Merveille A., Davis E. E., Becker-Heck A., Legendre M., Amirav I., Bataille G., Belmont J., Beydon N., Billen F., Clement A., Clercx C., Coste A., Crosbie R., Blic J., Deleuze S., Duquesnoy P., Escalier D., Escudier E., Fliegauf M., Horvath J., Hill K., Jorissen M., Just J., Kispert M., Lathrop M., Loges N., Marthin J., Momozawa Y., Montantin G., Nielsen K., Olbrich H., Papon J., Rayet I., Roger G., Schmidts M., Tenreiro H., Towbin J., Zelenika D., Zentgraf H., Georges M., Lequarre A., Katsanis N., Omran H. and Amselem S. (2011) CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nat Genet,* **43:** 72-78.

Mitchison H. M., Schmidts M., Loges N. T., Freshour J., Dritsoula A., Hirst R. A., O'Callaghan C., Blau H., Dabbagh M. A., Olbrich H., Beales P. L., Yagi T., Mussaffi H., Chung E. M., K., Omran H. and Mitchell D. R. (2012) Mutations in axonemal dynein assembly factor DNAAF3 cause primary ciliary dyskinesia. *Nat Genet,* **44:** 381-389.

Mitchison H. M. and Valente E. M. (2016) Motile and non-motile cilia in human pathology: from function to phenotypes. *The Journal of Pathology,* **241**: 294-309

Moore A., Escudier E., Roger G., Tamalet A., Pelosse B., Marlin S., Clement A., Geremek M., Delaisi B., Bridoux A., Coste A., Witt M., Duriez B. and Amselem S. (2006) RPGR is mutated in patients with a complex X linked phenotype combining primary ciliary dyskinesia and retinitis pigmentosa. *J Med Genet,* **43:** 326-333.

Morgan G. W., Denny P. W., Vaughan S., Goulding D., Jeffries T. R., Smith D. F., Gull K., Field M. C. (2005) An Evolutionarily Conserved Coiled-Coil Protein Implicated in Polycystic Kidney Disease Is Involved in Basal Body Duplication and Flagellar Biogenesis in *Trypanosoma brucei. Mol Cell Biol,* **25**: 3774-3783.

Morias R. D. S., Santo P. E., Ley M., Schelcher C., Abel Y., Plassart L., Desligniere E., Chagot M., Quinternet M., Paiva A. C. F., Hessman S., Morellet N., Sousa P. M. F., Vandermoere F., Bertrand E., Charpentier B., Bandeiras T. M., Chastang C. P., Verheggen C., Cianferani S. and

Manival X. (2022) Deciphering cellular and molecular determinants of human DPCD protein in complex with RUVBL1/RUVBL2 AAA-ATPases. *Journal of Molecular Biology,* **434**: 167760.

Nachury M. V., Loktev A. V., Zhang Q., Westlake C. J., Peranen J., Merdes A., Slusarski D. C., Scheller R. H., Bazan J. F., Sheffield V. C. and Jackson P. K. (2007) A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. *Cell,* **129**: 1201-1203.

Nauli S. M., Alenghat F. J., Luo Y., Williams E., Vassilev P., Li X., Elia A. E. H., Lu W., Brown E. M., Quinn S. J., Ingber D. E. and Zhou J. (2003) Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nat Genet,* **33:** 129-137.

Nevers Y., Prasad M. K., Poidevin L., Chenne K., Allot A., Kress A., Ripp R., Thompson J. D., Dollfus H., Poch O. and Lecompte O. (2017) Insights into Ciliary Genes and Evolution from Multi-Level Phylogenetic Profiling. *Molecular Biology and Evolution*, **34**: 2016-2034.

Niederlova V., Modrak M., Tsyklauri O., Huranova M. and Stepanek O. (2019) Meta-analysis of genotype-phenotype associations in Bardet-Biedl syndrome uncovers differences among causative genes. *Hum Mutat*, **40**: 2068-2087.

