Structural & Biochemical Characterisation of Cdc25C: A Dual Specificity Phosphatase

by

NAZIA AKHTAR

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Supervisors:
Dr Mark Jeeves and Professor Michael Overduin

School of Cancer Sciences
College of Medical and Dental Sciences
University of Birmingham
May 2014

Abstract

The dual specificity Cdc25 phosphatases regulate mitosis and are expressed in eukaryotes. The active site motif of the Cdc25 phosphatases is in common with other protein tyrosine phosphatases. However, unlike the classical tyrosine phosphatases, Cdc25 proteins can dephosphorylate phospho-threonine in addition to phospho-tyrosine and have a much shallower active site.

Increased expression of Cdc25 is correlated with poor prognosis in a range of cancers. In particular, increased expression of Cdc25C has been associated with prostate cancer making this protein an attractive target for drug discovery.

However, drug discovery for these proteins has been hampered due to the shallow nature of the active site, difficulty in identifying specific inhibitors and toxicity. The thesis aim was to structurally and biochemically characterize the Cdc25C protein in order to aid future drug design.

The regulatory domain was found to be flexible with limited secondary structure by nuclear magnetic resonance, circular dichroism spectroscopy, and small-angle X-ray scattering.

The catalytic domain had poor solubility at high concentrations therefore extensive construct and solution optimisations were carried out in order to improve the solubility of this domain. Different construct lengths of the catalytic domain were tested in different solution conditions. An optimal catalytic domain construct ($Cdc25C_{270-443}$) was identified and it was found cleavage of the N-terminal His tag further improved protein solubility. The additives 200 mM L-arginine and 200 mM sucrose were also shown to improve the long term solubility of this domain without affecting protein conformation. The ^{1}H, ^{15}N- HSQC spectrum revealed a folded protein with well dispersed peaks and it was also shown to be monomeric by AUC. Although, a Cdc25C-inhibitor complex structure could not be obtained a number of Cdc25C inhibitor compounds were identified by WaterLOGSY and ^{1}H, ^{15}N-HSQC.

Dedication

To my Family

Acknowledgements

Before I begin by individually acknowledging key people who have helped me complete my thesis I would like to thank everyone who has helped me in any shape or form to reach this stage of my life.

I would like to start by thanking my parents for their complete support throughout my education. If it was not for your sacrifices, encouragement, support, and dedication I would not be writing my acknowledgements section today! I would like to especially thank my mum from whom I learned never to give up. You have always been a great example to me; the way you approach every situation in life no matter how challenging with calm, patience, and a great deal of perseverance. Your positivity has helped me stay positive especially when experiments did not work which happened quite often. I would like to thank you for listening to my daily frustrations and for always believing in me.

I would like to thank my aunties who are more like my sisters for always cheering me up and making my dull days lively. I would like to acknowledge my Grandma who unfortunately passed away last year for keeping me on track by asking the question when I will finish (hopefully soon!). She was an amazing lady who I miss dearly.

I would like to thank Professor Janet Lord (University of Birmingham) for making my summer research project during my undergraduate studies so much fun that it motivated me to apply for a PhD! Your enthusiasm and passion for science was a great inspiration.

I would like to thank both my supervisors Professor Michael Overduin and Dr Mark Jeeves for being great mentors during my PhD. I would like to thank you for your support and unwavering patience, especially while I was writing my thesis. Thank you for the fast turnaround on my corrections and apologies for handing in late drafts. I would like to especially thank Dr Mark Jeeves for offering to do corrections while on his holiday! I would like to thank you for always being there and never refusing a question or a task no matter how inconvenient. I would also like to thank Dr Fiyaz Mohammed (University of Birmingham) for his help during my crystallization experiments.

I would like to acknowledge the Overduin group members for providing a positive and friendly environment and who have become like family over the years. I will really miss you all! I would especially like to thank Sandya for always asking if I had lunch and being concerned about my welfare. I would like to thank Pooja for always being happy to help. I would like to take a moment and thank Dr Caezar Al-Jassar (Zee) for his very helpful advice over the years and help with my SAXS analysis. I would like to thank Piv for listening to my frustrations while I was trying to finish my thesis over the last month. I would also like to thank Dr Mark Lenoir for putting up with me in the office. Finally, I would like to acknowledge the HWB-NMR staff for their help with the NMR spectrometers.

Contents

Abbreviations

AUC   Analytical ultracentrifugation

Cdc25   Cell division cycle 25

CD   Circular Dichroism

DMAB   Dimethylamine borane

EOM   Ensemble Optimization Method

FID   Free Induction Decay

GST   Glutathione S-transferases

HEPES   (4-(2-hydroxythyl)-1-peperazineethanesulphonic acid)

HMQC   Heteronuclear Multiple Quantum Correlation

HSQC   Heteronuclear Single Quantum Coherence

IPTG   Isopropyl-β-D-1-thiogalactopyranoside

NaCl   Sodium Chloride

NMR   Nuclear Magnetic Resonance

OD   Optical Density

PBS   Phosphate Buffered Saline

ppm   Parts per million

SAXS   Small angle X-ray scattering

SOFAST   Band-Selective Optimized Flip Angle Short Transient

TCEP   Tris(2-carboxyethyl)phosphine

TEV   Tobacco etch virus (protease)

T_m Melting temperature

WaterLOGSY   Water-ligand observed via gradient spectroscopy

Chapter 1.0 - Introduction

1.1 Cdc25 Biology

1.1.1 Phosphorylation

The phosphorylation of proteins is a key post translational event involved in cell signalling (Hunter, 2000; Pawson and Nash, 2003; Schlessinger, 2000; Hunter, 2009). This reversible process is regulated by the balancing actions between protein kinases and phosphatases. Phosphorylation mainly occurs on serine, tyrosine, and threonine residues. In the cell, serine (86 %) and threonine (12 %) phosphorylation is more common compared to tyrosine phosphorylation (2 %) (Olsen et al., 2006). However, this does not mean tyrosine phosphorylation is not important. Tyrosine phosphorylation is increased upon cell stimulation and has important signalling roles in cell communication, movement, and transport (Pao et al., 2007; Paul and Lombroso, 2003; Meng et al., 2000). The importance of tyrosine phosphorylation in human physiology is clear especially in cancer when the regulation of tyrosine phosphorylation becomes abnormal (Tonks, 2006; Rudolph et al., 2004).

1.1.2 Protein Tyrosine Phosphatase (PTP) Superfamily

Historically, it was thought that protein tyrosine phosphatases (PTPs) were "housekeeping" enzymes with no real specificities. Therefore, most of the research in the past in the protein phosphorylation field has been dedicated to protein tyrosine kinases (PTKs) (Hubbard and Till, 2000; Hunter, 2000; Robertson et al., 2000; Robinson et al., 2000) . It is now apparent that this is not the case and in hindsight protein phosphorylation is much more complicated than what was perceived. In fact, PTPs have a much more important and bigger role to play then was

previously thought. Similar to protein tyrosine kinases, PTPs are also tightly regulated. They are not indiscriminate and are considered to be highly selective and have specific substrates.

There are a total of 107 human genes which encode for all the phosphatases in the PTP superfamily (Alonso et al., 2004). In general, members of the PTP superfamily share common features in that they have a multi-domain architecture, conserved active site motif and a similar catalytic mechanism (Alonso et al., 2004; Tonks, 2006; Andersen et al., 2001). This PTP superfamily has been classified based on substrate specificity into four groups (Figure 1.1.2). The Class I group is the largest and consists of the classical PTPs and the VH1-like dual specificity phosphatases. The classical PTPs, which are tyrosine specific, can be further subdivided into the receptor-like PTPs (RPTPs) and the nonreceptor PTPs (NRPTPs). The VH1-like dual specificity phosphatases can also be further subdivided but into seven different families which includes the PRLs, atypical Dual Specificity Phosphatases (DSPs) and myotubularins; making this group very diverse in substrate specificity. The other three groups are the class II, III, and the aspartate based PTPs. Class II has only one member, the low molecular weight PTP (LMPTP) and class III is composed of the cell division cycle 25 (Cdc25) dual specificity phosphatases.
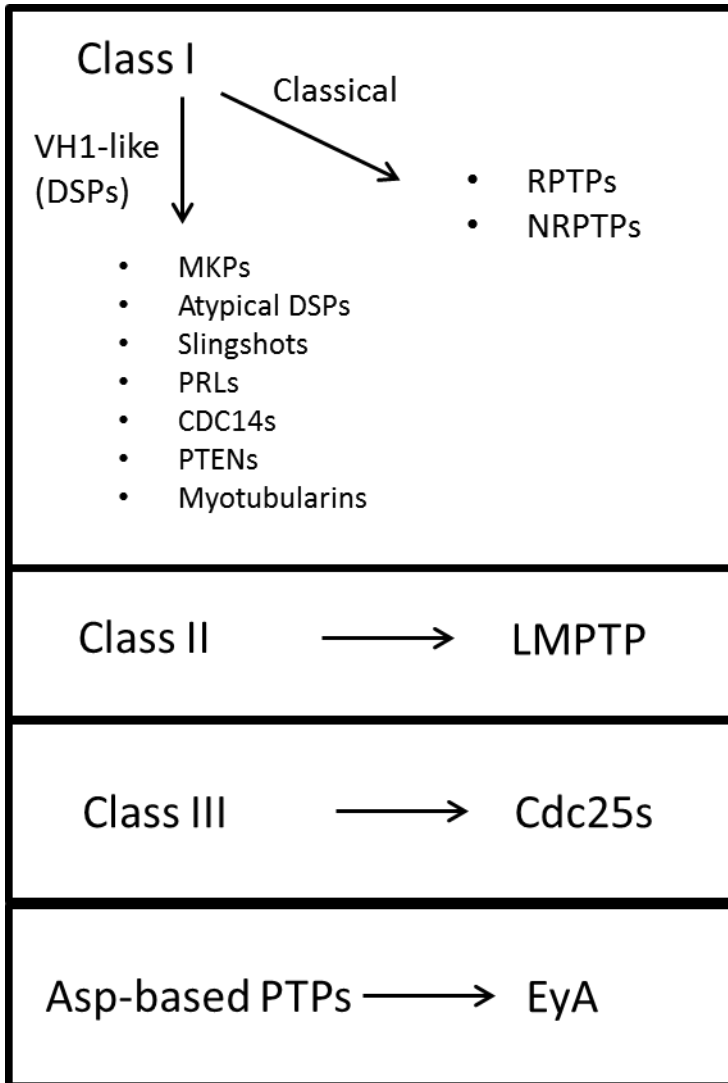
Figure 1.1.2 PTP Classification

Protein tyrosine phosphatases can be classified into four main groups. The first group known as Class I can be further divided into two groups, Vaccinia virus gene H1 protein (VH1) - like dual specificity phosphatases and the classical tyrosine phosphatases. Class II has one member, the low molecular weight protein tyrosine phosphatase (LMPTP). Class III contains the cell division cycle 25 (Cdc25) proteins and the last group is known as the aspartate-based protein tyrosine phosphatases group which has the Eya proteins. This Figure is adapted from (Alonso et al., 2004).

1.1.3 Cdc25 Phosphatases

The Cdc25 DSPs were some of the first dual specificity phosphatases to be discovered (Russell and Nurse, 1986; Galaktionov and Beach, 1991). Cdc25C was first discovered in yeast and it was not long before it was found that the *Cdc25* gene is well conserved in all eukaryotic organisms (Sadhu et al., 1990; Draetta and Eckstein, 1997; Khadaroo et al., 2004). In humans the Cdc25 DSP family consists of three genes located on chromosomes 3, 20, and 5 which express the homologues Cdc25A, B, and C respectively (Alonso et al., 2004).

1.1.4 Role in the Cell Cycle

The three Cdc25 homologues A, B, and C regulate cell cycle checkpoints by activating Cdk/cyclin complexes which are inactive due to the kinases WEE1 and MYT1 (Figure 1.1.4). Cdc25A is thought to promote entry into the S phase of the cell cycle by activating Cdk2 (Hoffmann et al., 1994; Blomberg and Hoffmann, 1999). Cdc25B accumulates in late S phase and is present during mitosis while Cdc25C is expressed in the late G_2 phase. Cdc25B and C regulate mitosis and therefore have roles in late G_2 and mitosis (Karlsson et al., 1999; De Souza et al., 2000). Both Cdc25B and Cdc25C dephosphorylate Cdk1 activating the Cdk1/CycB complex. Interestingly, Cdc25A has also been shown to activate Cdk1/CycB and the Cdc25 homologues B and C have been shown to be involved in promoting entry into S phase (Lindqvist et al., 2005; Boutros et al., 2006). Therefore, it has been suggested that the three Cdc25 phosphatases A, B, and C may work together in regulating the cell cycle and their role is not necessarily fixed to parts of the cell cycle. The fact that the *Cdc25* genes can undergo alternative splicing resulting in a number of splice variants for Cdc25 A, B, and C creates an

extra layer of complexity in defining their roles (Baldin et al., 1997; Wegener et al., 2000; Forrest et al., 1999).

1.1.5 Regulation

It is not surprising that the Cdc25 protein phosphatases are tightly regulated both spatially and temporally considering their integral role in the cell cycle. Their rate of production, localisation, and activity are all regulated.

A tight balance is kept between Cdc25 production and break down. The degradation of Cdc25A is mediated through ubiquitin proteolysis (Donzelli et al., 2002; Jin et al., 2003). This takes place when mitosis is complete. Another degradation process involving the Skp1/Cul1/F-box (SCF) complex is employed for the degradation of Cdc25A during the G_2 and S stages of the cell cycle (Busino et al., 2003). Cdc25B and C are also degraded via the ubiquitin degradation pathway.

The subcellular localisation of the Cdc25 phosphatases is regulated partly by interactions with the 14-3-3 proteins. 14-3-3 proteins are multifunctional regulatory proteins which are thought to be evolved from the TPR (Tetratrico Peptide Repeat) family (Fu et al., 2000; Sluchanko and Gusev, 2010). Phosphorylation by CHK1 and/or C-TAK1 kinases of a serine on the Cdc25 phosphatases enables binding to 14-3-3 proteins by exposing a 14-3-3 binding site (Conklin et al., 1995; Peng et al., 1998; Chen et al., 2003). Specifically, for Cdc25C, the phosphorylation of its serine residue 216 causes the protein to be sequestered in the cytoplasm (Peng et al., 1998). Cdc25C is then activated upon mitosis following the dissociation of 14-3-3. Cdc25 localisation is also regulated by nuclear localisation sequences (NLS) and nuclear export signals (NES) which are located in the N-terminal regulatory domain of the Cdc25 phosphatases (Dalal

et al., 1999; Kumagai and Dunphy, 1999; Uchida et al., 2004). These signals allow shuttling of the Cdc25 proteins from the cytoplasm to the nucleus and *vice versa*.

The N-terminal regulatory domains of the Cdc25 phosphatases possess a number of phosphorylation sites that are phosphorylated by kinases which regulate the catalytic activity of the Cdc25 proteins. Table 1.1.5 lists the kinases that phosphorylate known phosphorylation sites on Cdc25C. Cdc25C is hyperphosphorylated before the induction of mitosis increasing its activity for its substrate, the CDK1/CycB complex (Trunnell et al., 2011; Goulev and Charvin, 2011). Following activation this complex further activates Cdc25C via a positive feedback loop. When phosphorylated the Cdc25B phosphatase exhibits an increase in activity during cell cycle progression from the G_2 to M phase.

1.1.6 The G_2/M checkpoint and DNA damage

The ATM (ataxia-telangiectasia mutated) and ATR (ATM-related) signalling pathways regulate the G_2/M checkpoint response to DNA damage (Bulavin et al., 2001; Xiao et al., 2003; Liu et al., 2000). In response to DNA damage, for example UV damage, CHK1 and CHK2 kinases are activated. These kinases result in the phosphorylation and hence inactivation of the Cdc25 phosphatases. For example, phosphorylation by CHK1 can lead to the binding of 14-3-3 to Cdc25 and ultimately Cdc25 becomes inaccessible to CDK1/CycB. This allows the cell adequate time to repair the damage before the cell progresses into mitosis. In addition to the ATM and ATR pathways the p38 MAPK (mitogen activated protein kinase) pathway has also been described in mediating the G_2/M checkpoint response to DNA damage.
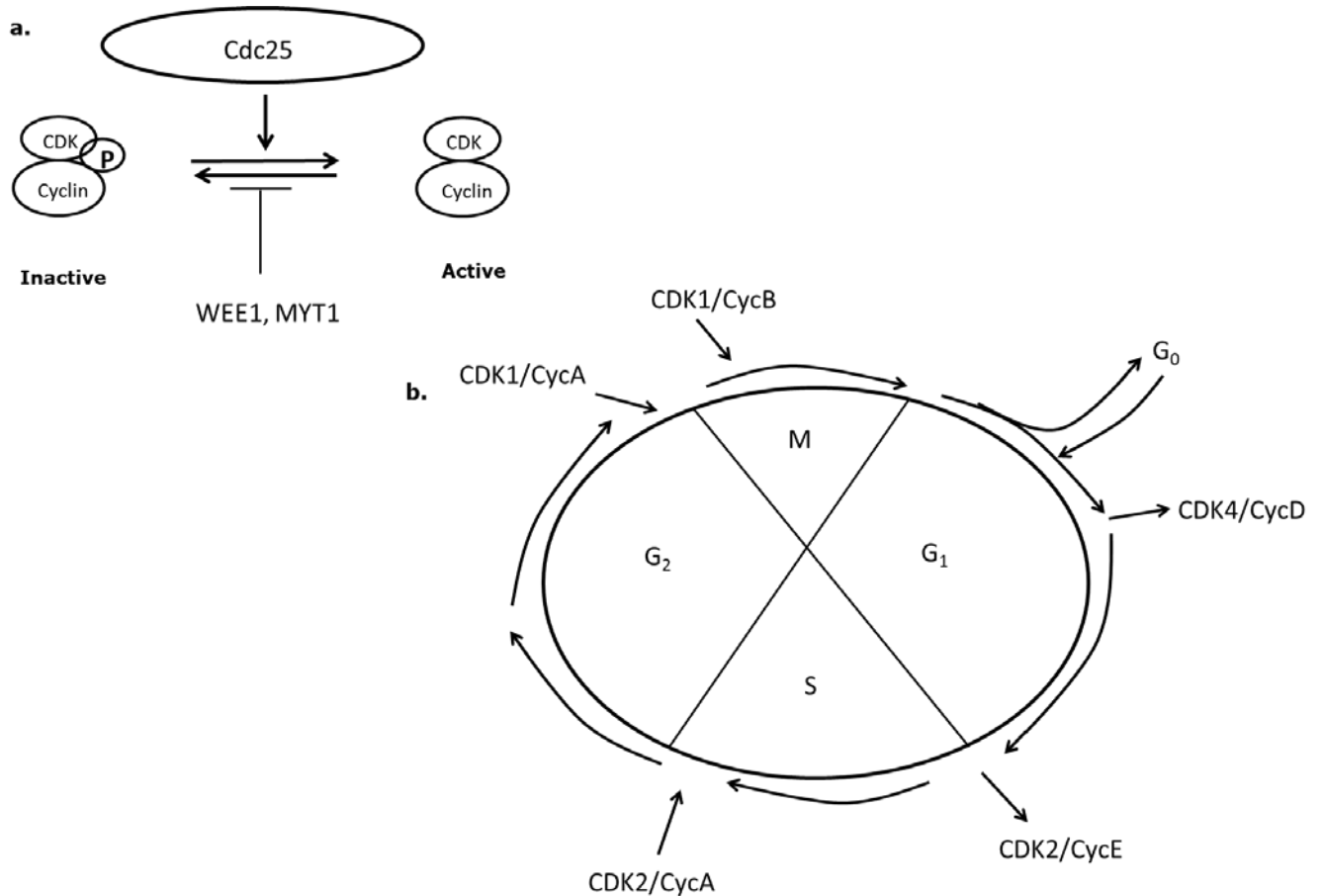
Figure 1.1.4 Activation of the cell cycle by the Cdc25 phosphatases

The CDK/cyclin complexes are regulated by phosphorylation. They are inactivated by WEE1 and MYT1 kinases and activated by the Cdc25 phosphatases (a). The cell cycle can be split into four phases (b). The cell prepares for S phase during the G_1 phase and prepares for mitosis during the G_2 phase. DNA is replicated in S phase and the cell divides during mitosis. Cells in G_1 can enter G_0 which is the resting state. CDK/cyclins promote the transition from one phase to the next of the cell cycle following activation by Cdc25 proteins. Cdc25A promotes entry into S phase while Cdc25B and C promote mitosis. However, evidence suggests that all three Cdc25 phosphatases co-operate together throughout the cell cycle in regulating cell cycle transitions.

Kinase	Phosphorylation site
CDK1/cyclin B	T48, T67, S122, T130, S168, S214
Plk1	S198
Plk3	S191, S198
Chk1	S216, S247, S263
Chk2	S216
C-TAK1	S216
MEK/ERK	S216

Table 1.1.5 Cdc25C phosphorylation sites

A list of some of the kinases which phosphorylate known phosphorylation sites on Cdc25C. Adapted from (Young, 2012).

1.2 Cdc25 Structure, catalytic mechanism, and interaction with substrate

1.2.1 Cdc25 structure

The Cdc25 proteins are 30 - 70 kDa in size. They can be split into two domains: the regulatory and catalytic (Figure 1.2.1.1). Between the homologues, the N- terminal regulatory domain is highly variable in comparison to the C- terminal catalytic domain which has a much more conserved amino acid sequence with a pairwise identity of ~ 60 % (Rudolph, 2007). The N-terminal domain is important for not only regulation but also protein localisation while the catalytic domain is important for the enzymatic function of the protein. Each Cdc25 homologue has splice variants because the N-terminal domain is subjected to alternative splicing; Cdc25C which was the first human homologue to be discovered has five splice variants. Little is known about the role of Cdc25 splice variants.

The crystal structures of the Cdc25 phosphatases A (PDB: 1C25), B (PDB: 1QB0), and C (PDB: 3OP3) have been solved (Fauman et al., 1998; Reynolds et al., 1999a). However, these crystal structures are only of the catalytic domains alone and not the full-length proteins. Also, none of these phosphatases have been solved with a ligand or an inhibitor. Overall, the Cdc25 catalytic domains have a similar structure which is basically a central layer of β-sheets surrounded by α-helices. The crystal structure of Cdc25B is shown in Figure 1.2.1.2a**.**

Interestingly, the catalytic domain of Cdc25A cannot bind to oxyanions while the catalytic domain of Cdc25B can easily bind to sulfate and tungstate (Reynolds et al., 1999). Another key difference between the crystal structures is in the C-terminal region. The C-terminal region of Cdc25A and C is undefined due to a lack of electron density. However, the C-terminal region (Cdc25B$_{531-547}$) of Cdc25B contains a well-defined α-helix.

The Cdc25 dual specificity phosphatases (DSPs) share the conserved active-site signature motif with the PTP superfamily and have a similar catalytic mechanism. However, they differ in that they can dephosphorylate two phosphorylated residues (in the case of Cdc25, p-Thr and p-Tyr) rather than one. They also have a much shallower active site which most likely accounts for the divergent specificity.

The phosphate binding loop of Cdc25, which contains the PTP signature motif, can be superimposed onto other PTPs (Figure 1.2.1.2b). The overall fold adopted by the catalytic domain remains unique to this protein phosphatase when compared to other phosphatases. Interestingly, it is very similar to the one adopted by the rhodanese enzyme. A structural alignment between a rhodanese protein (GlpE) and Cdc25A revealed an rmsd of 1.80 Å (> 96 Cα pairs) (Spallarossa et al., 2001). However, rhodanese shares weak sequence similarity with the Cdc25 DSPs and has a completely different function; rhodanese is engaged during cyanide detoxification (Hofmann et al., 1998).
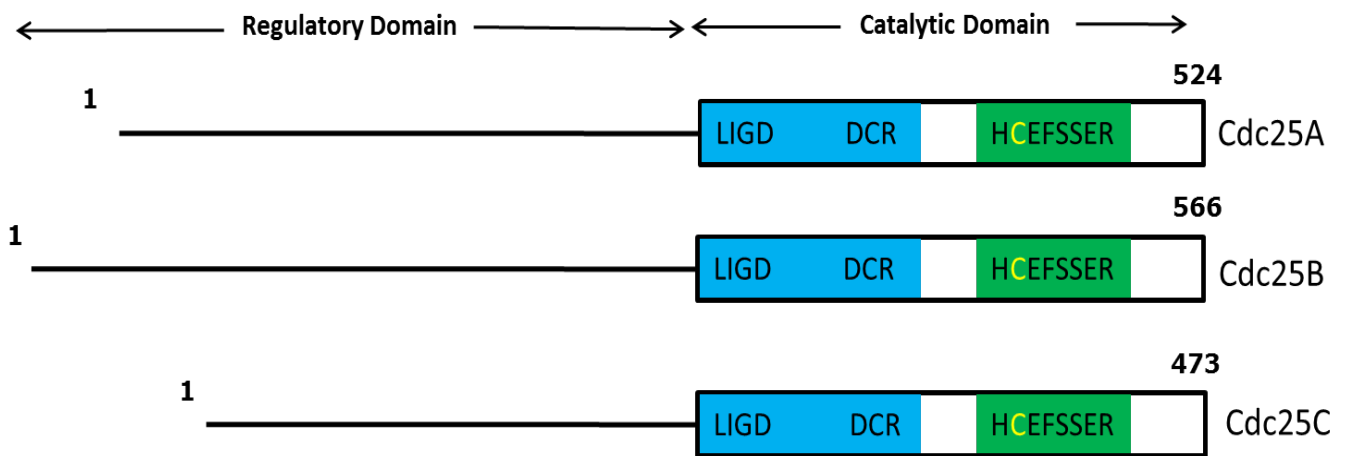
Figure 1.2.1.1 Cdc25 Proteins

The Cdc25 dual specificity phosphatases can be split into two domains. The regulatory domain is involved in the regulation of the protein and the catalytic domain is important for enzymatic activity. Each Cdc25 homologue has a different size; Cdc25C is the smallest among the homologues. This figure illustrates the conserved motifs in the catalytic domain. The highly conserved active site motif is in green and the other two motifs are in blue. For each homologue the catalytic cysteine residue in the active site motif is highlighted in yellow.
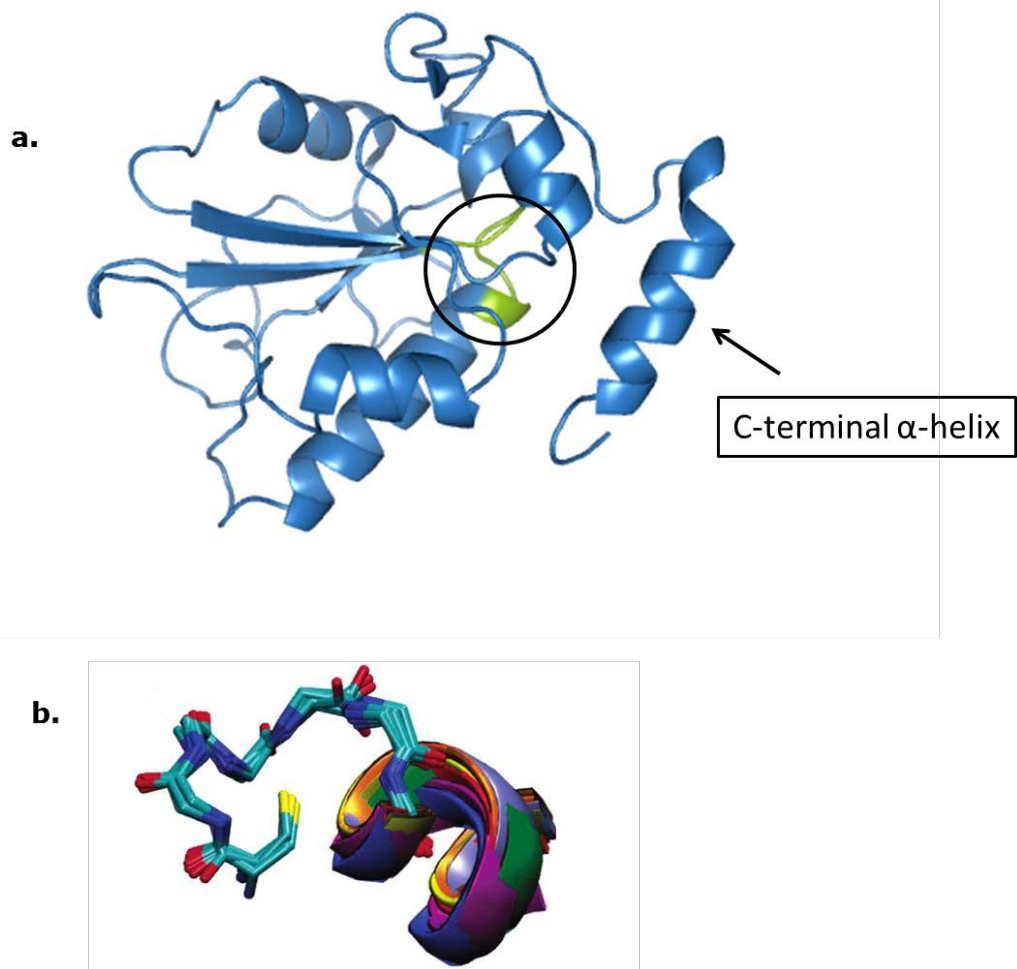
Figure 1.2.1.2 Crystal structure of the Cdc25B catalytic domain

The crystal structure of the Cdc25B (PDB: 1QB0) homologue is shown as a ribbon diagram (a). The protein is in blue and the active site residues (HCX$_5$R) are highlighted in green. The C-terminal α-helix is indicated by an arrow. The active site region (black circle) is shown in more detail (b). The active site region of Cdc25B is overlaid with other protein tyrosine phosphatases (PDBs: 1D5R, 1FPZ, 1PHR, 1I9S, 1OHC, 1I57, 1YTS, 1VHR, 1GWZ, and 2SHP). This Figure is adapted from (Rudolph, 2007)

1.2.2 Catalytic mechanism

The Cdc25 proteins share an active site signature motif (HCX$_5$R) with other PTPs. The cysteine which is critical for catalysis is followed by five amino acids and an arginine. The amide backbone of this arginine and the five amino acids collectively form the active site loop (Rudolph et al., 2004).

Essentially, the proposed catalytic mechanism (Figure 1.2.2) of Cdc25 is similar to that of the PTPs which requires the catalytic cysteine to have a low pKa in order to enhance its role as a nucleophile; Cdc25 has a pKa of 5.9 (Rudolph, 2002). This cysteine nucleophile attacks the phosphate of the substrate resulting in the break-down of the phosphorus-oxygen bond.

For PTPs (D181 in PTP1B), the aspartic acid located in the WPD loop is the general acid which provides a proton to the tyrosyl leaving group (Tonks, 2006). In the case of Cdc25 DSPs that do not have a WPD loop, the identity of the catalytic acid is unknown. There is speculation in the Cdc25 field that this elusive residue may be located on the target substrate rather than the protein itself (Chen et al., 2000). This could provide an explanation of Cdc25 specificity for its substrate taking into account the shallow and featureless nature of the active site (Fauman et al., 1998; Reynolds et al., 1999a).

Completion of the first step of catalysis results in the release of a substrate which has now been dephosphorylated and the formation of a thiophosphate intermediate. This intermediate is subsequently hydrolysed in the second step of the catalytic mechanism to release free phosphate. The Q-loop in the classical PTPs helps coordinate the water molecule involved in this process. The Cdc25 DSPs do not have a Q-loop. In fact the only structural feature they have in common to the classical PTPs is the active site loop (P-loop).

The invariant arginine in the active site loop helps in coordinating the phosphate group when the substrate binds and importantly stabilizes the thiophosphate intermediate. For PTPs, the second step is aided by the same aspartic acid mentioned above now taking the role of a general base.

1.2.3 Catalytic acid

Crystal structure studies of Cdc25A and B revealed the possibility of Glu431 or Glu435 for Cdc25A and Glu474 or Glu478 for Cdc25B of acting as the general acid due to their likely proximity to the leaving group (Fauman et al., 1998; Reynolds et al., 1999a). However, later studies either ruled these residues out or were not able to confidently confirm this (McCain et al., 2002; Chen et al., 2000). Despite many studies the identity of the catalytic acid which promotes the catalytic reaction by acting as both a general acid and a base during catalysis is still not known.
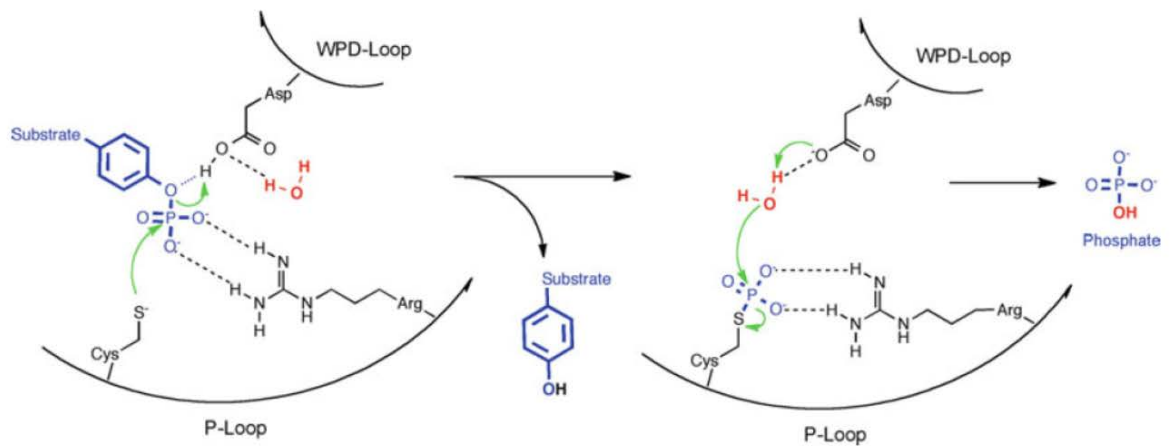
Figure 1.2.2 The Catalytic Mechanism

This catalytic mechanism is shared by all PTPs. Figure taken from (Tautz et al., 2013).

1.2.4 Substrate Recognition and the Cdc25: CDK/Cyclin Interface

Sohn et al. have completed a detailed study to describe the interaction of Cdc25B during catalysis with its substrate Cdk2-pTpY/Cyc A (Gottlin et al., 1996; Sohn and Rudolph, 2007; Sohn et al., 2004, 2005; Sohn and Rudolph, 2006). They have combined structural information with site directed mutagenesis and kinetic studies to not only shed light into this kinetic mechanism but they have also produced an experimentally validated docked model for this protein-protein interaction. It can be assumed that the other two Cdc25 phosphatases will recognize and interact with their biological substrates during enzyme catalysis in a similar manner even though no detailed studies on these phosphatases have yet been conducted.

The first step of the three step kinetic model described by Sohn et al. involves the formation of the enzyme-substrate complex. The rates of both association and dissociation dictate the formation of this complex. This association between Cdc25B and its substrate is key in providing substrate specificity. The flat nature of the active site suggests it contributes very little to the recognition of Cdc25 with its protein substrate.

Cdc25B makes use of a site 20-30 Å away from the active site (Sohn et al., 2004). Evolutionarily conserved arginine residues 488, 492, and the tyrosine residue 497 located at this "remote docking site" on Cdc25B play an important role in the specific recognition of this protein with its native substrate Cdk2-pTpY/CycA (Sohn et al., 2004, 2005). These amino acids are collectively known as the "remote hot-spot residues". These remote residues do not affect the activity of this Cdc25 towards non-native small molecule substrates. They interact with Asp206 and Asp210 on Cdk2, affecting the activity of Cdc25B towards its native substrate (Sohn et al., 2004, 2005). Arg492 of Cdc25B was shown to be central in this interaction (Sohn et al., 2005; Sohn and Rudolph, 2006). Sohn et al. believe that the remote docking site is important for

substrate association while the active site contributes to product dissociation hence both sites are engaged in substrate recognition. All the hot-spot residues when mutated had reduced rates of association (Sohn et al., 2007). Mutating the active site Cys473 to an aspartic acid resulted in a greater rate of dissociation. Interestingly, the crystal structure of this mutant showed that it had a different conformation of the active site loop. This conformation is not suited to phosphate binding. This weak phosphate binding is most likely due to the loss of interactions with the active site. The authors claim this reflects the dissociation of the product during enzyme catalysis. Dephosphorylation of the substrate will result in the loss of productive interactions with the active site and the substrate will therefore dissociate. The partially phosphorylated (Cdk2-TpY/CycA) substrate will then need to re-associate to complete the dephosphorylation reaction.
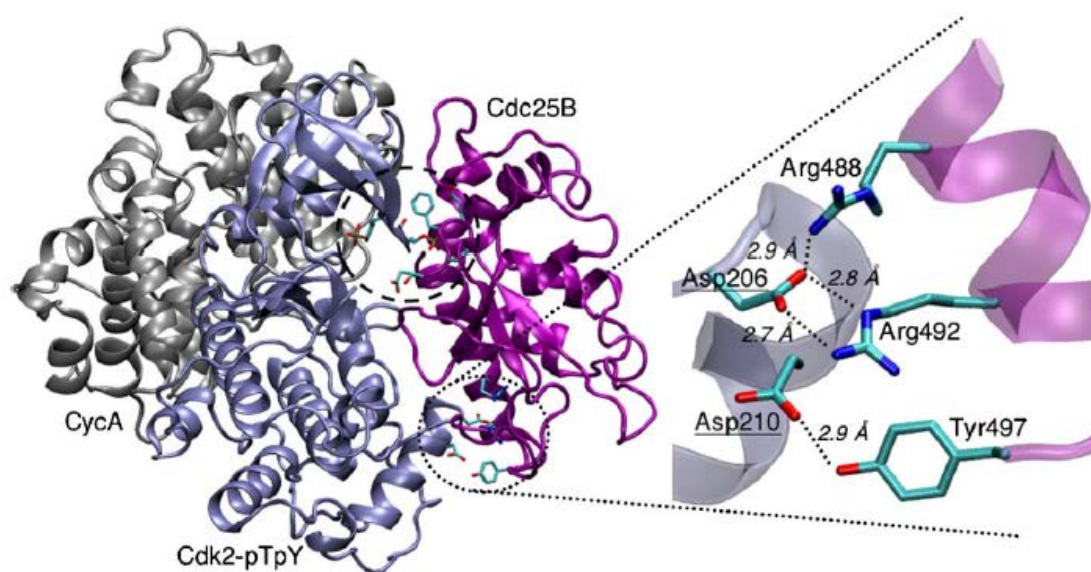
Figure 1.2.4 Recognition interface between Cdc25B and Cdk2-pTpY/CycA

This Figure which has been taken from (Sohn and Rudolph, 2006) *illustrates the remote docking site (see text for details). Cdc25B is in purple, Cdk2 is in blue, and CycA is in gray. The catalytic site (broken circle) and remote hotspot (dotted circle) are circled.*

1.2.5 Substrate Specificity

Cdc25 proteins exhibit poor specificity and activity with phospho-peptide substrates in comparison with their native substrate (Rudolph et al., 2001). This is probably not surprising considering the broad active site interface and the importance of engaging the alternate site discussed above for substrate recognition. For native substrates, pThr is preferred over pTyr. This preference is reversed with phosphopeptides as substrates.

P-nitrophenyl phosphate is a poor Cdc25 substrate with a k_{cat}/K_m of 15-25 M^{-1} s^{-1} for Cdc25A compared to 3-*O*-methylfluorescein phosphate (OMFP) which is a better artificial substrate with a k_{cat}/K_m of 1.1-1.3 x 10^4 M^{-1} s^{-1} (Rudolph et al., 2001).

1.3 Cdc25 and Disease

1.3.1 Cdc25 expression is up-regulated in Cancer

In the past, difficulties in detecting Cdc25 protein levels in tissues have resulted in different techniques and approaches being employed (Rudolph et al., 2004). This has resulted in controversy and confusion in the literature; mainly due to the difficulty in comparing studies employing different techniques. One of the approaches taken has been to look at RNA instead of protein levels. However, it was later concluded there was limited correlation between these two variables. Despite these challenges trends have emerged.

Overexpression of the Cdc25 phosphatases has been noted in a number of cancers. Some of the human cancers where Cdc25 overexpression has been reported are head and neck cancer, breast, ovarian, pancreatic, colorectal, prostate, and non-Hodgkin lymphoma (Gasparotto et al., 1997; Ito et al., 2004; Guo et al., 2004). A relationship between the expression level of

Cdc25 and clinical outcome has been observed where overexpression generally leads to poorer clinical outcomes.

Overexpression of Cdc25B has been linked with poor disease prognosis. It has been commonly correlated with advanced stages of tumours. In oesophageal cancer it was found cells over-expressing Cdc25B were more sensitive to radiotherapy (Kishi et al., 2002).

It does not only seem that one Cdc25 isoform may have a greater role in a cancer subtype compared to another. To complicate matters it has been noted that overexpression of Cdc25 may not only be isoform-specific but also splice variant specific. Increased expression of the splice variant Cdc25B2 has been associated with a poor disease prognosis in non-Hodgkin lymphoma (Hernández et al., 2000). The expression of the splice variant Cdc25B3 is up-regulated in many pancreatic cancers (Guo et al., 2004). In prostate cancer Cdc25C5 mRNA levels were found to be overexpressed (Ozen and Ittmann, 2005).

1.3.2 How is Cdc25 up-regulated in cancer?

How Cdc25 is up-regulated in cancer is still unclear and yet to be fully described. The overexpression of Cdc25 can potentially be as a result of a defect occurring anywhere from the genetic to the protein level.

A significant relationship has been identified between the proto-oncogene MYC and Cdc25 expression. Positive correlations have been noted in breast cancer, lung carcinoma as well as in non-Hodgkin lymphoma (Ben-Yosef et al., 1998; Sasaki et al., 2001; Hernández et al., 1998). *Cdc25A and B* genes have been reported to contain MYC/MAX binding sites (Galaktionov et al., 1996). The levels of mRNA were increased in response to the activation of MYC. This was associated with overexpression.

In breast cancer cell lines, where Cdc25A is overexpressed, it was shown that the half-life of this protein was increased (Löffler et al., 2003). This study indicates that rather than greater expression, greater stability of the Cdc25 proteins could account for the high-levels of Cdc25 protein seen in some cancer cell lines.

Mutations in the Cdc25 regulators, the ATR kinase and the CHK kinases have been noted. They have been shown to lead to low level expressions of βTRCP in many cancers which include prostate, lung and gastric cancer (Gerstein et al., 2002; He et al., 2005). The CHK kinases are either mutated or expressed at lower levels or both in several cancers including colon and breast cancer as well as many carcinomas including carcinomas of the colon, ovary, lung and breast (Bartek and Lukas, 2003).

1.3.3 Targeting the Cdc25 phosphatases

Given the role of the Cdc25 proteins in the cell cycle it is not surprising that these phosphatases are involved in a diverse range of cancers. One of the ways Cdc25 could therefore contribute to tumorigenesis is through the inappropriate activation of the cell cycle.

Transfecting cells with active site mutants of Cdc25 affects cell cycle progress; G1 arrest resulted because of the inactive Cdc25A mutant (Hoffmann et al., 1994; Jinno et al., 1994). G2 arrest was the result of inactive Cdc25B or Cdc25C mutants (Millar et al., 1991; Lammer et al., 1998; Gabrielli et al., 1996). Increased expression of Cdc25A has been shown to affect S phase and a speedy entry into the S phase was noted (Blomberg and Hoffmann, 1999; Sexl et al., 1999; Falck et al., 2001). Cdc25B has been shown to be important for the continuation of the cell cycle after repair to damage such as DNA damage occurs and overexpression of this protein results in cells entering mitosis without the necessary time for repair (Van Vugt et al., 2004; Albert et al., 2012; Karlsson et al., 1999).

As mentioned, increased expression and activity of the Cdc25 phosphatases has been noted in many cancers. The presence of high-levels of Cdc25 phosphatases will make it difficult for the cell's regulatory machinery to remove them resulting in sustained high-levels of these proteins in the nucleus. This could result in the bypass of essential cell cycle checkpoints, disregarding threats to the cell such as DNA damage and forcing the cell past check points without allowing time for any of the essential repairs to occur. Thus, the inappropriate activation of the CDK/cyclin complex could lead to the increased proliferation of cells that are damaged.

The direct interaction of the Cdc25 phosphatases with the CDK/cyclin complexes provides a unique approach in anti-cancer therapy. Targeting the Cdc25 phosphatases is an indirect way of targeting the CDK/cyclin complexes which activate the cell cycle. The inhibition of the Cdc25 phosphatases will inhibit these complexes and prevent the damaged cell from progressing into the next phase of the cell cycle which can ultimately lead to cell death by apoptosis.

1.4 Cdc25 Inhibitors

Over the years there has been a huge amount of interest in the development of inhibitors for the Cdc25 phosphatases. There has been a number of inhibitor scaffolds which have been identified from which derivative compounds have been developed. However, to date no compound has entered the clinic. Inhibitors of Cdc25 (Figure 1.4) can be split into two classes, inhibitors from natural sources and synthetic inhibitors.

1.4.1 Natural inhibitors

The dnacins (1) were one of the earliest inhibitors shown to inhibit Cdc25. Although, the antibiotics dnacin A and B were shown to have antitumor properties they were weak inhibitors of Cdc25 (Horiguchi et al., 1994). Among the natural inhibitors the most popular were the

dysidiolides. These compounds provided a base for the following generation of natural and synthetic compounds. The dysidiolide (2) compound which is extracted from marine sponge can inhibit Cdc25A selectively with an IC_{50} of 9.4 µM (Gunasekera et al., 1996). This compound can result in arrest of the cell cycle. It has been shown to prevent the proliferation of lung cancer and leukemia cell lines.

Cholestane derivatives can be traced back to the dysidiolide scaffold. These were shown to inhibit Cdc25 with IC_{50} values of < 10 µM. The seco-cholestane based compounds negatively affected the growth of the colon cancer cell line HT-29 (Zalkow et al., 2000).

Another cholestane derivative sulfircin (3) which can be isolated from deep water sponge was also shown to inhibit Cdc25 however not specifically (Cebula et al., 1997). It was initially shown to have antifungal properties before its inhibition effect on Cdc25 was known. It can inhibit Cdc25A with an IC_{50} of 7.8 µM. The aliphatic chain length was identified to be an important aspect in its inhibition ability.

1.4.2 Synthetic inhibitors
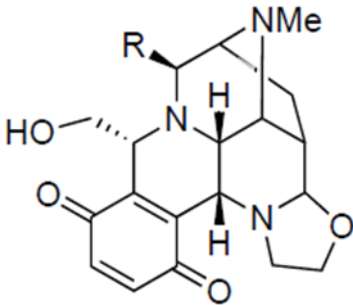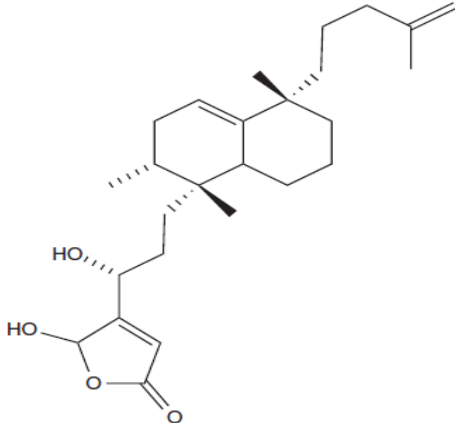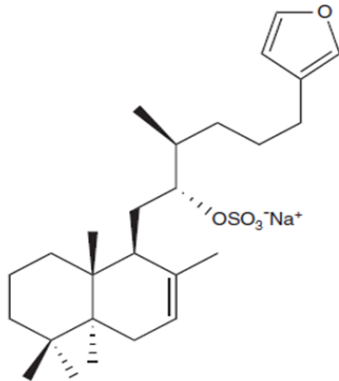
Synthetic inhibitors were developed in an effort to identify compounds with greater inhibition activities and specificities. Rhodanine based derivatives (4) were shown to have effective inhibition activity against all Cdc25 homologues (Ahn et al., 2007). Interestingly, a few of the compounds in this series selectively inhibited Cdc25B with IC_{50} up to 2.7 µM.

The compound TPY-835 (5) which is derived from cinnamic acid inhibits Cdc25A with an IC_{50} of 5.1 µM and Cdc25B with an IC_{50} of 5.7 µM (Aoyagi et al., 2005). It did not inhibit Cdc25C nor did it inhibit the serine/threonine phosphatases PP1 and PP2A. It also displayed some anticancer activity against lung cancer in mice.

Maleimide derivatives have also been identified having inhibition activity against Cdc25. The compound PM-20 (6) (Kar et al., 2006) was considered to be one of the best among this series. It can inhibit all the three homologues of Cdc25 with IC_{50} of 5, 10, and 40 µM for Cdc25 A, B, and C respectively.

Benzothiazole and benzoxazole-4-7-diones which were described by the Prevost group (Galcera Contour et al., 2007 & 2009) can inhibit Cdc25C with IC_{50} of 10 µM or less. Compounds in this series were effective against human pancreatic and prostate cancer cell lines. They also found that these compounds exhibited low levels of toxicity in non-cancerous cells. These compounds were based on a quinone derivative BN82685 (7). BN82685 has an IC_{50} of 0.250 µM for Cdc25A, 0.250 µM for Cdc25B, and 0.171 µM for Cdc25C making it a very potent compound. BN82685 was described as working irreversibly. Application of this compound resulted in a delay in the assembly of mitotic spindles as well as a negative effect on the dynamics of microtubules. Importantly, it was also shown to be active after administering orally (Brezak et al., 2005).

The invention of the benzothiazole and benzooxazole-4-7-diones led to the discovery of a novel potent inhibitor IRC-083864 (currently known as Debio 0931) with inhibition activity in the nanomolar range. IRC-083864 (8) was shown not only to irreversibly inhibit the three human Cdc25 homologues but also splice variants of Cdc25B (Brezak et al., 2009). This compound affected cell proliferation and cell cycle transition. It prevented mitotic entry and was effective against a range of tumour cell lines as well as human xenografts of prostate and pancreatic cancer in mice. Toxicity was not observed at low concentrations and at high concentrations a reduction in body weight was noted. IRC-083864 is currently the only inhibitor of Cdc25 which has progressed to clinical trial phase II.

Inhibitor Name	Structure	Cdc25 Target and IC_{50}
1. Dnacins		Cdc25B (dnacin A1, 141 µM and dnacin B1, 64.4 µM)
2. Dysidiolide		Cdc25A (9.4 µM)
3. Sulfircin		Cdc25A (7.8 µM)
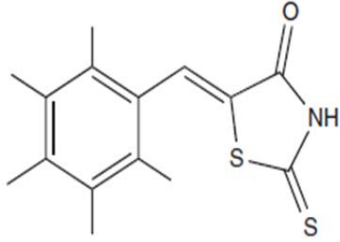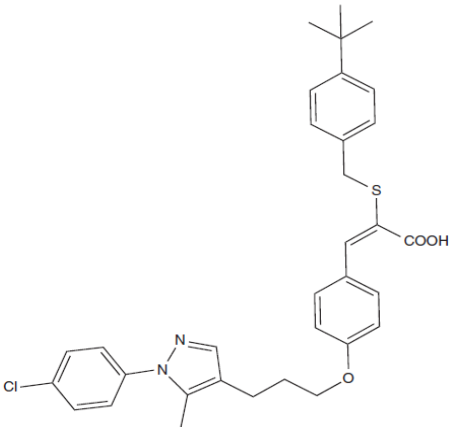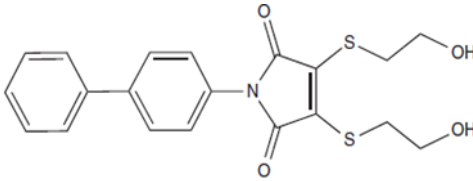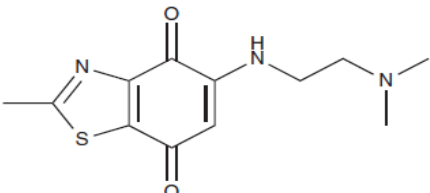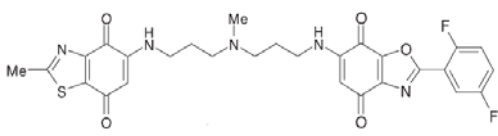
4. Rhodanine based derivatives		Cdc25B (2.7 µM)
5. TPY-835		Cdc25A (5.1 µM) Cdc25B (5.7 µM)
6. PM-20		Cdc25A (5 µM) Cdc25B (10 µM) Cdc25C (40 µM)
7. BN82685		Cdc25A (0.250 µM) Cdc25B (0.250 µM) Cdc25C (0.171 µM)
8. IRC-083864		Cdc25A (1.7 nM) Cdc25B (B2, 4.4 nM and B3, 9.4 nM) Cdc25C (0.4 nM)

Figure 1.4 Cdc25 Inhibitors

Name, chemical structures, Cdc25 target and IC$_{50}$ are tabulated. This Figure was adapted from (Lavecchia et al., 2010).

1.5 Summary

The Cdc25 proteins have been reported to be overexpressed in a variety of cancers. The role of the Cdc25 proteins is in the cell cycle where they function by activating their only known biological substrate: the cdk/cyclin complex. Their intimate relationship with Cdk makes them interesting targets for cancer therapy.

Research into the development of inhibitors for Cdc25 has made great strides over the past 13 years. One of the main difficulties in developing inhibitor compounds has been translating positive *in vitro* effects to *in vivo*. Although, the inhibitor compounds extracted from natural sources initially showed great potential they generally lack in specificity and their inhibition activity is limited to the micromolar range. Therefore, due to the limitations of these natural compounds more focus has been directed towards the development of synthetic inhibitors. Obtaining an NMR or X-ray crystal structure of the Cdc25 protein-inhibitor complex will lead to greater understanding of the Cdc25 structure and development of potent inhibitor compounds.

1.6 Studying Proteins

1.6.1 Protein Stability

A proteins thermodynamic stability is defined as the Gibbs free energy change between the two protein states: native (folded) and the denatured (unfolded) state (see eq. below). The size of the Gibbs free energy change of folding (ΔG_f) reflects the stability of the protein (Pace, 1990). The smaller this value of ΔG_f the less stable the protein is and it is less resistant to unfolding.

$$\Delta G_f = G_f - G_u$$

(ΔG_f = Gibbs free energy change for folding, G_f = Gibbs free energy of the folded state, and G_u = Gibbs free energy of the unfolded state).

Generally, there is a negative correlation with temperature and protein stability where an increase in temperature will result in a decrease in protein stability. This is of course not true for all proteins especially proteins that are heat resistant such as enolase and α-glucosidase of *Pyrococcus furiousus* which work best at temperatures over 90 °C (Costantino et al., 1990; Peak et al., 1994).

When ΔG_u is zero, the proportion of protein in the unfolded state is equal to the proportion of protein in the folded state. The temperature at which this occurs is known as the melting temperature (T_m). The melting temperature of a protein can indicate the stability of the protein. A change in T_m from one condition to another can be measured and this can help evaluate the effect of each condition on the thermal stability of the protein. A positive ΔT_m indicates greater stability and a more ordered structure while a negative ΔT_m is linked to destabilization probably due to the protein adopting a less structured conformation (Cho et al., 2011).