Noor A., Windpassinger C., Patel M., Stachowiak B., Mikhailov A., Azam M., Irfan M., Siddiqui Z. K., Naeem F., Paterson A. D., Lutfullah M., Vincent J. B. and Ayub B. (2008) CC2D2A, encoding a coiled-coil and C2 domain protein, causes autosomal- recessive mental retardation with retinitis pigmentosa. *Am. J. Hum. Genet.,* **82**: 1011-1018.

Ogi H., Sakuraba Y., Xiao L., Shen C., Cynthia M. A., Ohta S., Arnold M. A., Ramirez N., Houghton P. J. and Kitagawa K. (2015) The oncogenic role of the cochaperone Sgt1. *Oncogenesis,* **18:** e149.

Oguchi M. E., Okuyama K., Homma Y. and Fukada M. (2020) A comprehensive analysis of Rab GTPases reveals a role for Rab34 in serum starvation-induced primary ciliogenesis. *J Biol Chem*, **295:** 12675-12685.

Olbrich H., Haffner K., Kispert A., Volkel A., Volz A., Sasmaz G., Reinhardt R., Hennig S., Lehrach H., Konietzko N., Zariwala M., Noone P. G., Knowles M., Mitchison H. M., Meeks M., Chung E. M. K., Hildebrandt F., Sudbrak R. and Omran H. (2002) Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. *Nat Genet,* **30:** 143-144.

Olbrich H., Schmidts M., Werner C., Nielsen K. G., Mitchison H. M. and Omran H. (2012) Recessive *HYDIN* Mutations Cause Primary Ciliary Dyskinesia without Randomization of Left-Right Body Asymmetry. *Am J Hum Genet,* **91:** 672-684.

Ooi C. P., Benz C. and Urbaniak M. D. (2020) Phosphoproteomic analysis of mammalian infective Trypanosoma brucei subjected to heat shock suggests atypical mechanisms for thermotolerance. *Journal of Proteomics*, **219**: 103735.

Ostrowski L. E., Dutcher S. K. and Lo C. W. (2011) Cilia and Models for Studying Structure and Function. *Proceedings of the American Thoracic Society,* **8:** 423-429.

Pazour G. J., Dickert B. L., Vucica Y., Seeley E. S., Rosenbaum J. L., Witman G. B. and Cole D. G. (2000) Chlamydomonas IFT88 and its mouse homologue, polycystic kidney disease gene tg737, are required for assembly of cilia and flagella. *J Cell Biol*, **151:** 709-718.

Pazour G. J., Agustin J. T. S., Follit J. A., Rosenbaum J. L. and Witman G. B. (2002) Polycystin-2 localizes to kidney cilia and the ciliary level is elevated in orpk mice with polycystic kidney disease. *Curr Biol,* **12:** 378-380.

Praveen K., Davis E. E. and Katsanis N. (2015) Unique among ciliopathies: primary ciliary dyskinesia, a motile cilia disorder. *F1000Prime Rep,* **7:** 36.

Qin H., Wang Z., Diener D. and Rosenbaum J. (2007) Intraflagellar transport protein 27 is a small G protein involved in cell-cycle control. *Curr Biol,* **17:** 193-202.

Reiter J. F. and Leroux M. R. (2017) Genes and molecular pathways underpinning ciliopathies. *Nature Reviews Molecular Cell Biology,* **18:** 533-547.

Rosenbaum J. L. and Witman G. B. (2002) Intraflagellar transport. *Nature*, **3**: 813-825.

Satir P. and Christensen T. (2007) Overview of Structure and Function of Mammalian Cilia. *Annu. Rev. Physiol, 69***:** 377-400.

Satir P. and Christensen T. (2008) Structure and function of mammalian cilia. *Histochem Cell Biol*, **129:** 687-693.