1.6.2 Thermal Shift Assay (ThermoFluor®)

1.6.2.1 Introduction

ThermoFluor® is a technique that is used to identify buffer conditions which improve the thermal stability of a protein. It can also be used to identify ligands which can stabilize the protein. It is an advantageous assay because it is relatively quick and inexpensive. Essentially, any globular protein can be tested and only small amounts of protein are required. It is a useful technique to use before any biophysical or structural analysis takes place because buffers in which the protein is stable in can be identified before any other experiments are performed, which can save time and energy.

1.6.2.2 The ThermoFluor® technique

The type of dye used for ThermoFluor® is one which has a high fluorescence in hydrophobic environments (Niesen et al., 2007). In polar environments, there is little or no fluorescence since it is quenched. These dyes include SYPRO orange, Nile red, and Dapoxyl sulfonic acid. SYPRO orange is a popular choice because it has a high signal to noise ratio.

An RT-PCR machine is used to perform the ThermoFluor® experiment. During the experiment fluorescence is measured at each temperature ranging from 25 °C to 90 °C. As the temperature increases the protein will become unstable and will unfold. This exposes the protein's hydrophobic core to solution. The fluorescence dye has an affinity for these hydrophobic regions and will bind to them resulting in an increase in the fluorescence signal (Niesen et al., 2007). The results can be plotted with fluorescence against temperature and the result is a sigmoidal curve which displays a two state transition representing the unfolded and folded states of the protein.

T_m can be taken from the point of inflection of the melting curve. In this thesis the ThermoFluor® technique was used to assess the effect of different buffers and/or additive conditions on thermal stability by measuring changes in T_m in order to identify buffers that improved the stability of the Cdc25C catalytic domain.

1.6.3 Protein Solubility

There are many factors that influence protein solubility. These can include environmental factors such as changes in temperature, pressure, and the addition of chemical reagents.

Poor or inadequate protein solubility is one of the bottle-necks in structural biology. It is also a concern in the pharmaceutical industry because it can limit the shelf-life of drugs. Finding ways to improve protein solubility and hence increase the life-span of protein drugs is of great value in the pharmaceutical industry (Wang, 1999). For example, a lot of effort has been directed towards improving the longevity of protein pharmaceuticals such as insulin, calcitonin, and leptin (Bakaysa et al., 1996; Fowler et al., 2005; Eui et al., 2005).

The solubility of a protein is affected by the interactions the protein forms whether those are with its self, other proteins in the solution, and the interactions it forms with its aqueous environment.

The surface of a protein determines the type of interaction it makes with its surrounding buffer or molecules. Surfaces of water soluble proteins tend to have amino acids with charged or polar side chains. These hydrophilic amino acids allow the protein to have favourable relationships with the aqueous buffers.

A hydration layer surrounds proteins in aqueous solution. This hydration layer has special properties which makes it distinct from bulk water. It is much more ordered with a density of

more than 10 % and heat capacity greater than 15 % compared to bulk water (Creighton, 1992).

The solubility of a protein tends to be related to the interactions of its side chains with water. Proteins tend to precipitate out of solution to reduce unfavourable interactions between the proteins exposed hydrophobic amino acids and the surrounding water (Fields et al., 1992). Protein solubility can also be affected by disordered loops which extend from a proteins surface. These can be vulnerable in forming non-specific interactions which could result in the protein precipitating out of the solution.

It is difficult to predict what the best method is to improve the solubility of a protein. The tried and tested methods may not necessarily be applicable for a particular protein. There are a diverse range of methods that can be applied and therefore one should consider carefully which methods are suitable for the protein of interest.

1.6.4 Solubility Screening Methods

There are two methods available to use during screening of solution conditions when optimizing protein solubility. Generally, one of the two methods is used although both can be used. They are known as the microdialysis button test and the microdrop screen (named as the hanging-drop solubility assay in this thesis) (Bagby et al., 2001). These methods have been adapted from the protein crystallization field. The microdrop screen will only be discussed in detail here.

The protein first needs to be highly concentrated in a minimally complex buffer. The protein concentration is usually the highest concentration you will need for your experiment or alternatively if this is not possible it should be concentrated to the maximum concentration that can be obtained. For each buffer condition a small amount of protein which can range

from 100 nl to a few µl volumes is mixed with equal volume of the buffer being tested. Following the set-up of either technique, the protein is then visualized for precipitation after a certain time period whether it is hours or a few days.

1.6.4.1 Hanging-Drop Solubility Assay

The hanging-drop solubility assay (Figure 1.6.4.1) is based on vapour diffusion. This method was initially developed for crystallography where the protein drop is combined with a buffer drop from the well (the well buffer contains the precipitating reagent) on a coverslip which is then inverted and secured onto the well. Over time, the protein drop becomes concentrated because the solvent diffuses into the well buffer. This results in equilibrium between the buffer that the protein is in and the well buffer. Over time, the precipitating reagent will promote protein precipitation which can also lead to crystal formation. This method was adapted by Lepre and Moore (Lepre and Moore, 1998) to promote protein solubility rather than precipitation. Essentially the method is the same here. The only difference is that the buffers tested do not contain reagents which promote precipitation but stabilizers instead which promote protein solubility.

There are two main advantages of this assay. Firstly, many buffer conditions can be tested and secondly only a small amount of the precious protein sample is required for each condition. In this thesis, 96-well plates were used to do the initial screens followed by 24-well plates. Once the starting buffer and pH have been determined it is convenient to first screen as many conditions as possible. Using a robot such as a Mosquito liquid handler makes this task less demanding. This is one of the drawbacks of the microdialysis button test because the dialysis buttons have to be assembled manually which is labour intensive and due to the large volumes of buffer required for each condition the number of conditions that can be tested is limited.
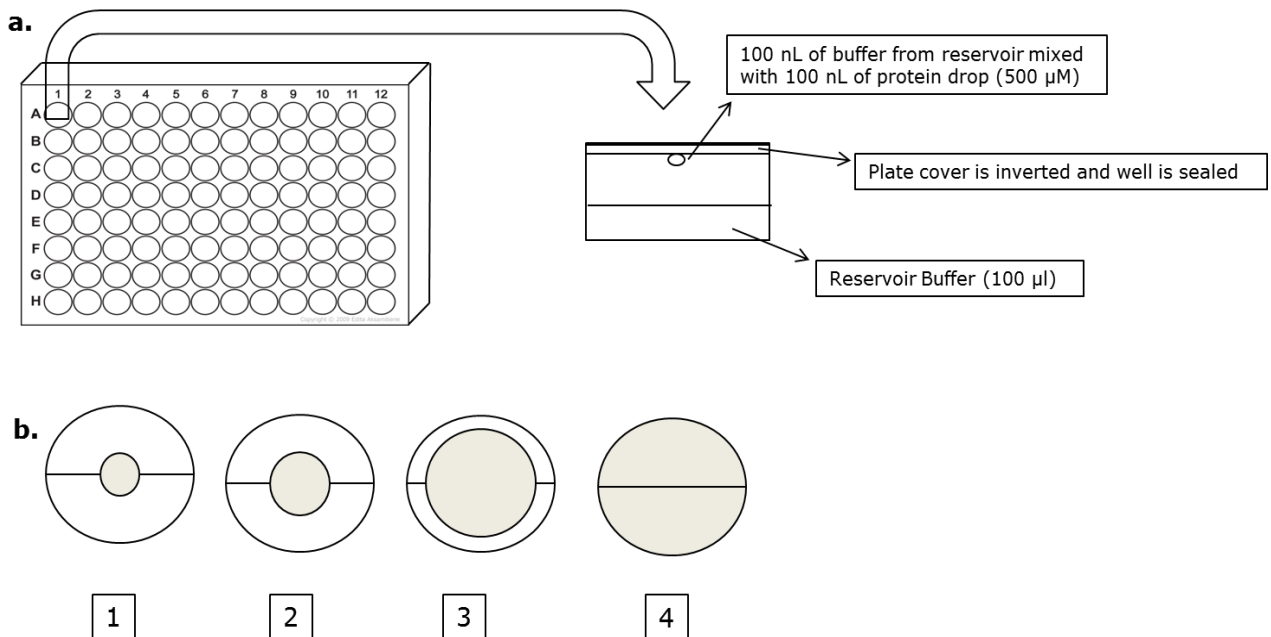
Figure 1.6.4.1 Hanging-drop solubility assay

Hanging-drop solubility assay set-up (a). For small scale screens a 96 well plate is used. An automated machine can be used to pipette equal volumes of the buffer and the protein drop. The standard drop volume is 100 nl but this can vary during optimisation screens. Following set-up of the plate, the plate is incubated at the desired temperature for a few days. The drop is then visualised for precipitation and the Lepre and Moore scoring system is used to analyse the results where maximum precipitation is assigned a score of four (Lepre and Moore, 1998).

When optimizing the protein of interest it is important to follow a logical order. Initially, it is best optimizing the pH and buffer. Once a suitable pH and buffer has been confirmed then one can move onto detailed screens. These screens can initially start with common additives from the wealth of additives available which can then be followed by using less common additives (Bagby et al., 2001). Common additives include: salts such as sodium and potassium chloride, reducing agents like TCEP and DTT, glycerol, and mild detergents like CHAPS.

The addition of a reducing agent in a protein's buffer is good practice especially when there are exposed cysteines in a protein. The cysteines can be oxidised and form disulphide bridges which can result in protein aggregation. Adding a reducing agent such as TCEP can avoid this problem.

1.6.5 Improving Protein Solubility

There are many options to consider when optimising protein solubility (Figure 1.6.5). It is important to ensure that the protein is folded before any detailed solubility optimization commences. This can be done by acquiring a ^{1}H - ^{15}N HSQC or Circular Dichroism spectrum.
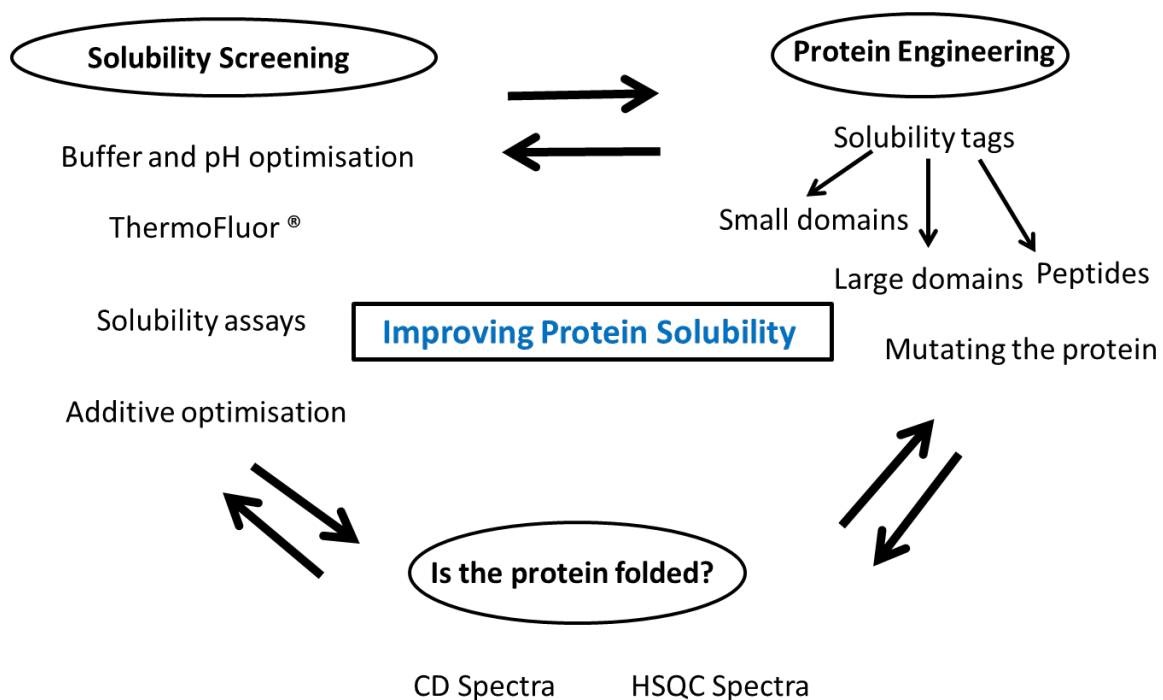
Figure 1.6.5 Improving protein solubility

There are many approaches one can try when optimising protein solubility. Each of these approaches can be followed or be preceded by the other as shown in the flow diagram. These approaches range from extensive buffer optimisations to improving construct design. The protein can be probed by CD or NMR (1D or 2D) to ensure the protein is folded before and after any optimisation.

One route of improving protein solubility is to mutate residues on the protein (Dale et al., 1994; Mosavi and Peng, 2003; Trevino et al., 2007). The drawback to this approach is that it is often difficult to decide which residue to mutate and the impact of such a mutation on the protein is not known until the protein is tested.

A common approach is to selectively mutate exposed hydrophobic amino acids to hydrophilic ones. This is easier when one has access to the 3D structure of the protein. It is better to mutate one or two residues and assess protein solubility because one generally does not want to make unnecessary changes to the native protein sequence. Not all hydrophilic residues have the same effect on protein solubility. A study which looked into the impact of all the amino acids on the RNase Sa protein solubility found that some hydrophilic amino acids favour protein solubility compared to others (Trevino et al., 2007). For example, they found that serine, aspartic acid, and glutamic acid were better at promoting protein solubility.

There are a number of technicalities which need to be addressed in order to have an optimized target ready for NMR study, in particular, if assignments and detailed structural studies are required. The target protein needs to be pure, stable at a pH of 7 or below, have a good expression level, and be stable at high concentrations for the duration of the experiment which could last for weeks. It is also very important for the target protein to be homogeneous meaning that in solution it should be "monodisperse", containing a single species of protein.

When deciding on the buffer of choice there are many factors which need to be taken into account which will allow the protein of interest to be stable. The buffer type, pH, salt, and additional additives will need to be optimised to make up the final buffer.

1.6.6 Optimization of Buffer and pH

When deciding on the type of buffer to use one must take into account several factors. Firstly, the chosen buffer needs to have the right buffering capacity. Meaning, it should be capable to buffer the chosen pH. This is an important factor which affects protein solubility (Green A, 1931).

When a protein is closer to its isoelectric point (pI) it will have the least solubility. The isoelectric point of a protein is the pH at which the protein has no net charge. Protein solubility is increased when the pH is away from the isoelectric pH in either direction. The concentration of the buffer should be enough to be able to allow adequate buffering. In order to maintain the pH, this is usually 50 – 100 mM, although in some cases 20 mM can also be used.

One also needs to be aware of the limitations of their chosen buffer as well as its strengths. For example, phosphate can inhibit some enzymes such as dehydrogenases and kinases. The Tris buffer is temperature sensitive.

Generally, it is preferred for experiments to be performed with conditions that are similar to physiological conditions. A pH of 7.4 is ideal for most cytosolic proteins like Cdc25C.

1.6.7 Optimization of Salt

Salts can affect protein stability through electrostatic shielding and by forming complexes with the protein. Electrostatic shielding can have a negative or positive impact on protein stability and the impact it has is protein dependent.

Based on the Hoffmeister series, the most stabilizing ions include ammonium, potassium, sodium, sulphate, and acetate ions. Salts stabilize proteins by preferentially hydrating them which results in the build-up of water surrounding the protein.

Sodium chloride is the most common salt present in most buffers. Varying the concentration of sodium chloride and hence the ionic strength of the buffer solution is known to affect protein solubility. For example, factor VIII SQ requires sodium chloride in order to stabilize it and increase its solubility in solution (Fatouros and Thomas, 1997).

Increasing the salt concentration further will reach a stage where the salt will start competing for water with the protein resulting in the exclusion of water molecules surrounding the protein. The protein will then no longer be soluble and precipitate out of solution. This is called "salting out".

1.6.8 Optimization of Additives

1.6.8.1 Osmolytes

Osmolytes include polysaccharides, sugars, and amino acids. Osmolytes are produced naturally when organisms experience denaturing conditions or an increase in osmotic pressure. Osmolytes can indirectly stabilize proteins by changing the buffer environment. The advantage of using an osmoltye as a stabilizer is that they in general do not affect the structure or function of the protein.

1.6.8.2 TMAO (osmolyte)

Trimethylamine *N*-oxide (TMAO) is an osmolyte that stabilizes proteins (Cho et al., 2011; Zou et al., 2002; Auton and Bolen, 2005). It is naturally found in sharks as well as saltwater fish where it counteracts the toxic effect of high concentrations of urea. It stabilizes the native state of proteins by increasing their thermodynamic stability.

1.6.8.3 Sucrose (osmolyte)

Sucrose was shown to limit aggregation and stabilize the recombinant human interferon-γ protein by making it favour a more compact form (Kendrick et al., 1998). The Timasheff

mechanism provides an insight into how this occurred (Lee and Timasheff, 1981). Lee and Timasheff observed that there is a preferential exclusion of sucrose on protein surfaces. This preferential exclusion of sucrose can result in a compact conformation of a protein species being favoured which is the case for the recombinant human interferon- γ protein.

1.6.8.4 Metal Ions

Metal ions can also stabilize proteins. They do this by making the protein less flexible and more compact.

1.6.8.5 Arginine

The use of arginine in buffers has been associated with a number of benefits for proteins. It is commonly used in protein refolding experiments to prevent aggregation and aid correct folding (De Bernardez Clark, 1998). Other benefits of arginine include greater thermal stability, inhibition in the build-up of partially folded protein intermediates as well as the prevention of non-specific interactions (Ghosh et al., 2009). Higher concentrations of arginine alone can greatly improve heat induced aggregation (Shiraki et al., 2002; Arakawa et al., 2007a). Concentrations of up to 0.5 M arginine have been shown to have little or no effect on protein secondary structure (Ghosh et al., 2009).

Arginine in combination with equimolar glutamic acid is also effective in improving protein solubility (Golovanov et al., 2004; Vedadi et al., 2006; Blobel et al., 2011). A 50 mM mixture of Glu/Arg has been shown to prevent proteolytic degradation as well as inhibit protein aggregation, ultimately increasing the longevity of the protein sample (Golovanov et al., 2004). Vedadi et al. showed that 16 % of all the proteins tested were stabilized against heat induced aggregation by the addition of 50 mM Glu/Arg and more than 4 $^{\circ}$C increases in T_m were noted (Vedadi et al., 2006).

1.6.9 Solubility Fusion Tags

In structural biology it is common to attach a tag onto a protein whether that is a solubility tag, a tag which aids protein purification or a tag that does both. There is a wide range of fusion tags available and the type of fusion tag used depends on the requirements.

Fusion tags come in all shapes and sizes. They can range from short stretches of amino acids to large tags such as the MBP (maltose binding protein) fusion tag (Table 1.6.9).

The MBP system is a well-known and well-studied system. The MBP protein has been shown to be a better solubilising agent then both the GST (Glutathione S- Transferase) and TRX (thioredoxin) systems. Kapust et al. (Kapust and Waugh, 1999) suggest the deep hydrophobic cleft in the MBP fusion protein aids protein solubility. They also claim the MBP fusion protein of having a chaperone role. Despite these advantages, the main limitation of using this tag for NMR is its size. Large tags such as the GST or MBP tags will need to be removed before any NMR studies take place. If the tag is being used as a solubility enhancing tag, removal of the tag can result in the loss of the solubilising effect hence poor solubility of the target protein. If these large fusion tags are not removed, signals from the fusion tags will also appear in the HSQC spectrum along with the desired protein resonances. This will result in complications in the data being collected. Therefore, the use of fusing such large tags to proteins for NMR studies is limited.

Tag	MBP	GST	His tag	Poly-Arg/Lys	GB1 (SET)
Size (kDa)	42	26	~ 0.90	~ 1-2	6.2
✓	• Improves solubility (better than the GST tag) • Has a chaperone role • Affinity tag	• Improves solubility • Affinity tag	• Small size • Affinity tag • Does not need to be cleaved for NMR study	• Small size • Improves solubility • Does not need to be cleaved for NMR study	• Small size • Improves solubility • Can choose acidic/basic tag based on the target protein pI • Does not need to be cleaved for NMR study
X	• Soluble aggregates • Size (cleave?) • Complex NMR spectrum if not cleaved	• Size (cleave?) • Complex NMR spectrum if not cleaved	• Poor at solubilising proteins	• Non-cleavable/difficult to cleave	• Complex NMR spectrum

Table 1.6.9 Solubility tags

A Table to summarise the pros and cons of the tags

1.6.9.1 Poly-Lys and Poly-Arg tags

Fusing short poly-Lys or poly-Arg tags to the N- or C- termini of the target protein has been shown to significantly improve protein solubility (Kato et al., 2007). This study showed a positive relationship with protein solubility and increasing the number of lysine or arginine residues for the bovine trypsin inhibitor, BPTI-22. There was no change seen in the structure or activity of the BPTI-22 protein upon fusion of these tags. Adding five or six consecutive Arg or Lys amino acids were shown to be optimal in improving protein long-term solubility. The solubility enhancing effect was greater when these tags were fused to the C- termini. There

was also an improvement seen in the ^{1}H-^{15}N HSQC spectra. Generally, for these types of tags, cleavage is not desired and can be problematic (Terpe, 2003).

1.6.9.2 His Tag

The most popular fusion tag used in structural studies is the His tag which is a stretch of at least six histidines (Hammarström et al., 2002). It is quite useful in that it allows proteins to be purified efficiently. This polyhistidine tag has an affinity for bivalent Nickel or Cobalt metal ions which are immobilized onto a resin. The advantages of a His tag is that it is not sensitive to changes in the purification buffers. Another advantage is that the His tag can be used to purify proteins under denaturing conditions making it a popular tag during such studies. Since the tag will not become denatured like other tags such as MBP it does not need a folded protein to work. It can easily be used to purify peptides. Its small size also provides an advantage in that the tag does not need to be removed for NMR studies. However, it is a poor solubility enhancement tag and in some cases can cause problems and needs to be removed (Woestenenk et al., 2004; Hammarström et al., 2002).

1.6.9.3 Other Solubility Tags

Another option is fusing smaller domains such as the thioredoxin (Trx) tag or the GB1 tag to the target protein (Yasukawa et al., 1995; Zhou and Wagner, 2010). Both tags have been shown to enhance protein solubility and long term stability. The GB1 tag has successfully been shown to improve protein solubility and the NMR HSQC spectrum of the target protein (Zhou and Wagner, 2010). It is a small domain ~ 6.2 kDa in size fused to the protein of interest. There is also flexibility in the type of GB1 tag used in that it can complement the charged state of the protein. The difficulty of using this type of tag is in the complication that results in the HSQC NMR spectrum. Since this tag is not cleaved signals of the GB1 protein and the target protein

will be present in the spectrum. However, this tag is useful if the protein has poor solubility, and a poor NMR spectrum; it can help resolve the signals of the target protein.

1.7 Small angle X-ray Scattering (SAXS)

1.7.1 Background

Small angle X-ray scattering (SAXS) was initially developed in the late 1930s by a French physicist named Andre Guinier (Guinier, 1955). It is a technique which provides useful insights into the size, shape and flexibility of a system (Putnam et al., 2007). It is particularly useful in structural biology where macromolecules can be investigated in solution.

With a resolution up to 1-2 nm, SAXS is a relatively low resolution protein 3D structure determination technique compared to NMR and X-ray crystallography. However, the ease of sample preparation and data analysis as well as the potential to complement NMR and X-ray crystallography makes it a useful and robust method particularly for flexible or multidomain systems (Mertens and Svergun, 2010).

From the SAXS scattering curve a low resolution 3D protein model can be generated using *ab-initio* techniques (Grant et al., 2011; Franke and Svergun, 2009). 3D protein atomic structures determined by NMR or X-ray crystallography can be fitted into the SAXS model; these can be either of the same protein or smaller domains of the full-length protein. This can help to obtain an overall model of a macromolecule complex. One of the key advantages of this technique over NMR is that SAXS is not restricted to protein size and large complexes can be studied.

If the macromolecule being investigated possesses high flexibility and disorder it becomes difficult to work with especially when trying to generate diffraction quality crystals since they need to be well-ordered and have minimum flexibility. However, this is not a problem for SAXS because protein crystals are not required and systems possessing high flexibility can be studied. Therefore, SAXS is useful in studying difficult protein targets. If X-ray crystallography

data is available this can be compared to the SAXS data which allows comparison of the protein in both solid and solution state. SAXS can therefore help to validate the crystal structure.

1.7.2 Technique

SAXS data is usually collected at synchrotron facilities. This is because proteins weakly scatter X-rays, therefore high intensity X-ray photon beams are required. The final scattering pattern of the protein which is generated is commonly collected at angles ≤ 3 Å which provides important information on the size, shape, and conformation of the protein (Svergun and Koch, 2002).

Figure 1.7.2 is a schematic of the SAXS experiment. The SAXS instrument is composed of an X-ray source, a component where the sample is held, a collimation system, beam stop and a detection component.

The collimation system is useful in that it narrows the X-ray beam. The beam stop is an important barrier which stops the resulting beam that has not been scattered by the sample from striking the detector. This would not only damage the detector but would also overlap the signal from the protein sample.

The data collected is in 2D form which is then transformed to a 1D format. Two scattering plots are obtained. One is from the protein sample and the other from the buffer. This allows the buffer subtraction to take place which then results in a scattering plot of just the protein alone. This is plotted as logarithmic intensity against q where q is the momentum transfer and is defined as: $q = 4\pi\sin\theta/\lambda$ (angle between incident and scattered beam is 2θ).
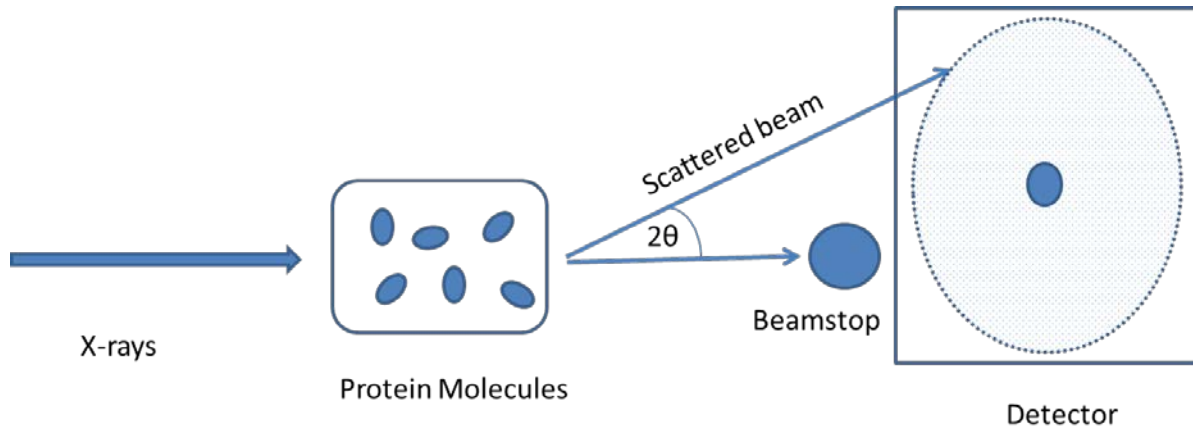
Figure 1.7.2 Small Angle X-ray Scattering (SAXS)

The sample is exposed to a highly focused X-ray. The sample will then scatter these X-rays which are then detected by the detector. The beamstop prevents the detector from being damaged by the primary beam.

1.7.3 Data Analysis

1.7.3.1 Data Quality

In order to obtain good quality SAXS data and make reliable conclusions it is important that the sample is of good quality with minimal aggregation. The sample needs to have high purity and this can be checked using an SDS-PAGE prior to the SAXS experiments. As well as high purity the sample should be monodisperse; AUC can be used to assess if the sample is monodisperse. Another important quality check of the SAXS data is looking at the lower q (Guinier) region of the scattering plot (Jacques and Trewhella, 2010). This is useful in detecting any concentration dependent effects when comparing different concentrations and also allows you to see if there are any undesired interactions. When an increase in intensity is seen at the Guinier region which is called a "smiling" Guinier this means that there are attractive forces most likely due to aggregation (Figure 1.7.3a). If there is a decrease in intensity and a "frowning" Guinier is seen at lower q this means there are repulsive forces. This occurs because of interparticle

interference. However, it is normal to see a sudden drop in intensity at very low q. This is due to the effect of the beam stop.

1.7.3.2 Radius of Gyration

The radius of gyration (R_g) provides information which is essential when computing the dimensions of a particle (Svergun and Koch, 2003). The R_g can be extrapolated from the Guinier plot by fitting a straight line in the Guinier region (Figure 1.7.3a). An R_g can only be determined using the Guinier approximation approach if a straight line can be fitted in the Guinier region. If a straight line cannot be fitted this implies interfering forces and one cannot obtain the R_g from this analysis.

1.7.3.3 P(R) graph

The P(R) function which is also known as the "pair distance distribution function" can be computed using the GNOM program from the ATSAS software suite (Jacques and Trewhella, 2010). R_g can also be calculated here. The R_g parameter taken from this analysis is more precise compared to the R_g value which is extrapolated from the Guinier region. This is because this analysis takes into account all the scattering data rather than being restricted to the Guinier region. However, it is good practice to compare the R_g derived from the Guinier approximation to the R_g obtained here for consistency.

The P(R) profile gives information on the shape of the protein. It can therefore be useful in discerning conformational changes.

The way the P(R) curve approaches zero gives insights into the nature of the protein. It allows one to obtain the D_{max}, also known as the "maximum particle dimension" which is a measure of the size of the protein (Svergun and Koch, 2003). Tailing of the P(R) peak can indicate two things; there is aggregation or there is flexibility in the sample. If adequate quality checks have

been carried out and aggregation is ruled out then this tailing can be attributed to the flexibility in the protein. The limitation of the P(R) analysis is that although it can tell you if there is flexibility in the protein it cannot tell you where this flexibility lies within the protein. The ensemble optimization method (EOM) can give you a better idea (Bernadó et al., 2007).

1.7.3.4 Kratky Plot

The Kratky plot is a good qualitative measure of the folded state of the protein (Putnam et al., 2007; Receveur-Brechot and Durand, 2012). It is plotted as logarithmic intensity multiplied by q^2 against q (Figure 1.7.3 b). A folded protein is represented by a bell-shaped peak which is then followed by constant values converging around zero intensity with increasing q. A completely unfolded protein does not have the characteristic bell-shaped curve which defines a folded protein and instead plateaus at high q. A partially disordered protein has a bell-shaped curve for the folded component and an increase at high q which is due to the disordered component.
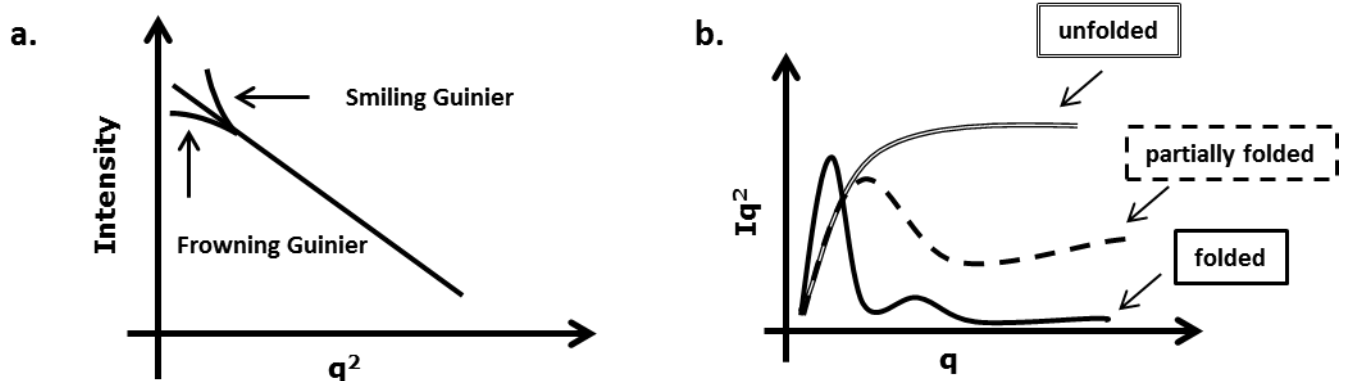
Figure 1.7.3 Data Analysis

The Guinier plot is used to assess data quality (a). If a straight line cannot be plotted in the Guinier region this indicates interparticle interference. The Kratky plot can be used to assess the folded state of the protein (b).

1.8 NMR and Crystallography theory

1.8.1 NMR: The Basics

Atomic nuclei such as ^{1}H, ^{15}N, and ^{13}C possess an overall nuclear spin. This is because they experience a magnetic moment. The spin quantum number (I) reflects the magnitude of the nuclear spin. The nuclei with a spin number of ½ are commonly used in the protein NMR field. These nuclei are ^{1}H, ^{15}N and ^{13}C. Nuclei which have a nuclear spin of zero are NMR silent, meaning they do not provide a magnetic signal because they do not have an inherent magnetic moment.

Because the NMR active nuclei possess nuclear spin when they are exposed to a magnetic field they will adopt a number of states. The nuclei that have a quantum number of ½ adopt two states known as α and β states (Figure 1.8.1). These states are either aligned with the magnetic field or against it. The lower energy state which is known as the α-state tends to align with the magnetic field and at equilibrium there will be more nuclei in this state compared to the high energy β-state. The difference in energy between the α- and β-state is very small. This energy difference is characterized by the Boltzmann distribution and correlates to the magnetic field strength and the gyromagnetic ratio (γ) (the gyromagnetic ratio is the ratio of the magnetic moment to the angular momentum) of the nuclei (Levitt, 2008).

Nuclear spins can be visualised as precessing around an axis. When a radio-frequency (RF) pulse is applied the equilibrium between α- and β- states is shifted and the α-spins will flip to the high energy β-state. The frequency of the RF pulse should be at the same frequency as the precessing frequency which is also known as the Larmor frequency. The Larmor frequency is defined by the gyromagnetic ratio and the magnetic field experienced by the nuclei. The resulting energy difference following the application of the RF pulse forms the basis of NMR spectroscopy.

$$\Delta E = h\nu$$

β

↑

α
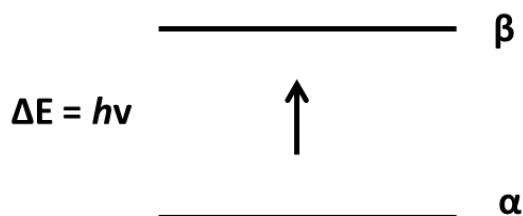
Figure 1.8.1 Transition from α to β state

The α-state is the low energy state. Following excitation by a radio-frequency (RF) pulse the α-state is shifted to the excited high energy β-state. The difference in energy is defined as ΔE = hv where h is the plank's constant and v is the frequency.

1.8.2 FID

The signal is recorded as an FID which known as the free-induction decay. The signal decays over time because of transverse relaxation (Levitt, 2008). The FID signal is plotted as RF intensity versus time and is Fourier transformed to RF intensity against frequency. The final 1D NMR spectrum is generally made up from a large number of FIDs to obtain the optimal signal to noise ratio.

1.8.3 Chemical Shifts

Chemical shifts are expressed in parts per million (ppm) rather than Hz. The resonance frequency of an internal standard such as tetramethylsilane (TMS) or deuterium oxide (D_2O) is

used for this conversion. The internal standard is used to determine the reference (zero) point.

Converting chemical shifts to ppm means experiments can be collected at different spectrometers and easily compared. For example, the same experiment collected at a 600 or 800 MHz spectrometer will result in the same chemical shifts (ppm) values recorded. This is because the chemical shift (ppm) is independent of the strength of the magnetic field.

The proton chemical environment determines the position of the proton signal on the proton spectrum. The proton nuclei which are surrounded by high electron density will experience high shielding and result in a proton signal at the lower end of the chemical shift scale (upfield). The protons which are surrounded by low electron density will experience low shielding and therefore will result in a proton signal at the higher end of the chemical shift scale (downfield).

1.8.4 Spin-Spin coupling

Nuclear spins can also interact with each other and this phenomenon is known as spin-spin coupling. Spin-spin coupling results in the splitting of the nuclei signals at a frequency which is specific to the nuclei and the distance between peaks is defined by the coupling constant J (1). The resulting splitting pattern and intensities can be predicted by Pascal's triangle.

1.8.5 NMR spectroscopy and compound screening

1.8.5.1 1D NMR

Protein ligand screening using NMR commonly involves two approaches. The first approach looks at the ligand using 1D NMR techniques such as WaterLOGSY (Water-ligand observed via gradient spectroscopy) (Ludwig and Guenther, 2009; Stockman and Dalvit, 2002). The second approach looks directly at the protein using 2D NMR techniques such as the HSQC.

Screening techniques involving 1D NMR ligand screening are popular because they require small amounts of protein at low concentrations which does not need to be labelled with expensive isotopes. These techniques can be used for initial screening as well as for the optimisation of lead compounds.

1D NMR ligand screening is based on the principle that proteins have a bigger size compared to the small molecule compounds. This size difference can be associated with relaxation properties which arise from the effects of tumbling (Ludwig and Guenther, 2009). Relaxation properties can be transferred from the protein to the small molecule compound changing the relaxation properties of the compound. This occurs when the compound binds to the protein. Provided that this interaction is transient and the small molecule compound comes off the protein within the time scale of the experiment this compound will have a different relaxation property compared to the compounds in solution which have not bound the protein and therefore will be detected. This is essentially the basis of the WaterLOGSY experiments.

NMR (1D) ligand screening is advantageous in picking up hits for ligands with low affinity and can enable characterization of the ligand binding epitope (Ludwig et al., 2008; Bhunia et al., 2012). The main limitation of 1D NMR ligand screening is that tight-binding ligands can result in false negatives. Therefore, protein-based NMR techniques, commonly HSQC experiments, are used to identify tight binders. These data can complement the 1D data. HSQCs can be used to map the binding onto the protein surface if assignments of the amide groups of the protein are available.

1.8.5.2 2D NMR

During the ^{1}H, ^{15}N-HSQC experiment magnetization is transferred by J-coupling from the amide proton to the attached labelled nitrogen. The chemical shift evolves on the nitrogen which is

then followed by a transfer of magnetization back to the original amide proton before the data is acquired.

The HSQC spectrum of a protein is essentially a finger print of the protein where each peak on the HSQC represents an amide group of a single amino acid. An exception is proline. Proline has an imino instead of an amino group. This means it does not possess a free proton that it can couple with nitrogen and hence will not produce a signal in the HSQC spectrum. Also, there are extra peaks from asparagine, glutamine, and tryptophan side chain amides.

A protein HSQC allows one to conclude if a protein is folded or not. If the peaks are well dispersed this indicates a folded protein. On the other hand if the peaks are not well dispersed and are all localised in the middle of the spectrum this indicates the protein is unfolded. If the protein is partially folded and contains a significant amount of disorder there will be a combination of dispersed peaks accounting for the folded component and a significant proportion of peaks clustered together in the middle of the spectrum which will account for the disordered component.

A protein HSQC can also be used to identify compounds or ligands which interact with the protein. These can be added and changes in the HSQC can then be monitored. If there is a binding event occurring this will be reflected by the chemical shift perturbations which will occur which can be in fast, slow, and/or intermediate exchange (Williamson, 2013). These chemical shifts changes can be mapped onto the protein structure.

The main limitation of protein NMR is the upper size limit which can range from 40-60 kDa. This is because the bigger the protein is the slower it will tumble. This slow tumbling results in broader line width in the NMR spectrum. The HSQC spectrum will have overlapping signals with poor resolution. Deuterating the large protein can help resolve these signals.

1.8.6 X-ray Crystallography

1.8.6.1 Introduction

X-ray crystallography is an alternative technique to NMR which is used to solve protein atomic structures. One of the advantages of using X-ray crystallography over NMR is that this technique is not restricted to the size of the protein.

The three-dimensional structures of proteins can be determined with their interacting partner such as a co-factor, ligand or inhibitor. Determining the atomic protein structure with the protein ligand/inhibitor allows greater understanding into the function of the protein. Also, this structural information can provide important insights in the development of drug compounds (Appelt et al., 1991; Sharff and Jhoti, 2003) allowing for the design of improved and more potent drugs.

1.8.6.2 The X-ray crystallography technique

Individual protein molecules in solution diffract X-rays weakly and therefore their resulting diffraction data cannot be detected. However, this can be overcome by using a protein crystal which is made up of a large number of protein molecules that are arranged in the same orientation in three dimensions. Hence, the crystal serves to amplify the signal and the resulting diffracted X-rays can be detected by the X-ray detector.

The protein crystal is placed between an X-ray beam and the detector. The crystal is then exposed to the X-ray beam. The direct X-ray beam is diffracted by the electrons which surround the atoms within the protein crystal. The diffracted X-rays are described as ''reflections'' which are detected as spots on the X-ray detector (Rhodes, 2000). Each of these spots represent a diffracted X-ray which is a simple wave defined by three parameters: frequency, amplitude, and phase. It is important to define these parameters for each diffracted

spot in order to successfully determine the three dimensional structure of a protein using X-ray crystallography.

1.8.6.3 Growing crystals

It is important that the protein has high purity and is freshly prepared before any crystallization trials commence (Dessau and Modis, 2011). This is because contaminants can interfere with the crystallisation process or generate crystals that diffract X-rays weakly leading to low resolution structures.

The crystallization process involves three steps (Durbin and Feher, 1996). The first step is nucleation in which the protein molecules associate together providing a foundation for the growth of a crystal. The next step is the growth of the crystal. The final step is when the crystal ceases to grow usually as a result of lack of free protein molecules within the solution or the presence of contaminants poisoning the growing faces of the crystal.

Crystals are grown under conditions which promote protein precipitation. It is essential that this precipitation process is tightly regulated and occurs very slowly to ensure maximum success for crystallising the target protein. Also, the precipitating condition should not induce the protein to denature or rapidly aggregate. Therefore, specialised crystallisation reagents are employed to promote precipitation. These crystallisation promoting reagents can range from salts, additives, and large polymers such as polyethylene glycol (PEG).

Commercial crystallisation screens are especially designed to help identify initial crystallisation conditions for the target protein (Wooh et al., 2003). The most popular method utilised to crystalize proteins is the 'hanging-drop' vapour diffusion technique. A subtle variation in this technique is the sitting drop method which can also be used. In either case equal volumes of purified protein is mixed with a crystallisation reagent condition (usually 100 + 100 nl drops)

with the Mosquito nano-litre crystallisation robot. Following set-up, the plate is incubated at the required temperature and later analysed for crystals using a light microscope.

Many variables such as the concentration of the protein, incubation temperature, pH and ionic strength of the solution as well as the type of precipitant can determine the success of a crystallisation experiment. Usually, crystallisation conditions for the target protein are screened using multiple commercial crystallisation screens. Hence, the identification of conditions suitable for crystal growth takes place using a trial and error approach and represents one of the biggest obstacles in X-ray crystallography. Once, the initial hits for the target protein have been identified the subsequent stage involves optimisation of the conditions to generate larger diffraction-grade crystals. Optimisation of the crystals can involve varying the drop volume of protein and buffer, slightly altering promising buffer conditions, and changing the incubation temperature. Once optimised conditions have been identified the screens will then be up-scaled so a large crystal can be generated which is then subjected to X-ray diffraction studies using an X-ray beam.

1.9 Thesis Aims

The main aim of this study was to structurally and biochemically characterise the Cdc25C full-length protein in order to aid future drug design. This aim was split into three aims:

1. Assessing the suitability of the Cdc25C full-length protein for small angle X-ray scattering (SAXS) experiments

2. Characterisation of the regulatory domain by determining if this domain is folded, determining the secondary structure content, and flexibility

3. Improving the solubility of the Cdc25C catalytic domain by optimising construct and solution conditions and assessing the potential for backbone assignments

4. Biophysical characterisation of the Cdc25C catalytic domain and testing inhibitor compounds in order to identify interaction sites on the protein

Chapter 2.0 - Materials and Methods

2.1 Construct Design

The Cdc25C constructs used in this thesis are listed in Tables 2.1.1, 2.1.2 and 2.1.3. These constructs were either His or GST tagged (Appendix (A) 1). The Cdc25C constructs were provided by our collaborator Professor Knapp (SGC, Oxford), purchased from the company ShineGene (Shanghai, China) or created using the QuikChange Lightning kit (Agilent Technologies, Cheshire, UK). The latter constructs (Tables 2.1.2 & 2.1.3) were designed using information gained from the disorder predictions' of Cdc25C using the 'DisMeta server' and insights gained from the literature regarding the addition of charged residues to improve protein stability (Huang et al., 2014; Kato et al., 2007)

Construct Name	Vector	Cell Line
His-Cdc25C$_{270-443}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-462}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-473}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{280-443}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{280-462}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)

Table 2.1.1
A list of constructs provided by SGC

Construct Name	Vector	Cell Line
GST-Cdc25C$_{1-473}$ (Full-length)	pGex6p-1	BL21 (DE3)
GST-Cdc25C$_{1-280}$ (N-terminal domain)	pGex6p-1	BL21 (DE3)
His-Cdc25C$_{1-280}$ (N-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{36-280}$	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{36-443}$	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{73-441}$	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{83-441}$	pNIC28-Bsa4	BL21 (DE3)

Table 2.1.2
A list of constructs purchased from ShineGene (Shanghai, China)

Construct Name	Vector	Cell Line
His-Cdc25C$_{270-429}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-432}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-436}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-449}$ (C-terminal domain)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-449}$ - 5Arg (C-terminal domain + Arg tag)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-449}$ - 5Lys (C-terminal domain + Lys tag)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-449}$-5Pro (C-terminal domain + Pro tag)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-449}$ -4Glu (C-terminal domain + Glu tag)	pNIC28-Bsa4	BL21 (DE3)
His-Cdc25C$_{270-449}$-GSSGS (C-terminal domain + GSSGS tag)	pNIC28-Bsa4	BL21 (DE3)

Table 2.1.3
A list of constructs synthesized using the Quikchange Lightning kit (Agilent Technologies)

2.2 Site Directed Mutagenesis

Agilent Technologies primer design program (www.agilent.com/genomics/qcpd) was used to design primers (Table 2.2.1) to be used in the mutagenesis reactions. The primers were then purchased from Invitrogen and the Agilent QuikChange lightning protocol was followed.

To synthesize constructs for the different C-terminal lengths of Cdc25C the relevant amino acid was mutated from the template vector to a stop codon.

Amino acids were mutated either in the N-terminal TEV cleavage site, the C- termini or both to generate constructs fused with short poly-amino acid solubility tag(s) (Table 2.2.1).

Construct	Template Vector	Primers (5'- 3')
His-270-429	His-270-443	+ tatggaactgtgtgaaccatagagctactgcccta - tagggcagtagctctatggttcacacagttccata
His-270-432	His-270-443	+ ctgtgtgaaccacagagctactaacctatgcatcatcagga - tcctgatgatgcataggttagtagctctgtggttcacacag
His-270-436	His-270-443	+ aaccacagagctactgccctatgcattgacaggaccacaagac - gtcttgtggtcctgtcaatgcatagggcagtagctctgtggtt
His-270-449	His-270-462	+ ctgaggtgtcgaagctagagcaaagtgcagg - cctgcactttgctctagcttcgacacctcag
His-270-449-5Arg	His-270-462	+ ttgctgaggtgtcgaagccggcgcagaaggcggtaaggggagcggcagctgcg - cgcagctgccgctcccccttaccgccttctgcgccggcttcgacacctcagcaa
His-270-449-5Lys	His-270-462	+ gagttgctgaggtgtcgaagcaagaagaaaaagaagtaaggggagcggcagctgcggg - cccgcagctgccgctcccccttacttctttttcttcttgcttcgacacctcagcaactc
His-270-449-5Pro	His-270-462 His-270-449-4Pro	+ gagttgctgaggtgtcgaagcccgcccccaccgtaggaaggggagcggcagctgcg - cgcagctgccgctcccccttcctacggtggggggcgggcttcgacacctcagcaactc + cccgcccccaccgccgtaaggggagcggca - tgccgctcccccttacggcggtgggggcggg
His-270-449-4Glu	His-270-462	+ tgagttgctgaggtgtcgaagcgaggaggaggagtaggaaggggagcggcagctgcg - cgcagctgccgctcccccttcctactcctcctcctcgcttcgacacctcagcaactca + cgaggaggaggaggagtaaggggagcggca - tgccgctcccccttactcctcctcctcctcg
His-270-449-GSSGS	His-270-462	+ agttgctgaggtgtcgaagcgggagcagcgggtcgtaaggggagcggcagctgcg - ccgcagctgccgctcccccttacgacccgctgctcccgcttcgacacctcagcaact

His-GSSGS-270-449	His-270-449	+cttctggtgtagatctgggtaccgagaacctgggcagctcaggcagcactcagatgctggagg aagattctaaccaggg
		-
		ccctggttagaatcttcctccagcatctgagtgctgcctgagctgcccaggttctcggtacccaga tctacaccagaag
His-GSSGS-270-449-GSSGS	His-270-449-GSSGS	+cttctggtgtagatctgggtaccgagaacctgggcagctcaggcagcactcagatgctggagg aagattctaaccaggg
		-
		ccctggttagaatcttcctccagcatctgagtgctgcctgagctgcccaggttctcggtacccaga tctacaccagaag

Table 2.2.1
A list of Cdc25C template vectors and primers used to synthesize Cdc25C constructs using the QuikChange Lightning Site Directed Mutagenesis Kit (Agilent Technologies, Cheshire, UK)

2.3 Small scale protein expression

2.3.1 Transformation

DH5α (Invitrogen, Paisley, UK) or BL21 (DE3) (Merck Chemicals, Nottingham, UK) cells and the desired vector were thawed on ice for 20 minutes. 1 µl of DNA (6-12 ng) was added to either 25 µl of DH5α cells or 50 µl of BL21 (DE3) cells. The cells were incubated on ice for a further 20 minutes. The cells were then subjected to heat-shock by incubating on a heating block for 30 seconds at 42 °C. This was followed by 2 minute incubation on ice. 80 µl or 200 µl of SOC media (Invitrogen, Paisley, UK) was added near the flame to the transformed DH5α and BL21 (DE3) cells respectively. The transformed cells were then incubated at 37 °C, 220 rpm for 1 hr before they were plated on LB or M9 Agar plates (A2) which contained the relevant antibiotic. The plates were incubated at 37 °C overnight.

2.3.2 DNA Extraction

A single colony of DH5α cells transformed with the desired vector was inoculated into 5 ml LB containing either ampicillin or kanamycin antibiotic. This culture was then incubated at 37 °C, 220 rpm for 16 hours. The cells were then pelleted at 4000 rpm. The QIAPREP Spin Miniprep Kit (Qiagen, Manchester, UK) was used to extract and purify the DNA. The DNA concentration

was determined using a Shimadzu UV-1700 Pharmaspec spectrophotometer with the DNA concentration (µg/ml) calculated to be equal to OD_{260} x Dilution factor x 50 µg/ml.

The $A_{260/280}$ ratio was used to check the purity of the DNA. The vector DNA was then sequenced by providing 300-400 ng DNA and 3.2 pmol sequencing primer in a final reaction volume of 10 µl to the Functional Genomics, Proteomics and Metabolomics Facility (University of Birmingham).

2.3.3 Glycerol Stocks (BL21 (DE3))

A BL21 (DE3) colony transformed with the Cdc25C expression vector of choice was inoculated into 3 ml LB containing the appropriate antibiotic. The culture was then incubated at 37 °C, 220 rpm in a shaking incubator for a few hours until the OD_{600} was ~ 1.0. 2 ml of this culture was inoculated into fresh 10 ml L.B containing the relevant antibiotic and this was then further incubated (as before) for a few hours. Sterile glycerol was then added to the 12 ml culture at a final concentration of 20 %. The culture containing glycerol was then mixed and aliquoted into sterile cryogenic tubes followed by flash-freezing in liquid nitrogen. The cryogenic tubes were then stored at -70 °C.

2.3.4 Small scale growth and protein expression trials

A 2 µl volume of the desired glycerol stock was inoculated into 2 ml of fresh LB media selecting for the appropriate antibiotic resistance. This overnight culture was incubated at 37 °C, 220 rpm.

The next day 300 µl of the overnight culture was inoculated into 15 ml LB or Terrific media (Sigma-Aldrich, Dorset, UK), selecting for antibiotic resistance and grown until the OD_{600} was ~ 0.9. The temperature was then reduced to 18 °C. After 20 minutes a 1 ml pre-induction sample was taken and then IPTG was added to a final concentration of 1 mM. The pre-induction

sample was centrifuged at 10 000 x g for 1 minute and the supernatant was discarded. The pellet was stored at -20 °C.

The next morning a 1 ml sample of the induced cells was taken. This sample was also centrifuged and the supernatant was discarded.

The pellets were resuspended in 100 µl of PBS (Phosphate Buffered Saline) and 100 µl 2x Laemmli buffer (Sigma-Aldrich, Dorset, UK) was added. The samples were then subjected to sonication using a Soniprep 150 (MSE) sonicator. Each sample was sonicated at 5 microns for 30 seconds followed by 1 minute incubation on ice. This was repeated three times for each sample. After sonication the samples were heated at 95 °C for 3 minutes. Following this, 10 µl of each sample was loaded and an SDS-PAGE was performed.

2.3.5 Separating soluble and insoluble fractions

From the induced cultures 1 ml samples were taken and spun down at 10 000 x g for 1 minute. The drained pellet was resuspended in 200 µl PBS and the samples were kept on ice for 15-20 minutes. The samples were then sonicated and from this the cell lysate was obtained. The remaining sonicated sample was then centrifuged at 24 000 x g, 4 °C, for 3 minutes. The soluble fraction (supernatant) was then isolated. Laemmli sample buffer (2x) was added to the cell lysate, soluble and insoluble fractions which were then heated at 95 °C for 3-5 minutes. The samples were then checked using SDS-PAGE.

2.3.6 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The precision plus protein [TM] standard marker (Bio-Rad, Hertfordshire, UK) was loaded into the first well of a 4-12 % Bis-Tris criterion XT precast gel (Bio-Rad, Hertfordshire, UK). The samples were then loaded into the consecutive wells. The SDS-PAGE gel was run in MES buffer (Bio-

Rad, Hertfordshire, UK) at 180V for 35 minutes. The gel was then stained for 1 hour with InstantBlue (Expedeon, Cambridge, UK).

2.4 LB and Terrific media Large scale protein expression

A volume of 10 µl from the appropriate glycerol stock was inoculated into 25 ml LB or Terrific media (Sigma-Aldrich, Dorset, UK), selecting for the relevant antibiotic and incubated overnight at 37 °C, 220 rpm.

The next day 25 ml of the LB or Terrific media overnight culture was inoculated into 1L of LB or Terrific media, respectively. Depending on the construct between 1 - 5 L of large scale cultures were grown. These were grown for at least 5 hours at 37 °C, 200 rpm until the OD_{600} was ~ 1.0. The temperature was then lowered to 18 or 25 °C for at least 20 minutes before IPTG was added to a final concentration of 1 mM. The cultures were then incubated at this temperature overnight.

2.5 M9 small scale protein expression trials

2.5.1 M9 minimal media optimization

A single colony of BL21 (DE3) transformed cells was picked from an M9 plate and inoculated into 2 ml M9 minimal media containing ^{15}N and ^{13}C labeled nutrient mix (A3). Kanamycin was added at a 30 µg/ml final concentration. This culture was incubated overnight at 37 °C, 220 rpm. The following day two sets (A and B) of 3x15 ml cultures were used. Each set contained 15 ml cultures of M9 media alone, M9 media supplemented with ^{13}C, ^{15}N labeled 10 % w/v ISOGRO® (Sigma-Aldrich, Dorset, UK) and M9 media supplemented with 1 % v/v ^{13}C, ^{15}N labeled BioExpress®-1000 (Cambridge Isotope Laboratories, Massachusetts, USA). All the cultures were inoculated with 300 µl of the overnight culture. Nutrient mix (A3) and Kanamycin were added

to all the 15 ml cultures. The cultures were then grown at 37 °C in a shaking incubator (220 rpm) until they had an OD_{600} of 0.9.

Readings at OD_{600} were taken from each culture. Starting from time 0, readings were taken every hour for 13 hours, to plot bacterial growth curves. When the set B 15 ml cultures reached an OD_{600} of 0.9 the temperature was lowered to 18 °C and the cells were induced overnight with 1 mM IPTG. The next day, 1 ml samples taken from the set B cultures were centrifuged and sonicated. Following this, 10 µl samples were loaded and subjected to SDS-PAGE.

2.5.2 Temperature Optimization

Three 15 ml cultures were grown as described in section 2.5.1. Once the OD_{600} reached approximately 0.9, the temperature was lowered and each 15 ml culture was incubated either at 18, 25 or 37 °C. Once the pre-induction sample was extracted, the cultures were induced overnight with 1 mM IPTG. The next day 1 ml samples were extracted, centrifuged, sonicated, and prepared for SDS-PAGE analysis. The total cell lysate was centrifuged to separate the soluble and insoluble fractions which were also analysed by SDS-PAGE. The gel was stained using the coomassie based stain InstantBlue (Expedeon, Cambridge, UK) and then analyzed to identify the optimum temperature to provide the best yield of soluble protein.

2.6 M9 large scale protein expression

A BL21 (DE3) colony from an M9 plate was inoculated into 2 ml M9 minimal media containing ^{15}N labeled nutrient mix and kanamycin (30 µg/ml). This culture was grown for 8 hours. 50 µl of the culture was inoculated into 25 ml fresh M9 minimal media containing ^{15}N labeled nutrient mix and kanamycin. This culture was incubated overnight at 37 °C, 220 rpm.

Each 1L M9 culture grown was inoculated with 25 ml of overnight culture and incubated at 37 °C until an OD_{600} of 0.9 was obtained. Depending on the construct the temperature was then lowered to either 18 or 25 °C. IPTG was then added at a final concentration of 1 mM and the cultures were incubated at the desired temperature, shaking overnight at 180 rpm.

2.7 Purification

1-5 L cultures of LB or M9 were centrifuged at 6000 rpm in a Beckman Coulter centrifuge (Avanti J-20XP, JLA 8.1 rotor) for 15 minutes. The supernatant was then discarded and the pellet resuspended in His or GST lysis buffer (A4) with EDTA-free protease inhibitor cocktail Tablets (Roche Applied Science, West Sussex, UK). The cells were then lysed by passing them three times through a C-3 EmulsiFlex cell homogenizer (Avestin, Mannheim, Germany). The lysed cells were centrifuged at high speed (25 000 x g, 45 min, and 4 °C) to separate the soluble and insoluble fractions. The soluble fraction (supernatant) was then filtered with a 0.45 μm filter (Sartorius, Surrey, UK) and kept on ice. The protein was purified via His or GSTrap purification dependent on whether the protein had a His_6 or GST purification tag.

2.7.1 His purification

The 5 ml His Trap HP (GE Healthcare, Buckinghamshire, UK) column was first equilibrated at a flow rate of 1 ml/min with equilibration buffer (A4). The flow rate was kept constant through-out the purification. The filtered supernatant was then passed onto the column, followed by 50 ml of wash buffer (A4). The protein was then eluted with elution buffer (A4) and 1 ml fractions of the protein were collected. 20 μl samples of the supernatant, flow-through, collected wash and eluted fractions were analysed by SDS-PAGE. Eluted fractions which contained the protein of interest were pooled for down-stream purification. Size exclusion chromatography was the next step of purification unless cleavage of the His_6 tag was required.

2.7.2 Cleavage by TEV (Tobacco Etch Virus) protease

TEV protease (0.4 mg), which was expressed and purified in-house, was added to the purified protein. SnakeSkin (Thermo Scientific, Massachusetts, USA) tubing with a molecular weight cutoff of 3.5 kDa and 35 mm diameter was used to dialyze the eluted protein into TEV cleavage buffer (A4). This was left at 4 °C for 48 hours; recommended by the SGC protocol (Savitsky et al., 2010). The His_6 tag was cleaved while the protein was being dialyzed into the cleavage buffer. Following cleavage the protein is passed again onto the His Trap column to separate the cleaved protein from the uncleaved protein.

2.7.3 GST purification

GST tagged proteins were purified in a similar manner as the His tagged proteins. The differences being that a GSTrap column (GE Healthcare, Buckinghamshire, UK) was used instead with GST purification buffers (A4). The flow rate was kept at 1 ml/min. The protein was eluted with 10 mM reduced glutathione.

2.7.4 Size exclusion chromatography

Size exclusion chromatography was the final purification step. Superdex 200 or 75 (GE Healthcare, Buckinghamshire, UK) size exclusion columns were used. The type of column used was dependent on the molecular weight of the protein being purified. The columns were equilibrated with 1.2 to 2 column volumes of buffer. The flow rate was kept at 2.5 ml/min and 3 or 4 ml fractions were collected.

2.8 N-terminal sequencing

N-terminal sequencing was used to confirm the identity and cleavage of Cdc25C proteins. Coomassie stained SDS-PAGE gel slices of the required proteins were submitted to AltaBioscience (Birmingham) for N-terminal sequencing.

2.9 X-ray crystallography

2.9.1 Reductive Methylation

Reductive methylation was performed to methylate solvent exposed lysine residues to promote protein crystallization. To methylate Cdc25C the protocol published by Shaw et al. was followed. Briefly, the protein was purified in 50 mM sodium phosphate (pH 7.0), 150 mM sodium chloride, and 1 mM TCEP and for each reaction 10 mg of purified protein in 1 ml volume was used. For each methylation reaction, 20 µl of 1M DMAB (dimethylamine borane complex) and 40 µl of 1M formaldehyde were added followed by incubation in the dark at 4 ^{o}C. Following this 2 hour incubation, the whole procedure was repeated two times. Finally, 10 µl of 1M DMAB was added to the solution and incubated overnight while shaking at 100 rpm. The next day, the reaction was quenched by eluting in 20 mM Tris pH 8.0, 150mM NaCl, and 1 mM TCEP following size exclusion chromatography.

2.9.2   Protein Crystallisation

Around 800 different crystallization conditions were tested from a wide range of commercial screens including PACT (Molecular Dimensions, Suffolk, UK), JCSG+ (Molecular Dimensions, Suffolk, UK), Structure screen I and II (Molecular Dimensions, Suffolk, UK), Index (Hampton Research, Aliso Viejo, California), PEG/Ion I and II (Hampton Research, Aliso Viejo, California), PEGRx I and II (Hampton Research, Aliso Viejo, California), and the Wizard I, II, III, and IV (Emerald Biosystems, Washington, USA). The Mosquito liquid handler (TTP labtech, Hertfordshire, UK) was used to set-up the 96 well plates. The same experimental set-up was used as described in section 2.16. Following, incubation the plates were assessed for crystals after at least 72 hours.

2.10 Protein Concentration

The proteins were concentrated down using Vivaspin (GE Healthcare, Buckinghamshire, UK) and Amicon (Merck Millipore, Nottingham, UK) concentrators with a molecular weight cut-off of 5 or 10 kDa.

The standard UV assay measuring absorbance at 280 nM was used to measure protein concentration using a BioMate UV spectrophotometer (Thermo Scientific, Massachusetts, USA). Although Cdc25C has no tryptophan amino acids in the protein sequence it has other residues: tyrosine, phenylalanine and cysteine, which absorb weakly at 280 nM. The Bradford assay was not used because of the high arginine content in Cdc25C which resulted in significant errors when measuring protein concentration. The ProtParam tool (Gasteiger et al., 2005) was used to obtain the theoretical extinction coefficients of the proteins. The extinction coefficient when all cysteine residues are assumed to be reduced was taken and the equation below was used to calculate the protein concentration:

Concentration (M) = (absorbance at 280 nm x dilution factor) / (extinction coefficient when all cysteines are reduced x path length)

2.11 Analytical gel filtration

The Superdex 200 5/150 3 ml GL column (GE Healthcare, Buckinghamshire, UK) was used with the flow rate at 0.5 ml/min kept constant throughout. The column was initially equilibrated with three column volumes with the buffer 50 mM sodium Phosphate (pH 7.0), 200 mM L-arginine, 200 mM sucrose and 1 mM TCEP. 60 µl of a 100 µM stock of the Cdc25C catalytic domain in this buffer was first injected and passed through the column. The standards from the Gel Filtration Low Molecular Weight Calibration Kit (GE Healthcare, Buckinghamshire, UK) were made up in the same buffer as the Cdc25C protein. A volume of 60 µl of 100 µg of each

standard from the kit was injected into the column. The injection loop (50 µl) and syringe were washed thoroughly between each run. Following collection of the data the 280 nM UV traces were compared and elution volumes (V_e) for the protein and standards were noted. A graph of log MW against V/V_0 (elution volume/void volume) was plotted and the molecular weight for the protein was then calculated using the linear equation $y = mx + c$.

2.12 Mass Spectrometry

The purified proteins were dialyzed into 50 mM Ammonium Acetate, pH 6.8. 50 µg of protein was then provided to Dr Cleidiane Zampronio (School of Biosciences, University of Birmingham) for mass spectrometry (ESI-MS) analysis.

2.13 Analytical Ultracentrifugation (AUC)

Protein samples of 30 µM in sodium phosphate buffer were given to Mrs Rosemary Parslow (School of Biosciences, University of Birmingham) for AUC data collection. Samples were centrifuged using a Beckman Coulter ultracentrifuge (Beckman XL-1). They were centrifuged at 30 000 or 40 000 rpm depending on the protein being analyzed. The experiment was conducted at 20 °C and recorded at the wavelength of 280 nm. The SEDFIT program (Brown and Schuck, 2006) was then used to analyze the data.

2.14 Circular Dichroism (CD)

The Cdc25C proteins were dialyzed into 50 mM sodium phosphate (pH 7.5), and 0.5 mM TCEP. Samples with a final concentration of 1.6 mg/ml were loaded onto a 0.1 mm path length cuvette which was placed into a JASCO J-715 spectropolarimeter. The experiments were conducted in collaboration with Dr. Raul Pacheco-Gomez (School of Biosciences, University of Birmingham). Data were recorded between 260 and 190 nm wavelengths. For each protein the

buffer CD spectrum was subtracted from the protein CD spectrum to obtain the final CD spectrum. The CDSSTR analysis program from the DichroWeb server (Whitmore and Wallace, 2004) was then used to deconvolute this data and obtain the protein predicted secondary structure.

2.15 Thermal Shift Assay

The thermal shift assay was used to assess Cdc25C protein stability and identify optimum buffers to improve thermal stability. Following purification, the proteins were dialyzed into 50 mM sodium phosphate/Hepes (pH 7/7.5), 50 mM sodium chloride and 0.5 mM TCEP. The pH was either 7 or 7.5 depending on the pI of the protein to be studied.

The protein was diluted to 40 μM in the same buffer as the stock protein and the SYPRO® Orange fluorophore dye (Invitrogen, Paisley, UK)) was added. The protein/dye solution was then mixed and 2 μL of the protein/dye was pipetted into each well followed by 18 μL of the buffer condition to be tested. The final concentration of the protein was 4 μM with 5x SYPRO® Orange dye. The Absolute QPCR seal (Thermo Scientific, Massachusetts, USA) was then used to seal the 96 well plates (Thermo Scientific, Massachusetts, USA). The plates were centrifuged at 1000 x g to allow mixing of the protein and buffer. The 96 well plates were heated in a RT-PCR machine (Mx3005P QPCR, Stratagene, La Jolla, California).

Initially, the Cdc25C proteins were screened against an in-house 96 well plate ThermoFluor® buffer screen to identify solution conditions. These were then further optimised and run during subsequent screens. They were tested either on their own or with the addition of a small selection of salts and additives.

Fluorescence was measured for the temperatures ranging from 25 °C to 90 °C while the excitation and emission wavelengths were 492 nm and 568 nm respectively. The melting

temperatures (T_m) were determined using the MxPro software (Stratagene, La Jolla, California). T_m was taken from the point of inflection of the melting curve. The melting temperature was recorded for each buffer condition tested and the conditions which resulted in the maximum T_m were then selected and further optimised.

2.16 Solubility Screen Assay

The solubility screen was used to build on the thermal shift assay in order to identify buffer and additive conditions in which the Cdc25C catalytic domain was the most stable by exhibiting little to no precipitation. The best conditions identified from the ThermoFluor® screens were used in the solubility screen. These conditions were used alone or combined with an expanded range of additives. Different pHs were also tested. In addition to additives and solubility polymers, metal compounds were also tested as well as in-house developed solubility peptides.

The Cdc25C protein was concentrated down to 500 μM. The relevant buffer or buffer/additive was pipetted at 100 μL volume into a 96 well plate (Iwaki). Concentrated protein drops (100 nl) along with 100 nl drops of the reservoir buffer were pipetted onto the viewdrop2 plate cover using the Mosquito liquid handler (TTP labtech Hertfordshire, UK). The plate cover was then inverted and the plate was sealed. The plates were then incubated for at least 48 hours at 23 °C. A light microscope was used to visualise the protein drops for precipitation. A scoring sheet was used to classify the extent of precipitation; 0 being no precipitation and 4 being maximum precipitation.

2.17 Small-angle X-ray Scattering (SAXS)

2.17.1 Data Collection

The SAXS data was collected from EMBL (DESY, Hamburg, Germany) and ESRF (Grenoble, France) using the X33 and BM29 BioSAXS beamlines respectively.

Depending on the Cdc25C protein analysed, it was either in 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, and 1 mM TCEP or 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine, 200 mM sucrose, and 1 mM TCEP. Data were collected for three different concentrations for each protein. The concentrations of the Cdc25C proteins were between 1 – 10 mg/ml. The facilities allowed automatic buffer subtraction from the scattering plots and hence this was conducted on site.

2.17.2 Data Analysis

The ATSAS 2.5.1 suite of programs which was downloaded from the EMBL Hamburg website was used to process and analyse the collected SAXS data.

The auto-subtracted scattering plots were initially checked for data quality by assessing radiation damage and signs of aggregation. In the case of the Cdc25C catalytic domain no aggregation was seen so the highest concentration was taken as the best scattering curve. For the His-Cdc25C$_{73\text{-}441}$ construct slight aggregation was noted at low scattering angles for the highest concentration. Therefore, the PRIMUS (Konarev et al., 2003) programme was used to create a merged data set using the scattering plots for the lower protein concentrations.

PRIMUS was used to obtain the value for the radius of gyration (Rg) using the Guinier approximation approach. GNOM was also used to obtain the Rg value which was compared for consistency to the Rg value obtained from PRIMUS. The pair-distance distribution function (P(R)) graph created using GNOM was used to ascertain the particle maximum dimension

(D_{max}). Where the pair-distance distribution function returned to zero was taken as the D_{max}. In cases where this was not obvious different D_{max} values were evaluated by manually inputting increasing D_{max} values and observing changes in the P(R) graph.

The envelope for the Cdc25C catalytic domain was created using programs from the ATSAS suite. Default parameters were used for all the programs used. The DAMMIF (Franke and Svergun, 2009) program was used to generate the initial *ab initio* models which were then evaluated, filtered and averaged using the program suite DAMAVER (Volkov and Svergun, 2003) to obtain a final envelope which best fitted the protein scattering plot.

The crystal structure of the Cdc25C catalytic domain (PDB: 3OP3) was then fitted into the envelope using the SUPCOMB program (Kozin and Svergun, 2001). This program provides a normalized spatial discrepancy (NSD) value for the fit where an NSD value < 1 is indicative of similarity between the two models and a value > 1 is due to differences.

The CRYSOL program (Svergun et al., 1995) was also used to compare the atomic resolution X-ray model to the low resolution SAXS solution model. This was done by using CRYSOL to generate a theoretical SAXS scattering curve for the Cdc25C catalytic domain (PDB: 3OP3) and fitting this to the experimental SAXS scattering curve. CRYSOL provided an χ^2 value for the fit. The smaller the value the better the fit and a value approaching zero indicates similarity between the two structures.

2.17.3 Determining flexibility

The Ensemble Optimisation Method (EOM) online program suite (Bernadó et al., 2007) was used to evaluate the flexibility of the proteins. This program suite requires at least two input files. These consist of the sequence file, SAXS experimental data file, and a PDB model of the folded domain. Since there is only one PDB model available of the Cdc25C catalytic domain

(PDB: 3OP3), ITASSER was used to generate additional models when required. The EOM suite was used using the default parameters. EOM initially generates 10, 000 models. A genetic algorithm is then used to select a sub-set of 20 models which most represent the experimental data. An averaged SAXS curve from the sub-set is obtained which is then fitted to the experimental SAXS curve. Flexibility was determined quantitatively with the calculated chi-squared value for the fit and qualitatively by observation.

2.18 Nuclear Magnetic Resonance (NMR) Spectroscopy

2.18.1 Sample preparation and Optimization of experimental parameters

Proteins to be studied by NMR were dialyzed into 50 mM sodium phosphate buffer. The pH, salt and additive concentrations as well as the type of additive varied depending on the protein sample. The sample volume was 600 µl which included 10 % v/v deuterium oxide (D_2O). Once the sample was prepared it was then positioned in the magnet. Following this, the lock signal was located and then centered, effectively locking the sample against the D_2O signal.

Automated gradient shimming was performed on the magnet in the z dimension to make the magnetic field homogeneous. Shimming was followed by tuning and matching of the spectrometer. Other parameters such as the transmitter offset (tof) and 90° pulse width (pw) were then optimized.

NMR experiments were run at the Biomolecular NMR facility (University of Birmingham) using the Agilent 600, 800 and 900 MHz spectrometers all equipped with triple resonance cryogenically cooled probes. Standard pulse sequences were taken from the Varian biopack. A Bruker 600 MHz BACS60 autosampler was used to collect waterLOGSY data with help from Dr Christian Ludwig (NMR Scientific Officer, University of Birmingham). All experiments were performed at 25 °C.

2.18.2 Protein 1D NMR

1D NMR spectra of the protein were initially collected to determine if the protein was folded before a HSQC/SOFAST-HMQC experiment commenced.

2.18.3 ^{1}H, ^{15}N – HSQC

The ^{15}N-labelled proteins were studied by Heteronuclear Single Quantum Coherence (HSQC) spectroscopy where 1670 points were collected in the direct dimension and 128 increments in the indirect dimension. The spectral widths were 13 ppm and 34 ppm in the direct and indirect dimensions respectively. The ^{1}H-^{15}N-HSQC pulse sequences from the Agilent biopack were used to conduct the experiments (Bodenhausen and Ruben, 1980; Kay et al., 1992).

2.18.4 ^{1}H, ^{15}N SOFAST- HMQC

Band-Selective Optimized Flip Angle Short Transient (SOFAST) - Heteronuclear Multiple Quantum Coherence (HMQC) spectroscopy is an alternative method used to acquire 2D spectra (Schanda et al., 2005). It provides the advantage of reducing the acquisition time while maintaining a good signal to noise ratio. The Agilent 600 or 800 MHz spectrometers were used for data collection with 768 points collected in the direct dimension and 128 increments in the indirect dimension. The spectral widths were 11.9 ppm in the ^{1}H dimension and 34 ppm in the ^{15}N dimension.

2.18.5 Compounds

Promising inhibitor compounds were identified for Cdc25C following an extensive research of the literature. These compounds were bought from Sigma-Aldrich (Dorset, UK), Atlantic Research Chemicals (Cornwall, UK) and ChemDiv (San Diego, USA). Where compounds were not available, similar compounds were purchased instead. The compounds which were tested

using the NMR techniques described in this section (2.19) are listed in a table in the appendix (A5).

2.18.6 WaterLOGSY

Compound stock solutions were made up in 100 % DMSO-d_6. Each compound at 0.5 mM concentration was tested via WaterLOGSY in the buffer 50 mM sodium phosphate (pH 7.0), 150 mM sodium chloride, and 1 mM TCEP. The Cdc25C catalytic domain protein was added at a final concentration of 5 µM. The WaterLOGSY (Dalvit et al., 2001) experiment was run for 15 minutes with a mixing time of 3 seconds using the Bruker 600 MHz BACS60 autosampler. The spectral width was set at 13 ppm and 256 scans were collected. The same phase correction was applied to the WaterLOGSY spectra +/- protein.

2.18.7 Data Processing

The NMRPipe (Delaglio et al., 1995) software was used to process the data. Data processing involved Fourier transformation followed by phase and base line correction of the spectra. The spectra were analyzed using the CcpNmr analysis software (Vranken et al., 2005).

Chapter 3.0 - Characterisation of the Cdc25C protein

In this chapter the Cdc25C full-length protein was analysed. Prediction tools were used to assess the secondary structure content and to aid design of constructs. The suitability of the Cdc25C full-length protein was assessed for SAXS. As mentioned earlier there are no 3D atomic structures of the Cdc25C full-length protein and therefore this was an attractive goal. Unfortunately, a SAXS envelope of the full-length protein could not obtained due to the lack of secondary structure in the N-terminal domain. The main focus in this chapter is the Cdc25C N-terminal (regulatory) domain while later chapters will focus on the Cdc25C catalytic domain.

3.1 Cdc25C Secondary Structure Prediction

In order to design expression constructs for the regulatory domain and full-length, secondary structure predictions of full-length Cdc25C were generated using DisMeta, a disorder prediction metaserver (Huang et al., 2014). DisMeta generated predictions from PROFsec and PSIPred predictors and also provided a consensus prediction as shown in Figure 3.1 (Rost and Sander, 1994; McGuffin et al., 2000). The N-terminal regulatory domain is predicted to have limited structure which is mainly helical content while the C-terminal catalytic domain is predicted to be much more structured.
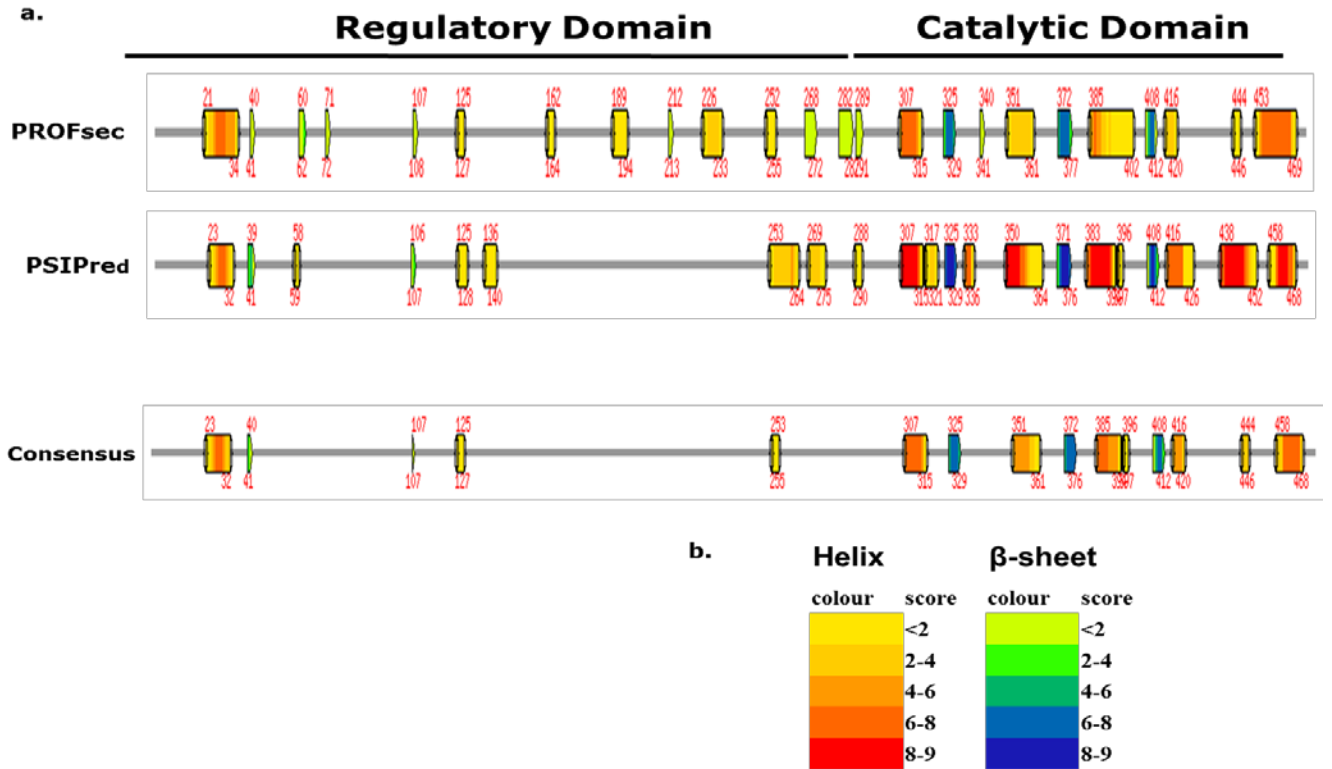
Figure 3.1 Secondary Structure Prediction

DisMeta secondary structure prediction output for the full-length Cdc25C protein (a). The scoring system is illustrated below where higher the score the stronger the prediction (b). Overall, the Cdc25C catalytic domain is likely to be more structured compared to the regulatory domain, which is predicted to be largely disordered.

3.2 Cdc25C Disorder Prediction

Disorder predictions for the Cdc25C full-length protein (Figure 3.2) were generated from the Regional Order Neural Network (RONN) predictor (Yang et al., 2005) for further analysis of likely domain boundaries and unstructured regions. The RONN predictor provides a probability value for the disorder for each amino acid residue.

The regions which were predicted to be disordered by RONN are amino acid residues 1 - 27, 77 - 110, 129 - 164, 171 - 233, and 436 – 468.
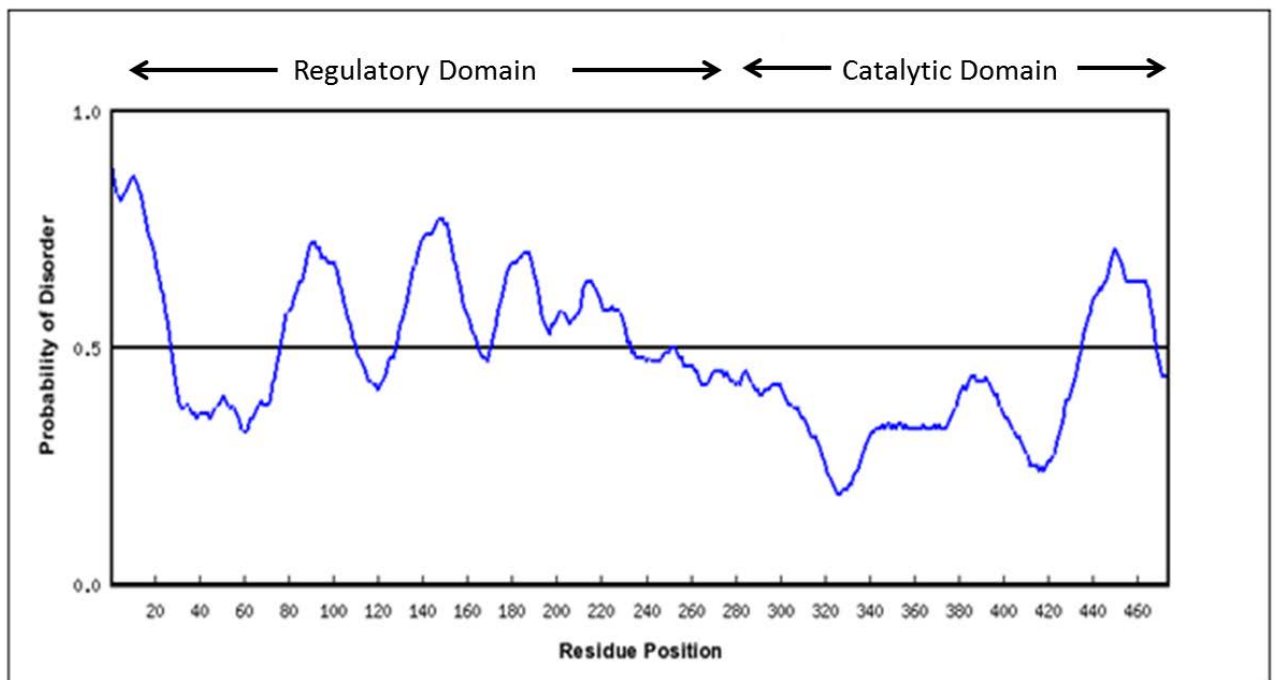
Figure 3.2 Disorder Prediction

Cdc25C disorder prediction by RONN (Yang et al., 2005). The black horizontal line signifies the threshold with a 0.5 probability cut-off. Amino acid residues with a probability over this threshold are predicted to be disordered.

3.3 Cdc25C constructs

Using information from a combination of the literature, secondary structure, and disorder predictions, full-length and N-terminal domain Cdc25C constructs were designed (Figure 3.3a). The expression level of these proteins were assessed using the SDS-PAGE gel from small scale expression and purification trials. Figure 3.3b summarizes the expression levels of each of the proteins. His-Cdc25C$_{1-280}$, GST-Cdc25C$_{1-280}$, and His-Cdc25C$_{36-443}$ expressed poorly while His-Cdc25C$_{73-441}$ and His-Cdc25C$_{83-441}$ had the best expression levels.
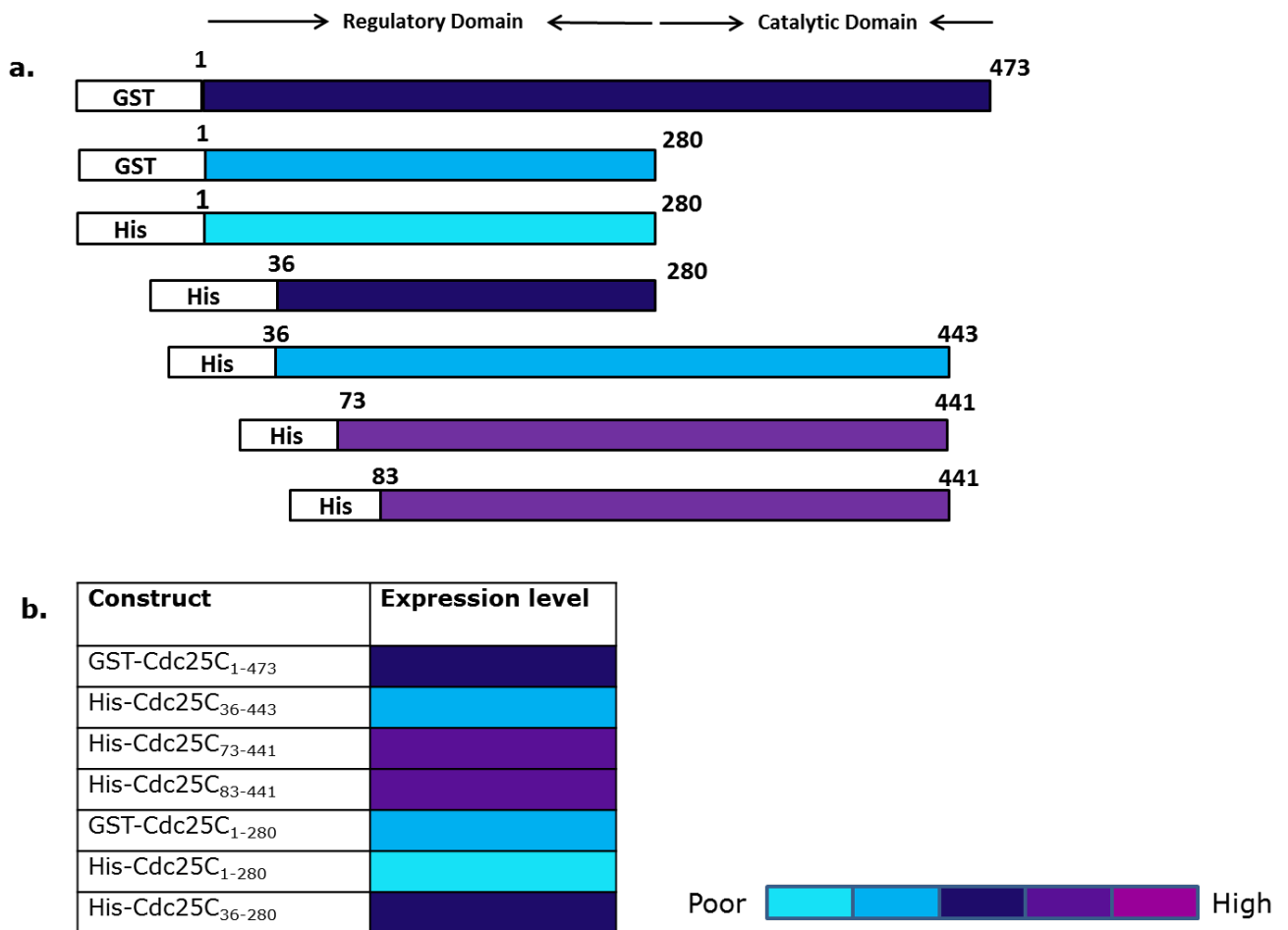
Figure 3.3 Cdc25C constructs

A schematic diagram illustrating the Cdc25C constructs (a). The proteins were constructed with either an N-terminal GST or His tag. The Table (b) provides a qualitative assessment of the expression level of each protein based on the SDS-PAGE gel from expression and small scale purification trials. The poor to high end of the scale corresponds to an expression level of around 30 µg up to 1 mg of purified protein per litre culture.

3.4 Characterisation of the N-terminal domain

3.4.1 Purification of His-Cdc25C$_{36-280}$

The His-Cdc25C$_{36-280}$ construct was chosen for the characterisation of the N-terminal domain because it had the best expression level compared to the other Cdc25C N-terminal domain constructs tested which were His-Cdc25C$_{1-280}$ and GST-Cdc25C$_{1-280}$ (Figure 3.3). Changing the His tag for a GST tag did not improve expression. However, removal of the first 35 amino acids from the Cdc25C amino acid sequence improved the level of expression.

Following, Histrap purification and cleavage of the N-terminal His tag a gel sample was subjected to N-terminal sequencing. N-terminal sequencing confirmed identity and cleavage of this protein. Cdc25C$_{36-280}$ was then subjected to further purification via size exclusion chromatography (Figure 3.4.1). Three peaks can be seen in the size exclusion profile (Figure 3.4.1a). Fractions 1-2 are from the first peak, fractions 3-5 are from the second peak, and fractions 6-13 are from the last peak (Figure 3.4.1b). The majority of the Cdc25C$_{36-280}$ protein was present in the third peak having a molecular weight below 37 kDa.
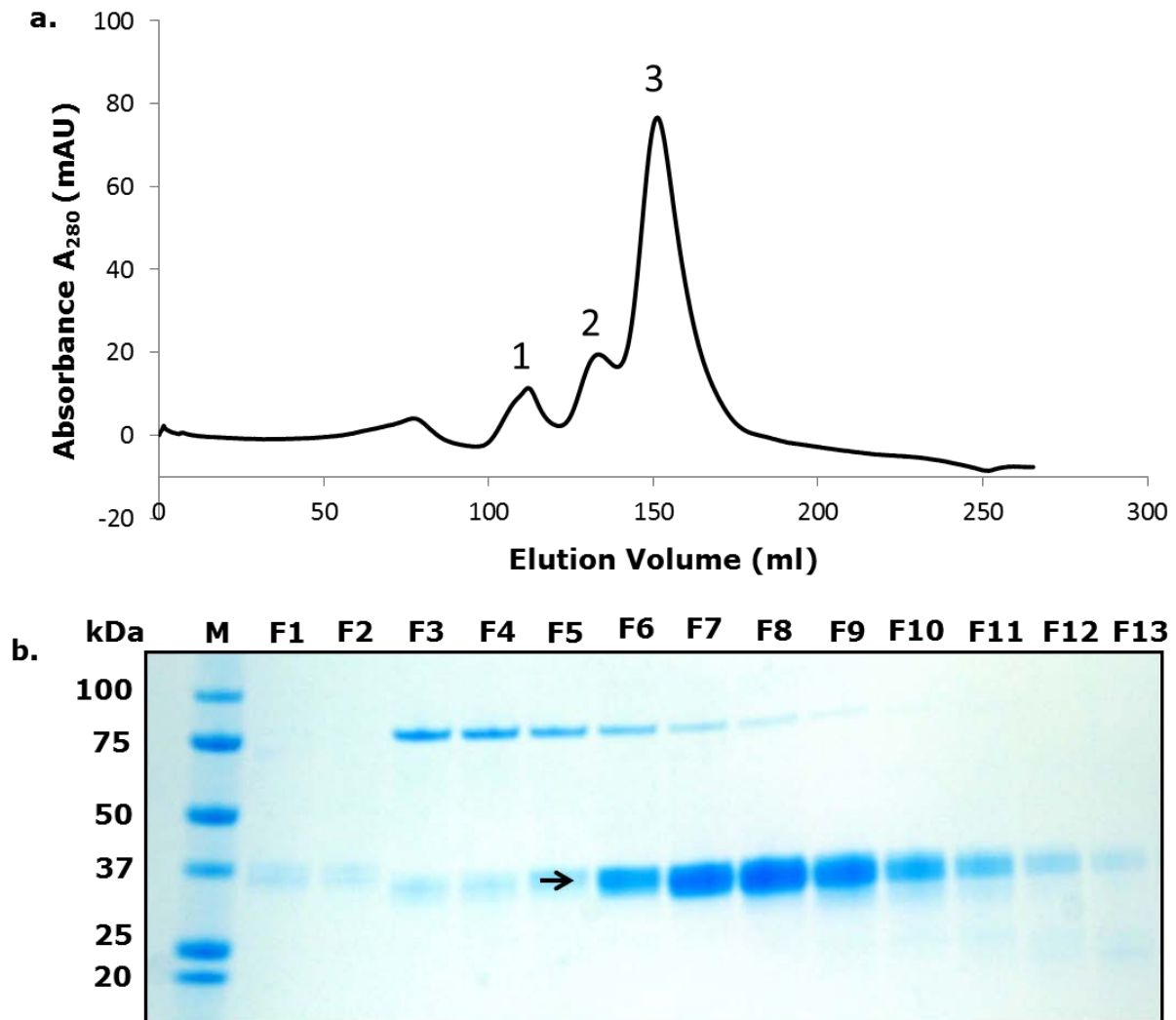
Figure 3.4.1 Purification of Cdc25C$_{36-280}$ by size exclusion chromatography

The Superdex 75 26/600 column was used to further purify Cdc25C$_{36-280}$ following Histrap purification. The protein was eluted in 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, and 1 mM TCEP. The UV A$_{280}$ size exclusion profile is shown (a). This profile has three peaks and Cdc25C$_{36-280}$ elutes in the third peak (140 – 175 ml). An SDS-PAGE was run with a 4-12 % Bis-Tris criterion XT precast gel (b). The first lane is the marker (M), followed by the fractions 1-2 which are from the first peak, fractions 3-5 are from the 2nd peak and fractions 6-13 are from the 3rd peak. Cdc25C$_{36-280}$ is depicted by a black arrow.

3.4.2 Proton NMR spectroscopy of Cdc25C$_{36-280}$

The folded state of Cdc25C$_{36-280}$ was assessed using 1D NMR experiments (Figure 3.4.2). The peaks in the 1D spectrum are not well dispersed which is generally indicative of disordered proteins (Page et al., 2005). Well-folded proteins usually have resolved peaks in the spectral regions between 11 to 1 ppm. The peaks for Cdc25C$_{36-280}$ range from 8.7 – 0.6 ppm and there are no peaks with negative ppm values. This indicates that Cdc25C$_{36-280}$ contains a significant amount of disorder.

The spread of peaks in the amide region (7-10 ppm) also provides a good indication on the folded state of a protein. The amide region for Cdc25C$_{36-280}$ is narrow ranging from 8.7 to 6.7 ppm and the peaks in this region are also not sharp which suggests that Cdc25C$_{36-280}$ may be unfolded.
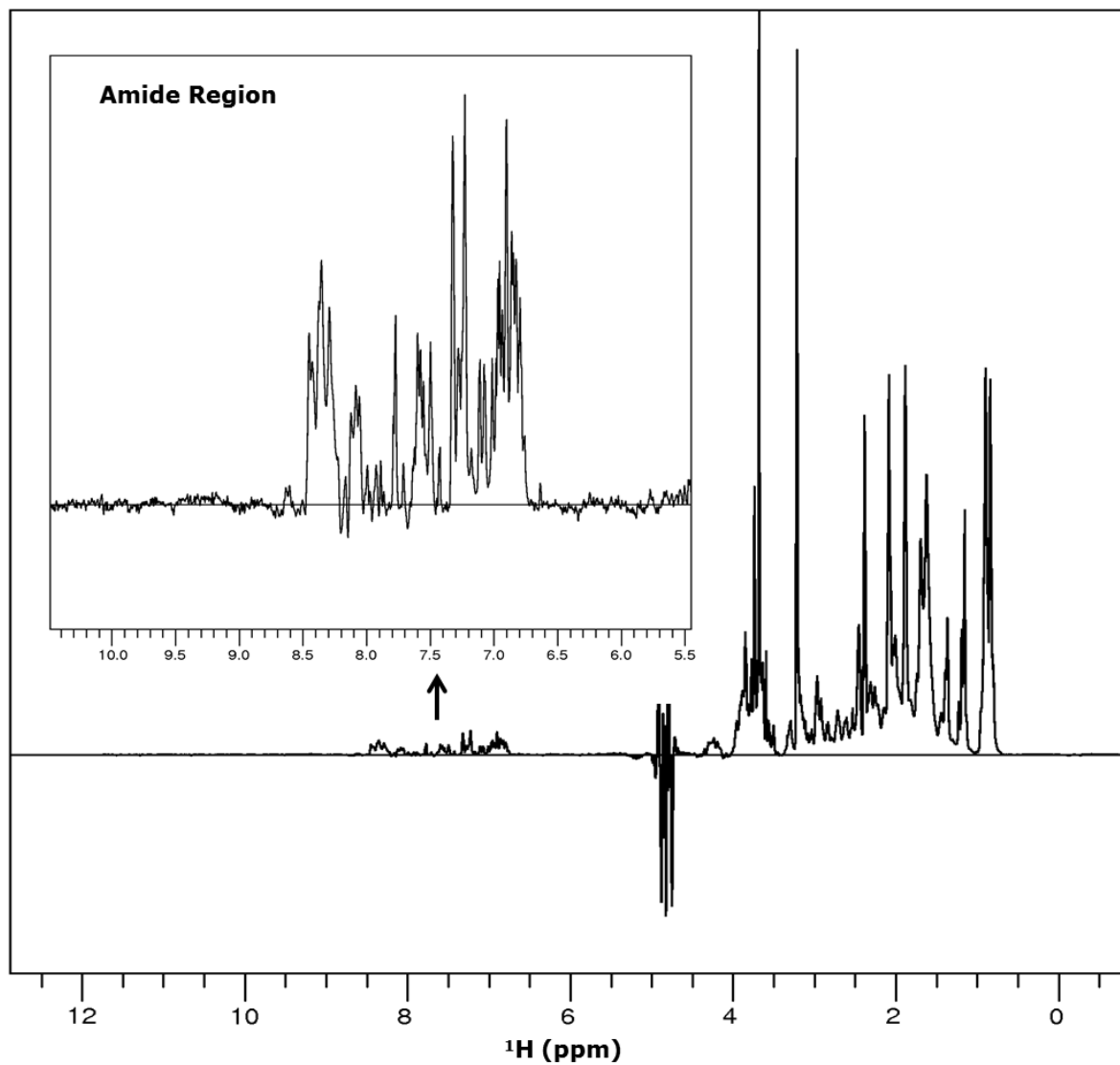
Figure 3.4.2 1D NMR of Cdc25C$_{36-280}$

A 50 µM sample of Cdc25C$_{36-280}$ in 50 mM sodium phosphate (pH 7.5), and 0.5 mM TCEP was subjected to 1D NMR. The Agilent 800 MHz spectrometer was used to perform the experiment. The poorly dispersed 1D NMR signals suggest that Cdc25C$_{36-280}$ is not well-folded.

3.4.3 Circular Dichroism

Circular Dichroism (CD) is a technique based upon the principle of how circularly polarised light is affected when passing through an "optically active" medium (Kelly et al., 2005). Molecules which are "optically active" are also known as being chiral. Biomolecules such as proteins possess chirality.

Only chiral molecules will have an effect on circularly polarised light because they absorb the two components of circularly polarised light in different proportions. The resulting difference in absorbance (ΔA) (see eq. below) is measured in a CD experiment and commonly expressed as ellipticity in degrees (Kelly and Price, 2000). Ellipticity in degrees is plotted as a function of wavelength.

$$\Delta A = A_L - A_R$$

(ΔA = change in absorbance, A_L = absorbance of the left component of circularly polarised light, and A_R = absorbance of the right component of circularly polarised light)

The Far-UV region (240-180 nm) of the CD spectrum is a good measure of the secondary structure of a protein (Kelly and Price, 2000). Each secondary structure type has its own characteristic CD signature. For example, the α-helix has two negative minima at 222 nm and 208 nm as well as a positive peak at 190 nm. The β-sheet CD spectrum shows a negative peak at 218 nm and a positive one at a wavelength of 196 nm. The random coil structure has a positive peak at 212 nm and a negative one around 195 nm. The main limitation of the CD technique is that it does not define which amino acids of the protein possess a particular secondary structure.

3.4.3.1 CD of Cdc25C$_{36-280}$

A CD spectrum was collected for Cdc25C$_{36-280}$ (Figure 3.4.3.1). The CD profile (experimental data) is negative at 190 nm which indicates the protein is unfolded. The CD spectrum shows

three negative bands at 196, 200 and around 203 nm respectively. The peak between 195 and 200 nm indicates the presence of the β-sheet component. The first negative band at 196 nm is typical of a random coil structure. The steep descent from -1 mdeg and negative bands at 200 nm and around 203 nm indicate some α-helix composition. The negative contributions of the α-helix and β-sheet components result in the CD profile not becoming positive which suggests that the N-terminal domain is not completely random coil and does contain some regular structural elements.

The CDSSTR program from DichroWeb was used to deconvolute the acquired CD data (Sreerama and Woody, 2000). It can be seen that the experimental data fits well with the reconstructed data (Figure 3.4.3.1). The CDSSTR program provided quantitative values for the different secondary structure elements. These values represent an average of the whole protein. Overall, 63 % of Cdc25C$_{36-380}$ appears to be disordered while 24 % is composed of α-helix and β-sheet.

b.

α-Helix (%)	β-strand (%)	Turns (%)	Unordered (%)
5	19	12	63
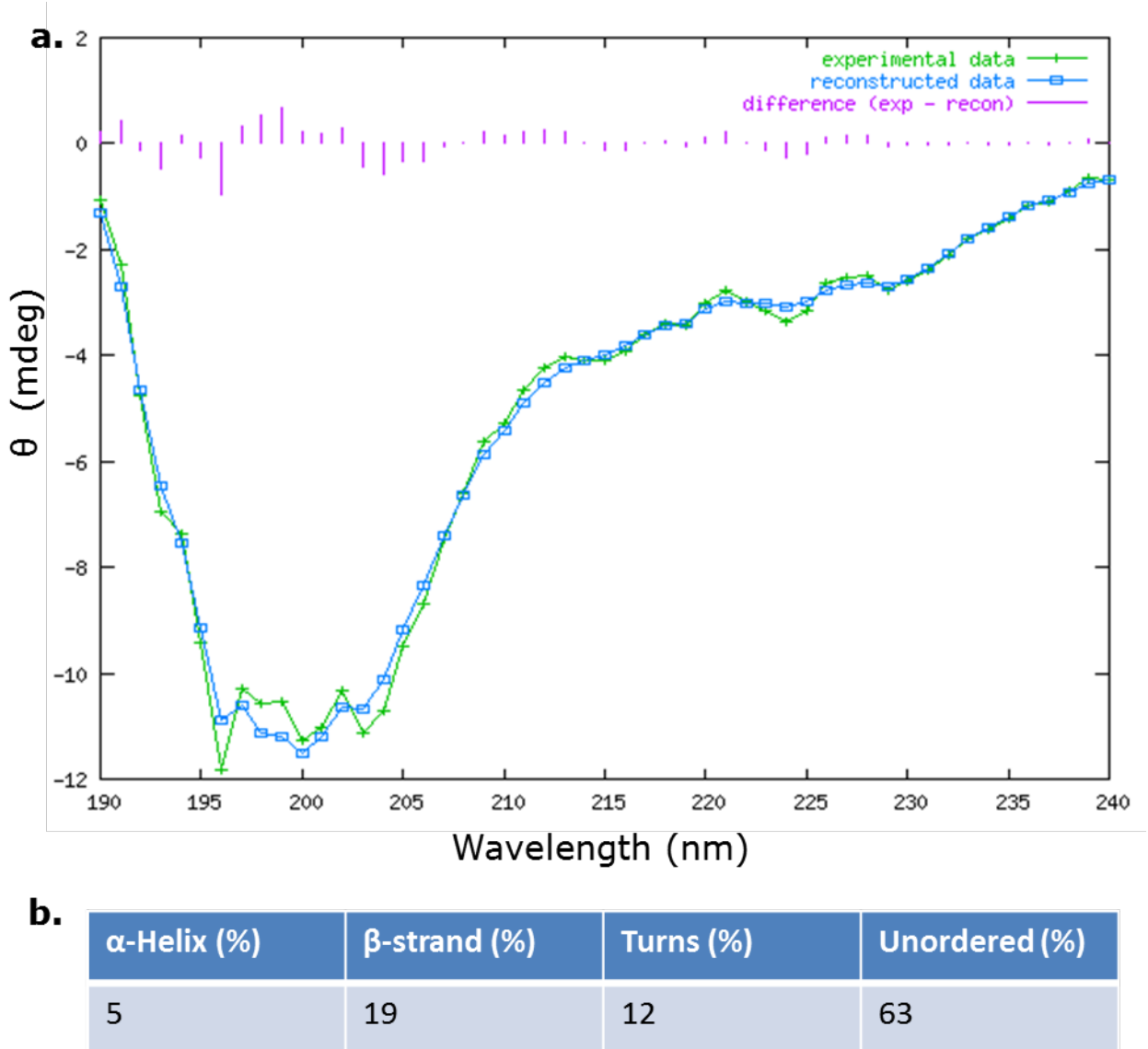
Figure 3.4.3.1 Far UV CD spectrum of Cdc25C$_{36-280}$

Cdc25C$_{36-280}$ was subjected to circular dichroism analysis in 50 mM sodium phosphate (pH 7.5), and 0.5 mM TCEP. The Dichroweb server was used for the CD data analysis. The data is shown graphically (a) and the Table (b) shows the secondary structure (%) content calculated by the CDSSTR program. The majority of the protein is unordered with only 36 % possessing secondary structure.

PROFsec and PSIPred (Rost and Sander, 1994; McGuffin et al., 2000) are tools which provide predictions of secondary structure (%). These predictors use a system of neural networks and predictions are based on the amino acid sequence. Both have been shown to have good levels of accuracy; PROFsec has an accuracy of 72 % and PSIPred has an accuracy of 76 %.

The regular secondary structure composition (%) derived from CD and from two secondary structure predictors were compared (Table 3.4.3.1). Based on the CD data $Cdc25C_{36-280}$ possesses 24 % regular secondary structure. In comparison, PROFsec predicted $Cdc25C_{36-280}$ to have 8.1 % of regular secondary structure and PSIPred predicted it to have 15.8 % regular secondary structure. Overall, CD indicates that $Cdc25C_{36-280}$ is unfolded but has more regular secondary structure than predicted.

	CD	PROFsec	PSIPred
% Regular secondary structure (α-helix + β-sheet)	24	8.1	15.8

Table 3.4.3.1 Regular secondary structure composition of Cdc25C$_{36-280}$

A comparison of the experimental secondary structure composition (%) derived from CD and the predicted secondary structure composition (%) using two popular prediction tools, PROFsec and PSIPred (Rost and Sander, 1994; McGuffin et al., 2000)*.*

3.5 Characterisation of full-length Cdc25C

3.5.1 Purification of GST-Cdc25C$_{1-473}$

In order to resolve the molecular envelope of full-length Cdc25C, the protein was expressed and purified for small angle X-ray scattering (SAXS) experiments. The His$_6$ tagged Cdc25C full-length has already been shown to be poorly soluble and the GST-tagged full-length is soluble (Fan et al., 2012). My aim therefore was to assess the suitability of this construct for SAXS.

The GST-tagged full-length Cdc25C protein was purified via GSTrap purification. The protein was ~ 83 kDa in size (Figure 3.5.1.1). The protein band around 25 kDa is likely to be the GST protein which has a molecular weight of 26 kDa.

GST-Cdc25C$_{1-473}$ was further purified by size exclusion chromatography using the Superdex 200 column (Figure 3.5.1.2a). The size exclusion profile shows two peaks. The first peak is a sharp narrow peak and the second peak is a small broad peak. The protein was eluted in the first peak which was straight after the 40 ml void volume. The gel fractions 1-6 were from this peak (Figure 3.5.1.2b). The second peak in the size exclusion profile is the GST peak with the gel fractions 7-10. Three different bands can be seen in these fractions. These are likely to be variants of the GST protein.