Satir P., Mitchell D. R. and Jekely G. (2008) How did the cilium evolve?. *Curr Top Dev Biol,* **85**: 63-82.

Satir P., Heuser T. and Sale W. S. (2014) A Structural Basis for How Motile Cilia Beat. *Bioscience,* **64**: 1073-1083.

Schroder J. M., Larsen J., Komarova Y., Akhmanova A., Thorsteinsson R. I., Grigoriev I., Manguso R., Christensen S. T., Pedersen S. F., Geimer S. and Pedersen L. B. (2011) EB1 and EB3 promote cilia biogenesis by several centrosome-related mechanisms. *J Cell Sci*, **124:** 2539-2551.

Shapiro A. J., Zariwala M. A., Ferkol T., Davis S. D., Sagel S. D., Dell S. D., Rosenfeld M., Olivier K. N., Milla C., Daniel S. J., Kimple A. J., Manion M., Knowles M. R. and Leigh M. W. (2016) Diagnosis, monitoring, and treatment of primary ciliary dyskinesia: PCD foundation consensus recommendations based on state-of-the-art review. *Pediatr Pulmonol,* **51**: 115-132.

Shoemark A., Dixon M., Beales P. L. and Hogg C. L. (2015) Bardet Biedl syndrome: motile ciliary phenotype. *Chest*, **147**: 764-770.

Shoemark A., Boon M., Brochhausen C., et al. (2020) International consensus guideline for reporting transmission electron microscopy results in the diagnosis of Primary Ciliary Dyskinesia (BEAT PCD TEM Criteria). *Eur Respir J,* **55**: 1900725.

Sinden R. E., Canning E. U. and Spain B. (1976) Gametogenesis and fertilization in Plasmodium yoeliinigeriensis: a transmission electron microscope study. *Proc R Soc Lond B Biol Sci*, **193**: 55-76.

Singla V., Romaguera-Ros M., Garcia-Verdugo J. M. and Reiter J. F. (2010) *Ofd1*, a human disease gene, regulates the length and distal structure of centrioles. *Dev Cell*, **18:** 410-424.

Sleiman S., Marshall A. E., Dong X., Mhanni A., Alidou-D'Anjou I., Frosk P., Marin S. E., Stark Z., Bigio M. R. D., McBride A., Sadedin S., Gallacher L., Christodoulou J., Boycott K. M., Dragon F. and Kernohan K. D. (2022) Compound heterozygous variants in SHQ1 are associated with a spectrum of neurological features, including early-onset dystonia*. Hum Mol Genet*, **31**: 614-624.

Smith U. M., Consugar M., Tee L. J., McKee B. M., Maina E. N., Whelan S., Morgan N. V., Goranson E., Gissen P., Lilliquist S., Aligianis I. A., Ward C. J., Pasha S., Punyashthiti R., Malik Sharif S., Batman P. A., Bennett C. P., Woods C. G., McKeown C., Bucourt M., Miller C. A., Cox P., Algazali L., Trembath R. C., Torres V. E., Attie-Bitach T., Kelly D. A., Maher E. R., Gattone V. H., Harris P. C. and Johnson C. A. (2006) The transmembrane protein meckelin (MKS3) is mutated in Meckel-Gruber syndrome and the wpk rat. *Nat Genet*, *38:*191-6.

Smith C. E. L., Lake A. V. R. and Johnson C. A. (2020) Primary Cilia, Ciliogenesis and the Actin Cytoskeleton: A Little Less Resorption, A Little More Actin Please. *Front. Cell Dev. Biol,* **17:** 622822.

Srivastava M., Simakov O. and Rokhsar D. S. (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature,* **466**: 720-726.

Sutto Z., Conner G. E. and Salathe M. (2004) Regulation of human airway ciliary beat frequency by intracellular pH. *The Journal of Physiology,* **560:** 519-532.