The majority of the GST-tagged full-length Cdc25C protein was present in fractions 2 and 3 (Figure 3.5.1.2b). The protein is not very pure since there are lower molecular weight bands present. This suggested that the protein contained low molecular weight contaminants or break-down products.
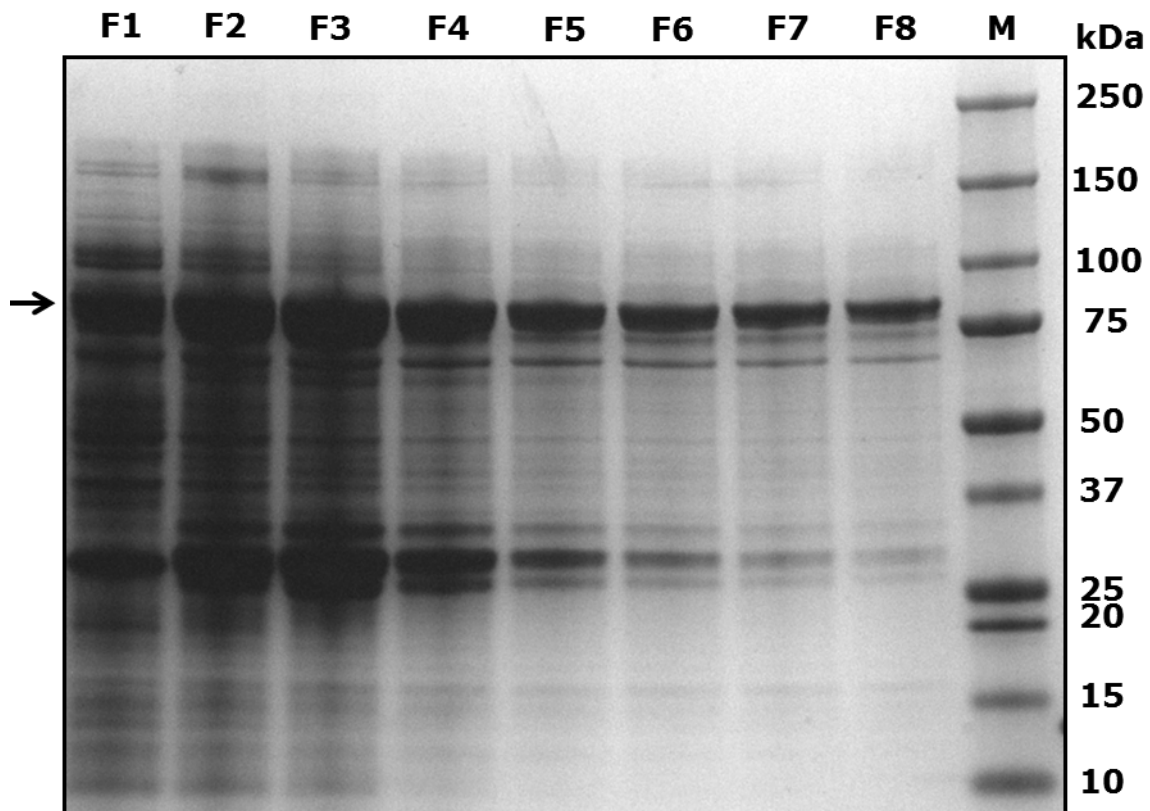
Figure 3.5.1.1 GSTrap purification of GST-Cdc25C$_{1-473}$

GST-tagged Cdc25C full-length (GST-Cdc25C$_{1-473}$) was purifed via GSTrap purification. The protein was eluted in 50 mM sodium phosphate (pH 7.5), 250 mM sodium chloride, 1 mM EDTA, 20 mM Glutathione, and 1 mM TCEP. Following this an SDS-PAGE was run with a 4-12 % Bis-Tris criterion XT precast gel. The last lane is the marker (M) which is preceded by the eight eluted fractions (F1-8). The position of GST-Cdc25C$_{1-473}$ is shown by the black arrow.
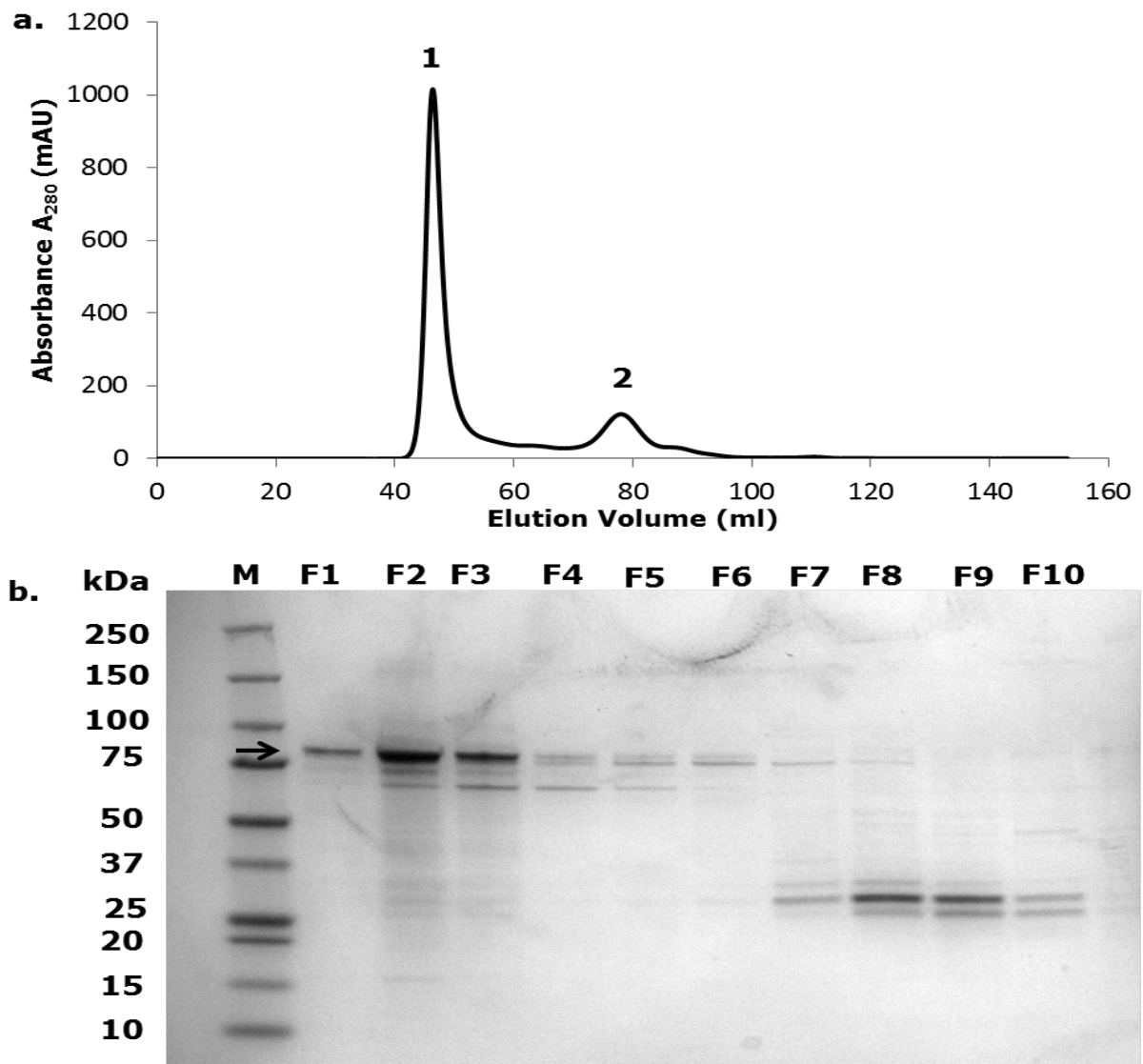
Figure 3.5.1.2 Size exclusion chromatography purification of GST-Cdc25C$_{1-473}$

UV A$_{280}$ size exclusion profile of GST-Cdc25C$_{1-473}$ following purification using the Superdex 200 16/600 column (a). The size exclusion profile shows two peaks. An SDS-PAGE was run with a 4-12 % Bis-Tris criterion XT precast gel. The first lane is the marker followed by fractions 1-6 which are from the first peak and fractions 7-10 are from the second peak (b). The first peak contains the ~ 83 kDa eluted protein which is depicted by the black arrow (b). The second peak contains the GST protein which is shown by protein bands above 25 kDa (b). The protein was eluted in 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, 1 mM EDTA, and 1 mM TCEP.

3.5.2 Analytical ultracentrifugation (AUC) analysis of GST-Cdc25C$_{1-473}$

Analytical ultracentrifugation (AUC) has many uses depending on the type of experiment being conducted, whether it is a sedimentation velocity or an equilibrium experiment (Cole et al., 2008). The sedimentation velocity experiment can be used to assess the oligomeric state of proteins. It can also provide a very good estimate of the size of the protein.

Here the sedimentation velocity experiment was used because the aim was to assess the oligomeric state and hence the quality of the GST-tagged full-length protein for SAXS experiments. The sedimentation velocity experiment measures the rate of movement of molecule boundaries after the application of the centrifugal force. When high centrifugal force is applied on the sample in the ultracentrifuge, the sample is exposed to a high gravitational field which results in the macromolecules sedimenting. This sedimentation is based upon the protein's hydrodynamic radius and accounts for the separation of the macromolecules in the sample.

A monodisperse species is represented by a single peak in the AUC profile. Here, there are two peaks followed by increasing c(s) distribution (Figure 3.5.2). The first peak is a sharp peak followed by a broader peak. This indicated the presence of a mixture of proteins with molecular weights ranging from 20 kDa onwards. The large increase in the c(s) distribution at the end suggested the presence of a large species indicating that the sample contained aggregated species.
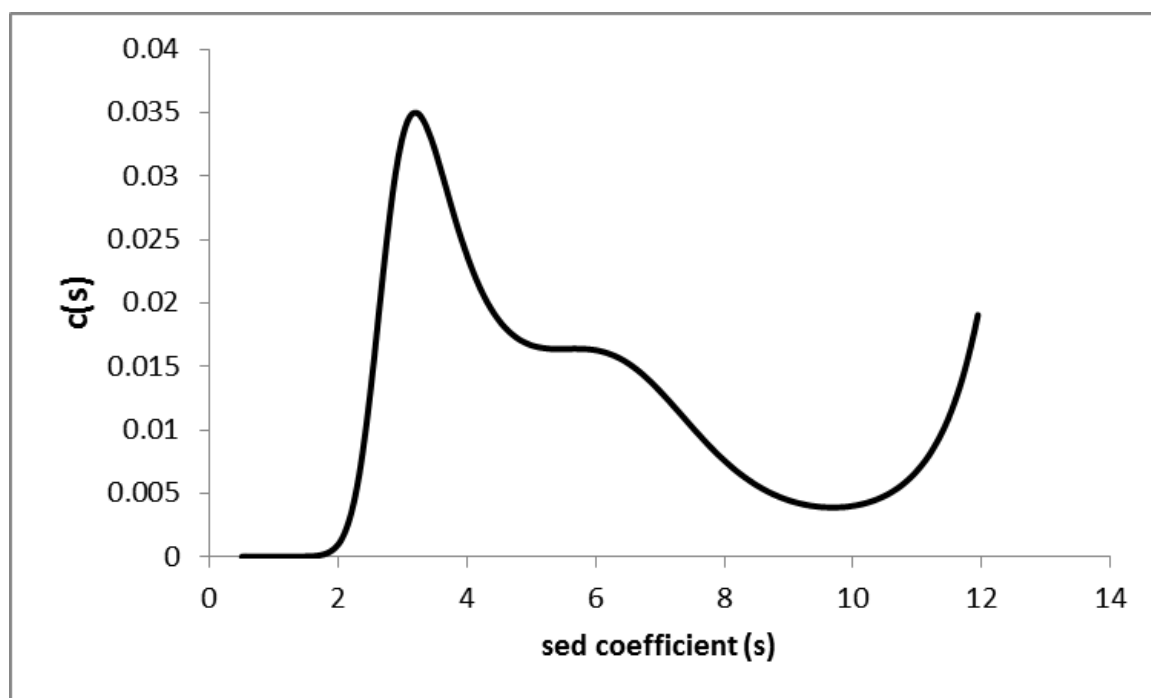
Figure 3.5.2 AUC profile of GST-Cdc25C$_{1-473}$

GST-Cdc25C$_{1-473}$ was subjected to AUC analysis in 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, 1 mM EDTA, and 1 mM TCEP. The SEDFIT (Brown and Schuck, 2006) *program was used to perform the c(s) analysis.*

3.5.3 Purification of His-Cdc25C$_{73-441}$

Purified GST-Cdc25C$_{1-473}$ protein was not monodisperse, presumably due to the lengthy disordered regions, and hence could not be used for SAXS experiments. Therefore, an alternative Cdc25C full-length construct was required. From the alternative Cdc25C constructs tested His-Cdc25C$_{73-441}$ was selected because it was the longest construct which also had the best expression level (Table 3.3, Figure 3.3).

The identity and cleavage of Cdc25C$_{73-441}$ was confirmed by N-terminal sequencing following Histrap purification and cleavage of the His$_6$ tag. The Cdc25C$_{73-441}$ protein was then further purified via size exclusion chromatography (Figure 3.5.3).

There are two peaks seen in the size exclusion profile. The first small peak is a higher order species and the second large peak contains the protein of interest (Figure 3.5.3a). Fractions 1-12 are from this peak (Figure 3.5.3b). It can be seen that Cdc25C$_{73-441}$ can be purified to homogeneity. The purest fractions were pooled and used for further experiments.
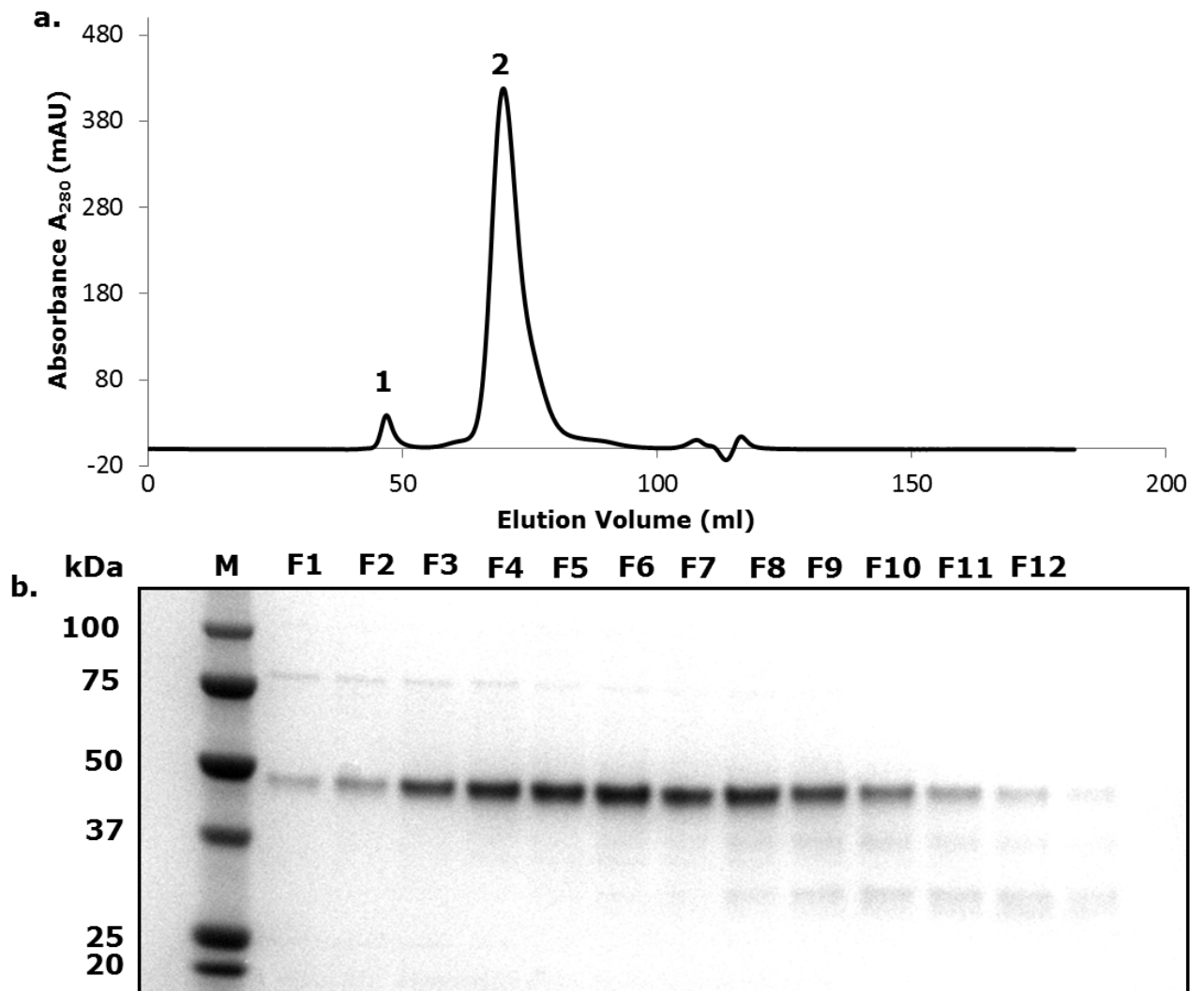
Figure 3.5.3 Purification of Cdc25C$_{73\text{-}441}$

Size exclusion profile of Cdc25C$_{73\text{-}441}$ with milli absorbance (mAU) measured at 280 nm against volume (ml) (a). The protein was eluted in 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, and 1 mM TCEP. An SDS-PAGE was run with a 4-12 % Bis-Tris criterion XT precast gel. Fractions 1-12 are from peak 2 (b).

3.5.4 CD of Cdc25C$_{73-441}$

The CD spectrum of Cdc25C$_{73-441}$ is positive at 190 nm suggesting a structured protein (Figure 3.5.4). The positive band around 190 nm indicates the presence of some α-helical and β-sheet content. The positive band is followed by a negative band at 206 nm which is probably due to the α-helix component in the protein. The shallow peak around 220 nm is most likely due to random coil influence. Based on the calculated secondary structure (%) by the CDSSTR program Cdc25C$_{73-441}$ has 45 % secondary structure and 56 % is unordered.

The regular secondary structure composition (%) of Cdc25C$_{73-441}$ derived by CD was also compared to the two popular predictors (Table 3). Based on the CD data Cdc25C$_{73-441}$ has 29 % of regular structure. According to PROFsec 23.5 % of Cdc25C$_{73-441}$ possesses regular secondary structure and 22.1 % according to PSIPred. These results are consistent. The CD data complements the predicted data.

	CD	PROFsec	PSIPred
% Regular secondary structure (α-helix + β-sheet)	29	23.5	22.1

Table 3.5.4 Regular secondary structure composition of Cdc25C$_{73-441}$

A comparison of the experimental secondary structure composition (%) derived from CD and the predicted secondary structure composition (%) using two popular prediction tools, PROFsec and PSIPred (Rost and Sander, 1994; McGuffin et al., 2000).

a.

b.
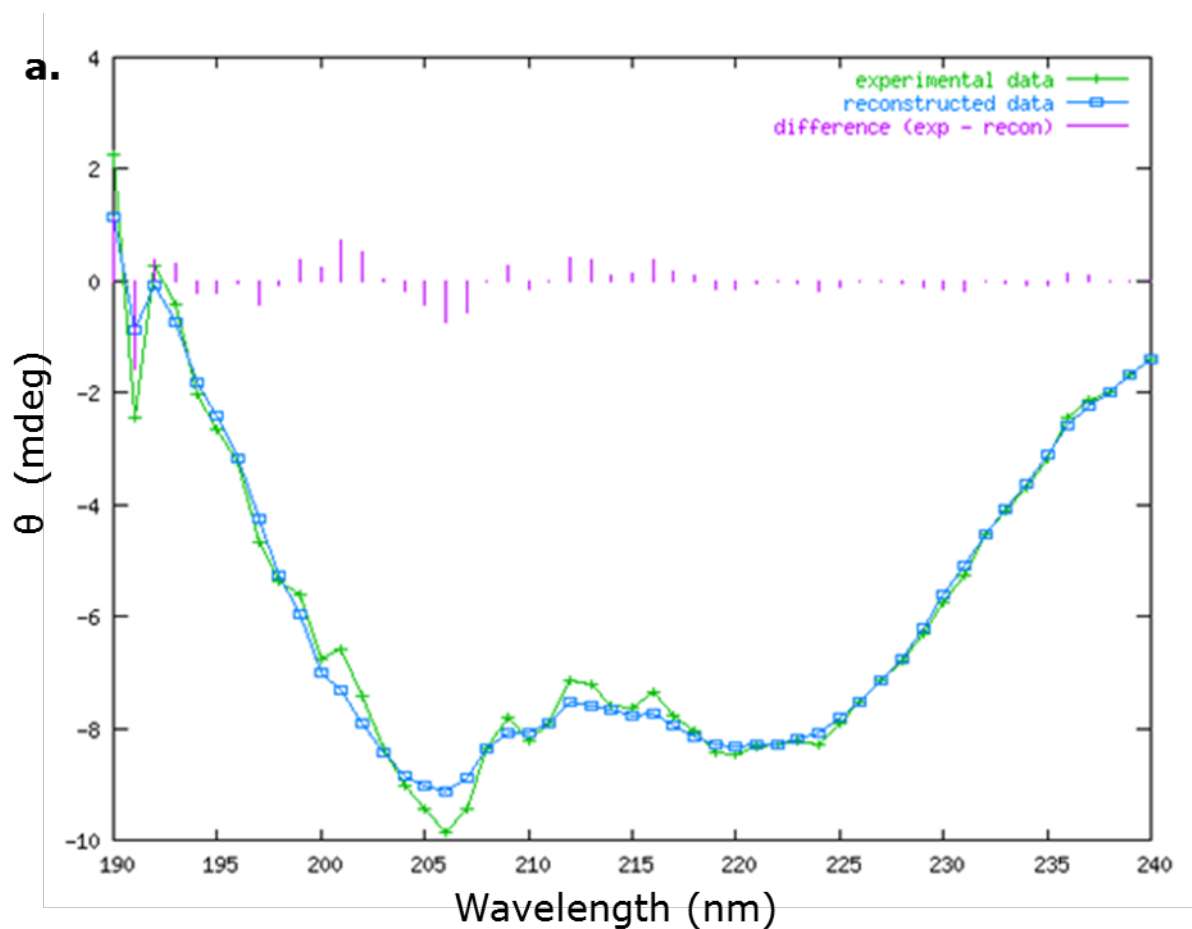
α-Helix (%)	β-strand (%)	Turns (%)	Unordered (%)
8	21	16	56

Figure 3.5.4 Far UV CD spectrum of Cdc25C$_{73-441}$

The Cdc25C$_{73-441}$ protein was subjected to circular dichroism analysis in 50 mM sodium phosphate (pH 7.5), and 0.5 mM TCEP. The Dichroweb server was used for the CD data analysis. The data is shown graphically (a) and Table (b) shows the calculated secondary structure (%) by the CDSSTR program (Sreerama and Woody, 2000). Overall, Cdc25C$_{73-441}$ contains 45 % secondary structure while 56 % is predicted to be unordered.

3.5.5 ^{1}H, ^{15}N-HSQC spectra of Cdc25C$_{73-441}$

A 2D-HSQC spectrum was acquired to assess the potential of Cdc25C$_{73-441}$ for its backbone amide resonance assignments (Figure 3.5.5). The 2D-HSQC spectrum which was collected for 2 hours showed few dispersed peaks. The majority of the peaks were overlapped in the middle of the spectrum (Figure 3.5.5a). The peaks which were dispersed were of low signal intensity. This was because of the low concentration of the protein and the presence of intense peaks from the random coil region (8.8 - 7.8 ppm) dominating the surrounding peaks. Therefore, the 2D-HSQC was run overnight to discern whether there were weaker dispersed resonances (Figure 3.5.5b). It can be seen that more peaks have appeared with moderate intensities and are better dispersed, indicating a folded domain.

The 2D-HSQC data complements the CD data suggesting that Cdc25C$_{73-441}$ is folded but contains a significant proportion of random coil structure. Cdc25C$_{73-441}$ was not suitable for backbone assignments because of the strong random coil peaks and lack of sufficiently intense resolved peaks.

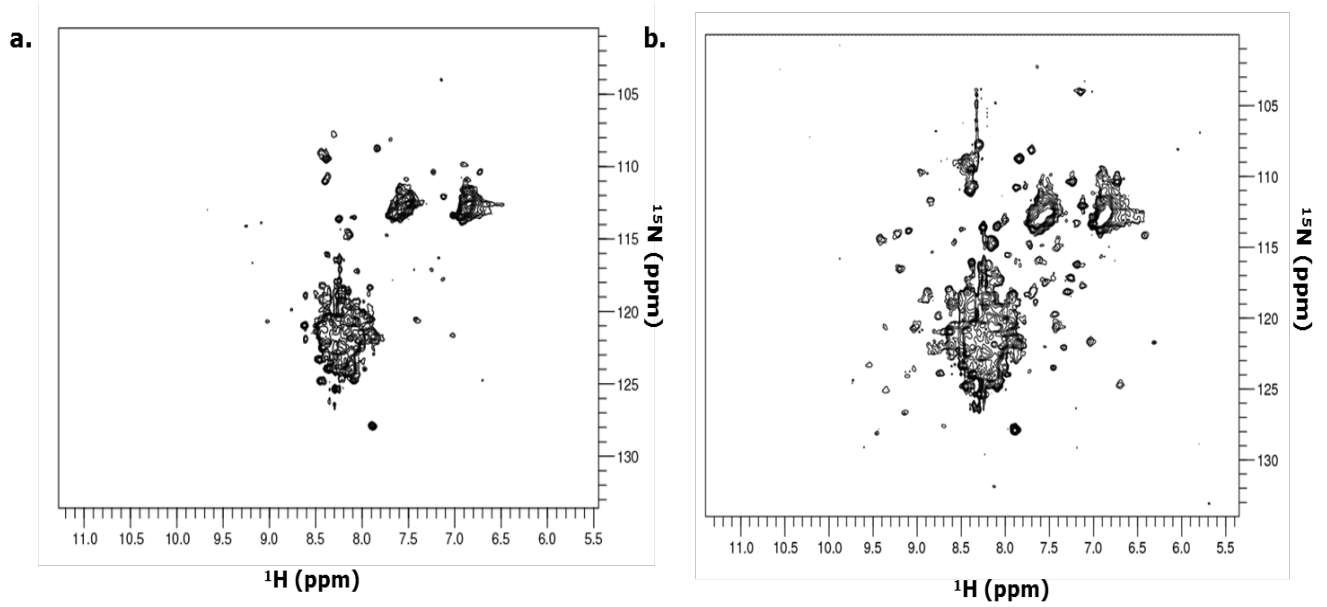**Figure 3.5.5 ^{1}H, ^{15}N-HSQC spectra of Cdc25C$_{73-441}$**
Spectra were collected of 50 µM Cdc25C$_{73-441}$ using an Agilent 600 MHz spectrometer after 2 hours (a) and overnight (b) in 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, and 1 mM TCEP.

3.5.6 SAXS of His-Cdc25C$_{73-441}$

3.5.6.1 AUC of His-Cdc25C$_{73-441}$

It is very important to have a monodisperse sample in order to reliably interpret the SAXS data. This is because the final SAXS scattering plot is an average of the protein in solution. Having different species of a protein will disrupt this average and will result in poor quality SAXS data. Here, AUC was used to assess the oligomeric state of His-Cdc25C$_{73-441}$.

A single peak in the AUC spectrum confirmed that His-Cdc25C$_{73-441}$ is monodisperse (Figure 3.5.6.1). The sedimentation coefficient obtained from the AUC analysis was 1.95 S which corresponded to a molecular weight of 41 kDa. Therefore, His-Cdc25C$_{73-441}$ was monomeric.

The predicted molecular weight based on the protein amino acid sequence is 44.7 kDa. The lower molecular weight calculated by AUC could be because the sedimentation velocity experiment does not give an exact molecular weight but provides an estimate that is influenced by molecular shape.

3.5.6.2 SAXS Parameters for His-Cdc25C$_{73-441}$

Data were collected at the protein concentrations of 2.8, 2.0, and 1.3 mg/ml (Figure 3.5.6.2.1a). Slight aggregation was noted in the Guinier regions of the higher concentrations. Therefore, a merged SAXS plot was created where the low concentration data (1.3 mg/ml) was used for the low angles and the 2.0 mg/ml data was used for the higher angles (Figure 3.5.6.2.1b). Also, the noise near the beam stop was removed to ensure the final SAXS plot used for the analysis was of optimal quality.

A straight line was fitted in the Guinier region of the merged SAXS plot and the R$_g$ was calculated using the primus program to be 4.19 nm (Figure 3.5.6.2.1c). This complements the R$_g$ calculated from the P(R) distribution which was 4.17 nm (Table 3.5.6.2).

The Kratky plot has a bell-shaped curve which corresponds to the folded component of the protein (catalytic domain) (Figure 3.5.6.2.2a). The broad nature of this bell-shaped curve followed by an increase in the intensity accounts for the disordered component in the protein. It was difficult to obtain a reliable D_{max} value from the P(R) distribution due to tailing of the P(R) curve. Without any adjustments the D_{max} was calculated to be 12.5 nm (Figure 3.5.6.2.2b) However, this is not necessarily accurate because the P(R) distribution is cut-off too early. Different D_{max} values were computed to obtain a P(R) distribution which smoothly approached zero. Following this process, D_{max} was estimated to be between 29 and 34 nm (Table 3.5.6.2). The P(R) distribution with a D_{max} of 29 nm is bumpy rather than smooth and is forced to end early which is seen by the way it curves towards higher R values (Figure 3.5.6.2.2c). The P(R) distribution with a D_{max} of 34 nm is also bumpy. However, it does not curve when it approaches zero (Figure 3.5.6.2.2d). Therefore, the D_{max} is likely to be closer to 34 nm than 29 nm.

Parameter (nm)	His-Cdc25C$_{73-441}$
R_g (Guinier)	4.19 +/- 0.066 ($qR_g < 1.3$)
R_g (P(R))	4.17
D_{max}	29 – 34

Table 3.5.6.2 SAXS parameters for His-Cdc25C$_{73-441}$

The SAXS parameters for His-Cdc25C$_{73-441}$ obtained from the SAXS analysis are tabulated (see text).
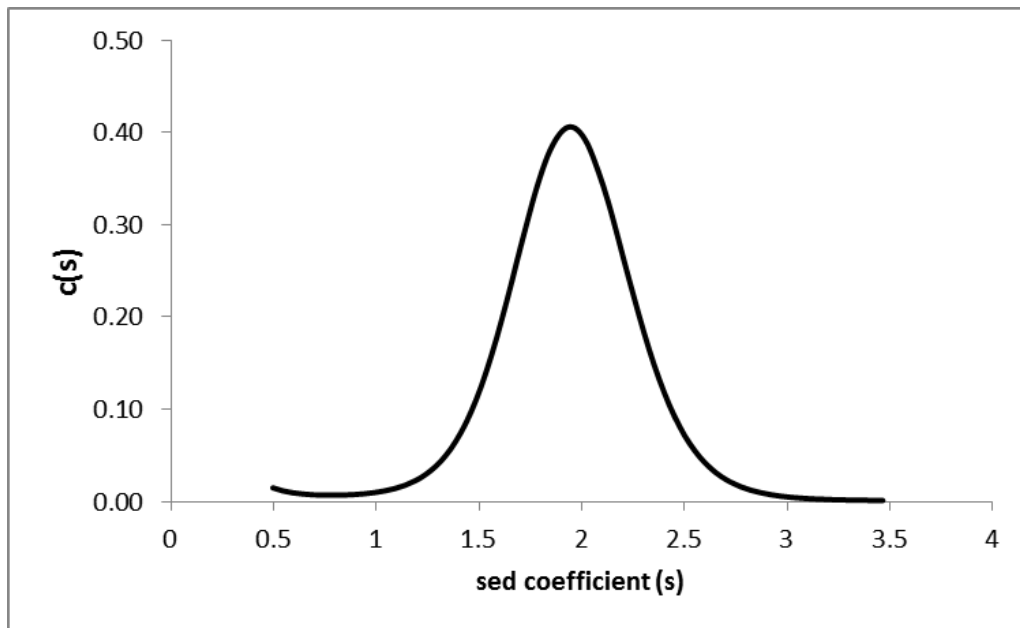
Figure 3.5.6.1 AUC of His-Cdc25C$_{73\text{-}441}$

His-Cdc25C$_{73\text{-}441}$ was subjected to AUC analysis in 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, and 1 mM TCEP. The SEDFIT (Brown and Schuck, 2006) *program was used to perform the c(s) analysis. Based on the AUC analysis His-Cdc25C$_{73\text{-}441}$ was determined to be monodisperse.*
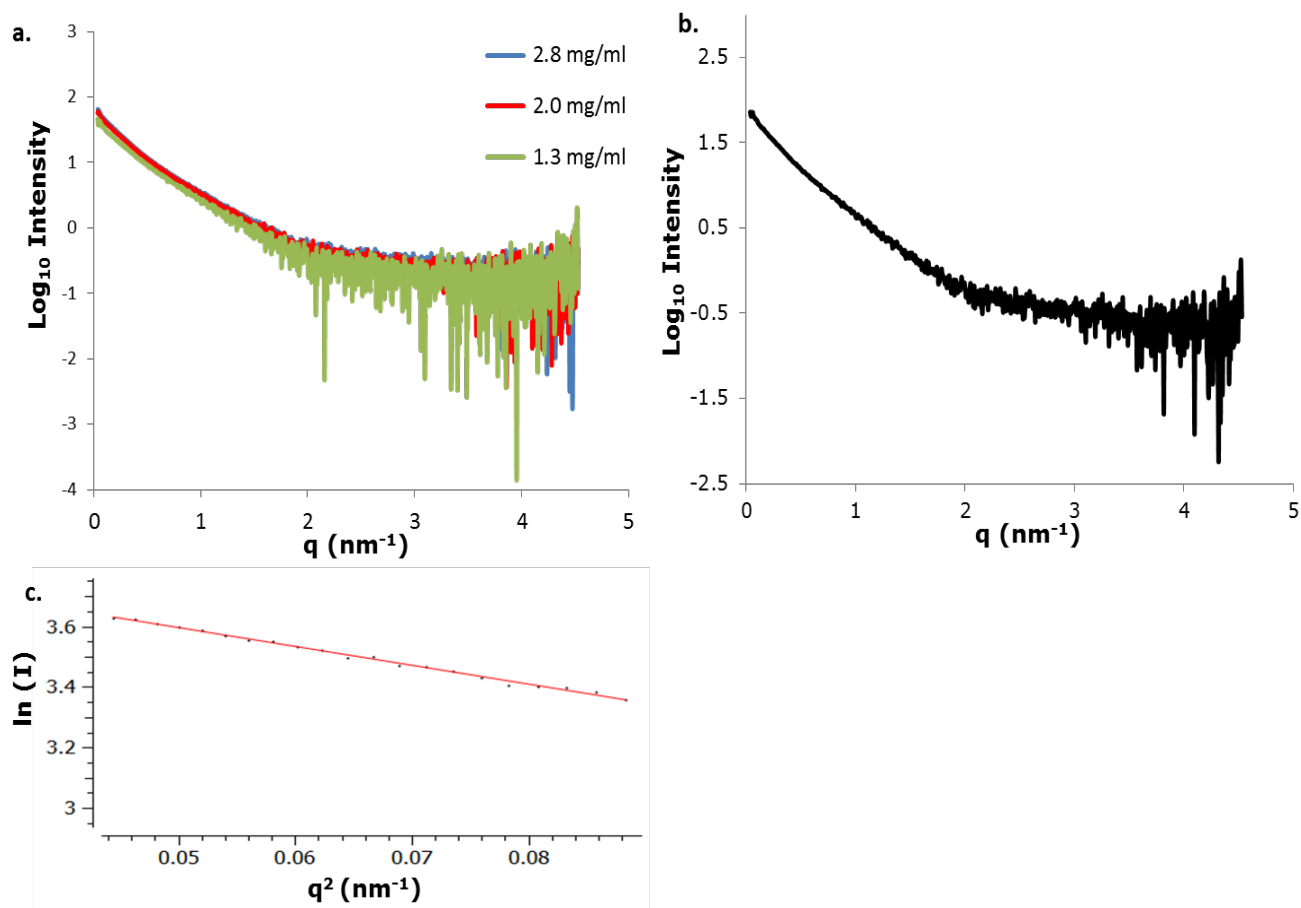
Figure 3.5.6.2.1 SAXS graphs of His-Cdc25C$_{73-441}$

SAXS data were collected at ESRF (Grenoble, France) using the BM29 beamline. Data were collected for three concentrations in the buffer 50 mM sodium phosphate (pH 7.5), 150 mM sodium chloride, and 1 mM TCEP. The three concentrations are overlaid and plotted with the logarithmic intensity against q (momentum transfer) where q is expressed in nm^{-1} (a). The merged plot is shown (b) and the Guinier plot which was determined from the merged plot is shown in (c).
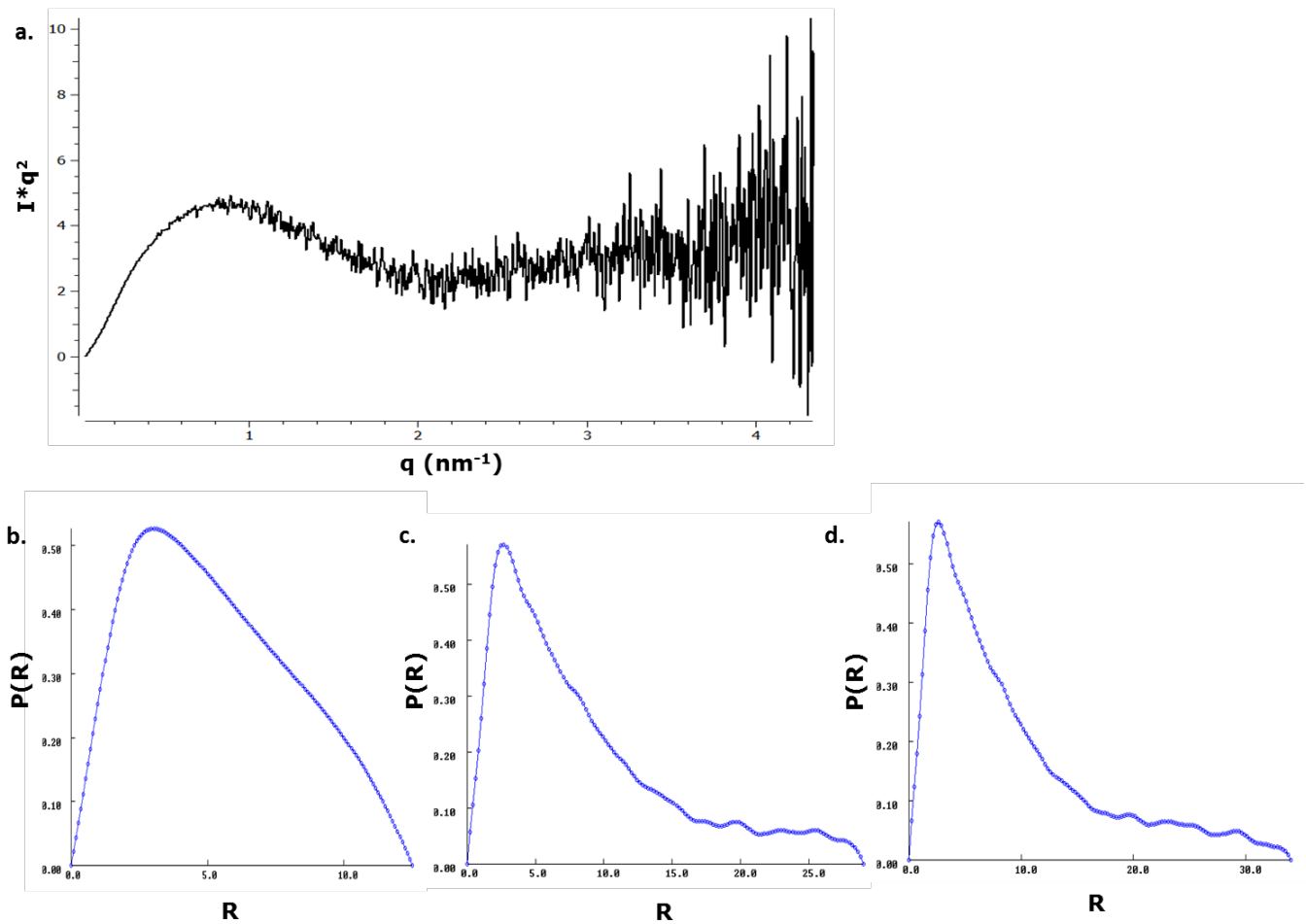
Figure 3.5.6.2.2 Kratky plot and P(R) graphs of His-Cdc25C$_{73-441}$

The merged plot was used to obtain the Kratky plot and the P(R) distributions. The Kratky plot plotted with intensity multiplied by q^2 ($I*q^2$) against q. The P(R) distributions are plotted P(R) against R where R is expressed in nm (D_{max} (nm) = 12.5 (b), 29 (c), and 34 (d)).

3.5.6.3 Selecting the best I-TASSER model

The tailing of the P(R) distribution suggested flexibility in the protein. Since earlier data described in this chapter (Figures 3.4.2 & 3.4.3.1) have shown the N-terminal domain to be predominantly unfolded it was thought the flexibility could be attributed to this region.

In order to assess flexibility a 3D protein model of His-Cdc25C$_{73-441}$ was generated. Since there is no PDB structure of the full-length protein, I-TASSER (Zhang, 2008) was used to generate a model. The I-TASSER online server uses the protein amino acid sequence input and creates 3D protein models based on a combination of *ab initio* modelling and information from the PDB database.

CRYSOL (Svergun et al., 1995) was then used to generate theoretical SAXS scattering plots of the 3D protein models that gave the best I-TASSER scores. The theoretical SAXS scattering plot for each model was then fitted to the SAXS data to select a model which best fit the data (Figure 3.5.6.3). Model 3 fitted the SAXS data best with an χ^2 value of 2.35 compared to the other two models which had χ^2 values of 6.88 (model 1) and 2.48 (model 2).

3.5.6.4 Assessing flexibility

The ensemble optimisation method (EOM) and CRYSOL were used to characterize this flexibility.

CRYSOL calculates a theoretical SAXS scattering plot for a 3D atomic model, fits this to the SAXS data, and provides a χ^2 value for the quality of the fit.

EOM is useful for proteins that have disordered flexible regions (Bernadó et al., 2007). EOM produces a number of models to account for the flexibility from which an ensemble is selected that is most representative of the data and an averaged SAXS scattering plot is produced. This

scattering plot is then fitted to the SAXS data and a quantitative value for the quality of the fit is given.

When the N-terminal domain (Cdc25C$_{73-279}$) was made rigid the theoretical SAXS scattering curve did not fit very well to the SAXS data compared to when this region was made flexible (Figure 3.5.6.4.1). Making this region flexible significantly improved the fit throughout with an χ^{2} of 0.87 compared to 2.35. The 20 models generated by EOM from which an averaged scattering plot was produced were overlaid using Chimera (Figure 3.5.6.4.2). It can be seen that the flexible region has many possible orientations. Overall, this suggests that at least part of Cdc25C$_{73-279}$ is very flexible.

a.

b.
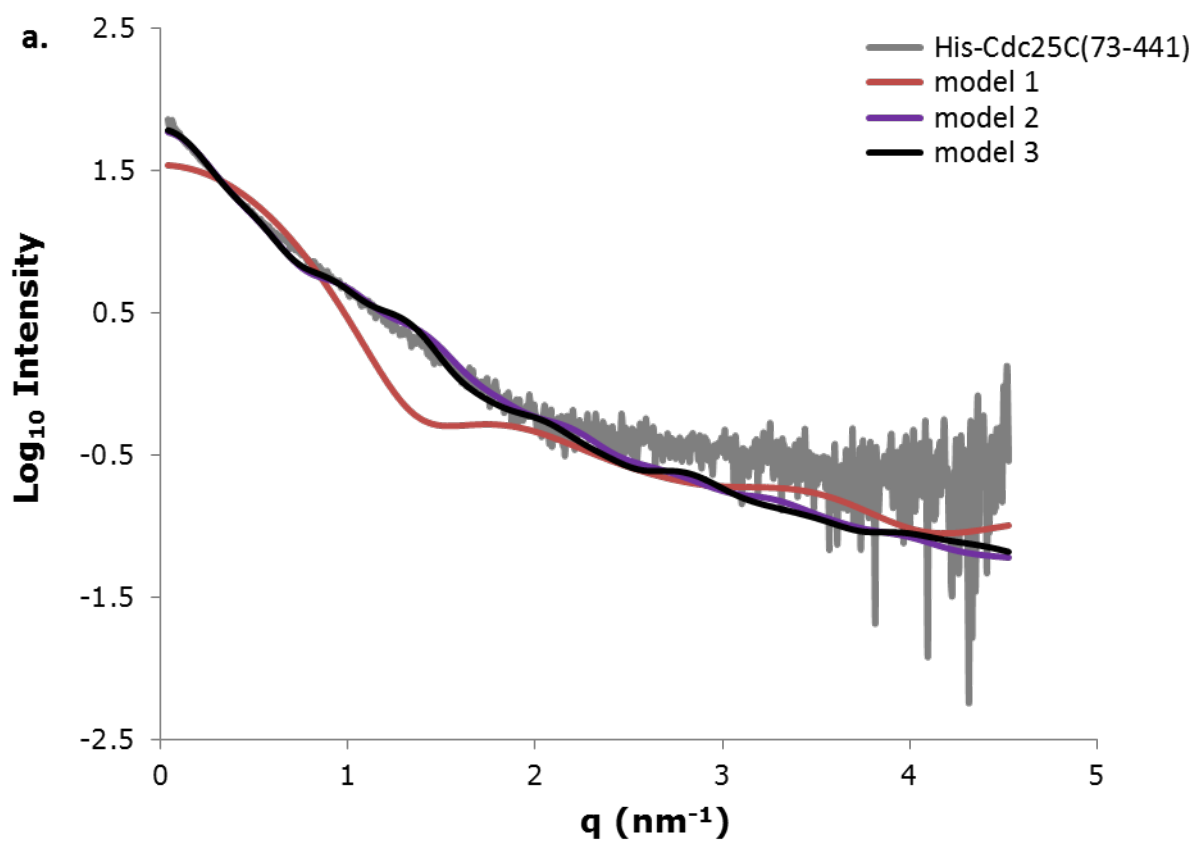
Model	χ^2
1	6.88
2	2.48
3	2.35

Figure 3.5.6.3 I-TASSER models

CRYSOL was used to generate theoretical SAXS scattering plots of the I-TASSER models which were fitted to the SAXS data (a). CRYSOL also calculated χ^2 values for the fits which are tabulated in (b).
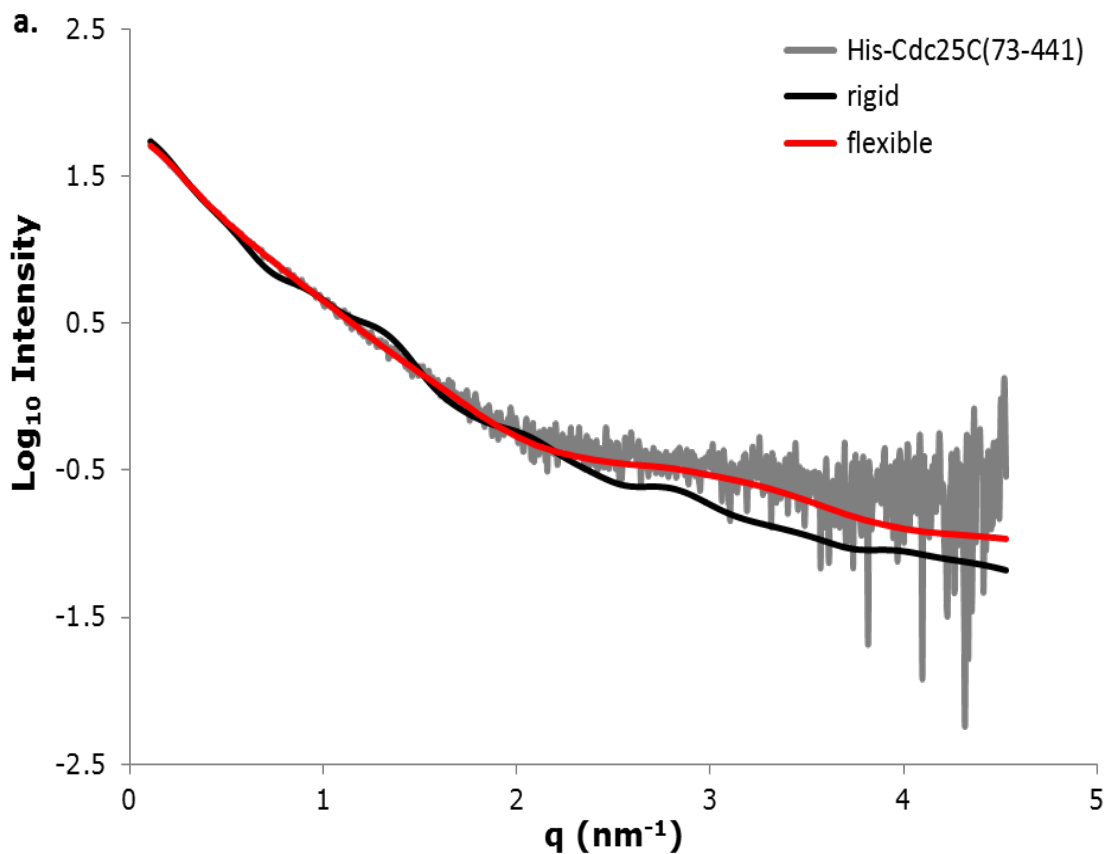
Figure 3.5.6.4.1 His-Cdc25C$_{73-279}$ is very flexible

I-TASSER model 3 was used to model His-Cdc25C$_{73-279}$ as a rigid or flexible protein. CRYSOL was used to obtain the fitted SAXS curve when this region was rigid and EOM was used to obtain the fitted SAXS curve when this region was flexible (a). The χ^2 values for the fits are tabulated in (b). The data are plotted with intensity (Log$_{10}$) against the momentum transfer (q). The scattering curve fits better to the data when His-Cdc25C$_{73-279}$ is flexible.
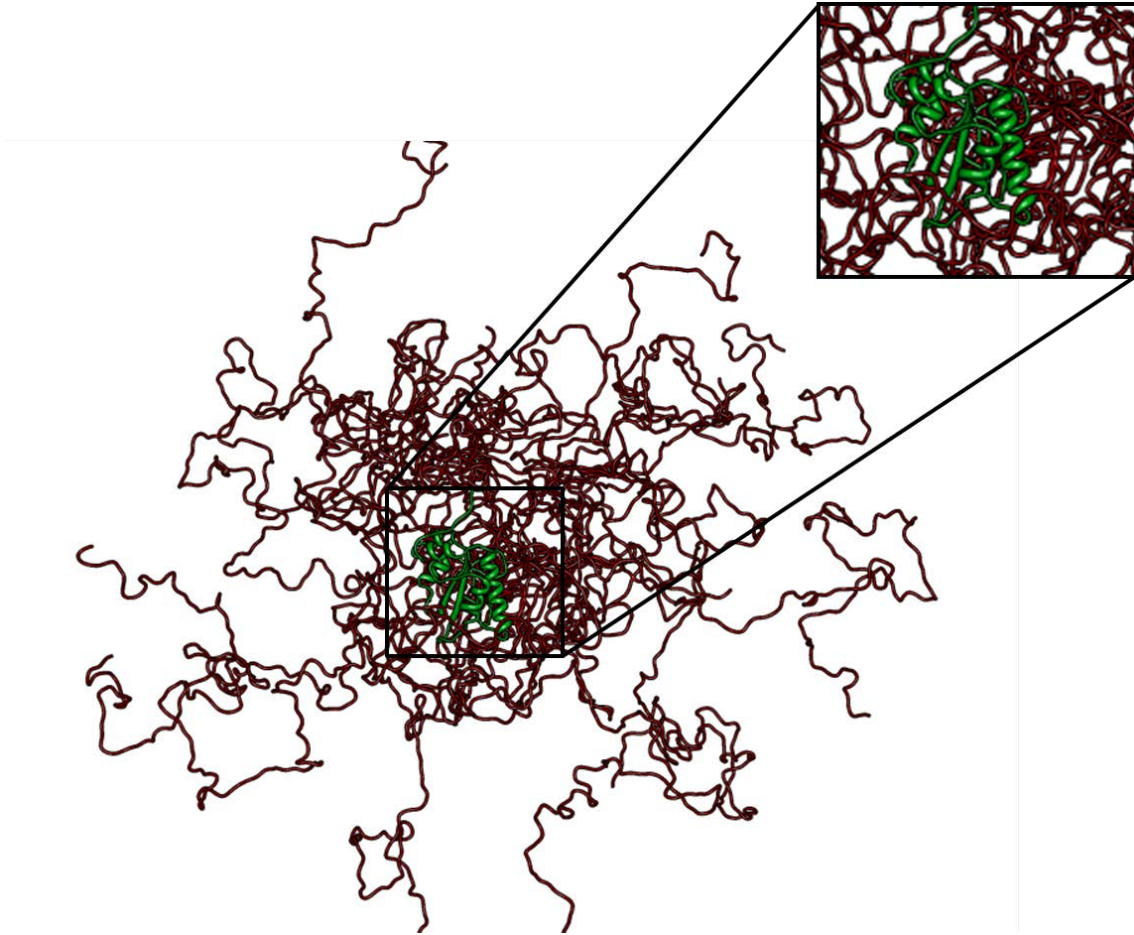
Figure 3.5.6.4.2 Overlay of the EOM ensemble

Chimera (Pettersen et al., 2004) was used to overlay the 20 models (ensemble) produced by EOM when Cdc25C$_{73-279}$ is flexible. The His-Cdc25C$_{73-279}$ region is shown in dark red and the Cdc25C catalytic domain (Cdc25C$_{280-441}$) is in green.

3.5.7 The 198 amino acids of the N-terminal domain affect the thermal stability of the full-length protein

Having established that the N-terminal domain was disordered and highly flexible its effect on the thermal stability of the full-length protein was investigated. Therefore, both the catalytic domain (Cdc25C$_{270-443}$) and Cdc25C$_{73-441,}$ which represents 78 % of the full-length protein, were tested. These proteins were prepared in 50 mM sodium phosphate (pH 7.5), 0.5 mM TCEP and were subjected to ThermoFluor® analysis. Each of the buffers tested had 150 mM sodium chloride.

The melting temperature range for Cdc25C$_{73-441}$ was 42.5 to 47.4 ^{o}C. Tris pH 8 (T$_m$ = 47.4 ^{o}C) and Hepes pH 7.5 (T$_m$ = 46.3 ^{o}C) gave the highest T$_m$ values compared to the other buffers tested. Bis-Tris pH 6.0 gave the lowest thermal stability (T$_m$ = 42.5 ^{o}C).

When the T$_m$ of Cdc25C$_{73-441}$ in the tested buffers was compared to that of the catalytic domain it was noted that overall the catalytic domain was 3 ^{o}C more thermally stable. The 198 amino acids of the N-terminal domain (Cdc25C$_{73-269}$) resulted in a consistent thermal destabilization of Cdc25C$_{73-441}$ suggesting the presence of the N-terminal domain reduces the stability of the full-length protein.

a.

Buffer	T_m (°C), Cdc25C$_{73-441}$	T_m (°C), Cdc25C$_{270-443}$ (Catalytic Domain)	*ΔT_m (°C)
Bis-Tris pH 6.0, 150 mM NaCl	42.5	46.2	-3.7
Sodium Cacodylate pH 6.5, 150 mM NaCl	43.8	47.3	-3.5
MES pH 6.5, 150 mM NaCl	44.0	47.2	-3.2
Sodium Phosphate pH 7.0, 150 mM NaCl	45.7	49.2	-3.5
MOPS pH 7.0, 150 mM NaCl	45.3	48.7	-3.4
HEPES pH 7.5, 150 mM NaCl	46.3	49.0	-2.7
Tris pH 8.0, 150 mM NaCl	47.4	50.0	-2.6

b.

Legend:
- Bis-Tris pH 6.0, 150 mM NaCl
- Sodium Cacodylate pH 6.5, 150 mM NaCl
- MES pH 6.5, 150 mM NaCl
- Sodium Phosphate pH 7.0, 150 mM NaCl
- MOPS pH 7.0, 150 mM NaCl
- HEPES pH 7.5, 150 mM NaCl
- Tris pH 8.0, 150 mM NaCl
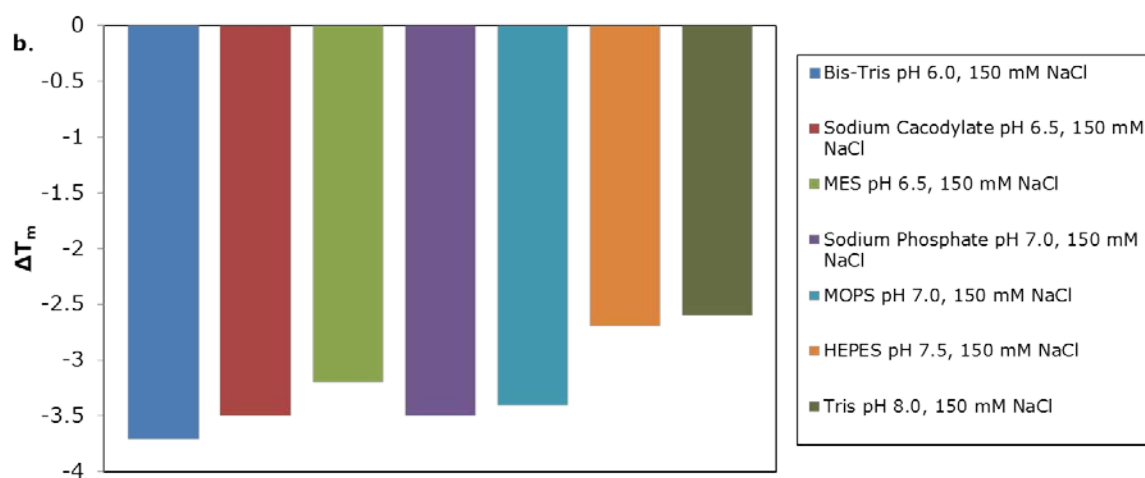
3.5.7 Comparing thermal stability of Cdc25C$_{73-441}$ and Cdc25C$_{270-443}$

The protein stability was tested 3x in each buffer and an average T_m was obtained. The T_m of Cdc25C$_{73-441}$ and Cdc25C$_{270-443}$ for each buffer condition and ΔT_m (ΔT_m = T_m of Cdc25C$_{73-441}$ − T_m of Cdc25C$_{270-443}$) are tabulated (a). ΔT_m is also plotted as a bar graph (b). The addition of 198 amino acids (from the N-terminal domain) of Cdc25C results in a consistent destabilization of Cdc25C$_{73-441}$.*

3.6 Discussion of Chapter 3.0

The data in this chapter indicates the N-terminal domain (regulatory) of the Cdc25C protein is significantly disordered and possesses limited regular secondary structure.

It was difficult to purify pure full-length Cdc25C. The AUC data suggested the purified GST-tagged Cdc25C full-length protein sample was not monodisperse. As well as containing a range of low molecular weight species it also had a larger species present. This indicates break down products and/or low molecular weight contaminants. A larger molecular weight species indicates the presence of an aggregate. Interestingly, a recent paper has also highlighted the difficulties of purifying full-length Cdc25C (Franckhauser et al., 2013). The untagged full-length protein purified from *E. coli* is insoluble. Purification from other systems such as yeast or insect cells is difficult because of the tight interaction of Cdc25C with 14.3.3 resulting in the co-purification with 14.3.3 and other proteins. Therefore, Cdc25C full-length is generally purified in *E. coli* tagged with GST. However, the data here and the recently published study indicate the GST-tagged Cdc25C full-length protein may not be appropriate for activity and cell based assays.

Here, a folded, monodisperse, His$_6$ tagged Cdc25C protein was purified successfully which represents ~ 78 % of the full-length. AUC also indicated it was monomeric. The same study described above also suggests the Cdc25C full-length is a tetramer. It may be that the oligomerization motif lies in the other 22 % of the Cdc25C full-length.

Although, a SAXS envelope of the full-length protein described in this chapter could not be resolved since a majority of the N-terminal domain is disordered, SAXS provided insight into the flexibility of Cdc25C.

The EOM data suggested the N-terminal domain is very flexible. This flexibility may be important for its function. It would be interesting to investigate the effect of phosphorylating known phosphorylation sites on the N-terminal domain and assessing the effect on flexibility and structure.

Many proteins involved in cell signalling and cancer have been described possessing a significant amount of intrinsic disorder. It has been suggested intrinsic disorder is important for the functional role of the protein. For example, histone tails are disordered (Jenuwein and Allis, 2001; Berger, 2002). These tails can be readily modified by post translational processes such as phosphorylation, acetylation, and methylation affecting the accessibility of DNA. The N-terminal domain of Cdc25C possesses a number of phosphorylation sites and is involved in many regulatory processes such as affecting catalytic activity, localisation and degradation. Intrinsic disorder in this region may allow the Cdc25C protein flexibility in associating with many partners and efficiently responding to the changing dynamics of the cell.

Some proteins can become structured upon binding to their partner in a process termed "coupled folding and binding". When purified with the catalytic domain the N-terminal domain did not become structured. This could be because the first 72 residues of the N-terminal domain and/or the last 32 residues of the catalytic domain are required for the N-terminal domain to become structured. It could also be that the N-terminal is not structured in the full-length protein and becomes structured upon binding with a small molecule ligand or a protein partner. For example the N-terminal domain of the protein DFF45 (DNA fragmentation factor 45) becomes structured when forming a dimer with the protein DFF40 (Zhou et al., 2001). Another example is the interaction of Cdk2 with its intrinsically disordered inhibitor p21 resulting in an ordered conformation of p21 (Kriwacki et al., 1996).

Chapter 4.0 - Purification, solubility optimization, and structural characterisation of the Cdc25C catalytic domain

4.1 Purification of His-Cdc25C$_{270-462}$

The constructs in this chapter have been cloned into the pNic28-Bsa4 vector. Each construct has a His tag, TEV cleavage site followed by the Cdc25C amino acid sequence.

Unlabeled His-Cdc25C$_{270-462}$ was expressed in BL21 (DE3) cells and purified by Ni-NTA affinity chromatography (Figure 4.1.1) followed by size exclusion chromatography using a Superdex 75 column (Figure 4.1.2). The third peak from the size exclusion column, which is the most prominent peak eluting between 155-185 ml, contained the protein of interest (Figure 4.1.2a). The two other smaller peaks are the protein aggregate and high order oligomer peaks respectively.

A high level of purity (> 95 %) was achieved for His-Cdc25C$_{270-462}$ (Figure 4.1.2b). The purest fractions eluted in the third peak (Figure 4.1.2a) were then concentrated. During concentration it was noted that the protein was exhibiting signs of precipitation. At this point concentration was halted. The protein concentration was ~ 500 µM. The protein sample was centrifuged and the precipitate was discarded. The soluble protein supernatant was stored at 4 °C prior to ThermoFluor® experiments.

His-Cdc25C$_{270-462}$ was dialyzed into 50 mM Hepes, pH 7.4, 50 mM NaCl, and 0.5 mM TCEP. The NaCl concentration was lowered to 50 mM because having a high salt concentration in the protein stock could influence the results. His-Cdc25C$_{270-462}$ was subjected to an in-house NMR buffer screen (Overduin lab) to identify the optimal buffer and additive conditions.
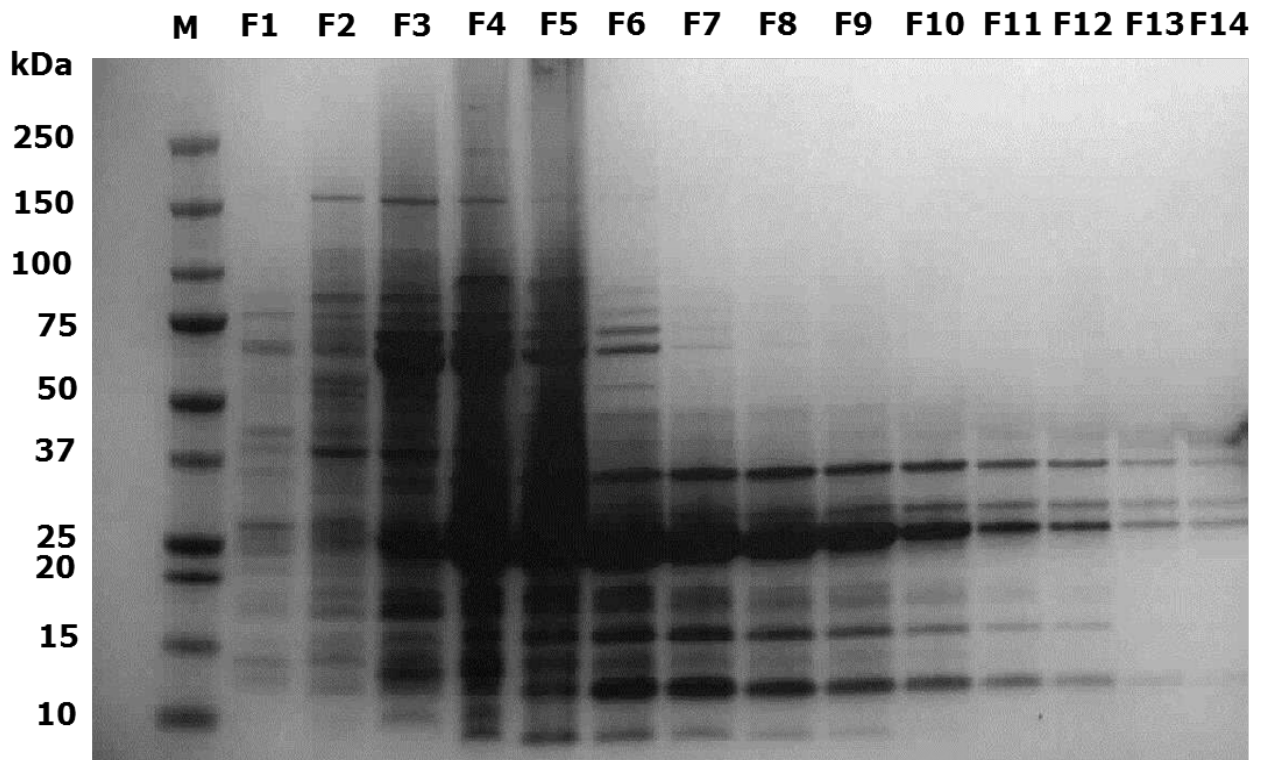
Figure 4.1.1 Histrap purification of His-Cdc25C$_{270-462}$

The first lane is the precision plus protein ^{TM} standard marker (Bio-Rad). This is followed by the 1 ml eluted fractions 1-14. The eluted fractions were first tested for the presence of protein using Bradford reagent (Bio-Rad). Fractions containing protein were then subjected to SDS-PAGE which was run with a 4-12 % Bis-Tris criterion XT precast gel.
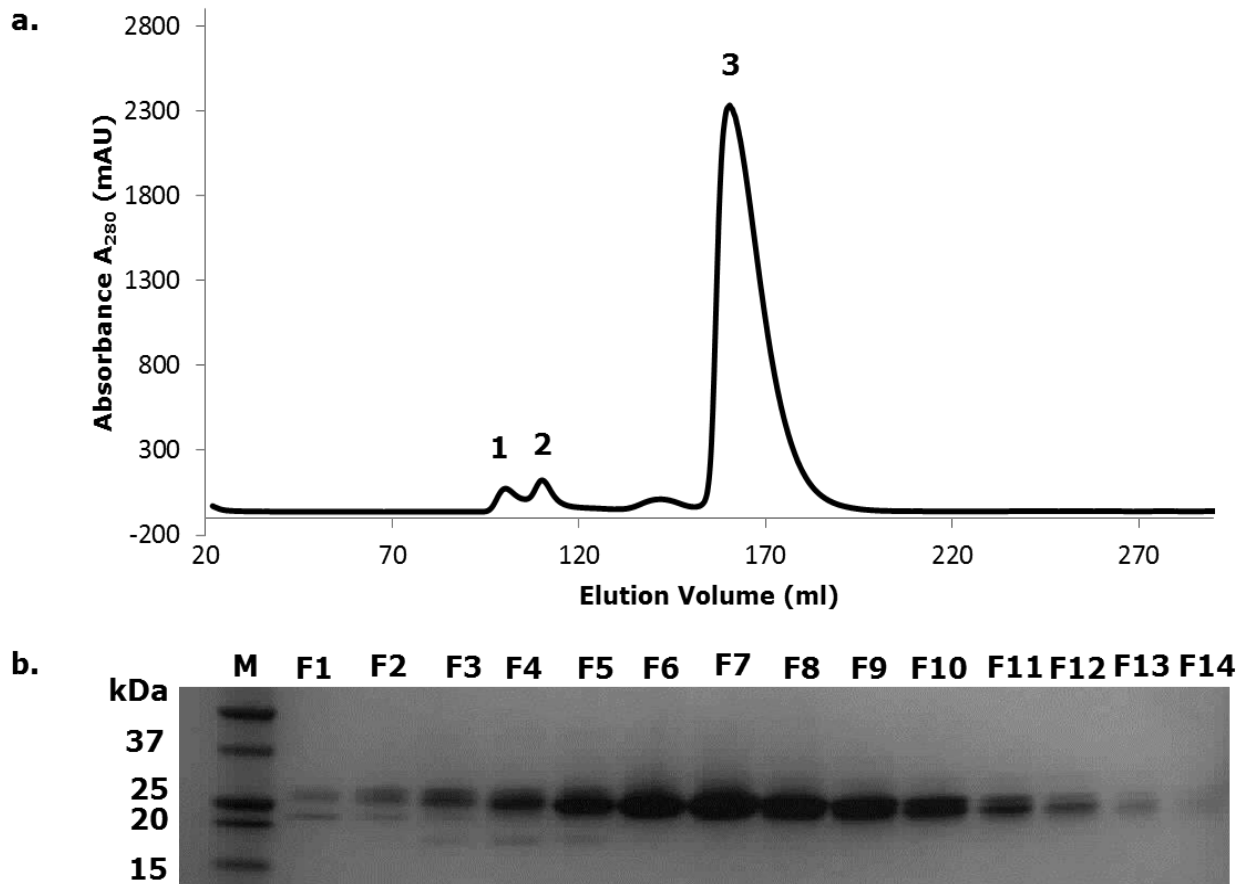
Figure 4.1.2 Purification of His-Cdc25C$_{270\text{-}462}$ using Superdex 75 column

Size exclusion profiles at UV A$_{280}$ of His-Cdc25C$_{270\text{-}462}$ where three peaks can be seen (a). The protein was eluted in 50 mM Hepes pH 7.4, 300 mM NaCl, and 0.5 mM TCEP. The first two peaks are the aggregate and high-order structures respectively. The third peak is the largest peak which contains the pure protein. (b) An InstantBlue stained SDS-PAGE which shows marker (M) and eluted fractions (F1-F14) from the third peak.

4.2 Buffer Optimization of His-Cdc25C$_{270-462}$

His-Cdc25C$_{270-462}$ gave melting temperatures ranging from 45.0 to 49.2 °C in the ThermoFluor®

assay of buffers (Figure 4.2). Each 50 mM buffer condition tested also contained 150 mM NaCl.

The average T$_m$ values for each buffer condition were compared with the T$_m$ of 150 mM NaCl

alone to obtain the ΔT$_m$ for each buffer condition.

Looking at the ΔT$_m$ values it can be seen that Hepes (pH 7.5) does not have any effect on the

thermal stability of His-Cdc25C$_{270-462}$. Sodium Cacodylate (pH 6.5), MES (pH 6.5), Bis-Tris (pH

6.0), and MOPS (pH 7.0) all had negative ΔT$_m$ where apart from MOPS (pH 7.0), ΔT$_m$ was large >

-3 °C. This large negative ΔT$_m$ indicates destabilization of the protein in those conditions which

is likely due to the buffer pH being close to the theoretical pI of the protein which is 6.10 as

estimated by the Protparam program (Expasy). The buffers which had positive ΔT$_m$ with respect

to 150 mM NaCl were Tris (pH 8) (ΔT$_m$ = + 0.1 °C) and Sodium phosphate (pH 7.0) (ΔT$_m$ = + 0.5

°C). Sodium phosphate (pH 7.0) was taken to be the optimal buffer.
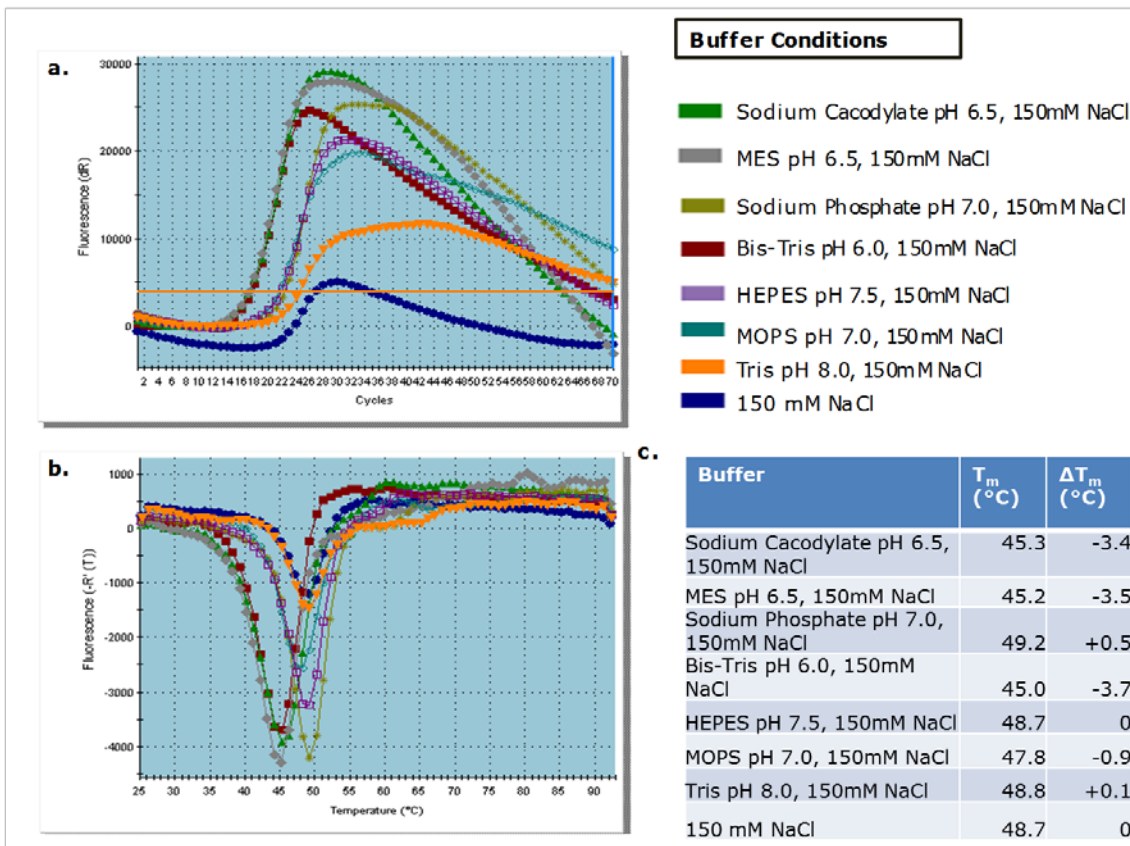
Figure 4.2 Thermofluor buffer profile for His-Cdc25C$_{270\text{-}462}$

Base-line corrected fluorescence (dR) plot which shows the fluorescence intensities at each thermal cycle (a). The inverse of the first derivative of the fluorescence as a function of temperature (-R' (T)) is plotted in (b) from which the T$_m$ (melting temperature) is derived. T$_m$ is the point of inflection of the first curve which corresponds to a minimum in the negative first derivative plot. The melting temperatures for each 50 mM buffer condition are tabulated along with ΔT$_m$ (c). 50 mM sodium phosphate (pH 7.0), 150 mM NaCl buffer gave the maximum T$_m$.

4.3 NaCl optimization of His-Cdc25C$_{270-462}$

Following, the identification of the optimal buffer and additive conditions, sodium chloride optimization was carried out. The reason for this was to see if a change in the salt concentration had an effect on the thermal stability of His-Cdc25C$_{270-462}$.

When conducting NMR experiments it is desirable to have as little salt as possible since having a high concentration of salt can remove the benefit of using a cryoprobe. However, having too little salt can result in the destabilization of a protein and lowered solubility. Therefore, it is necessary to optimize the salt concentration of a chosen buffer.

The NaCl concentration was tested up to 0.5 M in 50 mM sodium phosphate (pH 7.0) and 0.5 mM TCEP (Figure 4.3). The average T_m values obtained from three independent experiments were compared with the control containing no NaCl.

The NaCl concentrations from 0 – 75 mM did not result in any change in ΔT_m compared to 0 mM NaCl. There was an increase in ΔT_m of 1 °C when the NaCl concentration was increased to 100 mM. There was no further increase in ΔT_m from 100 – 250 mM NaCl. Further increases in ΔT_m were noted with higher NaCl concentrations where the maximum concentration tested, 500 mM NaCl gave the maximum ΔT_m of +3.1 °C. Therefore, increasing NaCl concentrations improved the thermal stability of His-Cdc25C$_{270-462}$.

The concentration of 100 mM was chosen as optimum because there was no change in ΔT_m from 100 – 250 mM NaCl. Although, 300 mM NaCl did provide a further 0.5°C increase in T_m compared to 100 mM NaCl it was not chosen as the ideal concentration because this concentration was deemed to be high for NMR experiments, especially if backbone assignments were to be attempted. The next step was to acquire a ^{1}H, ^{15}N – HSQC spectrum to see if the protein was assignable.
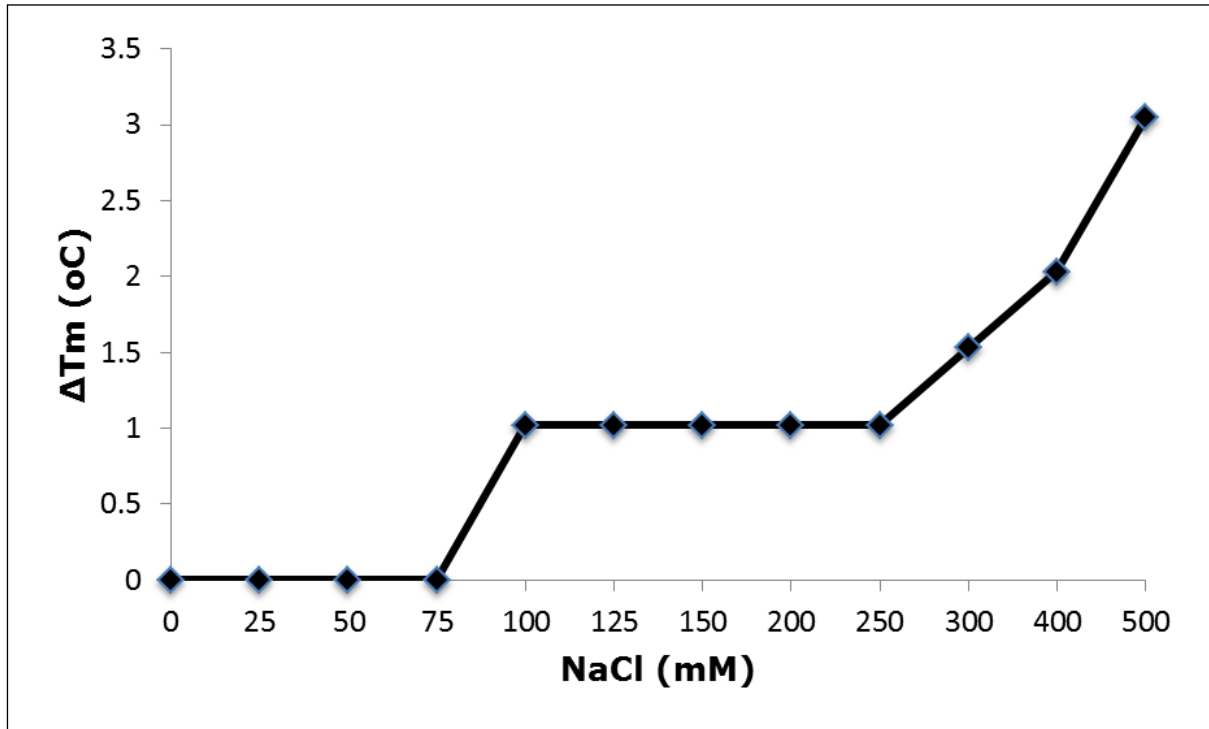
Figure 4.3 NaCl optimization of His-Cdc25C$_{270-462}$

Sodium chloride concentrations from 0 – 500 mM were tested in 50 mM sodium phosphate (pH 7.0) and 0.5 mM TCEP.

4.4 ^{1}H, ^{15}N – HSQC spectra of His-Cdc25C$_{270-462}$

Protein samples which were ^{15}N-labeled were prepared for HSQC experiments. During protein concentration, precipitation was again noted which confirmed that the protein had limited solubility. Despite this, protein samples ~ 500 µM were obtained.

A ^{1}H, ^{15}N – HSQC spectrum of His-Cdc25C$_{270-462}$ was acquired in 50 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 0.5 mM TCEP (Figure 4.4.1). The peaks are well dispersed and the protein is folded.

The expected peaks formula (see below) was used to calculate the number of expected peaks. From a total of 245 expected peaks, 213 peaks were observed ~ 87 % of expected peaks seen (total number of amino acids = 216, Pro (P) = 11, Asn (N) = 6, Trp (W) = 0, and Gln (Q) = 14).

Expected peaks formula = (total number of amino acids in the protein sequence – proline(s) + tryptophan(s)) + 2(Asparagine(s) + Glutamine(s))

^{1}H, ^{15}N-HSQC spectra of His-Cdc25C$_{270-462}$ in two different pH conditions, pH 7.0 and pH 6.5 were collected (Figure 4.4.2). This was done to see if lowering the pH improves the peak count. For NMR experiments, it is desirable to use a low pH, this is because at a higher pH the amide deprotonates more quickly. This means it will only be in the protonated state for a short time meaning that the peak from those resonances will be of low intensity. A compromise is needed between having a pH that gives good NMR peak intensity and one that is physiologically relevant and at which the protein is stable and soluble.

Lowering the pH by 0.5 unit resulted in only three extra peaks being seen (216 peaks observed) ~ 88 % of the expected peaks (Figure 4.4.2).

It was noted that subsequently after the ^{1}H, ^{15}N-HSQC experiments a white precipitate was visible in the NMR tubes. The majority of the protein had precipitated. Therefore, it was decided to stay at pH 7.0 rather than pH 6.5. This was done because the theoretical pI of the protein is 6.10. Having a buffer pH very close to the pI would increase the chances of precipitation.
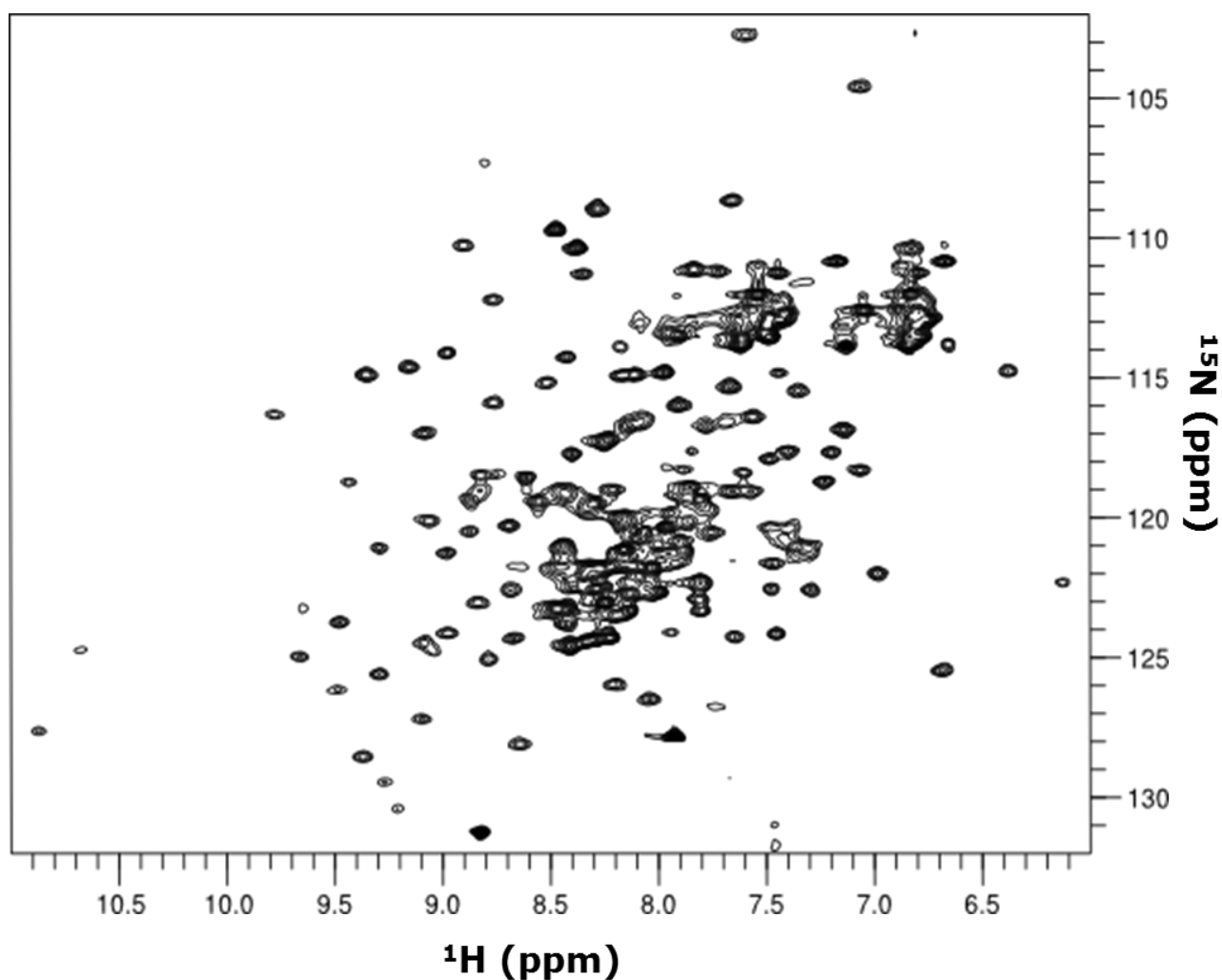
Figure 4.4.1: ^{1}H, ^{15}N – HSQC of His-Cdc25C$_{270\text{-}462}$

^{1}H, ^{15}N-HSQC spectrum of 500 µM His-Cdc25C$_{270\text{-}462}$ in 50 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 0.5 mM TCEP collected at 25 ^{o}C using a 600 MHz Agilent Direct Drive spectrometer.
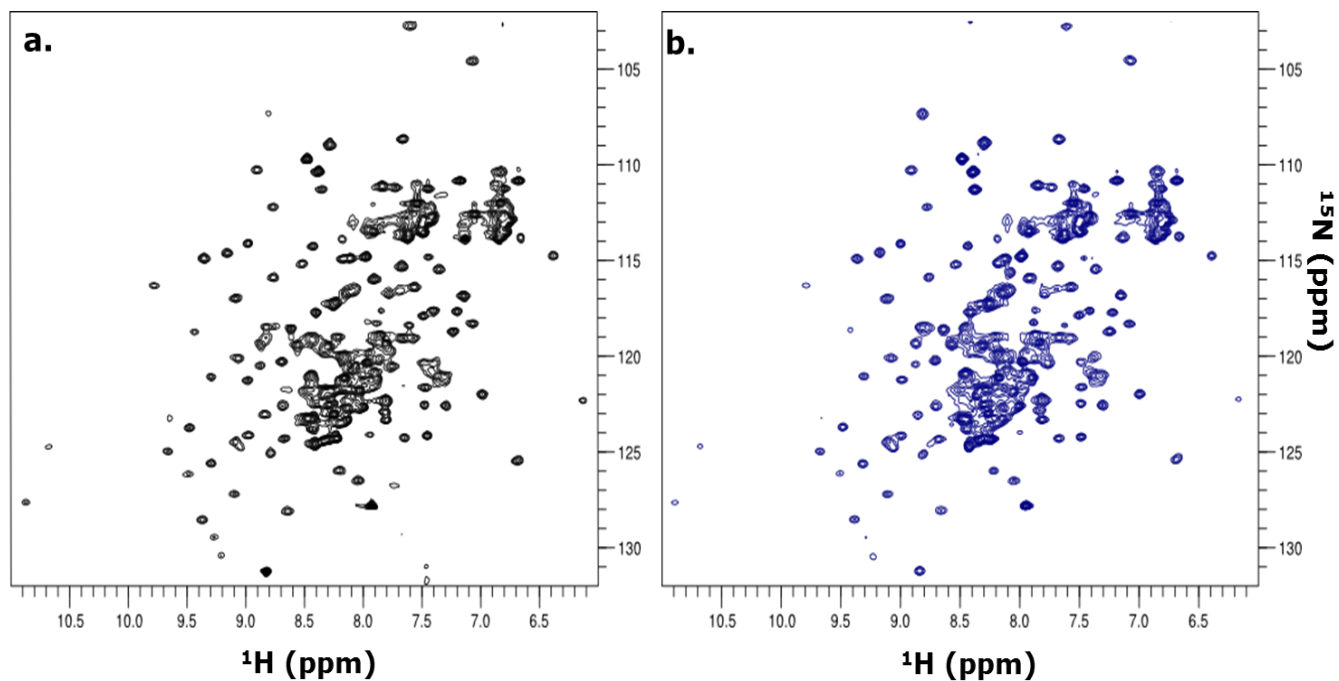
Figure 4.4.2: ^{1}H, ^{15}N – HSQC spectra of His-Cdc25C$_{270\text{-}462}$ in pH 7.0 and pH 6.5

^{1}H, ^{15}N-HSQC spectra of 500 µM His-Cdc25C$_{270\text{-}462}$ in 50 mM sodium phosphate (pH 7.0), 100 mM NaCl, 0.5 mM TCEP (a) and 50 mM sodium phosphate (pH 6.5), 100 mM NaCl, 0.5 mM TCEP (b). Both spectra were collected at 25 ^{o}C using a 600 MHz Agilent Direct Drive spectrometer.

4.5 Additive optimisation

4.5.1 Additive Optimization (ThermoFluor®) of His-Cdc25C$_{270\text{-}462}$

The next step was to improve the solubility of His-Cdc25C$_{270\text{-}462}$. Additives were tested by ThermoFluor®. The idea being that an improvement in thermal stability may improve the solubility of the protein.

Additives were tested to identify those which would help improve the thermal stability of the protein and may also improve the ^{1}H-^{15}N-HSQC spectrum (Figure 4.5.1).

The addition of 0.5 M TMAO or 50 mM Glu/Arg in the ThermoFluor® optimised buffer, 50 mM sodium phosphate (pH 7.0), and 150 mM NaCl, improved thermal stability by 1 °C. The addition of 2 mM CHAPS resulted in ΔT_m of -3.9 °C (Figure 4.5.1b). This can be clearly seen by the shift of the melting curve to the left (Figure 4.5.1a). This considerable decrease in thermal stability suggests that CHAPS destabilizes the protein.
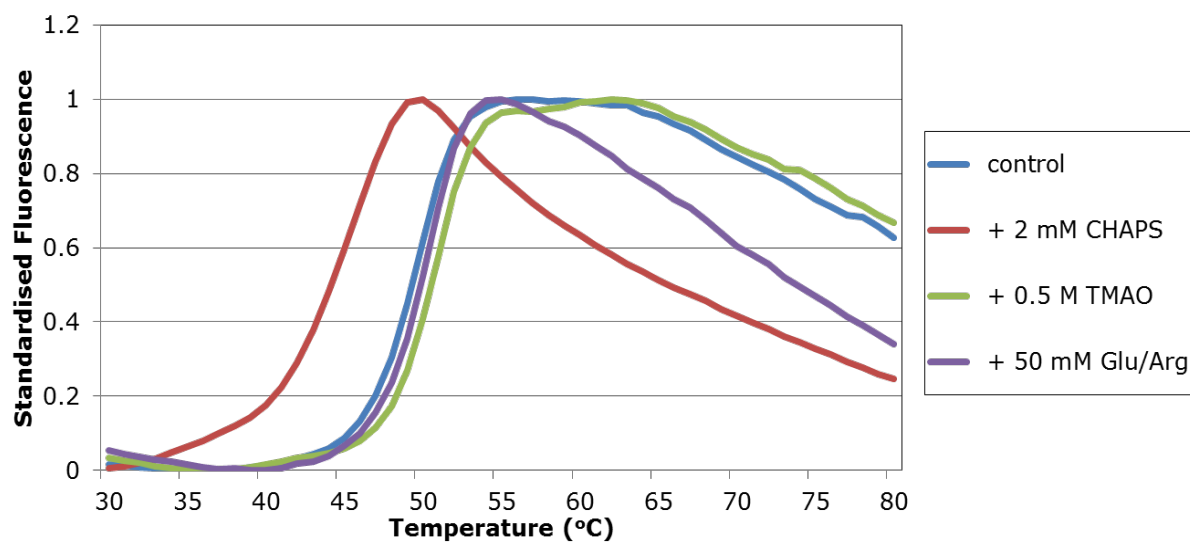
Buffer	T_m (°C)	ΔT_m (°C)
50 mM Sodium Phosphate, pH 7.0, 150 mM NaCl (control)	49.2	0
+ 2 mM CHAPS	45.3	-3.9
+ 0.5 M TMAO	50.2	+1
+ 50 mM Glu/Arg	50.2	+1

Figure 4.5.1 Thermofluor additive profile for His-Cdc25C$_{270-462}$

The sigmoidal melting curves for His-Cdc25C$_{270-462}$ with the different additives (a). The average melting temperature (T_m) and the change in T_m (ΔT_m) with respect to the control are listed in (b). The additives were tested in 50 mM sodium phosphate (pH 7.0), and 150 mM NaCl. The addition of 0.5 M TMAO or 50 mM Glu/Arg resulted in a 1 °C increase in melting temperature.

4.5.2 Solubility optimisation of His-Cdc25C$_{270-462}$

Fresh ^{15}N-labeled His-Cdc25C$_{270-462}$ was buffer exchanged into 50 mM sodium phosphate (pH 7.0), 100 mM sodium chloride, 0.5 M TMAO, 1 mM TCEP and 50 mM sodium phosphate (pH 7.0), 100 mM sodium chloride, 50 mM Glu/Arg, 1 mM TCEP. Following this, the protein in each additive condition was concentrated down to ~ 500 µM and kept at 4 ^{o}C for 1 day. Precipitation was noted in both samples. After centrifugation to remove the precipitate ^{1}H, ^{15}N − HSQC spectra were then collected.

With 0.5 M TMAO, the peaks in the HSQC are a lot less intense with some peaks being lost below the noise level. There was a greater amount of precipitation seen with this additive compared to 50 mM Glu/Arg. The intensity of peaks with 50 mM Glu/Arg was similar to the intensity of peaks seen in the HSQC spectrum with 50 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 0.5 mM TCEP (Figure 4.5.2). Therefore, His-Cdc25C$_{270-462}$ is more soluble in 50 mM Glu/Arg compared to 0.5 M TMAO.

From a total of 245 expected peaks, 196 peaks were observed with 0.5 M TMAO, and 218 peaks were observed with 50 mM Glu/Arg which is equivalent to 80 % and 89 % of the expected peaks observed respectively. 50 mM Glu/Arg improved the peak count of His-Cdc25C$_{270-462}$ by 9 %.

Although, 50 mM Glu/Arg improved the thermal stability, short-term solubility, and peak count for Cdc25C$_{270-462}$ it was not able to improve the long-term solubility of this protein. After the ^{1}H, ^{15}N-HSQC experiment which was run for 2 hrs, the majority of the protein precipitated.
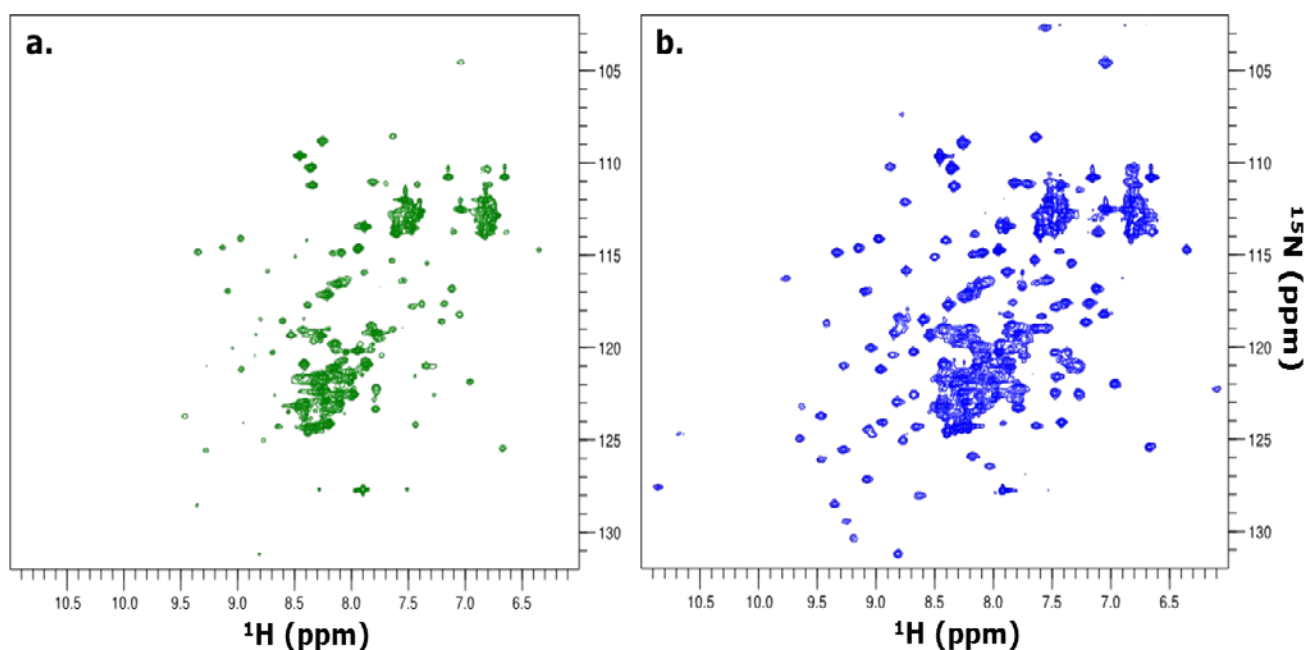
Figure 4.5.2: ^{1}H, ^{15}N – HSQC spectra of His-Cdc25C$_{270\text{-}462}$ in 0.5 M TMAO and 50 mM Glu/Arg

His-Cdc25C$_{270\text{-}462}$ was buffer exchanged into each additive buffer condition and concentrated to ~ 500 µM. The concentrated protein was kept at 4 °C for 1 day before HSQC spectra were collected. The ^{1}H, ^{15}N – HSQC spectra of His-Cdc25C$_{270\text{-}462}$ in 50 mM sodium phosphate (pH 7.0), 100 mM sodium chloride, 0.5 M TMAO, 1 mM TCEP (a) and in 50 mM sodium phosphate (pH 7.0), 100 mM Sodium Chloride, 50 mM Glu/Arg, 1 mM TCEP (b) were collected at 25 °C using the 600 MHz Agilent spectrometer. 50 mM Glu/Arg improved the solubility and peak count of His-Cdc25C$_{270\text{-}462}$.

4.6 Construct optimisation to improve protein solubility

It was critical to improve the long-term solubility of the protein in order to obtain NMR backbone assignment data. Therefore, the buffers with and without the additives listed in Table 4.6 were screened using the hanging-drop assay. In addition to this different Glu/Arg concentrations were tested. However, the 96 well plates which were observed after 48 hours at 23 ^{o}C showed extensive precipitation. Therefore, alternative constructs were tested with the aim to improve protein solubility and stability (the choice of constructs was based on personal communication: Professor Michael Overduin and Dr Mark Jeeves). The constructs generated (Tables 4.6.1 & 4.6.2) were tested using the hanging-drop solubility assay with the buffers and additives listed in Table 4.6. All the conditions were assessed after 48 hours and each protein was given an overall precipitaion score. The precipitation score was based on the Lepre and Moore system (Lepre and Moore, 1998) where a score of 0 was given for no precipitation and a score of 4 was given for extensive precipitation (see Figure 1.6.4.1).

Buffer Screen (+ 150 mM NaCl)	Additive Screen
Sodium Cacodylate, pH 6.5	2 mM CHAPS
MES, pH 6.5	0.5 M TMAO
Sodium phosphate, pH 7.0	50 mM Glu/Arg
Bis-Tris, pH 6.0	10 % Glycerol
HEPES, pH 7.5	200 mM Sucrose
MOPS, pH 7.0	
Tris, pH 8.0	

Table 4.6 Buffer and Additive list

All the buffers contained 150 mM NaCl and each of the buffers tested was also tested with the additives listed in the Table.

4.6.1 Different length constructs

The different length constructs tested are listed in Table 4.6.1. The constructs had molecular weights ranging from 22 to 26 kDa and a pI range from 5.8 to 6.6. This was because we wanted the chosen construct to be suitable for NMR study. The constructs which gave the best overall precipitation score were His-Cdc25C$_{270-449}$ and His-Cdc25C$_{270-443}$. Both of these constructs gave a precipitation score of 2.0. Shortening the C-terminal of His-Cdc25C$_{270-443}$ further from residue 443 resulted in a greater extent of precipitation observed. Removing ten amino acids from the N-terminal of His-Cdc25C$_{270-462}$ showed a slight improvement in the solubility. However, when these ten amino acids were removed from the N-terminal of His-Cdc25C$_{270-443}$ instead of a further improvement there was a decrease in protein solubility with an overall precipitation score of 4.0.

Construct	MW (kDa)	Number of amino acids	pI	Score
His$_6$-270-462	25	192	6.1	3
His$_6$-270-473	26	203	6.1	3/2
His$_6$-270-449	24	204	6.2	2
His$_6$-270-443	23	173	5.9	2
His$_6$-270-436	22	166	5.9	4
His$_6$-270-432	22	162	5.8	4
His$_6$-270-429	22	159	5.8	4
His$_6$-280-443	22	163	6.3	4
His$_6$-280-462	24	182	6.6	3/2

Table 4.6.1 Different length constructs

The proteins were purified from a 1L growth in sodium phosphate pH 7/7.5 (pH depended on the pI of the protein), 150 mM sodium chloride, 1 mM TCEP buffer, and concentrated to 400 μM. These proteins were then screened via the solubility screen assay using the buffer and additives listed in Table 4.6. Each protein was given an overall precipitation score after 48 hours incubation at 23 °C. The Table contains information on the molecular weight, number of amino acids, pI and overall precipitation score, for each protein.

4.6.2 Tagged constructs

Both His-Cdc25C$_{270-449}$ and His-Cdc25C$_{270-443}$ proteins had improved protein solubility compared to the other Cdc25C lengths tested. His-Cdc25C$_{270-449}$ was optimised further because it was the longer construct. It has been shown in the literature that adding a short stretch of amino acids to a protein terminal can significantly improve protein solubility. Therefore, a short stretch of 4 or 5 amino acids was added to the C-terminal of His-Cdc25C$_{270-449}$ (Table 4.6.2). The molecular weight range of these tagged proteins was from 23.7 to 24.7 kDa and the pI range was 5.8 - 7.0. The 4Glu tag was the best compared to the other tags. Overall, these tagged proteins did not further improve the protein solubility of His-Cdc25C$_{270-449}$ but resulted in a greater extent of precipitation seen with an overall precipitation score of 3.0.

Construct	MW (kDa)	Number of amino acids	Pi	score
His$_6$-270-449-5Arg	24.7	209	7.0	3
His$_6$-270-449-4Glu	24.5	208	5.8	3/2
His$_6$-270-449-5Lys	24.6	209	7.0	3
His$_6$-270-449-5Pro	24.4	209	6.2	3
His$_6$-270-449-GSSGS	24.3	209	6.2	3

Table 4.6.2 Tagged constructs

The proteins were purified from a 1L growth in Tris (pH 8)/Sodium phosphate (pH 7.5) (buffer was dependent on the pI of the protein), 150 mM sodium chloride, 1 mM TCEP buffer, and concentrated to 400 μM. These proteins were then screened via the solubility screen assay using the buffer and additives listed in Table 4.6. Each protein was given an overall precipitation score after 48 hours incubation at 23 ^{o}C. The Table contains information on the molecular weight, number of amino acids, pI, and overall precipitation score for each protein.

4.6.3 Cleavage of the His$_6$ tag

The N-terminal His tag was cleaved to assess its effect on Cdc25C protein solubility. The His tag was cleaved from both proteins, His-Cdc25C$_{270-449}$ and His-Cdc25C$_{270-443}$. These cleaved proteins were also tested with the solubility assay using the buffers and additives listed in Table 4.6. Cleaving the His tag resulted in a 48 hour precipitation score of 1.5 for Cdc25C$_{270-449}$ and a score of 1.0 for Cdc25C$_{270-443}$. Therefore, cleaving the His tag significantly improved protein solubility of both proteins and for Cdc25C$_{270-443}$ there was a greater improvement in protein solubility. The conditions which showed the most improvement contained 50 mM Glu/Arg.

^{1}H, ^{15}N-HSQC spectra of the untagged and His tagged Cdc25C$_{270-443}$ were performed (Figure 4.6.3). It can be seen that the two proteins overlay very well. This indicates that although the His tag negatively effects the solubility of Cdc25C it does not effect its conformation.
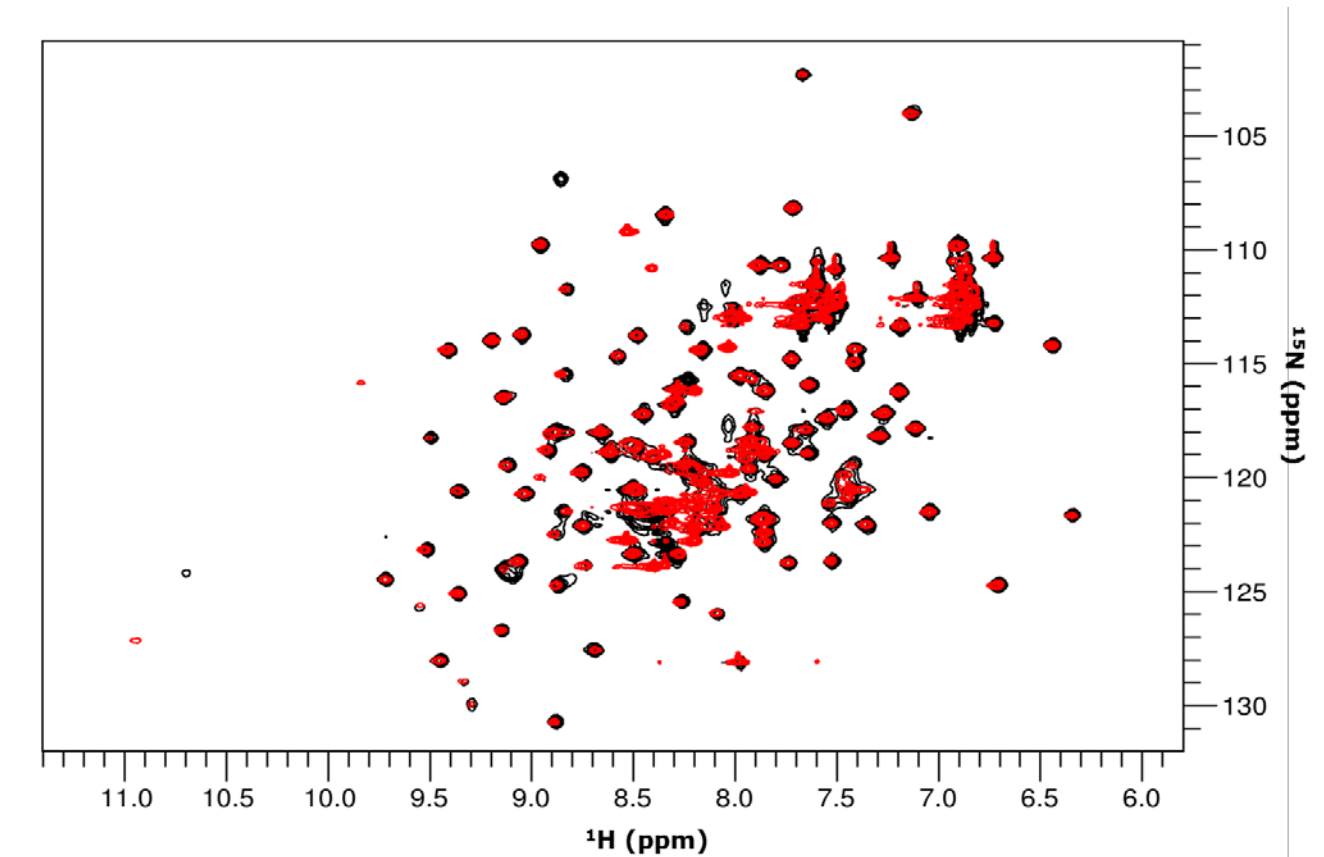
Figure 4.6.3 Cleavage of the His$_6$ tag

An overlay of ^1H, ^{15}N-HSQC spectra of 500 μM His-Cdc25C$_{270-443}$ (red) and Cdc25C$_{270-443}$ (black) in 50 mM sodium phosphate (pH 7.0), 150 mM NaCl, and 1 mM TCEP collected at 25 ^oC using a 600 MHz Agilent Direct Drive spectrometer.

4.7 Solution Optimisation of Cdc25C$_{270-443}$

The solubility screen was expanded in order to further improve the solubility of Cdc25C$_{270-443}$. Additional additives were tested using 50 mM sodium phosphate (pH 7.0) buffer. From this screen, it was concluded that 200 mM L-arginine and 200 mM sucrose further improved the solubility of the protein.

When the ^{1}H, ^{15}N-HSQC spectra of Cdc25C$_{270-443}$ are overlaid there is no significant change in protein conformation (Figure 4.7a). This indicated that these additives improved solubility of the protein without changing the protein structure and the protein also retained activity in this buffer.

To see if these additives affected the thermal stability of Cdc25C$_{270-443}$ they were tested by ThermoFluor® (Figure 4.7b). The addition of 200 mM L-arginine alone results in only a 0.1 ^{o}C increase in thermal stability while the addition of 200 mM sucrose resulted in a 1 ^{o}C increase in thermal stability. The addition of both L-arginine and sucrose did not further improve thermal stability. This suggests that sucrose could be improving protein solubility by enhancing the thermal stability of Cdc25C$_{270-443}$.