Tarkar A., Loges N. T., Slagle C. E., Francis R., Dougherty G. W., Tamayo J. V., Shook B., Cantino M., Schwartz D., Jahnke C., Olbrich H., Werner C., Raidt J., Pennekamp P., Abouhamed M., Hjeij R., Köhler G., Griese M., Li Y., Lemke K., Klena N., Liu X., Gabriel G., Tobita K., Jaspers M., Morgan L. C., Shapiro A. J., Letteboer S. J., Mans D. A., Carson J. L., Leigh M. W., Wolf W. E., Chen S., Lucas J. S., Onoufriadis A., Plagnol V., Schmidts M., Boldt K., Roepman R., Zariwala M. A., Lo C. W., Mitchison H. M., Knowles M. R., Burdine R. D., Loturco J. J. and Omran H. (2013) DYX1C1 is required for axonemal dynein assembly and ciliary motility. *Nat Genet*, **45:**995-1003.

Tilley A. E., Walters M. S., Shaykhiev R. and Crystal R. G. (2015) Cilia Dysfunction in Lung Disease. *Annu Rev Physiol,* **77:** 379-406.

Torruella G., Moreira D. and Lopez-Garcia P. (2017) Phylogenetic and ecological diversity of apusomonads, a lineage of deep-branching eukaryotes. *Environ Microbiol Rep,* **9**: 113-119.

Tsai C., Chao A., Jung S., Lin C. and Wang T. (2018) Stress-induced phosphoprotein 1 acts as a scaffold protein for glycogen synthase kinase-3 beta-mediated phosphorylation of lysine-specific demethylase 1. *Oncogenesis,* **7**: 31.

Tsang W. Y., Bossard C., Khanna H., Peranen J., Swaroop A., Malhorta V. and Dynlacht B. D. (2008) CP110 Suppresses Primary Cilia Formation through its Interaction with CEP290, a Protein Deficient in Human Ciliary Disease. *Dev Cell,* **15**: 187-197.

Turkyilmaz A., Geckinli B. B., Alavanda C., Ates E. A., Buyukbayrak E. E., Eren S. F. and Arman A. (2021) Meckel-Gruber Syndrome: Clinical and Molecular Genetic Profiles in Two Fetuses and Review of the Current Literature. *Genetic Testing and Molecular Biomarkers*, **24**: 6.

Vanaken G. J., Bassinet L., Boon M., Mani R., Honore I., Papon J., Cuppens H., Jaspers M., Lorent N., Coste A., Escudier E., Amselem S., Maitre B., Legendre M. and Christin-Maitre S. (2017) Infertility in an adult cohort with primary ciliary dyskinesia: phenotype–gene association. *European Respiratory Journal,* **50**: 1700314.

Varadi M., Anyango S., Deshpande M., et al. (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research*, **50:** 439-444.

Vasquez J., Wedel C., Cosentino R. O. and Siegel T. N. (2018) Exploiting CRISPR-Cas9 technology to investigate individual histone modifications. *Nucleic Acids Res*, **46:** e106.

Vaughan S. and Gull K. (2016) Basal body structure and cell cycle-dependent biogenesis in *Trypanosoma brucei. Cilia,* **5**: 5.

Wan K. Y. and Jekeley G. (2020) On the unity and diversity of cilia. *Philos Trans R Soc Lond B Biol Sci,* **375**: 1792.

Waters E. R., Lee G. J. and Vierling E. (1996) Evolution, structure and function of the small heat shock proteins in plants. *Journal of Experimental Botany,* **47:** 325-338.

Waters A. M. and Beales P. L. (2011) Ciliopathies: an expanding disease spectrum. *Pediatr Nephrol,* **26:** 1039-1056.

Werner C., Onnebrink J. G. and Omran H. (2015) Diagnosis and management of primary ciliary dyskinesia. *Cilia,* **4:** 2.

Wheway G., Nazlamova L. and Hancock J. T. (2018) Signalling through the Primary Cilium. *Front. Cell Dev. Biol,* **6:** 8.