Using a combination of construct and solution condition optimisation improved the solubility of the Cdc25C catalytic domain (at 1 mM concentration) to 4 days at 23 ^{o}C and one month at 4 ^{o}C.
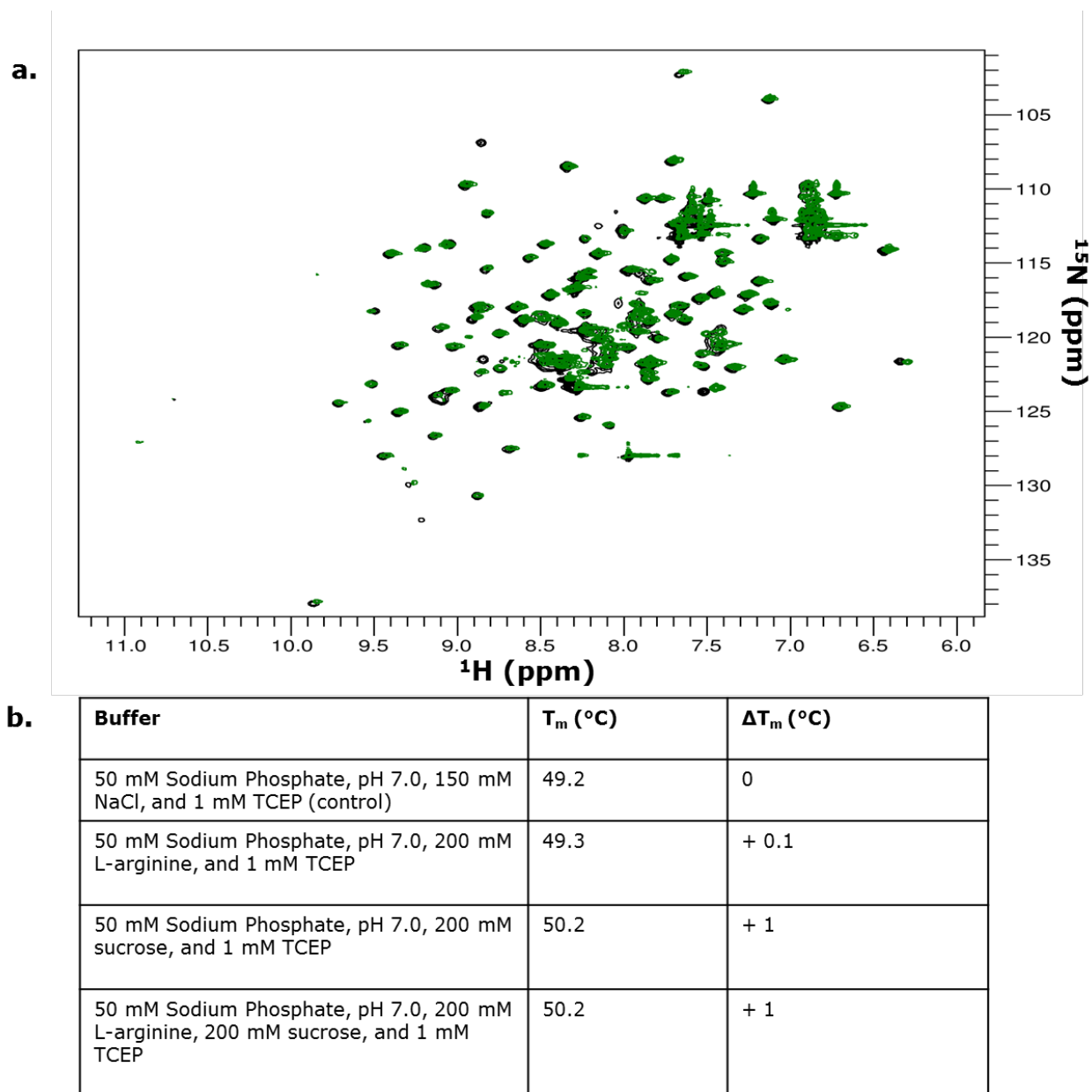
Figure 4.7 Solution Optimisation

An $^1H,^{15}N$-HSQC overlay of Cdc25C$_{270-443}$ in 50 mM sodium phosphate (pH 7.0), 150 mM NaCl, and 1 mM TCEP (black), and 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine, 200 mM sucrose, and 1 mM TCEP (green) (a). The Table (b) shows the effect of the additives on the T_m of Cdc25C$_{270-443}$.

Following solubility optimisation of the Cdc25C catalytic domain it was concluded that the Cdc25C$_{270-443}$ protein provided the optimal solubility. From now on this protein will be referred to as the Cdc25C catalytic domain.

4.8 Protein expression optimisation of the ^{13}C, ^{15}N labeled Cdc25C catalytic domain

4.8.1 M9 minimal media optimisation

The next stage was to optimise the expression of the Cdc25C catalytic domain. Since a primary aim was to obtain backbone assignment, for which experiments can generally take up to two weeks, it was desirable to maximise the amount of protein produced. There are many different ways one can improve the expression and hence yield of a protein. One of the strategies employed here was to optimise the M9 minimal media.

A bacterial growth curve is useful when deciding on the optimal induction OD$_{600}$. Ideally, induction should take place when the bacteria are still actively growing. This is to allow maximum growth of the culture and hence provide high protein yield. Therefore, usually, the culture is induced during middle of the exponential growth phase. However, it was noted that after induction of the *E. coli* expressing the Cdc25C catalytic domain there was very little growth of the bacterial biomass (data not shown). Therefore, it was decided to induce when the culture reached maximal biomass, before stationary phase is started which is ~ O.D$_{600}$ 0.9 (Figure 4.8.1a). This complements the SGC (Oxford) protocol for the growth of the Cdc25C catalytic domain, which includes induction at an OD$_{600}$ of 1.0.

The effect of adding supplements to the M9 minimal media on protein expression was investigated. The two media supplements used were ISOGRO and BioExpress. Both of these

supplements have been designed to enhance the M9 minimal medium and improve protein expression. There was also the benefit of potentially reducing the lag phase and therefore the total time taken to reach the high induction OD_{600}.

BioExpress® is provided as a 10x concentrated solution by CIL (Cambridge Isotope Laboratories, Inc) to be made up as a complete medium with pure water. However CIL have also shown that using 1 % of BioExpress in addition to the standard M9 minimal media can help improve *E. coli* growth rate and yield (Rhima et al., 2011). This is ideal because it is cheaper to supplement with BioExpress then use it to make a complete medium.

ISOGRO® is another supplement that can be added to M9 minimal media. This was used at a final concentration of 10 %, which was also according to the manufacturer's (Sigma) recommendations. Sigma state on their website that this concentration is optimal in improving *E. coli* growth rate and protein expression.

Compared to the cultures in M9 minimal media alone which had a lag phase of 3 hours, the lag phase was shortened by 2 hours with the addition of either supplement (Figure 4.8.1a).

The addition of either BioExpress or ISOGRO also reduced the total time taken to reach the induction OD_{600} (Figure 4.8.1a). With M9 minimal media alone, 7.5 hours were taken to reach an OD_{600} of 0.9. Supplemented with BioExpress it took 5.5 hours and supplemented with 10 % ISOGRO it took 5 hours to reach this OD_{600}. Therefore, the time taken to reach an induction OD_{600} of 0.9 was reduced with the supplements. With 1 % BioExpress, the induction OD_{600} was reduced by 2 hours and with 10 % ISOGRO the induction OD_{600} was reduced by 2.5 hours when compared to the M9 minimal media alone. The M9 culture reached the required OD_{600} the fastest when supplemented with 10 % ISOGRO.

Although, there was a significant improvement in the induction time, the protein expression levels did not change when compared to the M9 minimal media alone. Therefore, the addition of the supplements did not improve the expression and hence yield of the His-Cdc25C$_{270-443}$ protein (Figure 4.8.1b).
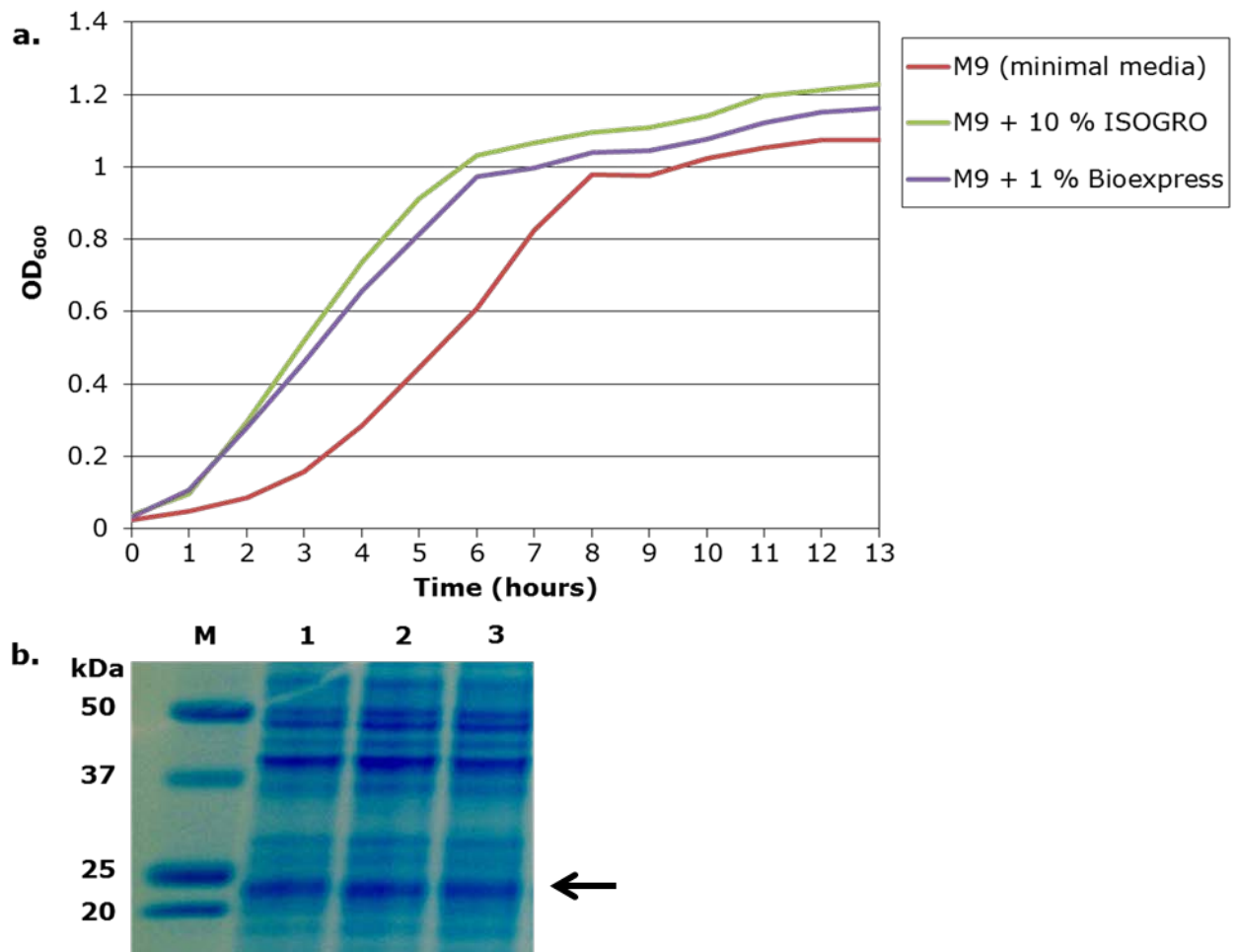
Figure 4.8.1 M9 minimal media optimisation

Bacterial growth curves (a) of uninduced, ^{13}C, ^{15}N labeled M9 minimal media alone (red), ^{13}C, ^{15}N labeled M9 minimal media supplemented with ^{13}C, ^{15}N 10 % w/v ISOGRO (green), and ^{13}C, ^{15}N M9 minimal media supplemented with ^{13}C, ^{15}N 1 % v/v BioExpress (purple). An SDS-PAGE of the induced cultures is shown (b). The media cultures were induced at the OD_{600} of 0.9 and incubated overnight at 18 ^{o}C (for methodology details, see section 2.5.1). The first lane contains the molecular weight marker (M), followed by total cell lysates, ^{13}C, ^{15}N labeled M9 minimal media (1), ^{13}C, ^{15}N labeled M9 minimal media supplemented with ^{13}C, ^{15}N 10 % w/v ISOGRO (2), and ^{13}C, ^{15}N labeled M9 minimal media supplemented with ^{13}C, ^{15}N 1 % v/v BioExpress (3). These experiments were repeated three times.

4.8.2 Temperature Optimisation

The aim here was to see if changing the temperature would improve the expression of Cdc25C. Generally, increasing the temperature means that protein will be produced at a faster rate. However, if the induction temperature is too high it can result in the protein not being folded correctly and hence present in inclusion bodies.

Three different temperatures were tested for an overnight induction. An induction temperature of 18 ^{o}C resulted in the least amount of protein being expressed (Figure 4.8.2). An induction temperature of 37 ^{o}C showed a greater increase in protein expression levels compared to 18, and 25 ^{o}C. However, none of this protein was present in the soluble fraction. This was further tested for a shorter induction time of 3 hours to determine if there was any expression in the soluble fraction; there was little expression in the soluble fraction (data not shown). Therefore, 37 ^{o}C is not a good induction temperature for Cdc25C.

An induction temperature of 25 ^{o}C had a greater level of Cdc25C expression compared to 18 ^{o}C and majority of the protein was expressed in the soluble fraction. This temperature was taken as the best temperature for overnight induction.
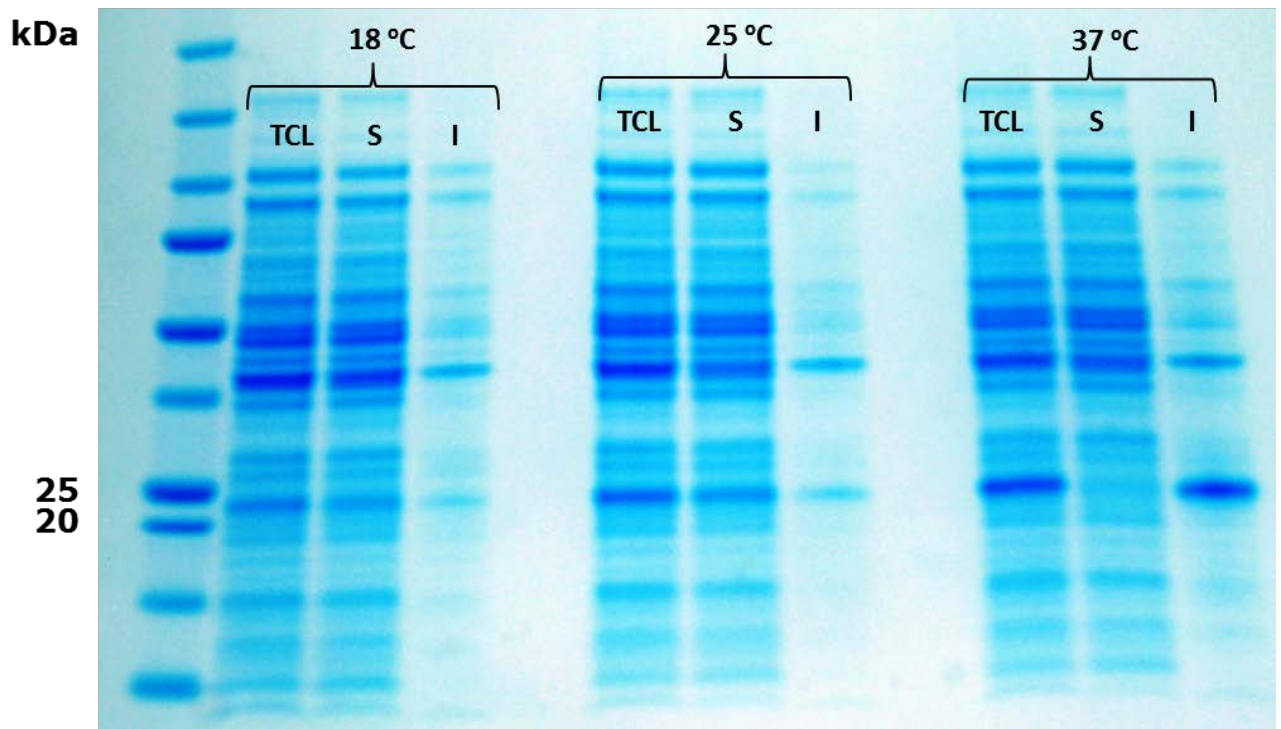
Figure 4.8.2 Temperature Optimisation

Cultures of Cdc25C expressing E. coli were grown in M9 minimal media. The cultures were induced at an OD_{600} of 0.9 with 1 mM IPTG and incubated overnight at 18, 25, and 37 ^{o}C (for methodology details, see section 2.5.2). The gel shows the marker in the first lane, total cell lysate (TCL), soluble (S), and insoluble (I) fractions. His-Cdc25C$_{270-443}$ has a molecular weight of ~ 23 kDa and therefore appears as protein bands between 20-25 kDa on the gel. An induction temperature of 25 ^{o}C was taken to be the optimal temperature for expression.

4.9 Biophysical and structural characterisation of the Cdc25C catalytic domain

The molecular weight and oligomeric state of the protein was investigated prior to preparing the protein sample for backbone assignment experiments. This was to ensure quality of the purified protein before structural analysis commenced.

4.9.1 Analytical size exclusion chromatography

The solution state of the Cdc25C catalytic domain was determined using analytical size exclusion chromatography (Figure 4.9.1). The profiles of all the standards which were run independently are overlaid with the elution profile of the Cdc25C catalytic domain. Following plotting of the calibration curve the calculated molecular weight was 24 kDa. The predicted molecular weight of the cleaved protein based on its amino acid sequence using the ProtParam tool (Gasteiger et al., 2005) was 20.6 kDa. This difference of 3.4 kDa is likely due to the assumption made when obtaining the molecular weight from the size exclusion chromatogram that the protein has a fully spherical shape. The Cdc25C catalytic domain may have a slightly bigger hydrodynamic radius which will affect the time taken to elute down the column.

4.9.2 Analytical ultracentrifugation (AUC)

The AUC analysis revealed a single peak indicating the presence of a single monomeric species. The sedimentation coefficient obtained from the AUC analysis was 1.3 S which corresponds to a molecular weight ~ 20 kDa (Figure 4.9.2). This corresponds to the results obtained from ProtParam and the analytical size exclusion chromatography analysis.
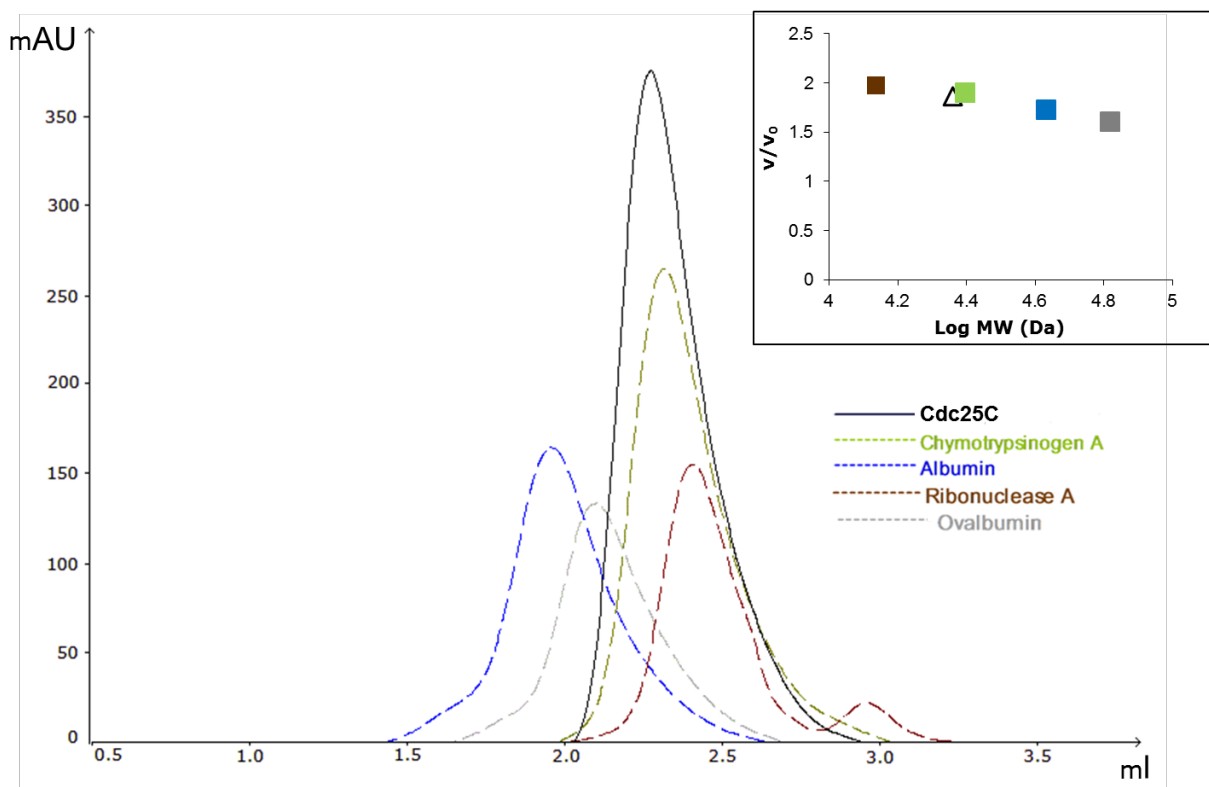
Figure 4.9.1 Molecular weight determination of the Cdc25C catalytic domain

The cleaved Cdc25C catalytic domain in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine, 200 mM sucrose, and 1 mM TCEP was injected onto the Superdex 200 5/150 column to determine its molecular weight (for methodology details, see section 2.11). The milli absorbance (mAU) was measured at 280 nm. The protein standards, Ribonuclease A (13.7 kDa), Chymotrypsinogen A (25 kDa), Albumin (66 kDa), and Ovalbumin (43 kDa) were used for the calibration. The box at the top right-hand corner shows a graph where the log molecular weight (MW) of these standards (squares) and the Cdc25C protein is plotted against their respective elution volume/void volume (V/V_0) ratios. The Cdc25C catalytic domain is depicted as a black triangle.
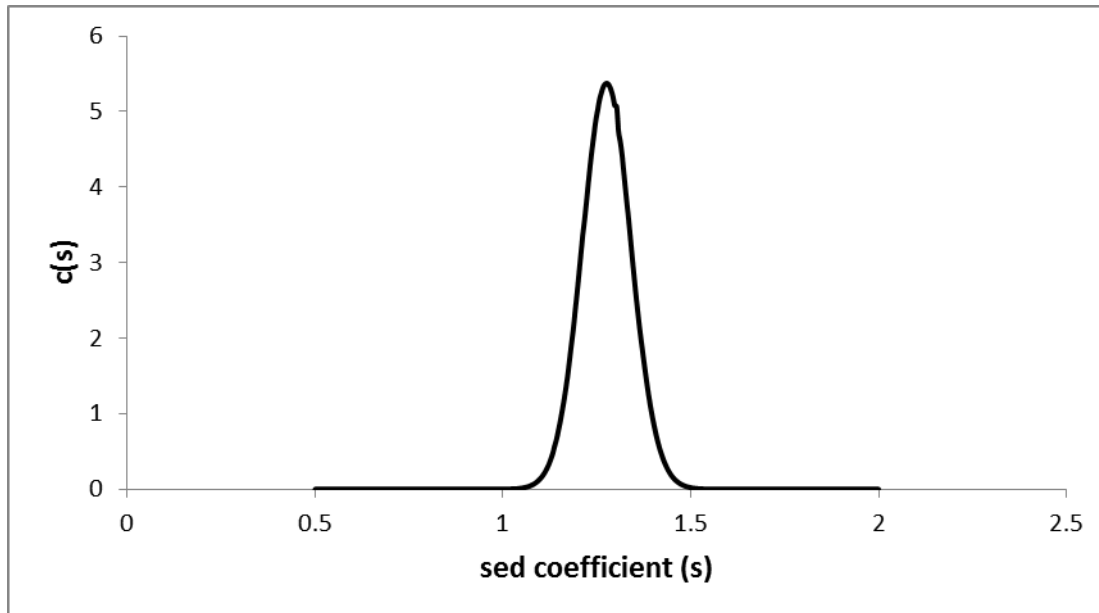
Figure 4.9.2 Determination of the oligomeric state of the Cdc25C catalytic domain

A 30 μM sample of the cleaved Cdc25C catalytic domain in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine, 200 mM sucrose, and 1 mM TCEP was subjected to AUC analysis. The program SEDFIT (Brown and Schuck, 2006) *was used to perform the c(s) analysis from which it was determined that the Cdc25C catalytic domain is a monomer.*

4.9.3 Backbone assignment

3D NMR experiments were attempted in order to obtain the backbone assignment of the Cdc25C catalytic domain (data not shown). However, this was not successful due to insufficient peaks in the carbon resolved experiments which precluded sequential assignment. There was precipitation which could account for the loss of the peaks. The peaks that were present had differential peak intensities suggesting the Cdc25C catalytic domain has a dynamic backbone.

4.10 Small angle X-ray scattering (SAXS)

4.10.1 SAXS Parameters

SAXS data for three different concentrations 1, 5, and 10 mg/ml were collected. Following the examination of the Guinier regions and comparing the different concentrations there was no aggregation noted in any of the SAXS plots. Therefore, the highest concentration, 10 mg/ml was taken forward for the analysis (Figure 4.10.1.1a). This was because the highest concentration provided the best signal to noise ratio. A few points at lower values of q were removed from this plot. This is noise due to the beam stop.

The radius of gyration (R_g) obtained from the Guinier plot was 1.9 nm which agrees with the R_g obtained from the P(R) analysis which was 1.96 nm (Figure 4.10.1.1b).

The bell-shaped curve in the Kratky plot is typical of a folded protein (Figure 4.10.1.1c). It can clearly be seen that although the initial tailing is constant there is an increase at higher q. This suggests there is a disordered element in the protein.

The P(R) graph shows tailing at high R which suggests that the protein is elongated (Figure 4.10.1.1d). This elongation could be because of the fact that the Cdc25C catalytic domain is not fully globular or because there is flexibility in the protein.

Visually inspecting how the curve ends while inputting different D_{max} values with the aim of obtaining a smooth P(R) graph placed the D_{max} between 7.2 – 8.0 nm where 7.7 nm was a good estimate. D_{max} is usually around three times the R_g value. This is not the case here which indicates Cdc25C is elongated or there is flexibility within the protein. The SAXS parameters for the catalytic domain of Cdc25C are summarised in Table 4.10.1.

The SAXS envelope of the Cdc25C catalytic domain was generated using the DAMMIF and DAMAVER programs from the ATSAS suite (Figure 4.10.1.2). The crystal structure of the Cdc25C

catalytic domain was then fitted into this envelope using SUPCOMB. The normalized spatial

discrepancy (NSD) value for the fit which was provided by SUPCOMB was 1.5 suggesting a

possible difference between the SAXS envelope and the crystal structure.

Parameter (nm)	Cdc25C catalytic domain
R_g (Guinier)	1.90 +/- 0.017 (qR_g < 1.3)
R_g (P(R))	1.96
D_{max}	7.7 (7.2 – 8.0)

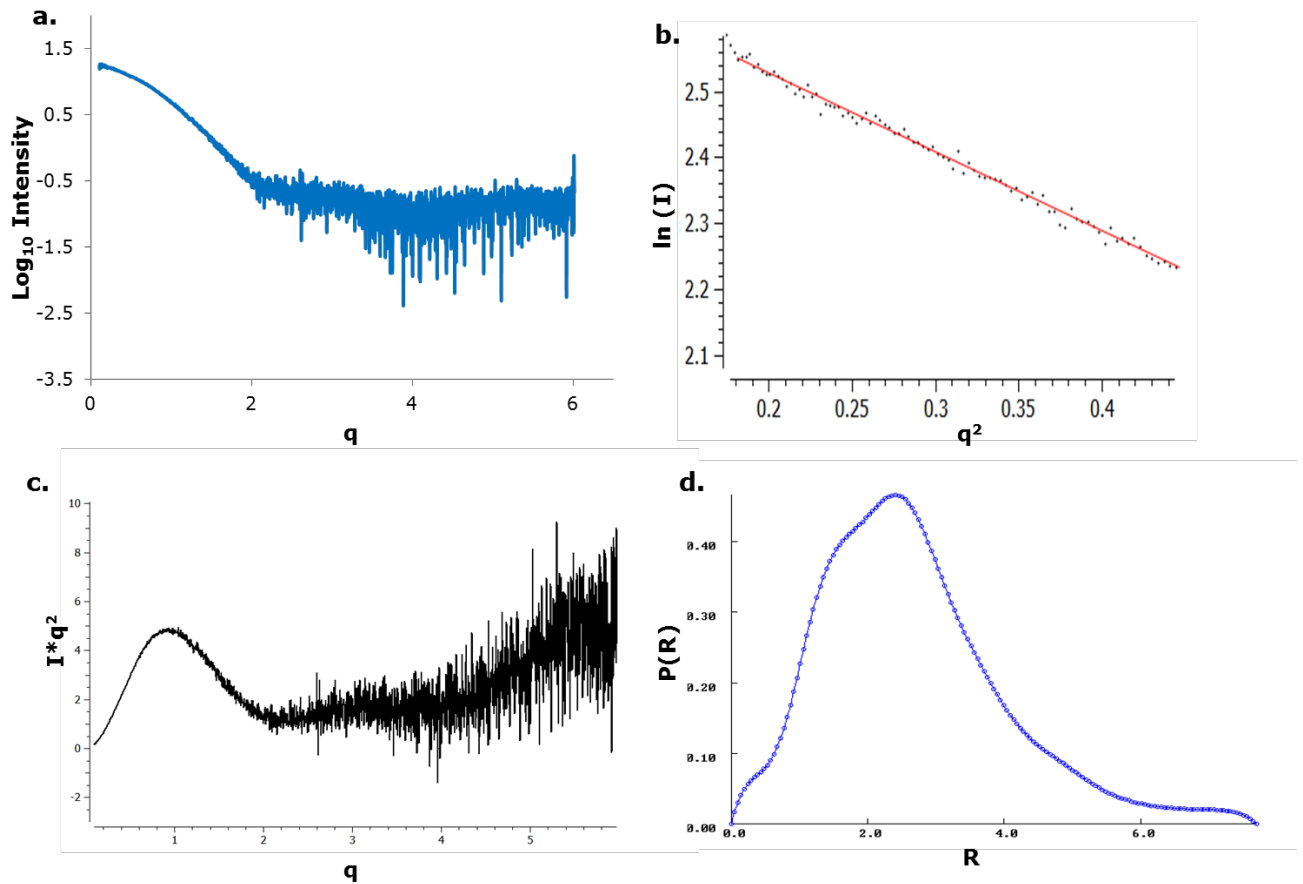**Table 4.10.1 SAXS parameters for the Cdc25C catalytic domain**

Figure 4.10.1.1 SAXS graphs of the Cdc25C catalytic domain

SAXS data were collected at EMBL (DESY, Germany). The ATSAS software suite was used to analyse the data. SAXS scattering plot of 10 mg/ml Cdc25C catalytic domain in the buffer 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine, 200 mM sucrose, and 1 mM TCEP (a). The logarithmic intensity is plotted against q (momentum transfer) which is expressed as nm^{-1}. The Guinier plot which is plotted as ln (I) against q^2 shows a straight line can be fitted in the Guinier region (b). The Kratky plot is plotted as intensity multiplied by q^2 against q. The profile of the Kratky plot suggests a folded protein with disordered regions. The P(R) graph which is plotted as P(R) against R shows tailing at high R. R is expressed in nm.
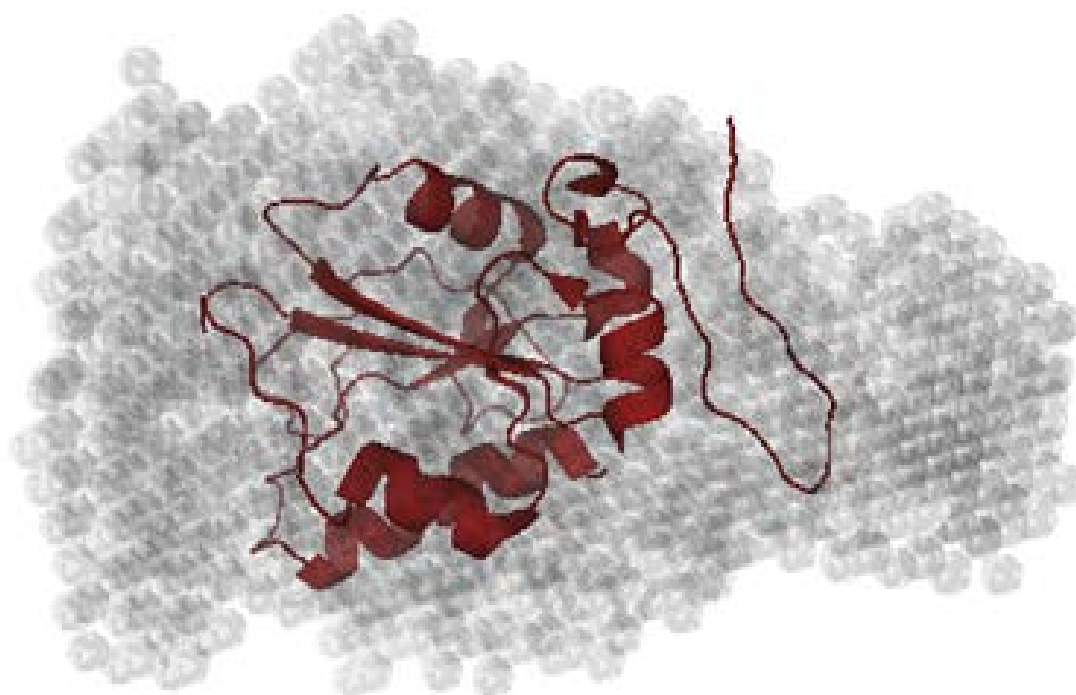
Figure 4.10.1.2 SAXS envelope of the Cdc25C catalytic domain

The crystal structure of the Cdc25C catalytic domain (PDB: 3OP3) was fitted to the SAXS envelope using SUPCOMB. The NSD value was 1.5 suggesting a difference between the low resolution SAXS envelope and the atomic 3D models.

4.10.2 Comparing solution data to the crystal data

CRYSOL was used to create a theoretical SAXS scattering curve for the crystal structure of the Cdc25C catalytic domain which was fitted to the solution SAXS data. Although, the fit is reasonable with an χ^2 value of 1.57 the SAXS scattering curve does not fully fit the data (Figure 5.10.2). This could be because the construct used for SAXS is different to the construct used for the crystal structure. The differences between the constructs are in the N-terminal region, the

catalytic region, and the C-terminal. The crystal structure has no electron density for the N-terminal residues 269-279 and electron density for two serine residues are missing in the catalytic site. Also, there are a few extra residues in the C-terminal where there is electron density until residue 448 while the SAXS construct ends at residue 443. Therefore, I-TASSER was used to generate a model which would be more representative of the construct used to collect the SAXS data.

CRYSOL was used to generate a scattering curve for the I-TASSER model and this was fitted to the solution SAXS data giving a χ^2 value of 1.65. This scattering curve also does not completely fit the SAXS data (Figure 4.10.2). However, it overlaid very well to the simulated SAXS scattering curve of the crystal structure. Overall, both SAXS scattering plots generated for the crystal structure and the I-TASSER model do not fully fit the SAXS solution data. Therefore, even after accounting for the differences between the solution and crystal constructs there is still a difference between the solution and crystal models of the Cdc25C catalytic domain.
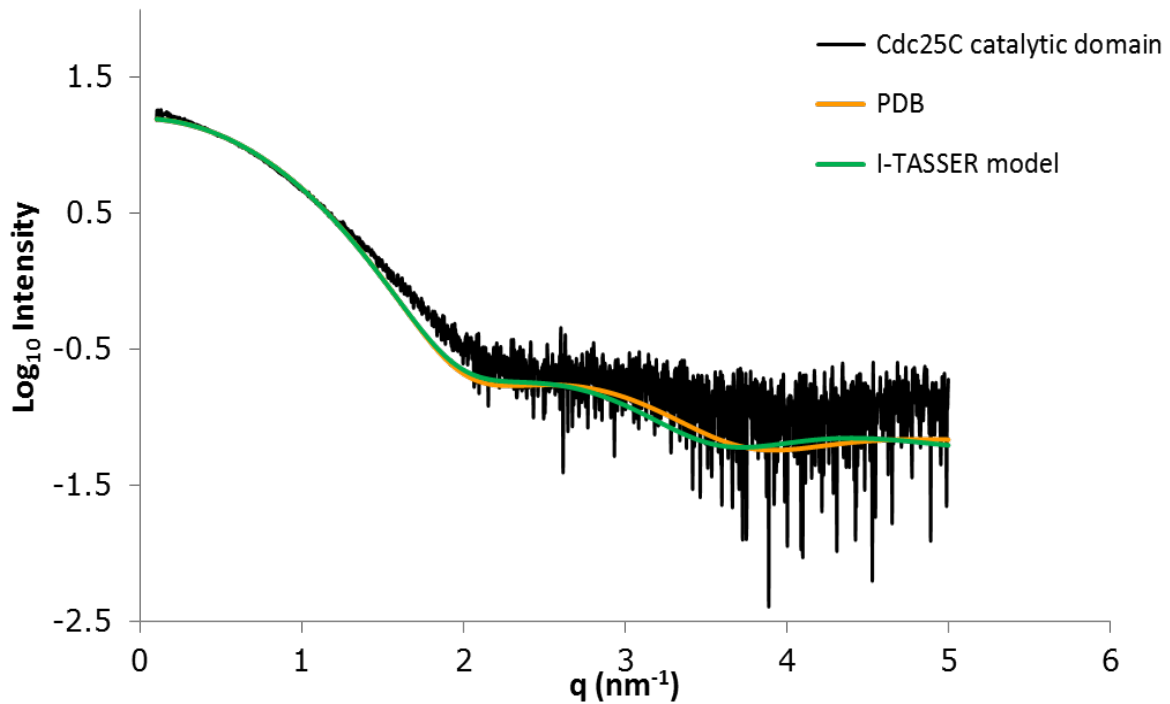
Figure 4.10.2 A comparison of the solution data to the crystal data

The rigid model obtained from I-TASSER and the crystal structure (PDB: 3OP3) were fitted to the SAXS data using CRYSOL. Both models do not fully fit the SAXS solution scattering curve.

4.10.3 Ensemble Optimisation method (EOM)

A combination of the results such as the shape of the Kratky plot, the tailing of the P(R) graph

and the large D_{max} indicated flexibility within the Cdc25C catalytic domain which could explain

the difference seen between the solution and crystal models. These methods cannot identify the regions in which the flexibility lies. Therefore, EOM was used to characterise this flexibility (for details on the EOM method, see methodology, section 2.17.3).

The 3D NMR experiments (data not shown) suggested that the backbone of the Cdc25C catalytic domain is dynamic. This could be because the catalytic domain of Cdc25C contains mobile loops or flexible termini.

4.10.3.1 Assessing flexibility of the termini

Since there is no X-ray electron density for the N-terminal residues 269 – 279 I-TASSER was used to generate a more complete model of the Cdc25C catalytic domain. I-TASSER (Zhang, 2008) uses the protein amino acid sequence and information from the PDB database to generate 3D atomic models. Where no data are available either because there is no matching PDB or there is no electron density in certain regions I-TASSER utilizes ab initio modelling to generate the final 3D model. The model which had the best I-TASSER score was chosen for the EOM analysis. The scattering curve when both termini are made rigid was determined from the CRYSOL program. The sequences which were assessed for flexibility are shown in Table 4.10.3.1.

Region	Amino acid sequence
N-termini$_{269-279}$	SMTQMLEEDSNQ
C-termini$_{431-443}$	SYCPMHHQDHKTE

Table 4.10.3.1 Termini sequences assessed by EOM for flexibility

With both termini flexible the EOM generated scattering plot fits very well to the SAXS data compared to when both termini are rigid (Figure 4.10.3.1.1). The fit has a χ^2 value of 1.32 when

both termini are flexible compared to when both termini are rigid which has an χ^2 value of 1.65. Chimera was used to overlay the final 20 models generated by EOM (Figure 4.10.3.1.2).

The termini were then assessed independently to try and identify a terminal which could account for the flexibility (Figure 4.10.3.1.3). The χ^2 values for the fits show no significant difference when the N-terminal is flexible (χ^2= 1.77) or when the C-terminal is flexible (χ^2= 1.77).

Overall, the EOM analysis suggests there is flexibility in the catalytic domain and both termini contribute to this flexibility. However, the total flexibility is not accounted for by the termini. This is because when both termini are flexible the χ^2 value is 1.32 (the closer the χ^2 value is to zero the better the fit). This suggests there is some inherent flexibility in the Cdc25C catalytic domain which cannot be fully accounted for by the termini.
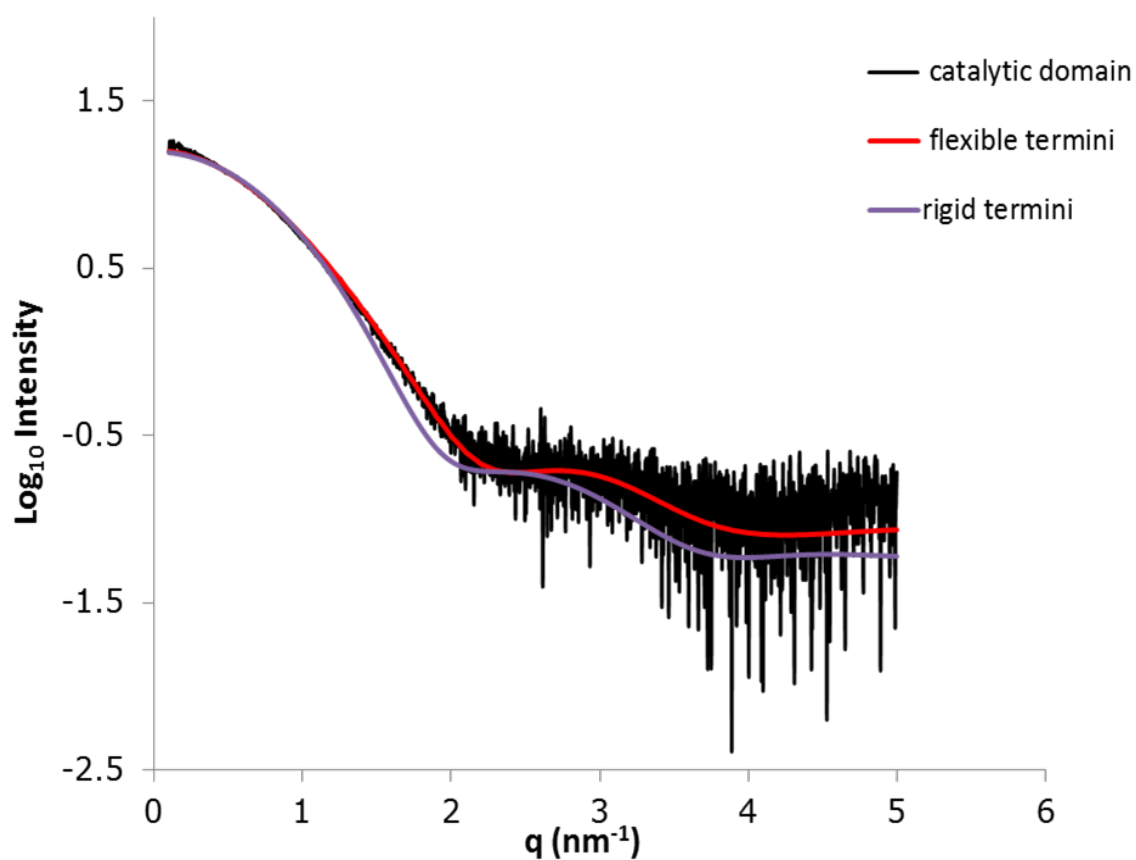
Figure 4.10.3.1.1 Assessing flexibility of both termini

Models with flexible or rigid termini were generated using the I-TASSER model. EOM was used to obtain the fitted SAXS scattering curve when both termini were flexible (χ^2 = 1.32) and CRYSOL was used to do this when both termini were rigid (χ^2 = 1.65). The data is plotted with intensity (Log$_{10}$) against the momentum transfer (q). The scattering curve fits better to the data when both termini are flexible.
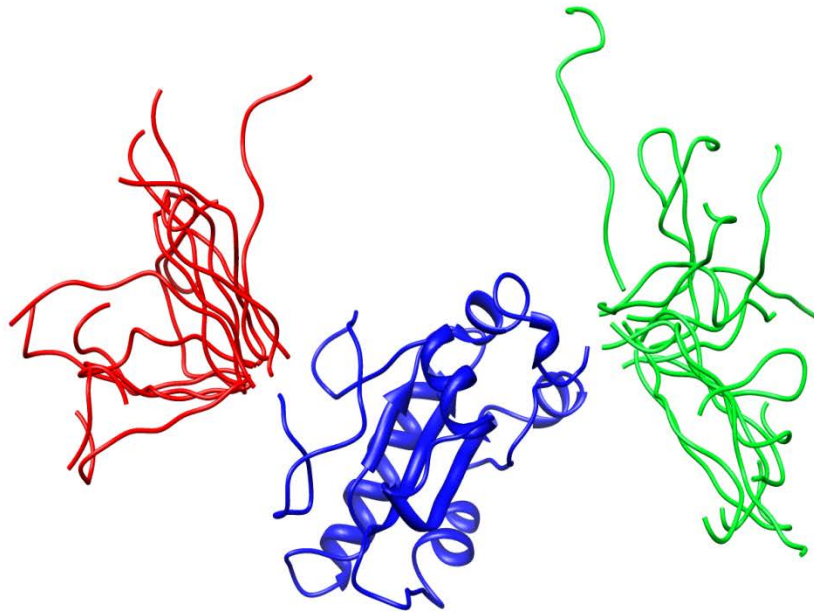
Figure 4.10.3.1.2 Overlay of the EOM ensemble when both termini are flexible

An overlay of the 20 models (ensemble) produced by EOM when both termini are defined as being flexible. This Figure shows there is a moderate degree of flexibility predicted in the termini. The program Chimera (Pettersen et al., 2004) was used to overlay the models. In blue is the model used for the catalytic domain and in red and green are the N and C termini respectively.
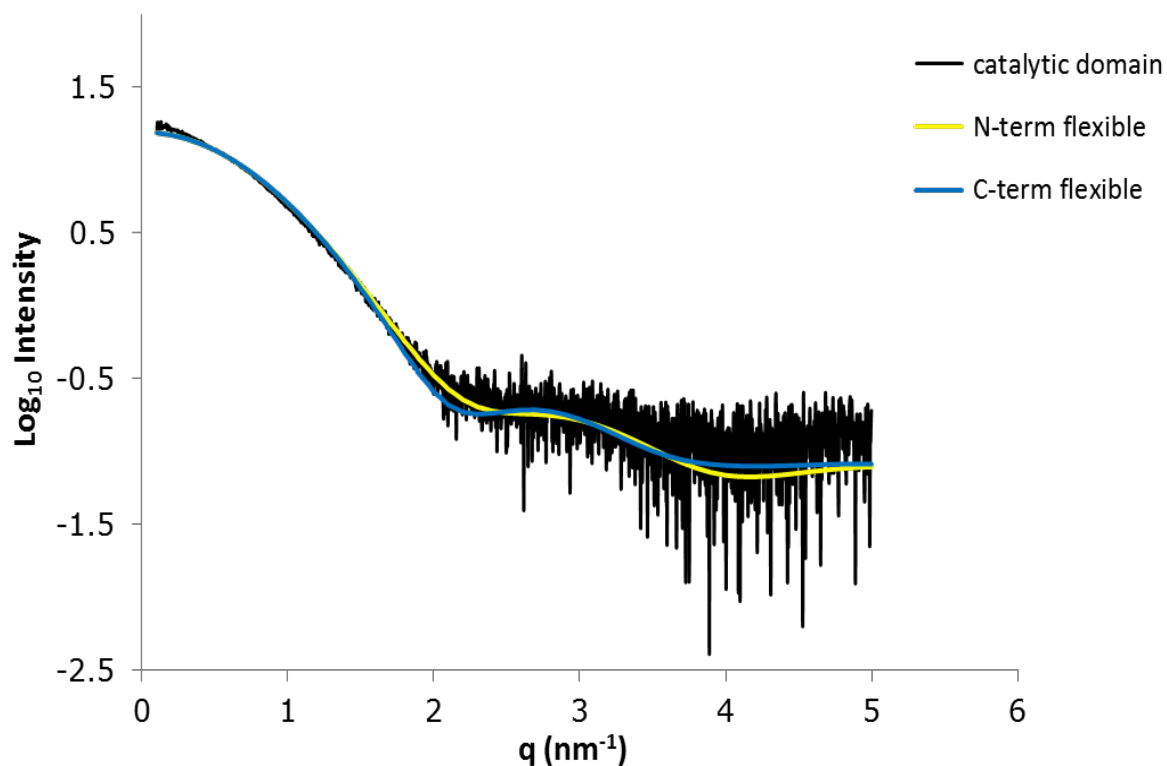
Figure 4.10.3.1.3 Assessing flexibility of individual termini

The programme EOM was used to obtain the fitted SAXS scattering curves when either the N-terminal or C-terminal was made flexible. EOM generates a pool of 10 000 models from which it selects 20 best models. It generates an averaged scattering curve of the best models which is then fitted to the SAXS data. With either the N- or C- terminal flexible the χ^2 was 1.77. The data are plotted with intensity (Log_{10}) against the momentum transfer (q).

4.11 Cdc25C B-factor analysis

B-factors are obtained from X-ray crystal structures. They can indicate the extent of mobility of the atoms in the crystal structure (Radivojac et al., 2004). Regions with high B-factors tend to experience high mobility compared to regions with low B-factors. The B-factors associated with the crystal structure of the Cdc25C catalytic domain (3OP3) were examined to identify potentially dynamic regions.

It can be seen that apart from the termini three other regions (a, b, and c) have high B-factors (Figure 4.11). These correspond to residues 299-301 (a), 365-369 (b), and 319-321 (c). These regions could therefore also contribute to the flexibility.
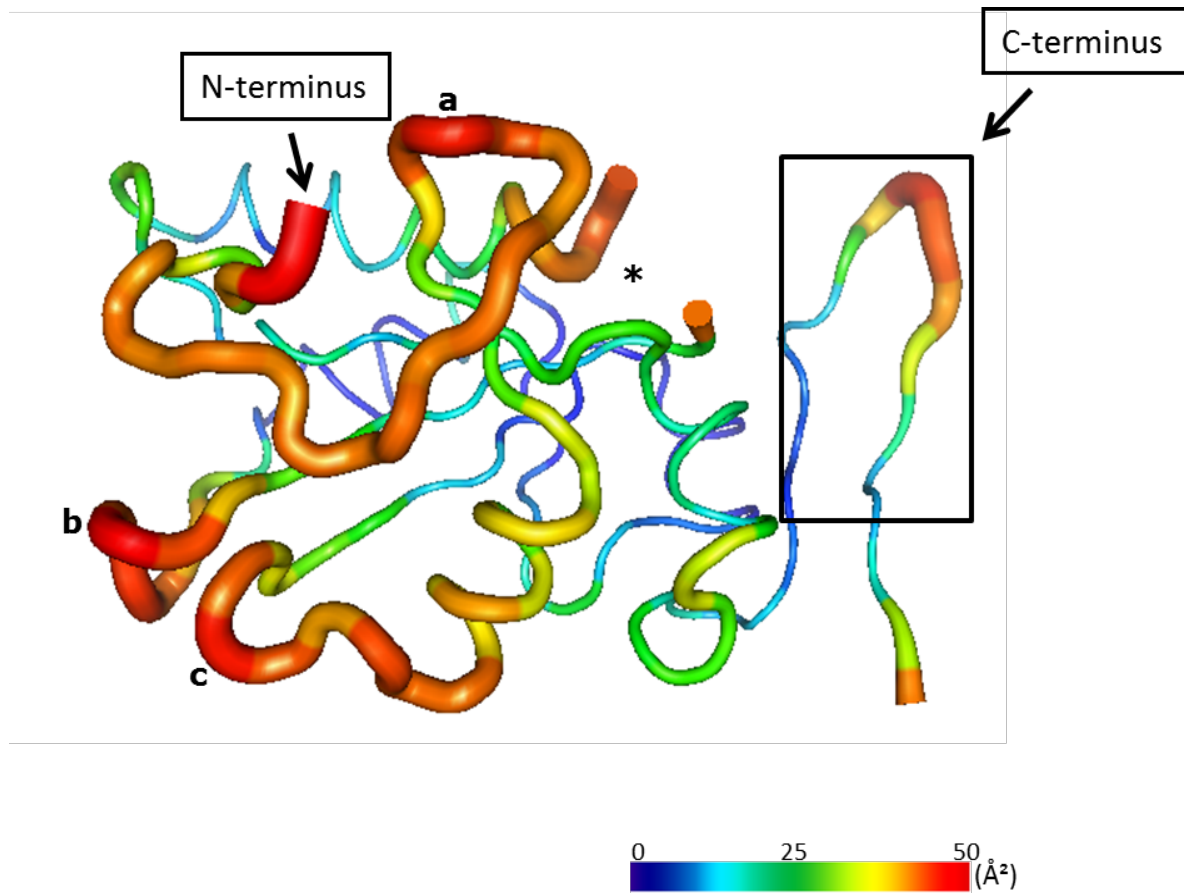
Figure 4.11 Cdc25C B-factor analysis

*The Cdc25C crystal structure in conjunction with PyMOL was used to generate the Figure. The B-factors are expressed in $Å^2$ with a scale ranging from 10 to 50 $Å^2$. The N- and C-termini are annotated. Additional regions indicated by the B-factors to experience high mobility are labelled a, b, and c. The catalytic site is marked ***

4.12 Discussion of Chapter 4.0

The main aim of this chapter was to identify an optimal Cdc25C catalytic domain construct and solution conditions in order to obtain backbone assignments. Sodium phosphate and increasing NaCl concentrations improved the thermal stability of the protein. Initially, 50 mM sodium phosphate (pH 7.0), 100 mM NaCl, and 0.5 mM TCEP was identified as a good starting buffer condition. The ^{1}H ^{15}N-HSQC collected in this condition showed the protein was folded with well dispersed peaks. Around 87 % of peaks were observed in the protein HSQC spectrum making assignments possible. However, significant precipitation was noted during these initial experiments. Therefore, different additives were screened in order to identify conditions that would improve the solubility of the protein and may also improve the HSQC peak count.

The addition of 50 mM Glu/Arg improved the short term solubility of the protein (His-Cdc25C$_{270-462}$) and modestly improved the HSQC spectrum. It has been previously reported that 50 mM Glu/Arg can significantly improve protein stability and solubility. For example, a study reported an increase in T$_m$ by 1.4 K for chymotrypsinogen A (Blobel et al., 2011). It was considered that the effect of 50 mM Glu/Arg in improving protein stability was reserved to the disordered protein loop regions. It was thought that these loops may collapse onto the protein core resulting in a compact structure with significantly reduced non-specific interactions of the disordered loops. This could also provide an explanation of why the addition of 50 mM Glu/Arg improved the short term solubility of His-Cdc25C$_{270-462}$.

Despite the short term improvement in solubility by 50 mM Glu/Arg there was still a significant amount of precipitation. Increasing the concentration of Glu/Arg did not further improve protein solubility.

Alternative Cdc25C catalytic domain constructs were generated and their solubility was tested. The shorter constructs His-Cdc25C$_{270-443}$ and His-Cdc25C$_{270-449}$ had the highest 48-hour

solubility score. His-Cdc25C$_{270\text{-}449}$ was further optimised by the addition of short solubility peptide tags. Unfortunately, there was no improvement in solubility noted.

Cleaving the N-terminal His tag resulted in an improvement in protein solubility. Interestingly, in support of this finding a study showed some proteins with a His tag have lower solubilities in comparison with no tag (Woestenenk et al., 2004). Although, the His tag had a negative effect on the solubility of the Cdc25C catalytic domain it did not affect its conformation since there was no significant difference in the ^{1}H, ^{15}N-HSQC overlay of the untagged and His tagged protein.

The findings in this study suggest the termini of the Cdc25C catalytic domain had a negative effect on protein solubility since shortening the termini improved solubility. There is no electron density in the crystal structure for these termini suggesting they are disordered. These disordered termini may be very flexible and vulnerable to forming non-specific interactions with each other and ultimately resulting in the protein to precipitate out of solution.

The protein solubility of the cleaved protein (Cdc25C$_{270\text{-}443}$) was further optimised by screening with an expanded range of additives. The additives 200 mM sucrose and 200 mM L-arginine further improved protein solubility. The solubility was improved to 4 days at 23 $^{\circ}$C and 1 month at 4 $^{\circ}$C. Therefore, a combination of shortening the termini and addition of these additives improved the long term solubility of the Cdc25C catalytic domain. The optimised buffer condition 50 mM sodium phosphate (pH 7.0), 200 mM sucrose, 200 mM L-arginine, and 1 mM TCEP did not affect the protein conformation and the protein retained activity in this condition. Sucrose has been described in the literature as a protein stabilising additive (see chapter 1.0). Sucrose is preferentially excluded from protein surfaces. Due to this preferential exclusion proteins become compact. The benefits of arginine are also well documented (see chapter

1.0). These include an increase in thermal stability, reduction of partially folded protein intermediates and a reduction of non-specific interactions. Although for Cdc25C an increase in thermal stability was not noted with L-arginine an increase in protein solubility was noted. It would be interesting to obtain SAXS data without the additives sucrose and L-arginine to see if there is indeed protein compaction. However, the solubility of the Cdc25C catalytic domain will need to be considered. On site purification at the synchotron facility may allow the collection of good quality data.

Although the solubility of the Cdc25C catalytic domain was significantly improved this was not sufficient to obtain backbone assignments. Additionally, it was noted that the Cdc25C catalytic domain backbone may be dynamic which could also have played a role in the lack of success in assigning the backbone. Interestingly, a recent study published last month has assigned the backbone of the Cdc25B catalytic domain (Lund and Cierpicki, 2014). The authors also found it difficult to obtain assignments of the native protein Cdc25B and used a mutation approach to stabilize the protein. They were able to stabilize the protein by mutating the active site cysteine. A similar approach could be tested for the Cdc25C catalytic domain.

Chapter 5.0 – Identifying inhibitors for Cdc25C

5.1 Introduction

The twenty compounds which were tested by NMR are listed in Table A6 (Appendix A6) which also contains their structure, molecular formula, molecular weight, and predicted LogP values. The LogP values were obtained from the molinspiration cheminformatics website (http://www.molinspiration.com/cgi-bin/properties). LogP is a partition co-efficient of octanol/water and measures hydrophobicity of molecules. The predicted LogP was used to short list the compounds where the compounds which had a LogP higher than 2.0 were not selected.

A few of the compounds included in the Table A6 (Appendix) have already been shown to interact with Cdc25C or are general phosphatase inhibitors. Sodium orthovanadate and BVT.948 are general protein tyrosine phosphatase inhibitors which have been shown to inhibit phosphatases competitively or noncompetitively respectively (Gordon, 1991; Liljebris et al., 2004). NSC 663284 and NSC 95397 have previously been shown to inhibit Cdc25C via phosphatase activity assays (Lazo et al., 2001). The rest of the compounds were selected based on insights gained from the literature. This study assessed the potential for NMR spectroscopy in identifying compound hits for the catalytic domain of Cdc25C.

5.2 Effect of DMSO on the Cdc25C catalytic domain

DMSO was used here to solubilise the compounds so that high concentration stock solutions could be prepared and also to ensure the compounds remained in solution during the experiment. Although DMSO is a popular solvent which is generally used to prepare stock solutions of compounds it has been shown for some proteins to negatively affect protein structure and solubility (Arakawa et al., 2007b; Tjernberg et al., 2006). Since the compounds tested here were prepared in DMSO it was important to investigate the effect of increasing

DMSO concentrations on the Cdc25C catalytic domain before NMR interaction studies commenced.

It can be seen from the ^{1}H, ^{15}N-HSQC spectra overlay of the Cdc25C catalytic domain with 0 % DMSO compared with the DMSO concentrations 5, 10, and 15 % v/v that the protein is folded in all these concentrations (Figure 5.2a). In these DMSO concentrations the peaks are well dispersed and can be mapped to the peaks with 0 % DMSO. A significant number of peaks show chemical shift perturbations suggesting the catalytic domain of Cdc25C interacts with DMSO. These chemical shift changes are in fast exchange indicating a weak interaction (see arrows on spectrum). Since DMSO has been shown to form favourable interactions with hydrophobic amino acids it is likely that the peaks which show chemical shift perturbations are hydrophobic residues (Arakawa et al., 2007b).

The ThermoFluor® assay was used to test the thermal stability of the Cdc25C catalytic domain with DMSO concentrations up to 20 % (Figure 5.2b). The thermal stability did not significantly change until 8 % of DMSO was added which resulted in a decrease in thermal stability by 1 ^{o}C. There was no further significant decrease in thermal stability. In addition, the solubility of the Cdc25C catalytic domain was tested with DMSO concentrations up to 20 % and it was noted that there was no negative effect on protein solubility (the Cdc25C catalytic domain also retained activity in the presence of DMSO). Therefore, the small decrease in thermal stability did not affect the solubility of the Cdc25C catalytic domain and it is hence stable in high concentrations of DMSO.
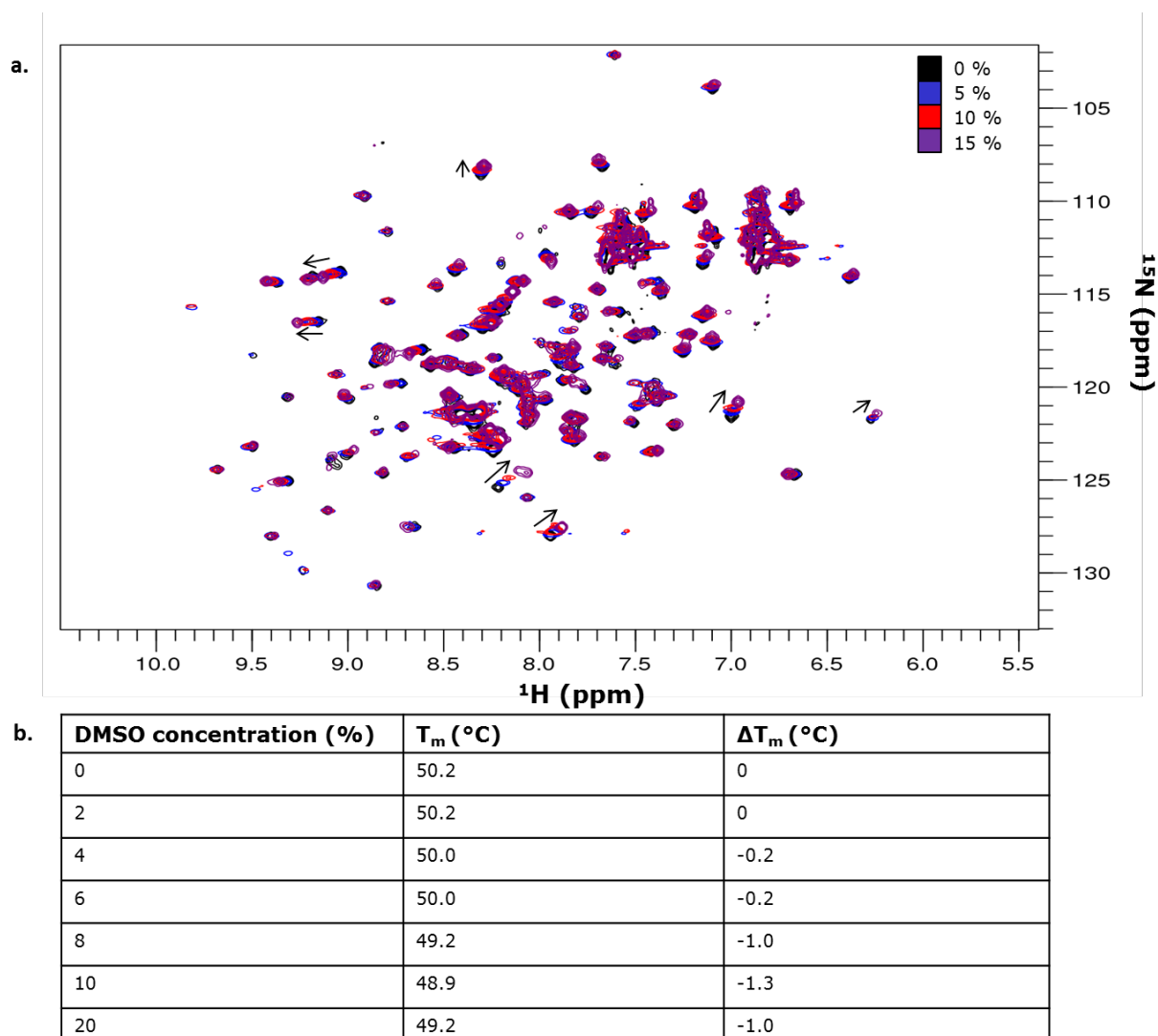
Figure 5.2 Effect of DMSO on the Cdc25C catalytic domain

A ^{1}H, ^{15}N SOFAST-HMQC overlay of 200 μM Cdc25C catalytic domain in 0, 5, 10, and 15 % v/v DMSO concentrations in the buffer 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP (a). The ^{1}H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using a 800 MHz Agilent spectrometer. The Table (b) shows the effect of increasing DMSO concentrations on the thermal stability of the Cdc25C catalytic domain. The T_m and ΔT_m are tabulated for each DMSO concentration.

5.3 WaterLOGSY results

DMSO concentration (%)	T_m (°C)	ΔT_m (°C)
0	50.2	0
2	50.2	0
4	50.0	-0.2
6	50.0	-0.2
8	49.2	-1.0
10	48.9	-1.3
20	49.2	-1.0

All the compounds except for sodium orthovanadate were first tested by WaterLOGSY to identify compounds which interact with the Cdc25C catalytic domain (Appendix A6). Sodium orthovanadate was not tested because this inhibitor compound has no protons and therefore is not applicable for this assay. The compounds which showed any changes via WaterLOGSY were compounds 2, 3, 4, 14, 15, 16, and 19. The compounds 4, 14, 15, and 16 gave the best WaterLOGSY results and are described here.

A positive phase change upon addition of the Cdc25C catalytic domain protein indicated bound protons suggesting an interaction. Although for some of the peaks in the WaterLOGSY spectra there was not a clear change in phase changes in peak intensities were observed upon the addition of protein.

For compound 4 (BVT.948) there were changes in peak intensities observed when the Cdc25C catalytic domain was added. One resonance displayed a clear phase change (Figure 5.3.1). This proton resonance had a chemical shift value of 7.81 ppm. Three resonances, 6.87, 7.27, and 7.30 ppm showed a positive phase change for compound 14 (Figure 5.3.2). For compound 15 eight out of nine proton resonances displayed a clear phase change (Figure 5.3.3). These proton resonances had the chemical shifts of 7.71, 7.73, 7.79, 7.8, 7.81, 7.99, 8.01, and 8.03 ppm. Finally, all the proton resonances of compound 16 had a phase change after addition of the protein (Figure 5.3.4). The chemical shifts of these resonances were 6.35, 7.3, 7.4, 7.5, 7.75, 8.53, and 8.80 ppm.
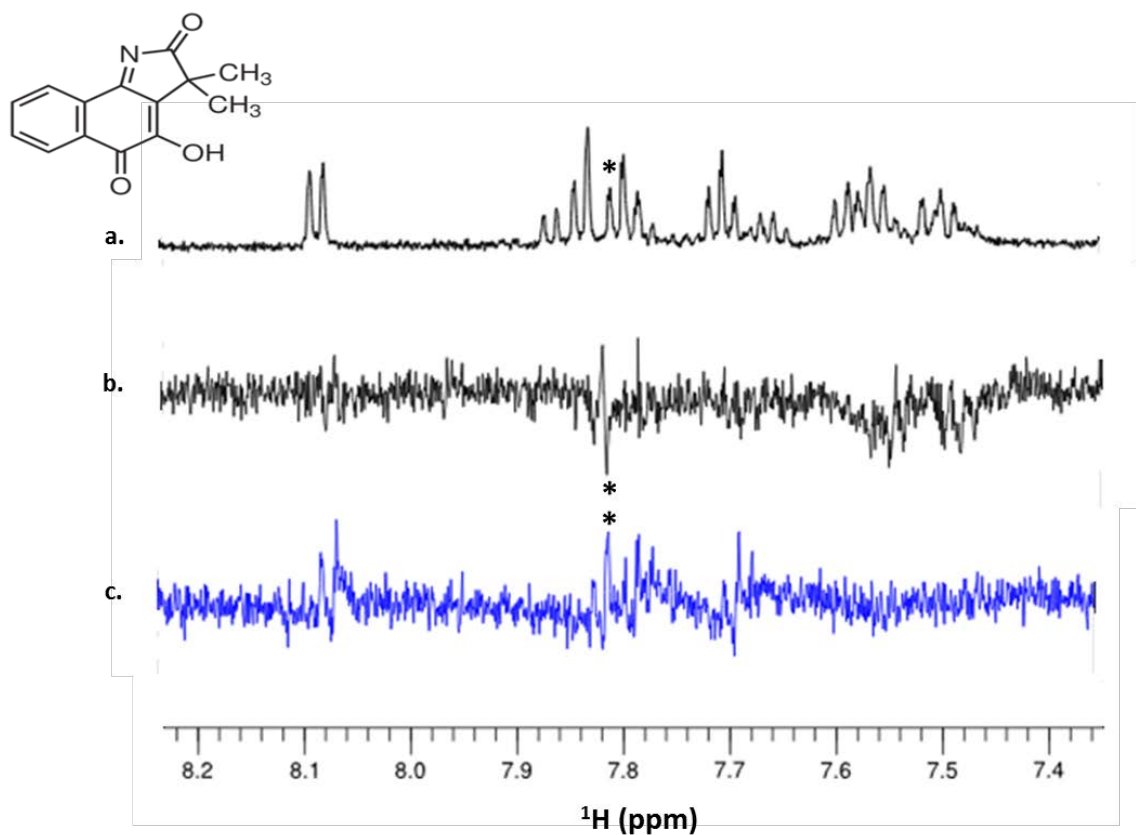
Figure 5.3.1 Compound 4: BVT.948

The proton 1D of the compound is shown (a), followed by the WaterLOGSY with no protein (b), and the acquired WaterLOGSY after the addition of 5 μM protein (c). The resonances which showed phase changes are marked by an asterisk. The Bruker 600 MHz BACS60 autosampler was used to collect the data.
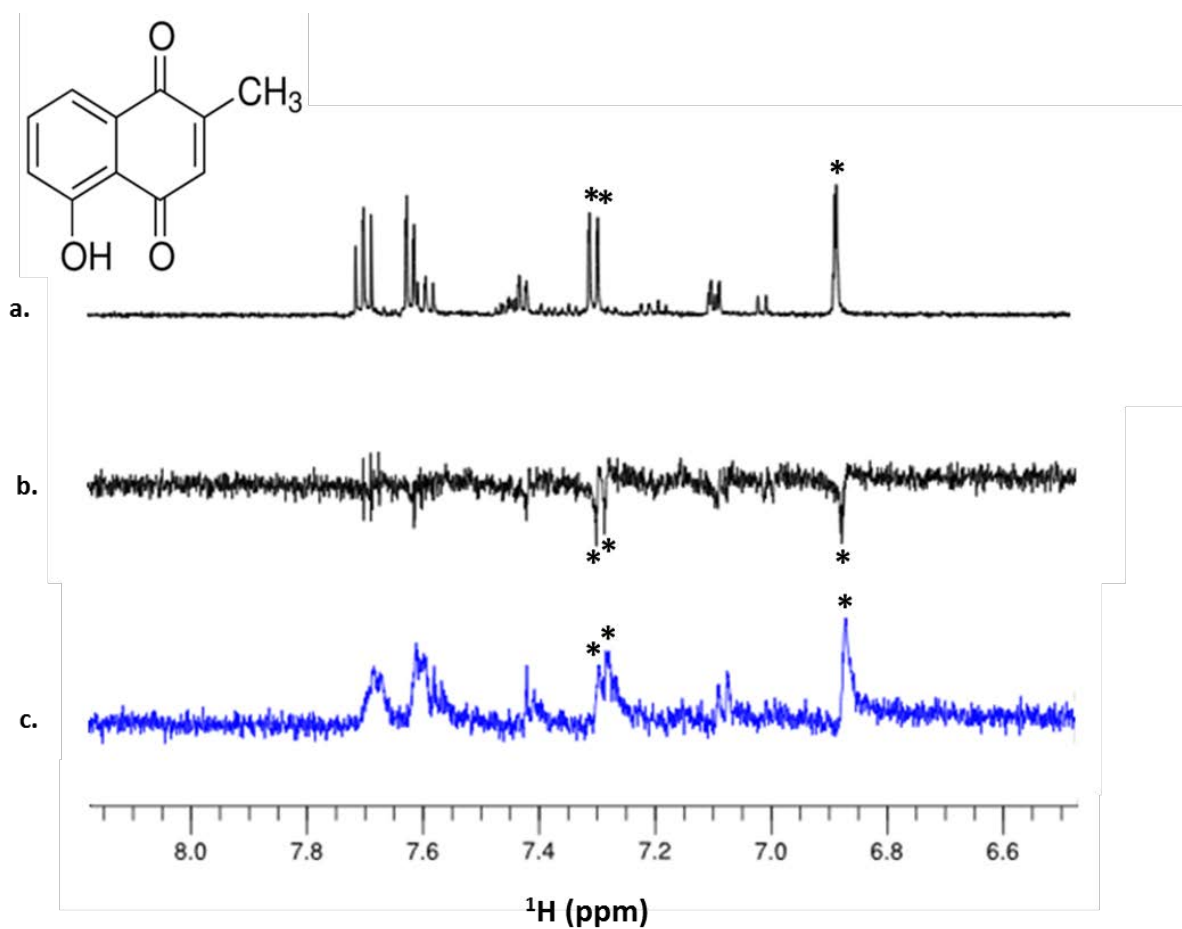
Figure 5.3.2 Compound 14

The proton 1D of the compound is shown (a), followed by the WaterLOGSY with no protein (b), and the acquired WaterLOGSY after the addition of 5 μM protein (c). The resonances which showed phase changes are marked by an asterisk. The Bruker 600 MHz BACS60 autosampler was used to collect the data.
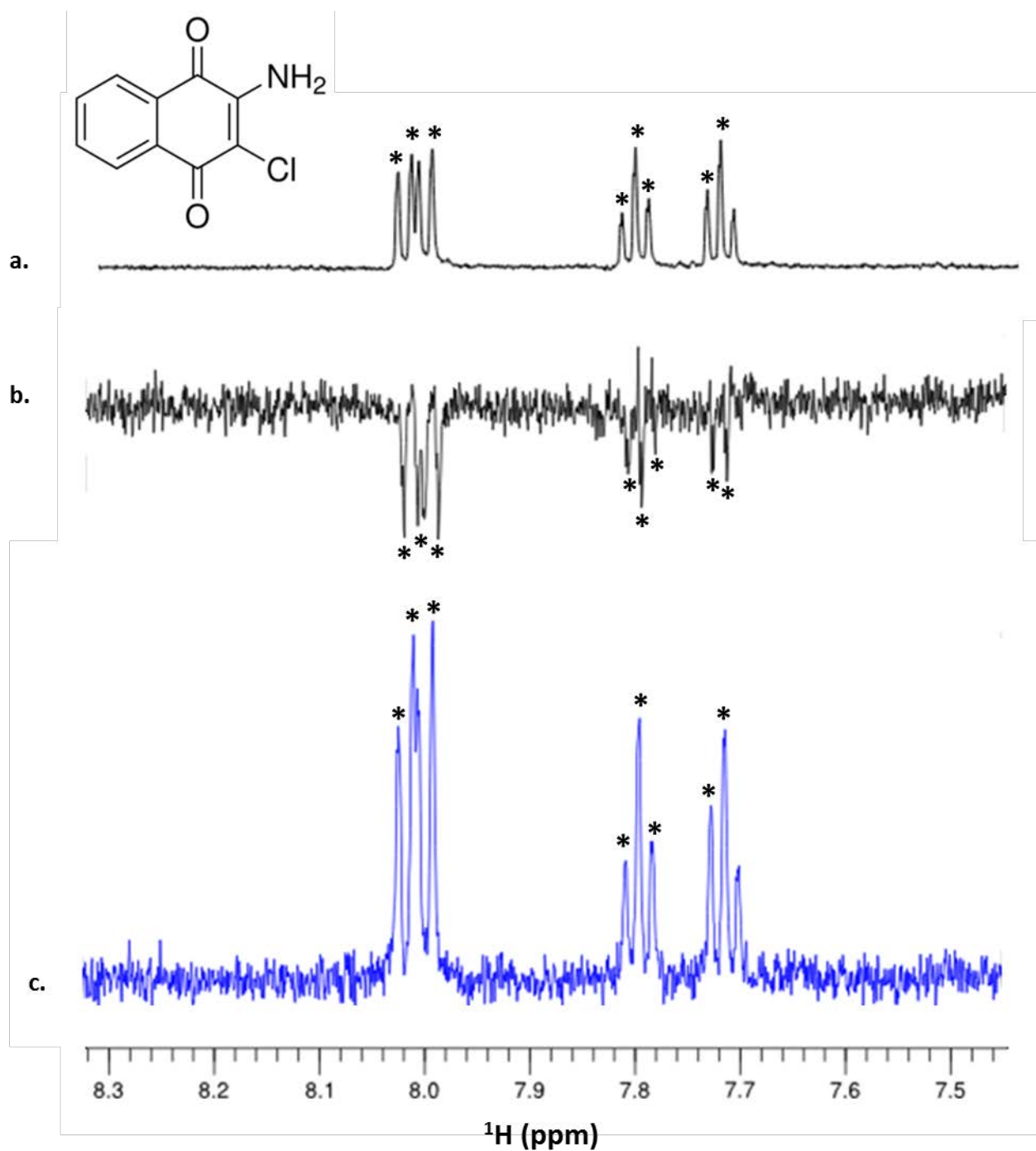
Figure 5.3.3 Compound 15

The proton 1D of the compound is shown (a), followed by the WaterLOGSY with no protein (b), and the acquired WaterLOGSY after the addition of 5 µM protein (c). The resonances which showed phase changes are marked by an asterisk. The Bruker 600 MHz BACS60 autosampler was used to collect the data.
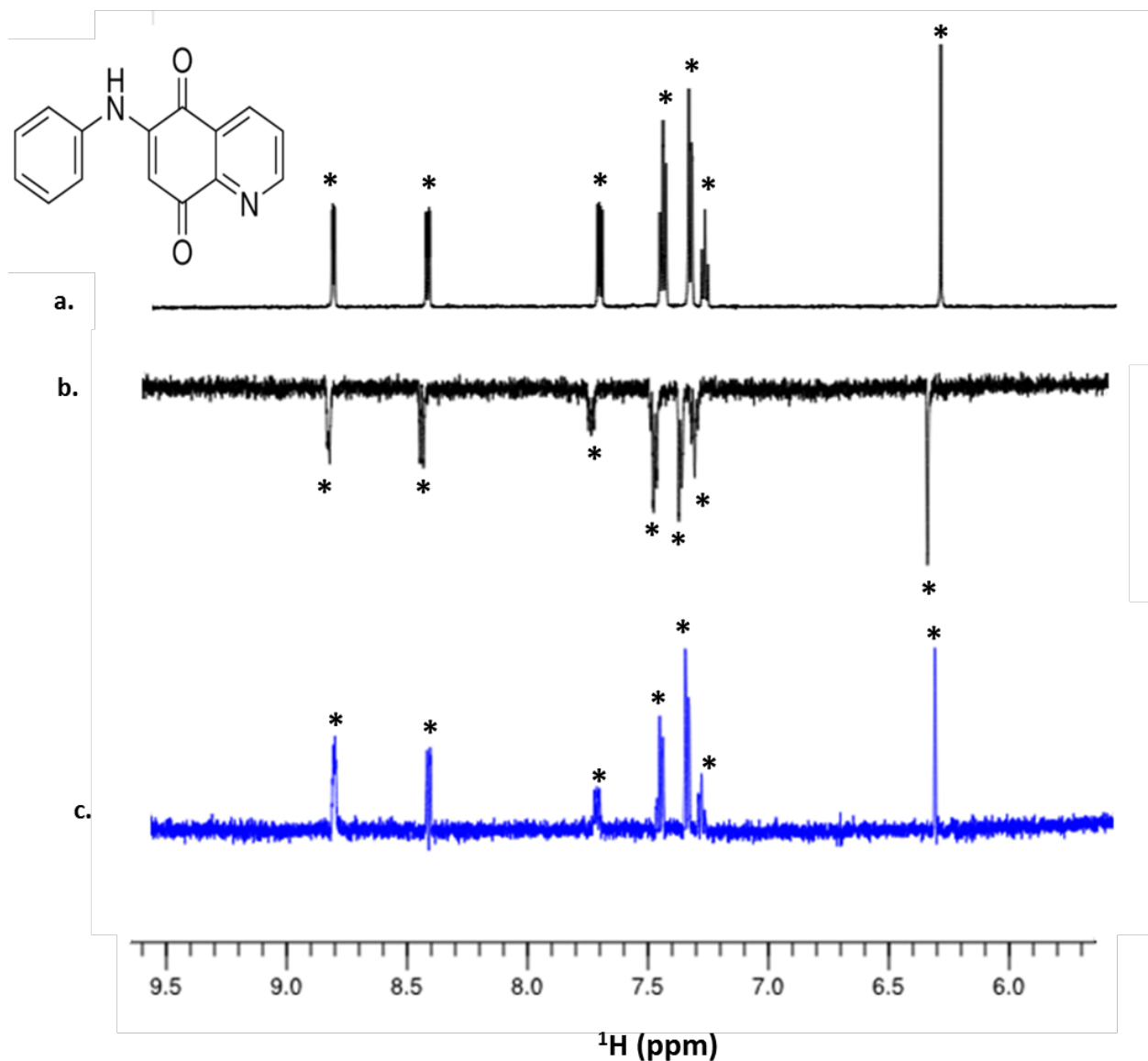
Figure 5.3.4 Compound 16

The proton 1D of the compound is shown (a), followed by the WaterLOGSY with no protein (b), and the acquired WaterLOGSY after the addition of 5 μM protein (c). The resonances which showed phase changes are marked by an asterisk. The Bruker 600 MHz BACS60 autosampler was used to collect the data.

5.4 2D results (^{1}H, ^{15}N SOFAST-HMQCs)

All the compounds which showed changes via WaterLOGSY were also tested by ^{1}H, ^{15}N SOFAST-HMQC experiments (compounds 2, 3, 4, 14, 15, 16, and 19 from Table A6). The chemical shift changes were compared to the general phosphatase inhibitors, BVT.948 (C4), and sodium orthovanadate (C1). The compounds were added in excess with minimum concentrations of 0.4 or 0.5 mM. An exception was the compound NSC 95397 (C3). This compound had poor solubility so it was only tested at 0.3 mM concentration.

5.4.1 Compound 4: BVT.948

There was a widespread loss of peaks observed. A number of peaks also showed weakening signal intensities. Therefore, these chemical shift perturbations are in intermediate exchange. This explains the weakening and loss of peaks upon the addition of the non-competitive phosphatase inhibitor BVT.948. The peaks which do not change are from regions which are not affected by the compound.

5.4.2 Compounds 2, 3, and 14

The compounds 2 (NSC 663284, Figure 5.4.2.1), 3 (NSC 95397, Figure 5.4.2.2), and 14 (Figure 5.4.2.3) resulted in a similar change in the 2D HMQC spectrum as BVT.948. These compounds resulted in a significant loss of peaks in the 2D HMQC spectrum. In addition to the significant loss of peaks there was a weakening of peaks. This suggested the peaks were in intermediate exchange indicating a strong interaction. A change in a significantly large number of chemical shifts implied a global conformational change of the protein upon binding of these compounds. The peaks which did not change are likely to be peaks which do not interact with the compounds or are not affected by them.
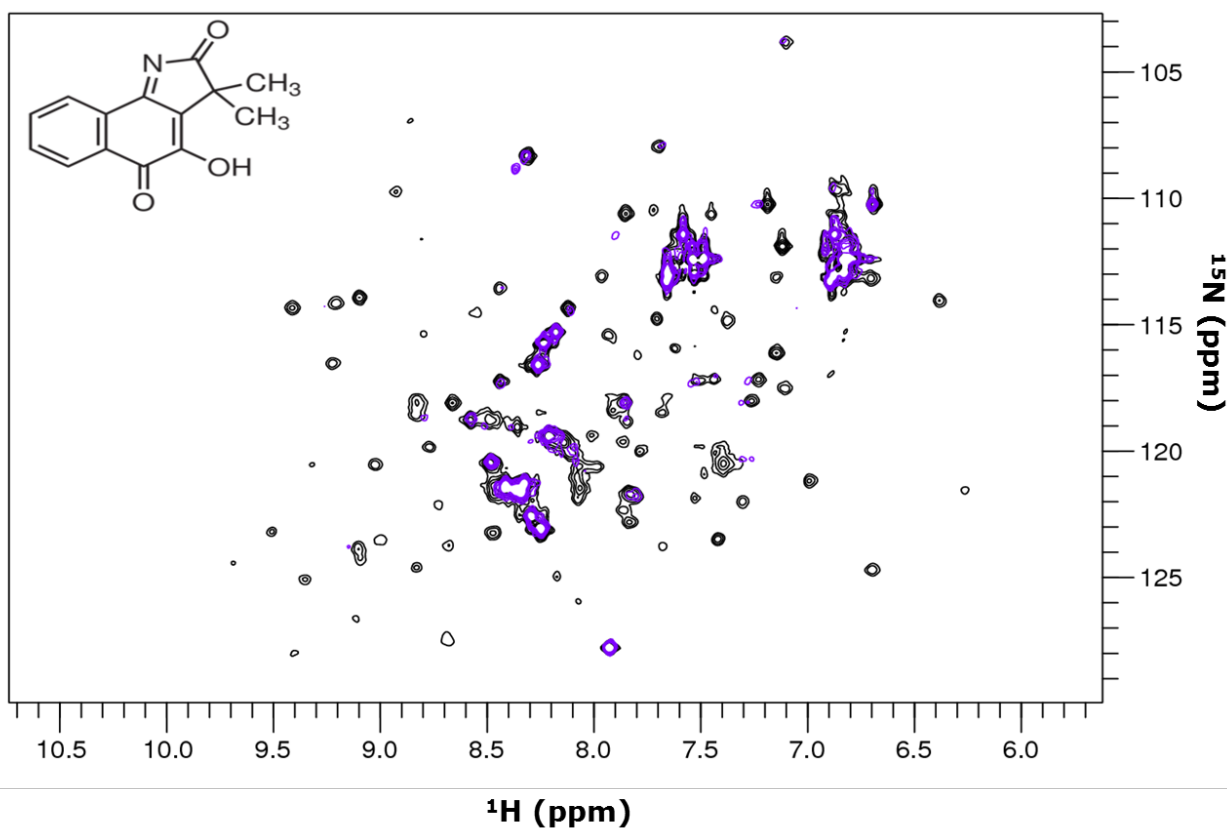
Figure 5.4.1 Compound 4:BVT.948

A ^{1}H, ^{15}N SOFAST-HMQC overlay of 200 μM Cdc25C catalytic domain in 10 % DMSO-d$_6$ (black) and 0.4 mM BVT.948 (purple) in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP. The ^{1}H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using an 800 MHz Agilent spectrometer.
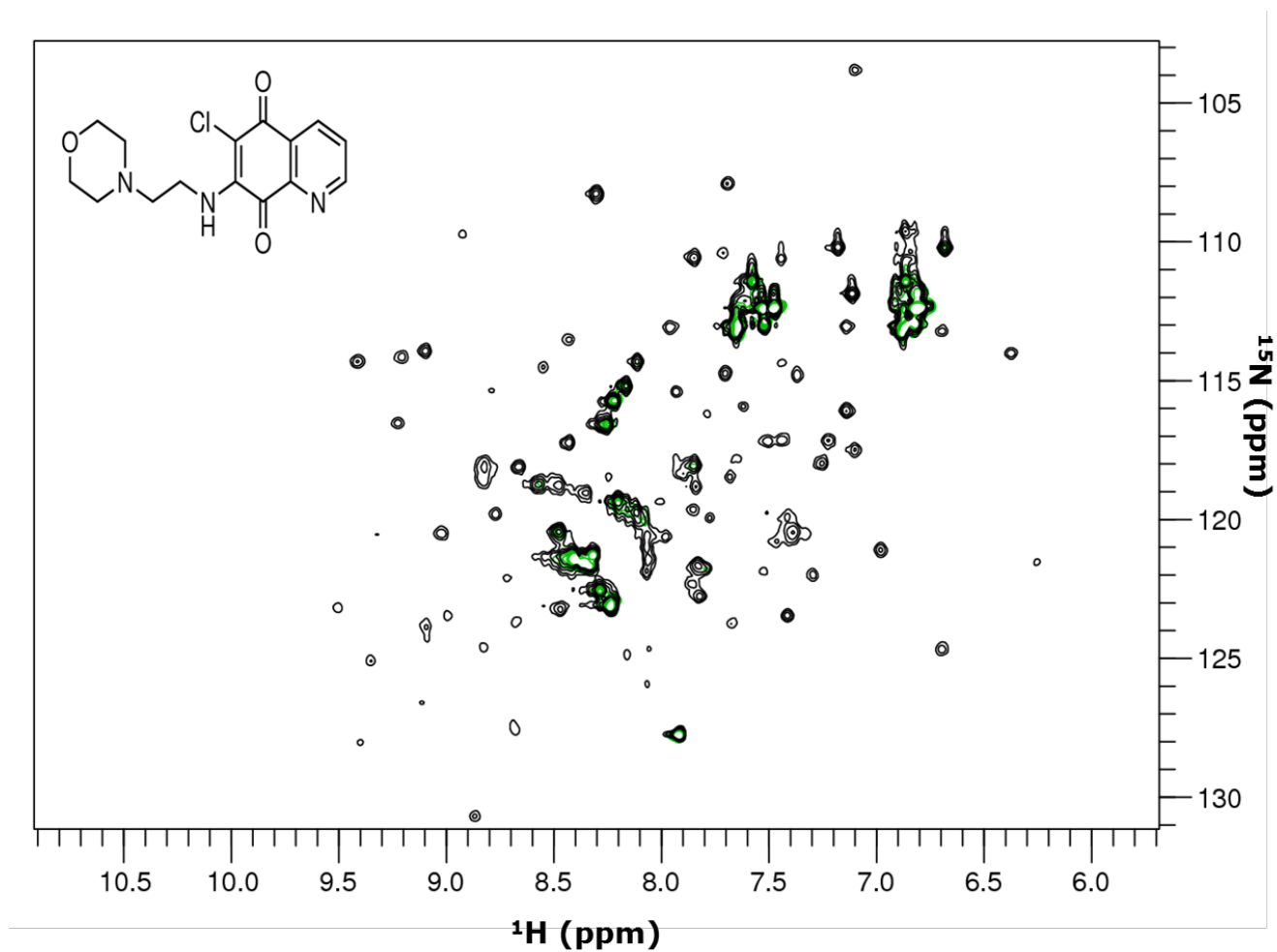
Figure 5.4.2.1 Compound 2: NSC 663284

A ^{1}H, ^{15}N SOFAST-HMQC overlay of 200 μM Cdc25C catalytic domain in 10 % DMSO-d$_6$ (black) and 0.4 mM NSC 663284 (green) in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP. The ^{1}H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using an 800 MHz Agilent spectrometer.
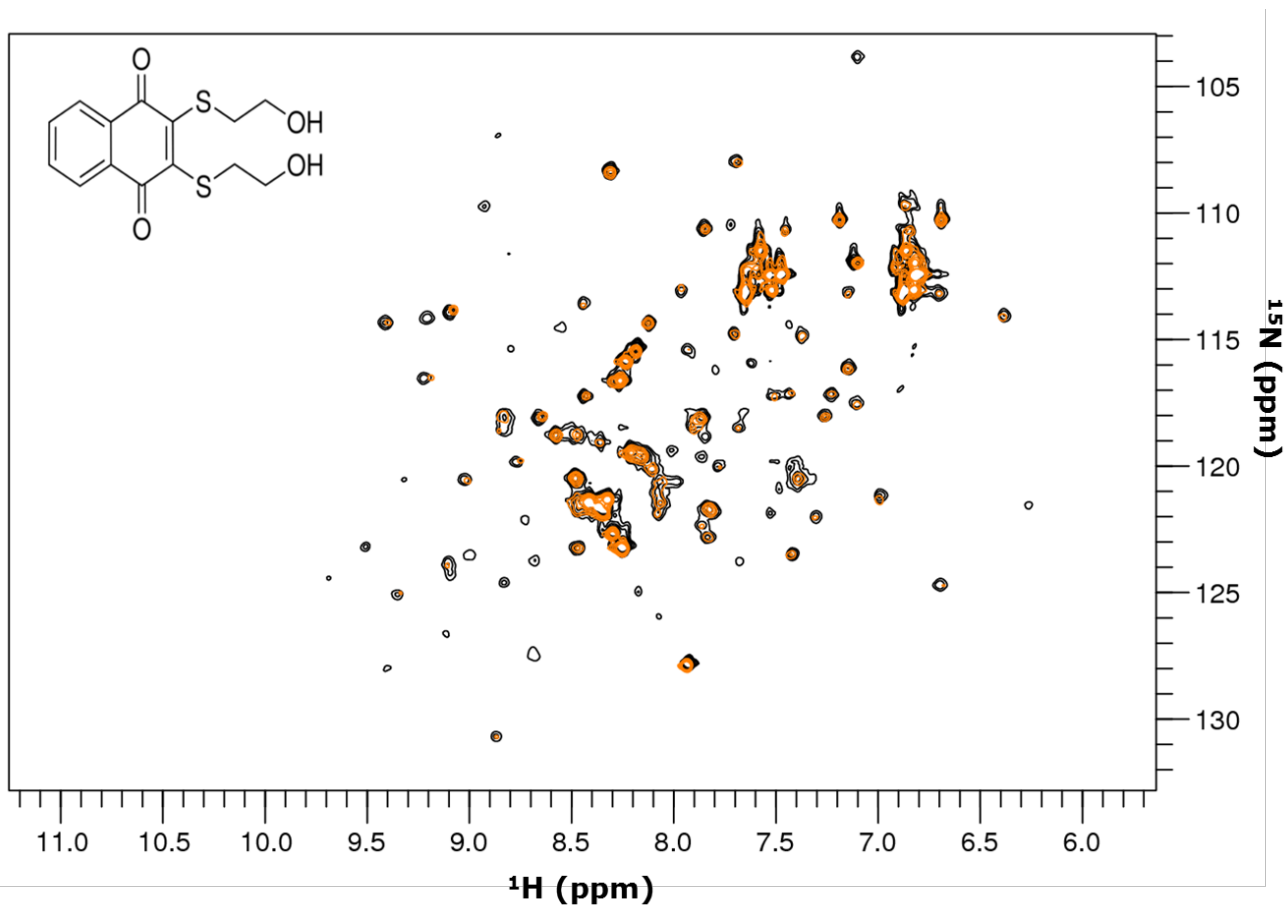
Figure 5.4.2.2 Compound 3: NSC 95397

A ^{1}H, ^{15}N SOFAST-HMQC overlay of 200 μM Cdc25C catalytic domain in 10 % DMSO-d$_6$ (black) and 0.3 mM NSC 95397 (orange) in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP. The ^{1}H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using an 800 MHz Agilent spectrometer.
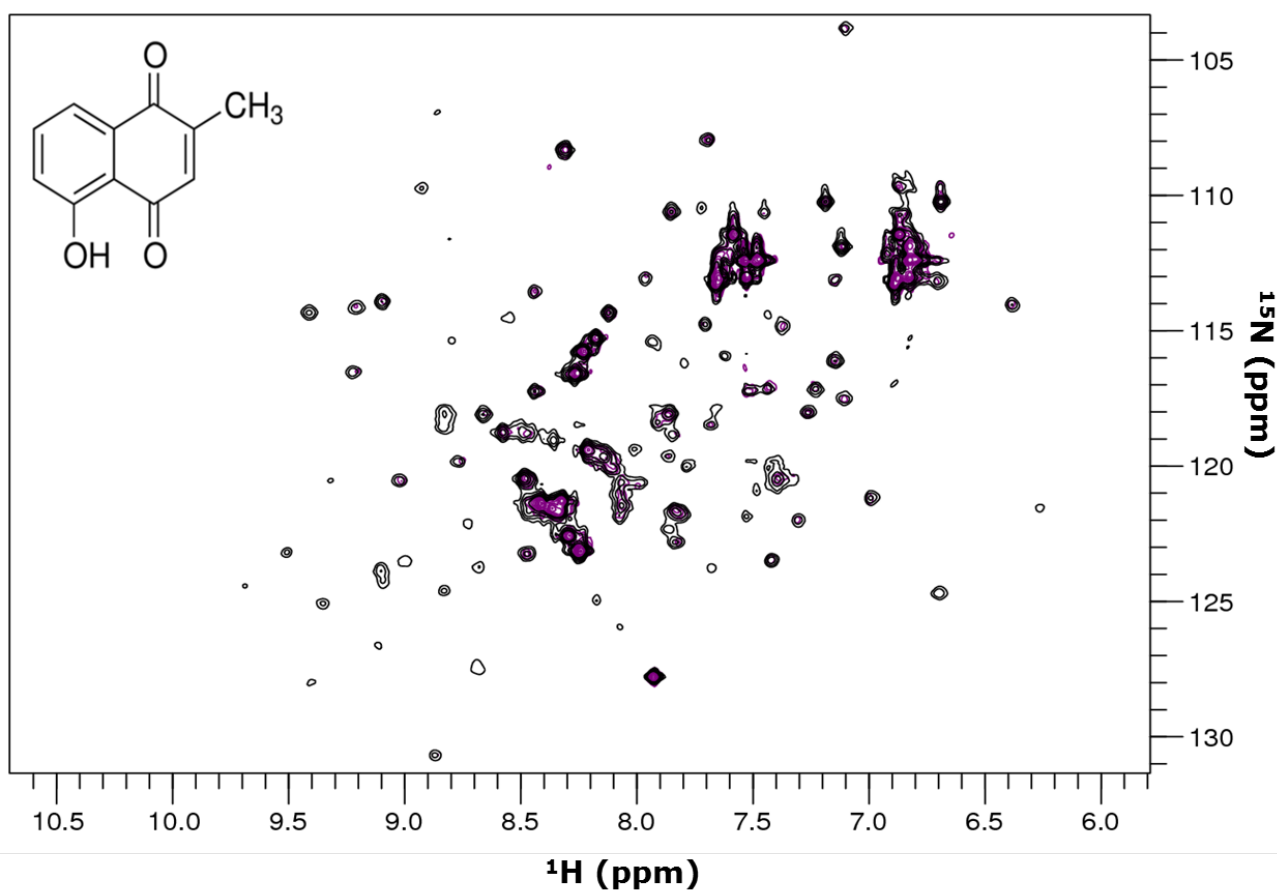
Figure 5.4.2.3 Compound 14

A ^{1}H, ^{15}N SOFAST-HMQC overlay of 200 µM Cdc25C catalytic domain in 10 % DMSO-d$_6$ (black) and 0.5 mM compound 14 (purple) in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP. The ^{1}H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using an 800 MHz Agilent spectrometer.

5.4.3 Compound 1: Sodium Orthovanadate

Sodium Orthovanadate is a general phosphatase inhibitor. It has a structure which mimics a phosphate group and therefore can bind competitively in the active site of phosphatases. The addition of sodium orthovanadate resulted in stronger peak intensities compared with no inhibitor. There were small chemical shift perturbations seen that displayed fast exchange characteristics which suggested a weak interaction with the catalytic domain of Cdc25C. Compounds 15, 16, and 19 showed a similar effect as sodium orthovanadate.

5.4.4 Compound 15

Peaks showing chemical shift perturbations were seen which were localised to two regions on the 2D HMQC spectrum. These chemical shift perturbations followed the fast exchange dynamic suggesting a weak interaction. New peaks, weak in intensity, also appeared. These peaks may be from residues at the interaction site.

5.4.5 Compound 16

Weak chemical shift perturbations were seen that followed a fast exchange phenomenon. They were localised to two regions on the NMR spectrum. Interestingly, one of these regions was the same region noted for compound 15.

5.4.6 Compound 19

This compound also displayed weak chemical shift perturbations. A few of these were also localised to the same region which was observed for the compounds 15 and 16. There were also additional chemical shift perturbations as well as new peaks observed (arrows). The new peaks could be from the binding site. The peak intensities overall were stronger with the compound compared to without.
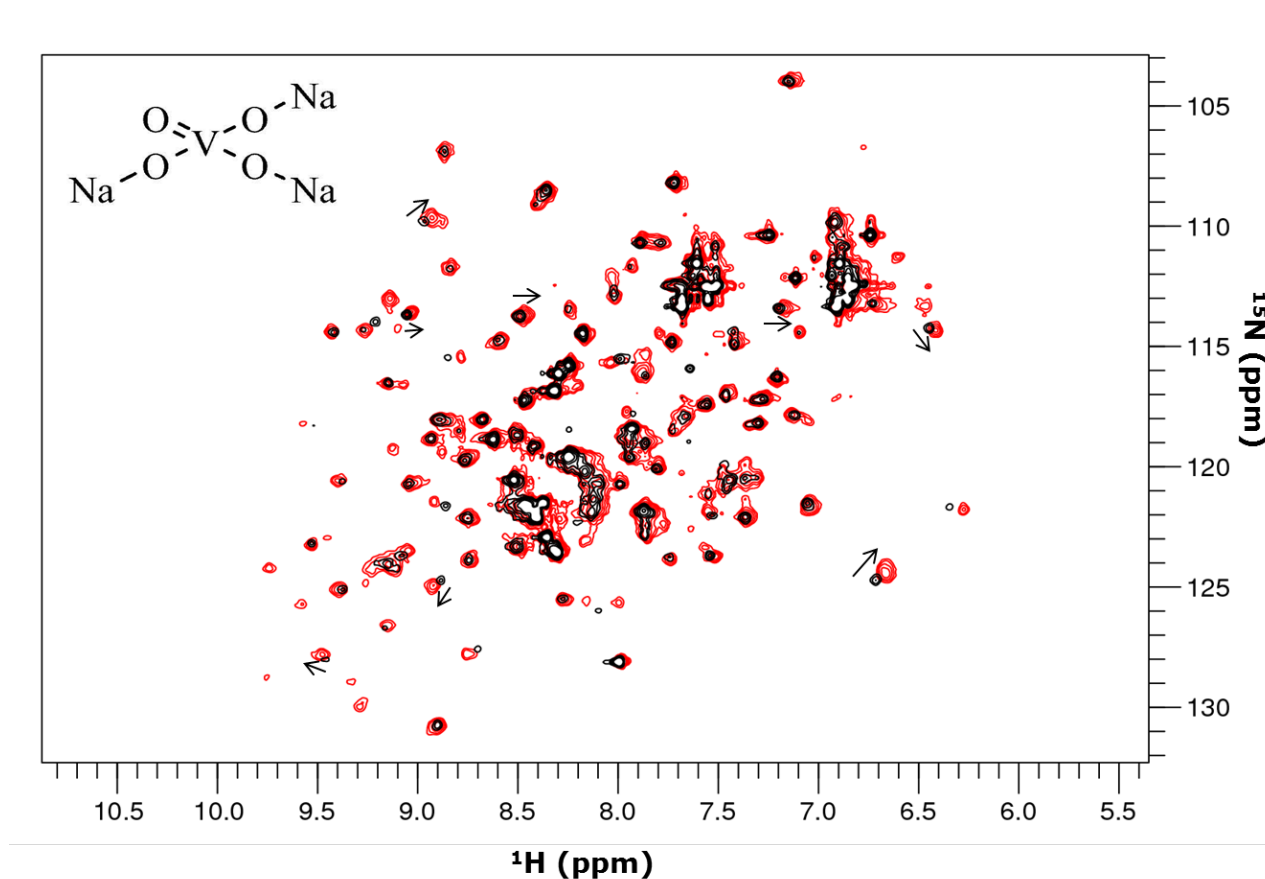
Figure 5.4.3 Compound 1: Sodium Orthovanadate

A ^{1}H, ^{15}N SOFAST-HMQC overlay of 200 µM Cdc25C catalytic domain (black) and 0.5 mM sodium orthovanadate (red) in 50 mM sodium phosphate (pH 7.0), 150 mM sodium chloride, and 1 mM TCEP. The ^{1}H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using a 600 MHz Agilent spectrometer. Overall, weak chemical shift perturbations were observed and some of these are marked by black arrows.

Figure 5.4.4 Compound 15

A ^1H, ^{15}N SOFAST-HMQC overlay of 200 μM Cdc25C catalytic domain in 15 % DMSO-d$_6$ (black) and 1 mM compound 15 (blue) in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP (a). The ^1H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using an 800 MHz Agilent spectrometer. The chemical shift perturbations were localised to two regions (b + c). Also new weak peaks appeared which are shown by black arrows.
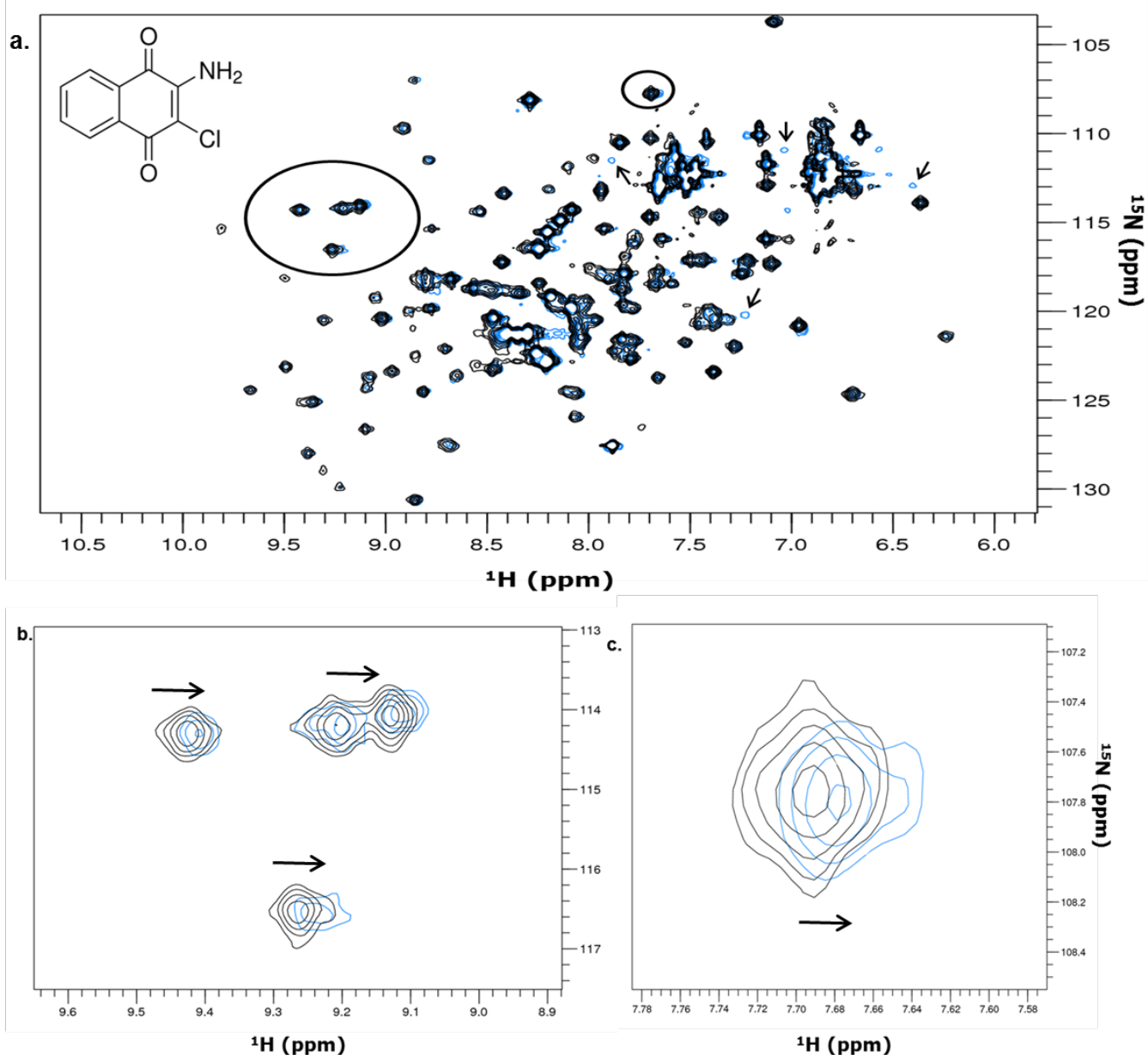
Figure 5.4.5 Compound 16

A ^1H, ^{15}N SOFAST-HMQC overlay of 200 µM Cdc25C catalytic domain in 15 % DMSO-d_6 (black) and 1 mM compound 16 (pink) in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP (a). The ^1H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^oC using an 800 MHz Agilent spectrometer. The chemical shift perturbations were localised to two regions (b + c).
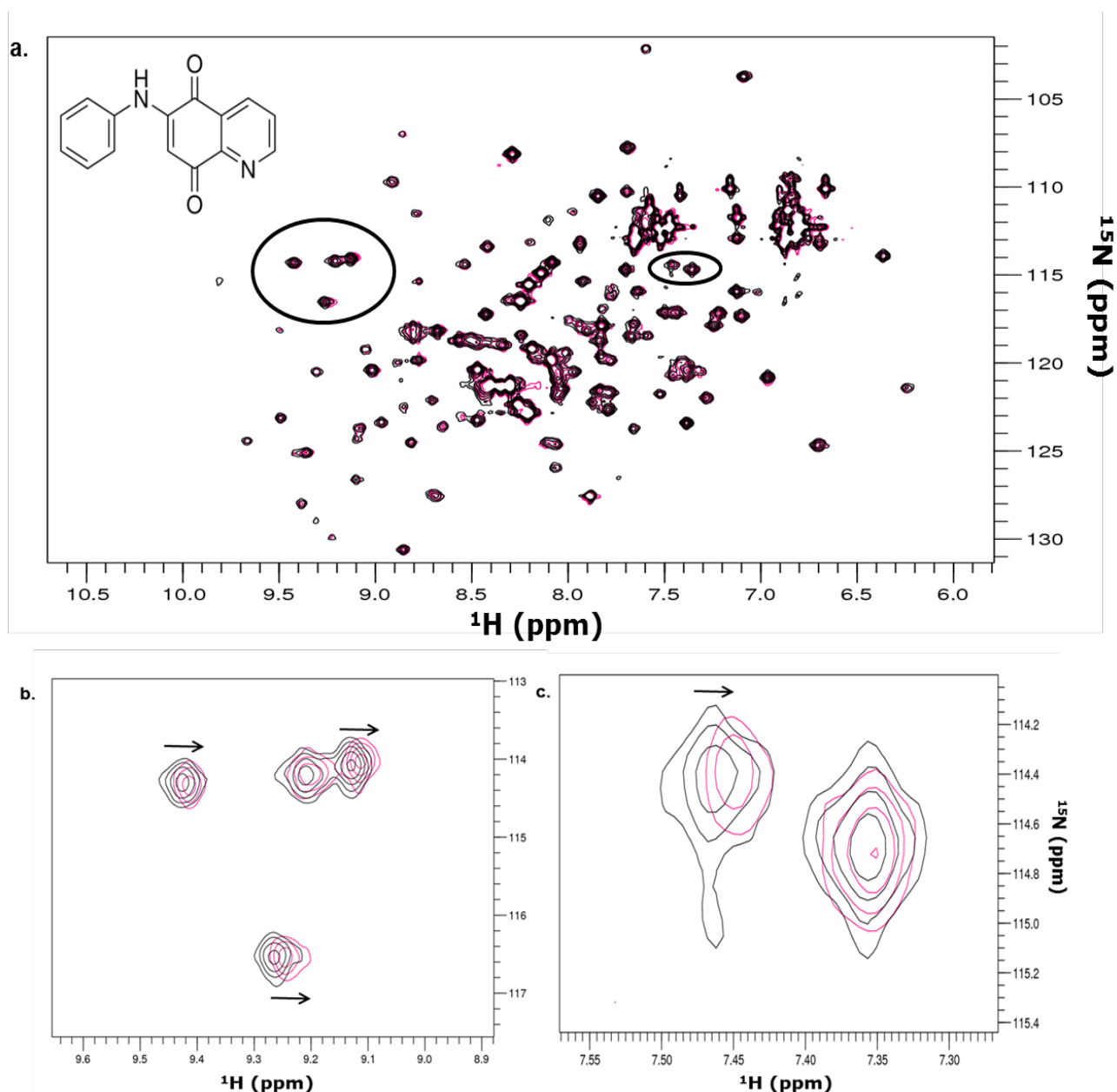
Figure 5.4.6 Compound 19

A ^{1}H, ^{15}N SOFAST-HMQC overlay of 200 μM Cdc25C catalytic domain in 10 % DMSO-d$_6$ (black) and 0.5 mM compound 19 (brown) in 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine 200 mM sucrose, and 1 mM TCEP (a). The ^{1}H, ^{15}N SOFAST-HMQC spectra were collected at 25 ^{o}C using an 800 MHz Agilent spectrometer.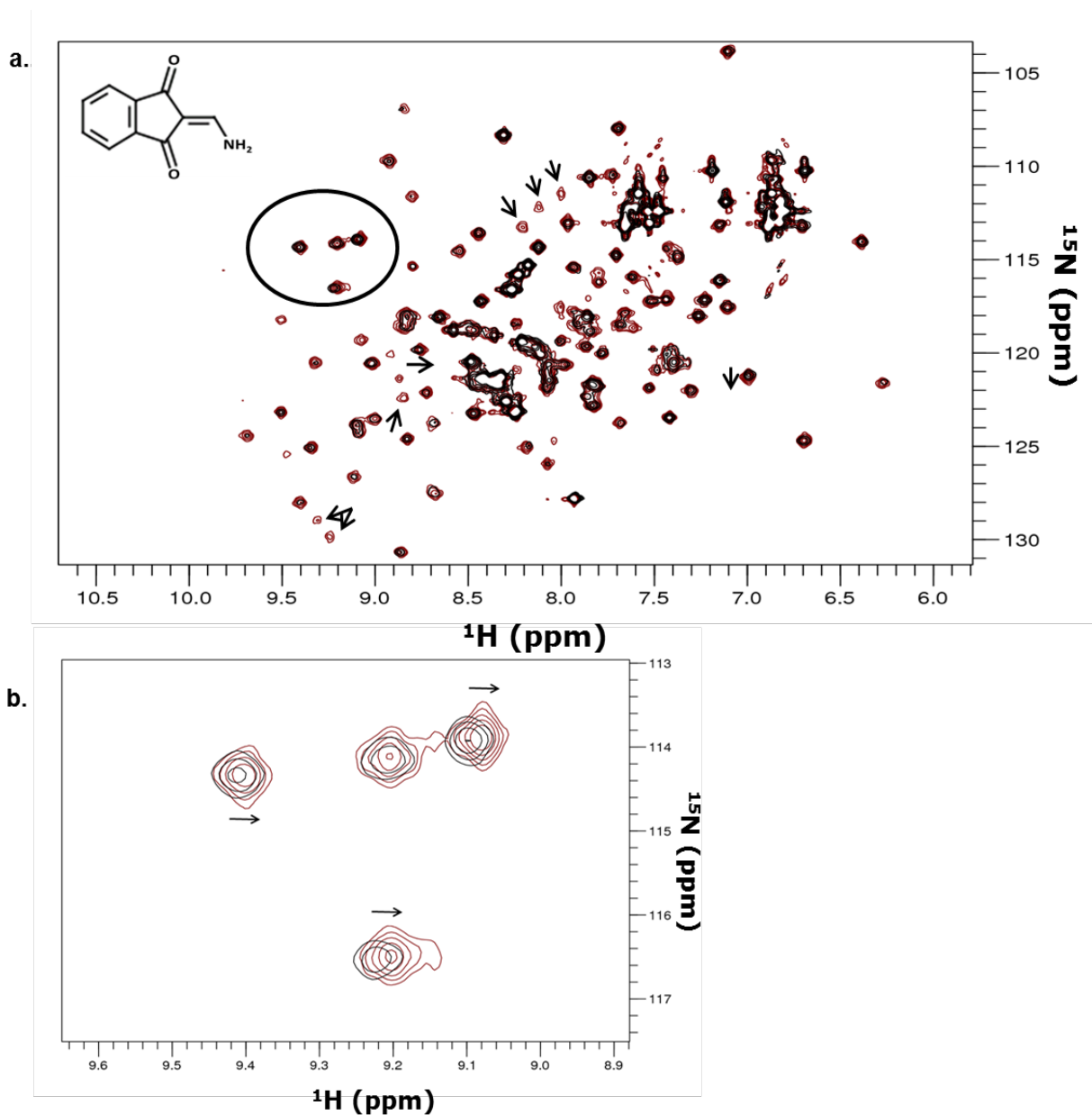 The chemical shift perturbations localised to the region observed for compounds 15 and 16 (b). Additional chemical shift perturbations and new peaks were observed which are shown by black arrows.