Wickstead B., Ersfeld K. and Gull K. (2002) Targeting of a tetracycline-inducible expression system to the transcriptionally silent minichromosomes of *Trypanosoma brucei*. *Molecular and Biochemical Parisitology,* **125**: 211-216.

Yates A.D., Allen J., Amode R. M., Azov A. G., Barba M., Becerra A., Bhai J., Campbell L. I., Martinez M. C., Chakiachvili M., Chougule K., Christensen M., Contreras-Moreira B., Cuzick A., Fioretto L. D. R., Davis P., Silva N. H. D., Diamantakis S., Dyer S., Elser S., Filippi C. V., Gall A., Grigoriadis D., Guijarro-Clarke C., Gupta P., Hammond-Kosack K. E., Howe K. L., Jaiswal P., Kaikala V., Kumar V., Kumari S., Langridge N., Le T., Luypaert M., Maslen G. L., Maurel T., Moore B., Muffato M., Mushtaq A., Naamati G., Naithani S., Olson A., Parker A., Paulini M., Pedro H., Perry E., Preece J., Quinton-Tulloch M., Rodgers F., Rosello M., Ruffier M., Seager J., Sitnik V., Szpak M., Tate J., Tello-Ruiz M. K., Trevanion S. J., Urban M., Ware D., Wei S., Williams G., Winterbottom A., Zarowiecki M., Finn R. D. and Flicek P. (2022) Ensembl Genomes 2022: an expanding genome resource for non-vertebrates. *Nucleic Acids Research*, **50**: 996-1003.

Ye Y. and Godzik A. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, **19**: 246-255.

Yoshida H., Ishida S., Yamamoto T., Ishikawa T., Nagata Y., Takeuchi K., Ueno H. and Imai Y. (2022) Effect of cilia-induced surface velocity on cerebrospinal fluid exchange in the lateral ventricles. *J. R. Soc. Interface,* **19**: 20220321

Yoshimura S., Egerer J., Fuchs E., Haas A. K. and Barr F. A. (2007) Functional dissection of Rab GTPases involved in primary cilium formation. *J Cell Biol,* **178:** 363-369.

Yuan S., Wang Z., Peng H., Ward S. M., Hennig G. W., Zheng H. and Yan W. (2021) Oviductal motile cilia are essential for oocyte pickupbut dispensable for sperm and embryo transport. *PNAS,* **118:** e2102940118.

Zabka M., Lesniak W., Prus W., Kuznicki J. and Filipek A. (2008) Sgt1 has co-chaperone properties and is up-regulated by heat shock. *Biochemical and Biophysical Research Communications*, **370**: 179-183.

Zariwala M., Noone P. G., Sannuti A., Minnix S., Zhou Z., Leigh M. W., Hazucha M., Carson J. L. and Knowles M. R. (2001) Germline mutations in an intermediate chain dynein cause primary ciliary dyskinesia. *Am J Respir Cell Mol Biol,* **25**: 577-583.

Zariwala M., O'Neal W. K., Noone P. G., Leigh M. W., Knowles M. R. and Ostrowski L. E. (2004) Investigation of the Possible Role of a Novel Gene, DPCD, in Primary Ciliary Dyskinesia. *Am J Respir Cell Mol Biol*, **30:** 428-434.

Zariwala M. A., Knowles M. R. and Leigh M. W. (2007) Primary Ciliary Dyskinesia. *GeneReviews® [Internet].*

Zariwala M. A., Knowles M. R. and Omran H. (2007) Genetic defects in ciliary structure and function. *Annu Rev Physiol,* **69**: 423-450.

Zuo X., Fogelgren B. and Lipschutz J. H. (2011) The small GTPase Cdc42 is necessary for primary ciliogenesis in renal tubular epithelial cells. *J Biol Chem,* **286:** 22469-22477.