5.5 Crystallisation of the Cdc25C catalytic domain

The Cdc25C catalytic domain proved to be refractory to backbone assignments therefore crystallography was tried as an alternative method for determining its structure. The aim was to crystallize the Cdc25C catalytic domain with the inhibitors which showed binding via NMR spectroscopy. Although, the structure of the catalytic domain of all the Cdc25 homologues has previously been determined, none has been crystallised in complex with an inhibitor or ligand. The aim was to first crystallise the Cdc25C catalytic domain and then soak in the inhibitor.

The purified Cdc25C catalytic domain was dialyzed into 50 mM sodium phosphate (pH 7.0), 200 mM L-arginine, 200 mM sucrose, and 1 mM TCEP. The Hepes buffer was selected in place of sodium phosphate because the phosphate exhibits a greater propensity to cross-react with metal ions present in commercial crystallisation reagents and yield false positive hits in the form of salt crystals.

Protein concentrations 5, 10, 20, and 24 mg/ml were tested with all the commercial crystallization screens listed in the methods. Initially, the hanging-drop screens were set up with the 24 mg/ml concentration at 23 ^{o}C. Inspection of the plates following several days of incubation revealed heavy precipitation in the majority of drops indicative of high protein concentrations. Hence, subsequent crystallisation experiments were performed at 23 ^{o}C with reduced protein concentrations. In addition, all screens were repeated for the 20 mg/ml concentration at 4 ^{o}C. After several weeks, crystals appeared in a range of protein concentrations (10, 20, and 24 mg/ml) at 23 ^{o}C. No crystals appeared in the drops incubated at 4 ^{o}C. However, all these crystals appeared in a single condition from the PEGRx II screen which comprised of 0.2 M magnesium chloride hexahydrate, 0.1 M sodium citrate tribasic dehydrate, pH 5.0, 10 % w/v polyethylene glycol (PEG) 20, 000 (Figure 5.5.1a). These crystals were square

shaped with some depth. At this stage no crystals were visible in drops containing the buffer alone.

An attempt was then made to optimise the conditions to obtain larger diffraction-grade crystals. This was carried out by increasing the protein:buffer drop volumes (400:400 nl and 1:1 µl) and reducing the concentration of PEG in the buffer (10 – 1 %). However, this proved unsuccessful. Also, a few weeks after the crystals had been identified, the control drop with buffer alone contained crystals (Figure 5.5.1b). These crystals were of similar morphology to previously observed crystals suggesting that they comprised of salt rather than protein.

Since the Cdc25C catalytic domain was methylated prior to crystallization to generate the crystal which is deposited in the protein database (PDB:3OP3) it was thought that methylating the protein could improve the number of crystal hits. Therefore, the Cdc25C catalytic domain was methylated (Figure 5.5.2).

The reductive methylation chemical reaction results in the dimethylation of the amine group of lysines as well as the N-terminal amino acid. This results in an increase of 28 Da per residue methylated. Since the Cdc25C catalytic domain (Cdc25C$_{270-443}$) contains ten lysines a 300 Da increase in size (includes the N-terminal amino acid) was expected. The mass spectrometry (Figure 5.4.2) data show that the unmethylated Cdc25C catalytic domain has a molecular weight of 20.6 kDa whereas the methylated protein had a molecular weight of 20.9 kDa. This indicated that the methylation reaction for the Cdc25C catalytic domain was successful.

The methylated protein was then dialyzed into 20 mM HEPES pH 8.0, 250 mM sodium chloride, and 2 mM TCEP in readiness for crystallisation trials. This buffer was used for the previous crystallisation of the Cdc25C catalytic domain (SGC, Oxford). The Cdc25C catalytic domain (5, 10, 20, and 24 mg/ml) was screened against commercial crystallisation screens and incubated

at 23 °C. Crystals again only appeared in the same single condition from the PEGRx II screen. These crystals appeared after two weeks and crystals were also present in the control drops suggesting that they comprised of salt.

The condition in which Cdc25C has already been crystallized (0.1 M Bis Tris, pH 5.5, 0.2 M ammonium sulphate, 23 % PEG 3350) was also tested. The concentration of the crystallisation reagents was also varied. Unfortunately, despite extensive efforts, there were no crystallisation hits observed.

The final strategy consisted of co-crystallising Cdc25C protein with an inhibitor. Combining protein with ligand has been an attractive strategy for generating crystals as often the complex is more stable than protein alone and therefore increasingly amenable to crystallisation. With this in mind the inhibitor NSC 663284 (at 0.5 mM) was incubated with methylated Cdc25C protein (10 mg/ml) and screened against several commercial screens at 23 °C. This also resulted in no crystal hits.
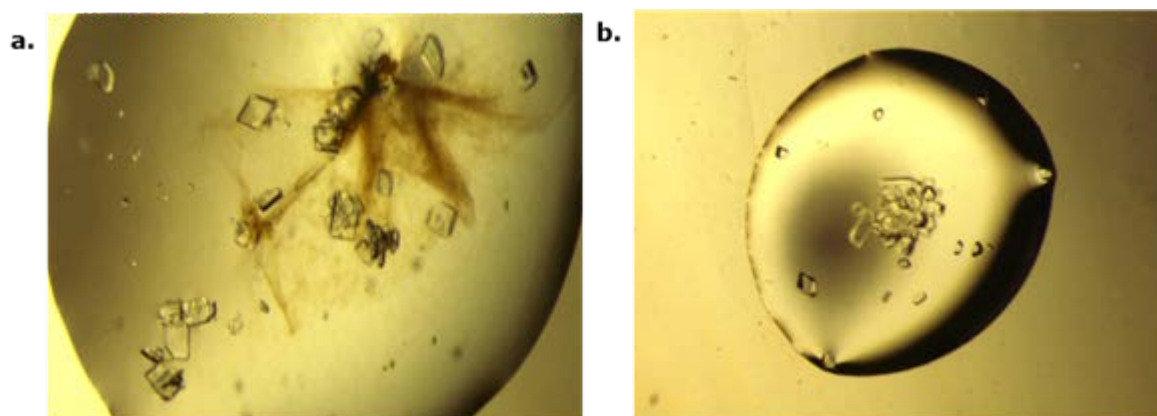
Figure 5.5.1 Crystallisation of the Cdc25C catalytic domain

Crystals appeared (a) after two weeks incubation (10 mg/ml) at 23 °C in the condition 0.2 M magnesium chloride hexahydrate, 0.1 M sodium citrate tribasic dehydrate, pH 5.0, 10 % w/v polyethylene glycol (PEG) 20, 000 from the PEGRx II screen. Crystals appeared in the control condition which had no protein after 4 weeks (b). This indicated the crystals were salt crystals.
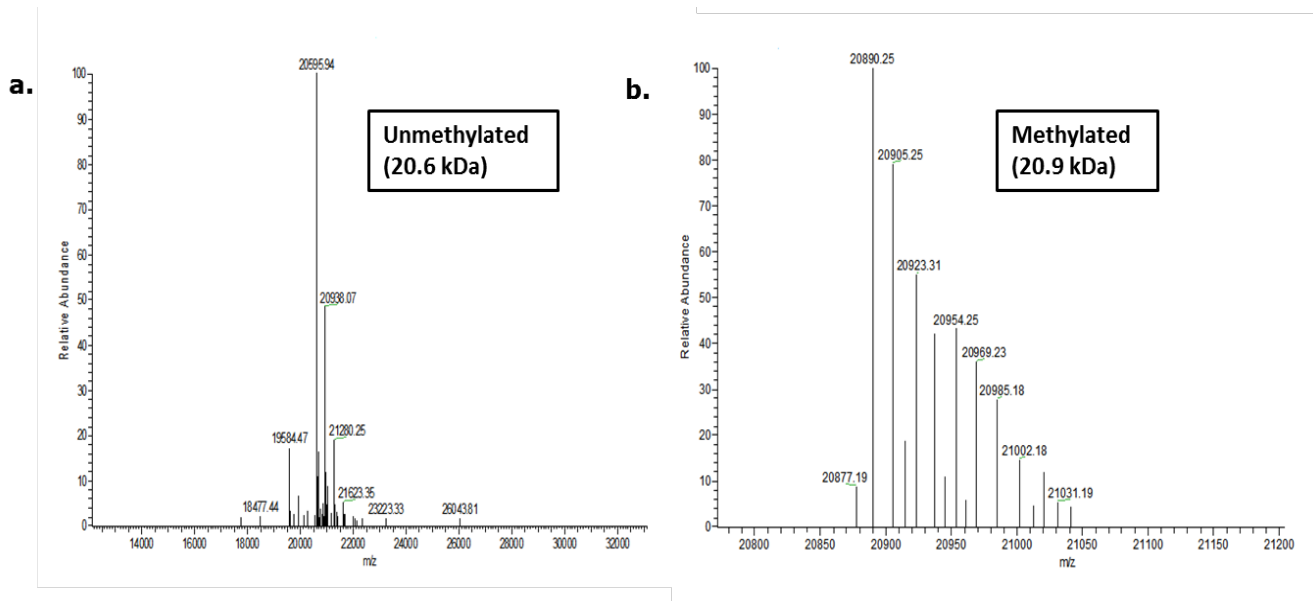
Figure 5.5.2 Reductive methylation of the Cdc25C catalytic domain

Mass spectrometry analysis of unmethylated (a) and methylated (b) Cdc25C catalytic domain. There was a mass increase of 300 Da.

5.6 Chapter 5 Discussion

The Cdc25C catalytic domain is soluble and stable in high concentrations of DMSO. From all the compounds tested, compounds 1, 2, 3, 4, 14, 15, 16, and 19 showed changes by NMR spectroscopy which suggested an interaction with the Cdc25C catalytic domain (Table A6). The compounds 4, 14, 15, and 16 gave the best WaterLOGSY results. Clear phase changes were observed in the WaterLOGSY spectra and where peaks did not show a definite change in phase changes in peak intensities were observed.

Compounds 1, 2, 3, 4, 14, 15, 16, and 19 (Table A6) were tested by the 2D SOFAST-HMQC method. Compounds 2, 3, and 14 resulted in a significant loss of peaks suggesting an intermediate exchange dynamic and overall showed a similar effect as compound 4 (BVT.948). The compounds 15, 16, and 19 generally resulted in weak chemical shift perturbations suggesting a fast exchange dynamic. Interestingly, a few of the chemical shift perturbations were localised to a common region on the spectrum. These compounds affected the Cdc25C catalytic domain 2D spectra similarly to compound 1 (sodium orthovanadate). Since sodium orthovanadate is a phosphate analogue and is considered to bind to the active site of protein tyrosine phosphatases therefore the compounds 2, 3, and 14 may be binding to the active site of the Cdc25C catalytic domain. The compounds 15, 16, and 19 may be binding at an allosteric site. However, backbone assignments are needed to confirm this.

Despite the large number of crystallization conditions tested there was only one promising condition identified with the native Cdc25C catalytic domain. Unfortunately, the crystals in this condition were later identified as salt crystals. Communication with the SGC in Oxford (Professor Knapp) who have solved the crystal structure of the Cdc25C catalytic domain (PDB: 3OP3, His-Cdc25C$_{270-462}$) in collaboration with the Midwest Centre for Structural Genomics

(MCSG) led us to understand that the native protein was a difficult target to crystallize. They used a 'reductive methylation' approach before crystallization to successfully crystallize the protein. This procedure has been shown to improve the crystallization success rate of stubborn proteins while retaining their native structure and function (Walter et al., 2006; Kim et al., 2008). The reductive methylation procedure results in the methylation of solvent exposed lysine residues. It is well known in the crystallography field that flexible side chains can hinder crystallization. Therefore, methylation of these side chains and hence lowering the conformational entropy may have played a significant role in promoting the crystallization of the Cdc25C catalytic domain.

The advantage of employing the reductive methylation procedure here was that the purified protein could be altered chemically without re-designing the construct. However, despite successfully methylating the Cdc25C catalytic domain (Cdc25C$_{270-443}$) there was no significant improvement observed in the number of preliminary crystallisation hits. Crystals were found in the same crystallization condition initially identified for the native protein and these later proved to be salt crystals. It is interesting that the same precipitant condition resulted in crystals for both the unmethylated and methylated proteins. Initially, it was thought that the high concentrations of arginine and sucrose in the protein buffer could be cross-reacting with the precipitant condition and inducing the formation of salt crystals. However, the purification buffer of the methylated protein lacked arginine or sucrose. The common condition between the two protein buffers was Hepes. It could be possible that the Hepes cross-reacted with the precipitant and led to the formation of salt crystals. It could also be possible that the magnesium chloride in the precipitant condition was crystallising.

Crystallization may not have been successful despite following the SGC (Oxford) crystallization protocol because the construct used to crystallize the protein here (Cdc25C$_{270\text{-}443}$) was different to the one used by the SGC (His-Cdc25C$_{270\text{-}462}$). The construct used here lacked the N-terminal His tag and was shorter in length. One of the key findings in Chapter 4 was that the removal of the His tag improved the solubility of Cdc25C$_{270\text{-}443}$. Therefore, retaining the His tag would reduce the solubility and may promote crystallization or it could exacerbate the precipitation that was already observed in the crystallization trials. Another possibility is that the residues 444 – 462 are important in promoting crystallization of this protein. However, this is unlikely considering there is no electron density for these residues indicating that this region is highly disordered.

Co-crystallization trials of the Cdc25C catalytic domain with an inhibitor identified from NMR titration experiments also yielded no crystals. It is possible that for future experiments, additional Cdc25C inhibitors could be used in the co-crystallisation trials.

An alternative approach that could be employed to improve the number of initial crystallisation hits for the Cdc25C catalytic domain is 'surface entropy reduction (SER)' (Cooper et al., 2007). This strategy could be used as an addition to the reductive methylation procedure. This technique promotes crystallisation of difficult protein targets by reducing the flexibility of surface exposed flexible side chains. This technique differs from that of methylation as it is not limited to lysine residues and requires a complete re-design of the construct. A specific cluster of 2, 3, or more flexible amino acids such as lysine, glutamic acid, and glutamine are mutated to alanine amino acids which can promote crystal contacts hence providing support to the crystal lattice. However, care needs to be taken in deciding which

amino acids to mutate to ensure they do not inadvertently affect the stability or solubility of the protein.

Another approach that could promote the crystallization of the Cdc25C catalytic domain is to fuse this domain with a protein tag (Derewenda, 2010). The protein tag is usually easy to crystallize. Indeed, large protein tags such as the MBP or GST tag have been shown to improve the crystallization success rate of difficult proteins (Kuge et al., 1997; Ullah et al., 2008). An additional benefit of these tags is that they are commonly used to purify and promote protein solubility therefore they may improve the solubility of the Cdc25C catalytic domain. It would be ideal to fuse these tags at the N-terminal of the Cdc25C catalytic domain replacing the N-terminal His tag with either an MBP or GST tag. This could greatly improve the solubility of the protein, promote crystallization, and therefore increase the number of crystallisation hits

6.0 Concluding Remarks

The findings from chapter 3 showed that the Cdc25C regulatory domain is intrinsically disordered. This provides insight into why crystallisation of the full-length protein has been difficult. Future studies need to focus on identifying a binding partner or a compound that may induce structure of this domain and the structure of this complex can potentially be solved. Also, the GST-tagged Cdc25C full-length protein which is generally used for activity assays may not be appropriate. It may be better to use the catalytic domain tagged to GST instead.

In chapter 4 construct and solution conditions were optimised for the Cdc25C catalytic domain. However, despite these optimisations it was difficult to obtain backbone assignments. Therefore, in the future alternative strategies will need to be employed for example a mutagenesis approach may be used. This has been shown to work for the Cdc25B homologue.

Chapter 5 has demonstrated that NMR spectroscopy can be used as an additional tool in identifying drug compounds. The first ^{1}H, ^{15}N HSQC of the Cdc25C homologue was obtained. The ability of Cdc25C to tolerate high concentrations of DMSO provides an advantage since many drug compounds have poor solubility.

References

Ahn, J.H., Korea Research Institute of Bioscience and Biotechnology; Korea Research Institute of Chemical Technology (2007) Rhodanine derivarives, a process for the preparation thereof and pharmaceutical composition containing the same. Patent WO2007081091

Albert, H., Battaglia, E., Monteiro, C., et al. (2012) Genotoxic stress modulates CDC25C phosphatase alternative splicing in human breast cancer cell lines. **Molecular oncology** [online], 6 (5): 542–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22871320 [Accessed 21 May 2014]

Alonso, A., Sasin, J., Bottini, N., et al. (2004) Protein tyrosine phosphatases in the human genome. **Cell**. 117 pp. 699–711

Andersen, J.N., Mortensen, O.H., Peters, G.H., et al. (2001) Structural and evolutionary relationships among protein tyrosine phosphatase domains. **Molecular and cellular biology** [online], 21 (21): 7117–36. Available from: http://mcb.asm.org/content/21/21/7117 [Accessed 23 October 2014]

Aoyagi, Y., Masuko, N., Ohkubo, S., et al. (2005) A novel cinnamic acid derivative that inhibits Cdc25 dual-specificity phosphatase activity. **Cancer Science**, 96: 614–619

Appelt, K., Bacquet, R.J., Bartlett, C.A., et al. (1991) Design of enzyme inhibitors using iterative protein crystallographic analysis. **Journal of medicinal chemistry** [online], 34 (7): 1925–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2066965 [Accessed 27 May 2014]

Arakawa, T., Ejima, D., Tsumoto, K., et al. (2007a) Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects. **Biophysical Chemistry**. 127 pp. 1–8

Arakawa, T., Kita, Y. and Timasheff, S.N. (2007b) Protein precipitation and denaturation by dimethyl sulfoxide. **Biophysical Chemistry**, 131: 62–70

Auton, M. and Bolen, D.W. (2005) Predicting the energetics of osmolyte-induced protein folding/unfolding. **Proceedings of the National Academy of Sciences of the United States of America**, 102: 15065–15068

Bagby, S., Tong, K.I. and Ikura, M. (2001) Optimization of protein solubility and stability for protein nuclear magnetic resonance. **Methods in Enzymology**, 339: 20–41

Bakaysa, D.L., Radziuk, J., Havel, H.A., et al. (1996) Physicochemical basis for the rapid time-action of LysB28ProB29-insulin: dissociation of a protein-ligand complex. **Protein science : a publication of the Protein Society**, 5: 2521–2531

Baldin, V., Cans, C., Superti-Furga, G., et al. (1997) Alternative splicing of the human CDC25B tyrosine phosphatase. Possible implications for growth control? **Oncogene**, 14: 2485–2495

Bartek, J. and Lukas, J. (2003) Chk1 and Chk2 kinases in checkpoint control and cancer. **Cancer Cell**. 3 pp. 421–429

Ben-Yosef, T., Yanuka, O., Halle, D., et al. (1998) Involvement of Myc targets in c-myc and N-myc induced human tumors. **Oncogene**, 17: 165–171

Berger, S.L. (2002) Histone modifications in transcriptional regulation. **Current opinion in genetics & development**, 12: 142–148

Bernadó, P., Mylonas, E., Petoukhov, M. V, et al. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. **Journal of the American Chemical Society**, 129: 5656–5664

De Bernardez Clark, E. (1998) Refolding of recombinant proteins. **Current Opinion in Biotechnology**. 9 pp. 157–163

Bhunia, A., Bhattacharjya, S. and Chatterjee, S. (2012) Applications of saturation transfer difference NMR in biological systems. **Drug discovery today** [online], 17 (9-10): 505–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22210119 [Accessed 28 April 2014]

Blobel, J., Brath, U., Bernad??, P., et al. (2011) Protein loop compaction and the origin of the effect of arginine and glutamic acid mixtures on solubility, stability and transient oligomerization of proteins. **European Biophysics Journal**, 40: 1327–1338

Blomberg, I. and Hoffmann, I. (1999) Ectopic expression of Cdc25A accelerates the G(1)/S transition and leads to premature activation of cyclin E- and cyclin A-dependent kinases. **Molecular and cellular biology**, 19: 6183–6194

Bodenhausen, G. and Ruben, D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. **Chemical Physics Letters**. 69 pp. 185–189

Boutros, R., Dozier, C. and Ducommun, B. (2006) The when and wheres of CDC25 phosphatases. **Current opinion in cell biology** [online], 18 (2): 185–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16488126 [Accessed 15 May 2014]

Brezak, M.-C., Quaranta, M., Contour-Galcera, M.-O., et al. (2005) Inhibition of human tumor cell growth in vivo by an orally bioavailable inhibitor of CDC25 phosphatases. **Molecular cancer therapeutics**, 4: 1378–1387

Brezak, M.-C., Valette, A., Quaranta, M., et al. (2009) IRC-083864, a novel bis quinone inhibitor of CDC25 phosphatases active against human cancer cells. **International journal of cancer. Journal international du cancer** [online], 124 (6): 1449–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19065668 [Accessed 21 May 2014]

Brown, P.H. and Schuck, P. (2006) Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. **Biophysical journal**, 90: 4651–4661

Bulavin, D. V, Higashimoto, Y., Popoff, I.J., et al. (2001) Initiation of a G2/M checkpoint after ultraviolet radiation requires p38 kinase. **Nature**, 411: 102–107

Busino, L., Donzelli, M., Chiesa, M., et al. (2003) Degradation of Cdc25A by beta-TrCP during S phase and in response to DNA damage. **Nature**, 426: 87–91

Cebula, R.E., Blanchard, J.L., Boisclair, M.D., et al. (1997) Synthesis and phosphatase inhibitory activity of analogs of sulfircin. **Bioorganic and Medicinal Chemistry Letters**, 7: 2015–2020

Chen, M.-S., Ryan, C.E. and Piwnica-Worms, H. (2003) Chk1 kinase negatively regulates mitotic function of Cdc25A phosphatase through 14-3-3 binding. **Molecular and cellular biology**, 23: 7488–7497

Chen, W., Wilborn, M. and Rudolph, J. (2000) Dual-specific Cdc25B phosphatase: In search of the catalytic acid. **Biochemistry**, 39: 10781–10789

Cho, S.S., Reddy, G., Straub, J.E., et al. (2011) Entropic stabilization of proteins by TMAO. **Journal of Physical Chemistry B**, 115: 13401–13407

Cole, J.L., Lary, J.W., P. Moody, T., et al. (2008) Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. **Methods in Cell Biology**. 84 pp. 143–179

Conklin, D.S., Galaktionov, K. and Beach, D. (1995) 14-3-3 proteins associate with cdc25 phosphatases. **Proceedings of the National Academy of Sciences of the United States of America**, 92: 7892–7896

Cooper, D.R., Boczek, T., Grelewska, K., et al. (2007) Protein crystallization by surface entropy reduction: Optimization of the SER strategy. **Acta Crystallographica Section D: Biological Crystallography**, 63: 636–645

Costantino, H.R., Brown, S.H. and Kelly, R.M. (1990) Purification and characterization of an alpha-glucosidase from a hyperthermophilic archaebacterium, Pyrococcus furiosus, exhibiting a temperature optimum of 105 to 115 degrees C. **Journal of bacteriology**, 172: 3654–3660

Dalal, S.N., Schweitzer, C.M., Gan, J., et al. (1999) Cytoplasmic localization of human cdc25C during interphase requires an intact 14-3-3 binding site. **Molecular and cellular biology**, 19: 4465–4479

Dale, G.E., Broger, C., Langen, H., et al. (1994) Improving protein solubility through rationally designed amino acid replacements: solubilization of the trimethoprim-resistant type S1 dihydrofolate reductase. **Protein engineering**, 7: 933–939

Dalvit, C., Fogliatto, G., Stewart, A., et al. (2001) **WaterLOGSY as a method for primary NMR screening : Practical aspects and range of applicability**., pp. 349–359

Delaglio, F., Grzesiek, S., Vuister, G.W., et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. **Journal of biomolecular NMR**, 6: 277–293

Derewenda, Z.S. (2010) Application of protein engineering to enhance crystallizability and improve crystal properties. **Acta Crystallographica Section D: Biological Crystallography**, 66: 604–615

Dessau, M.A. and Modis, Y. (2011) Protein crystallization for X-ray crystallography. **Journal of visualized experiments : JoVE**

Donzelli, M., Squatrito, M., Ganoth, D., et al. (2002) Dual mode of degradation of Cdc25 A phosphatase. **EMBO Journal**, 21: 4875–4884

Draetta, G. and Eckstein, J. (1997) Cdc25 protein phosphatases in cell proliferation. **Biochimica et biophysica acta** [online], 1332 (2): M53–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9141461 [Accessed 26 May 2014]

Durbin, S.D. and Feher, G. (1996) Protein crystallization. **Annual review of physical chemistry**, 47: 171–204

Eui, N.L., Young, M.K., Hye, J.L., et al. (2005) Stabilizing peptide fusion for solving the stability and solubility problems of therapeutic proteins. **Pharmaceutical Research**, 22: 1735–1746

Falck, J., Mailand, N., Syljuåsen, R.G., et al. (2001) The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. **Nature**, 410: 842–847

Fan, Z., Xu, X., Cao, J., et al. (2012) [Prokaryotic expression and purification of human GST-Cdc25C fusion protein and preliminary detection of its function]. **Xi bao yu fen zi mian yi xue za zhi = Chinese journal of cellular and molecular immunology** [online], 28 (3): 251–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22394631 [Accessed 24 May 2014]

Fatouros, A. and Thomas, O. (1997) Recombinant factor VIII SQ — influence of oxygen , metal ions , pH and ionic strength on its stability in aqueous solution. **International Journal of Pharmaceutics**, 155: 121–131

Fauman, E.B., Cogswell, J.P., Lovejoy, B., et al. (1998) **Crystal Structure of the Catalytic Domain of the Human Cell Cycle Control Phosphatase , Cdc25A**., 93: 617–625

Fields, G.B., Alonso, D.O. V, Stigter, D., et al. (1992) Theory for the Aggregation of Proteins and Copolymerst. **Journal of Psysical Chemistry**, 96: 3974–3981

Forrest, A.R., McCormack, A.K., DeSouza, C.P., et al. (1999) Multiple splicing variants of cdc25B regulate G2/M progression. **Biochemical and biophysical research communications**, 260: 510–515

Fowler, S.B., Poon, S., Muff, R., et al. (2005) Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. **Proceedings of the National Academy of Sciences of the United States of America**, 102: 10105–10110

Franckhauser, C., Fernandez, A. and Lamb, N.J.C. (2013) Purification and biochemical analysis of catalytically active human cdc25C dual specificity phosphatase. **Biochimie** [online], 95 (7): 1450–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23567337 [Accessed 21 May 2014]

Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. **Journal of Applied Crystallography**, 42: 342–346

Fu, H., Subramanian, R.R. and Masters, S.C. (2000) 14-3-3 proteins: structure, function, and regulation. **Annual review of pharmacology and toxicology**, 40: 617–647

Gabrielli, B.G., De Souza, C.P., Tonks, I.D., et al. (1996) Cytoplasmic accumulation of cdc25B phosphatase in mitosis triggers centrosomal microtubule nucleation in HeLa cells. **Journal of cell science**, 109 (Pt 5: 1081–1093

Galaktionov, K. and Beach, D. (1991) Specific activation of cdc25 tyrosine phosphatases by B-type cyclins: evidence for multiple roles of mitotic cyclins. **Cell**, 67: 1181–1194

Galaktionov, K., Chen, X. and Beach, D. (1996) Cdc25 cell-cycle phosphatase as a target of c-myc. **Nature**, 382: 511–517

Gasparotto, D., Maestro, R., Piccinin, S., et al. (1997) Overexpression of CDC25A and CDC25B in head and neck cancers. **Cancer research**, 57: 2366–2368

Gerstein, A. V, Almeida, T.A., Zhao, G., et al. (2002) APC/CTNNB1 (beta-catenin) pathway alterations in human prostate cancers. **Genes, chromosomes & cancer**, 34: 9–16

Ghosh, R., Sharma, S. and Chattopadhyay, K. (2009) **Effect of Arginine on Protein Aggregation Studied by Fluorescence Correlation Spectroscopy and Other Biophysical Methods †**., pp. 1135–1143

Golovanov, A.P., Hautbergue, G.M., Wilson, S.A., et al. (2004) A simple method for improving protein solubility and long-term stability. **Journal of the American Chemical Society**, 126: 8933–8939

Gordon, J.A. (1991) Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. **Methods in enzymology**, 201: 477–482

Gottlin, E.B., Xu, X., Epstein, D.M., et al. (1996) Kinetic analysis of the catalytic domain of human cdc25B. **The Journal of biological chemistry** [online], 271 (44): 27445–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8910325

Goulev, Y. and Charvin, G. (2011) Ultrasensitivity and Positive Feedback to Promote Sharp Mitotic Entry. **Molecular Cell**. 41 pp. 243–244

Grant, T.D., Luft, J.R., Wolfley, J.R., et al. (2011) Small angle X-ray scattering as a complementary tool for high-throughput structural studies. **Biopolymers**, 95: 517–530

Gunasekera, S.P., McCarthy, P.J., Kelly-Borges, M., et al. (1996) Dysidiolide: A novel protein phosphatase inhibitor from the Caribbean sponge Dyside A etheria de Laubenfels. **Journal of the American Chemical Society**, 118: 8759–8760

Guo, J., Kleeff, J., Li, J., et al. (2004) Expression and functional significance of CDC25B in human pancreatic ductal adenocarcinoma. **Oncogene**, 23: 71–81

Hammarström, M., Hellgren, N., van Den Berg, S., et al. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. **Protein science : a publication of the Protein Society**, 11: 313–321

He, N., Li, C., Zhang, X., et al. (2005) Regulation of lung cancer cell growth and invasiveness by beta-TRCP. **Molecular carcinogenesis**, 42: 18–28

Hernández, S., Hernández, L., Bea, S., et al. (2000) cdc25a and the splicing variant cdc25b2, but not cdc25B1, -B3 or -C, are over-expressed in aggressive human non-Hodgkin's lymphomas. **International journal of cancer. Journal international du cancer**, 89: 148–152

Hernández, S., Hernández, L., Beà, S., et al. (1998) cdc25 cell cycle-activating phosphatases and c-myc expression in human non-Hodgkin's lymphomas. **Cancer research**, 58: 1762–1767

Hoffmann, I., Draetta, G. and Karsenti, E. (1994) Activation of the phosphatase activity of human cdc25A by a cdk2-cyclin E dependent phosphorylation at the G1/S transition. **The EMBO journal**, 13: 4302–4310

Hofmann, K., Bucher, P. and Kajava, a V (1998) A model of Cdc25 phosphatase catalytic domain and Cdk-interaction surface based on the presence of a rhodanese homology domain. **Journal of molecular biology** [online], 282 (1): 195–208. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9733650

Horiguchi, T., Nishi, K., Hadoda, S., et al. (1994) Dnacin A1 and dnacin B1 are antitumor antibiotics that inhibit cdc25B phosphatase activity. **Biochemical Pharmacology**, 48: 2139–2141

Huang, Y.J., Acton, T.B. and Montelione, G.T. (2014) DisMeta: a meta server for construct design and optimization. **Methods in molecular biology (Clifton, N.J.)** [online], 1091: 3–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24203321 [Accessed 22 May 2014]

Hubbard, S.R. and Till, J.H. (2000) Protein tyrosine kinase structure and function. **Annual review of biochemistry**, 69: 373–398

Hunter, T. (2000) Signaling--2000 and beyond. **Cell**, 100: 113–127

Hunter, T. (2009) Tyrosine phosphorylation: thirty years and counting. **Current Opinion in Cell Biology**. 21 pp. 140–146

Ito, Y., Yoshida, H., Uruno, T., et al. (2004) Expression of cdc25A and cdc25B phosphatase in breast carcinoma. **Breast Cancer**, 11: 295–300

Jacques, D.A. and Trewhella, J. (2010) Small-angle scattering for structural biology--expanding the frontier while avoiding the pitfalls. **Protein science : a publication of the Protein Society**, 19: 642–657

Jenuwein, T. and Allis, C.D. (2001) Translating the histone code. **Science (New York, N.Y.)**, 293: 1074–1080

Jin, J., Shirogane, T., Xu, L., et al. (2003) SCFbeta-TRCP links Chk1 signaling to degradation of the Cdc25A protein phosphatase. **Genes & development**, 17: 3062–3074

Jinno, S., Suto, K., Nagata, A., et al. (1994) Cdc25A is a novel phosphatase functioning early in the cell cycle. **The EMBO journal**, 13: 1549–1556

Kapust, R.B. and Waugh, D.S. (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. **Protein science : a publication of the Protein Society**, 8: 1668–1674

Kar, S., Wang, M., Yao, W., et al. (2006) PM-20, a novel inhibitor of Cdc25A, induces extracellular signal-regulated kinase 1/2 phosphorylation and inhibits hepatocellular carcinoma growth in vitro and in vivo. **Molecular cancer therapeutics** [online], 5 (6): 1511–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16818510 [Accessed 21 May 2014]

Karlsson, C., Katich, S., Hagting, A., et al. (1999) Cdc25B and Cdc25C differ markedly in their properties as initiators of mitosis. **The Journal of cell biology**, 146: 573–584

Kato, A., Maki, K., Ebina, T., et al. (2007) Mutational analysis of protein solubility enhancement using short peptide tags. **Biopolymers**, 85: 12–18

Kay, L.E., Keifer, P. and Saarinen, T. (1992) Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. **J Am Chem Soc**, pp. 10663–10665

Kelly, S.M., Jess, T.J. and Price, N.C. (2005) **How to study proteins by circular dichroism**., 1751: 119–139

Kelly, S.M. and Price, N.C. (2000) The use of circular dichroism in the investigation of protein structure and function. **Current protein & peptide science**, 1: 349–384

Kendrick, B.S., Carpenter, J.F., Cleland, J.L., et al. (1998) A transient expansion of the native state precedes aggregation of recombinant human interferon-gamma. **Proceedings of the National Academy of Sciences of the United States of America**, 95: 14142–14146

Khadaroo, B., Robbens, S., Ferraz, C., et al. (2004) **The First Green Lineage cdc25 Dual-Specificity Phosphatase nd io sc en ce . N ot fo is tr ib ut en ce . N ot fo tr ib ut**., (April): 513–518

Kim, Y., Quartey, P., Li, H., et al. (2008) Large-scale evaluation of protein reductive methylation for improving protein crystallization. **Nature methods**. 5 pp. 853–854

Kishi, K., Doki, Y., Miyata, H., et al. (2002) Prediction of the response to chemoradiation and prognosis in oesophageal squamous cancer. **British Journal of Surgery**, 89: 597–603

Konarev, P. V., Volkov, V. V., Sokolova, A. V., et al. (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. **Journal of Applied Crystallography**, 36: 1277–1282

Kozin, M.B. and Svergun, D.I. (2001) Automated matching of high- and low-resolution structural models. **Journal of Applied Crystallography**, 34: 33–41

Kriwacki, R.W., Hengst, L., Tennant, L., et al. (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. **Proceedings of the National Academy of Sciences of the United States of America**, 93: 11504–11509

Kuge, M., Fujii, Y., Shimizu, T., et al. (1997) Use of a fusion protein to obtain crystals suitable for X-ray analysis: crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. **Protein science : a publication of the Protein Society**, 6: 1783–1786

Kumagai, A. and Dunphy, W.G. (1999) Binding of 14-3-3 proteins and nuclear export control the intracellular localization of the mitotic inducer Cdc25. **Genes & development**, 13: 1067–1072

Lammer, C., Wagerer, S., Saffrich, R., et al. (1998) The cdc25B phosphatase is essential for the G2/M phase transition in human cells. **Journal of cell science**, 111 (Pt 1: 2445–2453

Lavecchia, A., Di Giovanni, C. and Novellino, E. (2010) Inhibitors of Cdc25 phosphatases as anticancer agents: a patent review. **Expert opinion on therapeutic patents** [online], 20 (3): 405–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20166845

Lazo, J.S., Aslan, D.C., Southwick, E.C., et al. (2001) Discovery and biological evaluation of a new family of potent inhibitors of the dual specificity protein phosphatase Cdc25. **J Med Chem** [online], 44: 4042–4049. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11708908

Lee, J.C. and Timasheff, S.N. (1981) The stabilization of proteins by sucrose. **The Journal of biological chemistry**, 256: 7193–7201

Lepre, C.A. and Moore, J.M. (1998) Microdrop screening: a rapid method to optimize solvent conditions for NMR spectroscopy of proteins. **Journal of biomolecular NMR**, 12: 493–499

Liljebris, C., Baranczewski, P., Björkstrand, E., et al. (2004) Oxidation of protein tyrosine phosphatases as a pharmaceutical mechanism of action: a study using 4-hydroxy-3,3-dimethyl-2H-benzo[g]indole-2,5(3H)-dione. **The Journal of pharmacology and experimental therapeutics**, 309: 711–719

Lindqvist, A., Källström, H., Lundgren, A., et al. (2005) Cdc25B cooperates with Cdc25A to induce mitosis but has a unique role in activating cyclin B1-Cdk1 at the centrosome. **The Journal of cell biology**, 171: 35–45

Liu, Q., Guntuku, S., Cui, X.S., et al. (2000) Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. **Genes & development**, 14: 1448–1459

Löffler, H., Syljuåsen, R.G., Bartkova, J., et al. (2003) Distinct modes of deregulation of the proto-oncogenic Cdc25A phosphatase in human breast cancer cell lines. **Oncogene**, 22: 8063–8071

Ludwig, C. and Guenther, U.L. (2009) Ligand based NMR methods for drug discovery. **Frontiers in bioscience : a journal and virtual library**, 14: 4565–4574

Ludwig, C., Michiels, P.J. a, Wu, X., et al. (2008) SALMON: solvent accessibility, ligand binding, and mapping of ligand orientation by NMR spectroscopy. **Journal of medicinal chemistry** [online], 51 (1): 1–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18062662

Lund, G. and Cierpicki, T. (2014) Solution NMR studies reveal no global flexibility in the catalytic domain of CDC25B. **Proteins** [online]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24740794 [Accessed 27 May 2014]

McCain, D.F., Catrina, I.E., Hengge, A.C., et al. (2002) The catalytic mechanism of Cdc25A phosphatase. **The Journal of biological chemistry**, 277: 11190–11200

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

Meng, K., Rodriguez-Peña, A., Dimitrov, T., et al. (2000) Pleiotrophin signals increased tyrosine phosphorylation of beta beta-catenin through inactivation of the intrinsic catalytic activity of the receptor-type protein tyrosine phosphatase beta/zeta. **Proceedings of the National Academy of Sciences of the United States of America**, 97: 2603–2608

Mertens, H.D.T. and Svergun, D.I. (2010) Structural characterization of proteins and complexes using small-angle X-ray solution scattering. **Journal of structural biology**, 172: 128–141

Millar, J.B., Blevitt, J., Gerace, L., et al. (1991) p55CDC25 is a nuclear protein required for the initiation of mitosis in human cells. **Proceedings of the National Academy of Sciences of the United States of America**, 88: 10500–10504

Mosavi, L.K. and Peng, Z.-Y. (2003) Structure-based substitutions for increased solubility of a designed protein. **Protein engineering**, 16: 739–745

Niesen, F.H., Berglund, H. and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. **Nature protocols**, 2: 2212–2221

Olsen, J. V., Blagoev, B., Gnad, F., et al. (2006) Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. **Cell**, 127: 635–648

Ozen, M. and Ittmann, M. (2005) Increased expression and activity of CDC25C phosphatase and an alternatively spliced variant in prostate cancer. **Clinical cancer research : an official journal of the American Association for Cancer Research**, 11: 4701–4706

Pace, C.N. (1990) Measuring and increasing protein stability. **Trends in biotechnology** [online], 8 (4): 93–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1367432 [Accessed 23 May 2014]

Page, R., Peti, W., Wilson, I.A., et al. (2005) NMR screening and crystal quality of bacterially expressed prokaryotic and eukaryotic proteins in a structural genomics pipeline. **Proceedings of the National Academy of Sciences of the United States of America**, 102: 1901–1905

Pao, L.I., Badour, K., Siminovitch, K.A., et al. (2007) Nonreceptor protein-tyrosine phosphatases in immune cell signaling. **Annual review of immunology**, 25: 473–523

Paul, S. and Lombroso, P.J. (2003) Receptor and nonreceptor protein tyrosine phosphatases in the nervous system. **Cellular and Molecular Life Sciences**. 60 pp. 2465–2482

Pawson, T. and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. **Science (New York, N.Y.)**, 300: 445–452

Peak, M.J., Peak, J.G., Stevens, F.J., et al. (1994) The hyperthermophilic glycolytic enzyme enolase in the archaeon, Pyrococcus furiosus: comparison with mesophilic enolases. **Archives of biochemistry and biophysics**, 313: 280–286

Peng, C.Y., Graves, P.R., Ogg, S., et al. (1998) C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and promotes 14-3-3 protein binding. **Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research**, 9: 197–208

Pettersen, E.F., Goddard, T.D., Huang, C.C., et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of computational chemistry**, 25: 1605–1612

Putnam, C.D., Hammel, M., Hura, G.L., et al. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. **Quarterly reviews of biophysics** [online], 40 (3): 191–285. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18078545 [Accessed 30 April 2014]

Radivojac, P., Obradovic, Z., Smith, D.K., et al. (2004) Protein flexibility and intrinsic disorder. **Protein Science** [online], 13: 71–80. Available from: <Go to ISI>://000187587700008

Receveur-Brechot, V. and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. **Current protein & peptide science** [online], 13: 55–75. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3394175&tool=pmcentrez&rendertype=abstract

Reynolds, R.A., Yem, A.W., Wolfe, C.L., et al. (1999a) **Crystal Structure of the Catalytic Subunit of Cdc25B Required for G 2 / M Phase Transition of the Cell Cycle**.

Reynolds, R.A., Yem, A.W., Wolfe, C.L., et al. (1999b) Crystal structure of the catalytic subunit of Cdc25B required for G2/M phase transition of the cell cycle. **Journal of molecular biology**, 293: 559–568

Robertson, S.C., Tynan, J.A. and Donoghue, D.J. (2000) RTK mutations and human syndromes - When good receptors turn bad. **Trends in Genetics**. 16 pp. 265–271

Robinson, D.R., Wu, Y.M. and Lin, S.F. (2000) The protein tyrosine kinase family of the human genome. **Oncogene**, 19: 5548–5557

Rost, B. and Sander, C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. **Proteins: Structure, Function and Genetics**, 19: 55–72

Rudolph, J. (2002) Catalytic mechanism of Cdc25. **Biochemistry** [online], 41 (49): 14613–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12463761

Rudolph, J. (2007) Cdc25 phosphatases: Structure, specificity, and mechanism. **Biochemistry**. 46 pp. 3595–3604

Rudolph, J., Epstein, D.M., Parker, L., et al. (2001) Specificity of natural and artificial substrates for human Cdc25A. **Analytical biochemistry**, 289: 43–51

Rudolph, J., Kristja, K. and Carolina, N. (2004) **Cdc25 Phosphatases and Cancer**., 11: 1043–1051

Russell, P. and Nurse, P. (1986) cdc25+ functions as an inducer in the mitotic control of fission yeast. **Cell**, 45: 145–153

Sadhu, K., Reed, S.I., Richardson, H., et al. (1990) Human homolog of fission yeast cdc25 mitotic inducer is predominantly expressed in G2. **Proceedings of the National Academy of Sciences of the United States of America**, 87: 5139–5143

Sasaki, H., Yukiue, H., Kobayashi, Y., et al. (2001) Expression of the cdc25B gene as a prognosis marker in non-small cell lung cancer. **Cancer Letters**, 173: 187–192

Schanda, P., Kupce, E. and Brutscher, B. (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. **Journal of biomolecular NMR**, 33: 199–211

Schlessinger, J. (2000) Cell Signaling by Receptor Tyrosine Kinases. **Cell**. 103 pp. 211–225

Sexl, V., Diehl, J.A., Sherr, C.J., et al. (1999) A rate limiting function of cdc25A for S phase entry inversely correlates with tyrosine dephosphorylation of Cdk2. **Oncogene**, 18: 573–582

Sharff, A. and Jhoti, H. (2003) High-throughput crystallography to enhance drug discovery. **Current Opinion in Chemical Biology**. 7 pp. 340–345

Shiraki, K., Kudou, M., Fujiwara, S., et al. (2002) Biophysical effect of amino acids on the prevention of protein aggregation. **Journal of biochemistry**, 132: 591–595

Sluchanko, N.N. and Gusev, N.B. (2010) 14-3-3 proteins and regulation of cytoskeleton. **Biochemistry. Biokhimiia**, 75: 1528–1546

Sohn, J., Buhrman, G. and Rudolph, J. (2007) Kinetic and structural studies of specific protein-protein interactions in substrate catalysis by Cdc25B phosphatase. **Biochemistry**, 46: 807–818

Sohn, J., Kristjánsdóttir, K., Safi, a, et al. (2004) Remote hot spots mediate protein substrate recognition for the Cdc25 phosphatase. **Proceedings of the National Academy of Sciences of the United States of America** [online], 101 (47): 16437–41. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=534539&tool=pmcentrez&rendertype=abstract

Sohn, J., Parks, J.M., Buhrman, G., et al. (2005) Experimental validation of the docking orientation of Cdc25 with its Cdk2-CycA protein substrate. **Biochemistry** [online], 44 (50): 16563–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16342947

Sohn, J. and Rudolph, J. (2006) The Energetic Network of Hotspot Residues between Cdc25B Phosphatase and its Protein Substrate. **Journal of Molecular Biology**, 362: 1060–1071

Sohn, J. and Rudolph, J. (2007) Temperature dependence of binding and catalysis for the Cdc25B phosphatase. **Biophysical Chemistry**, 125: 549–555

De Souza, C.P., Ellem, K.A. and Gabrielli, B.G. (2000) Centrosomal and cytoplasmic Cdc2/cyclin B1 activation precedes nuclear mitotic events. **Experimental cell research**, 257: 11–21

Spallarossa, A., Donahue, J.L., Larson, T.J., et al. (2001) Escherichia coli GlpE is a prototype sulfurtransferase for the single-domain rhodanese homology superfamily. **Structure**, 9: 1117–1125

Sreerama, N. and Woody, R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. **Analytical biochemistry**, 287: 252–260

Stockman, B.J. and Dalvit, C. (2002) NMR screening techniques in drug discovery and drug design. **Progress in Nuclear Magnetic Resonance Spectroscopy**. 41 pp. 187–231

Svergun, D., Barberato, C. and Koch, M.H. (1995) CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. **Journal of Applied Crystallography**, 28: 768–773

Svergun, D.I. and Koch, M.H.J. (2002) Advances in structure analysis using small-angle scattering in solution. **Current Opinion in Structural Biology**. 12 pp. 654–660

Svergun, D.I. and Koch, M.H.J. (2003) Small-angle scattering studies of biological macromolecules in solution. **Reports on Progress in Physics**. 66 pp. 1735–1782

Tautz, L., Critton, D.A. and Grotegut, S. (2013) Protein tyrosine phosphatases: structure, function, and implication in human disease. **Methods in molecular biology (Clifton, N.J.)** [online], 1053: 179–221. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23860656 [Accessed 26 May 2014]

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

Tjernberg, A., Markova, N., Griffiths, W.J., et al. (2006) DMSO-related effects in protein characterization. **Journal of biomolecular screening** [online], 11: 131–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16490773

Tonks, N.K. (2006) Protein tyrosine phosphatases: from genes, to function, to disease. **Nature reviews. Molecular cell biology**, 7: 833–846

Trevino, S.R., Scholtz, J.M. and Pace, C.N. (2007) Amino Acid Contribution to Protein Solubility: Asp, Glu, and Ser Contribute more Favorably than the other Hydrophilic Amino Acids in RNase Sa. **Journal of Molecular Biology**, 366: 449–460

Trunnell, N.B., Poon, A.C., Kim, S.Y., et al. (2011) Ultrasensitivity in the Regulation of Cdc25C by Cdk1. **Molecular Cell**, 41: 263–274

Uchida, S., Ohtsubo, M., Shimura, M., et al. (2004) **Nuclear export signal in CDC25B**., 316: 226–232

Ullah, H., Scappini, E.L., Moon, A.F., et al. (2008) Structure of a signal transduction regulator, RACK1, from Arabidopsis thaliana. **Protein science : a publication of the Protein Society**, 17: 1771–1780

Vedadi, M., Niesen, F.H., Allali-Hassani, A., et al. (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. **Proceedings of the National Academy of Sciences of the United States of America**, 103: 15835–15840

Volkov, V. V. and Svergun, D.I. (2003) "Uniqueness of ab initio shape determination in small-angle scattering." In **Journal of Applied Crystallography**. **2003**. pp. 860–864

Vranken, W.F., Boucher, W., Stevens, T.J., et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. **Proteins**, 59: 687–696

Van Vugt, M.A.T.M., Br??s, A. and Medema, R.H. (2004) Polo-like kinase-1 controls recovery from a G2 DNA damage-induced arrest in mammalian cells. **Molecular Cell**, 15: 799–811

Walter, T.S., Meier, C., Assenberg, R., et al. (2006) Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. **Structure**, 14: 1617–1622

Wang, W. (1999) Instability, stabilization, and formulation of liquid protein pharmaceuticals. **International Journal of Pharmaceutics**. 185 pp. 129–188

Wegener, S., Hampe, W., Herrmann, D., et al. (2000) Alternative splicing in the regulatory region of the human phosphatases CDC25A and CDC25C. **European journal of cell biology**, 79: 810–815

Whitmore, L. and Wallace, B.A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. **Nucleic acids research**, 32: W668–W673

Williamson, M.P. (2013) Using chemical shift perturbation to characterise ligand binding. **Progress in Nuclear Magnetic Resonance Spectroscopy**. 73 pp. 1–16

Woestenenk, E.A., Hammarstr??m, M., Van Den Berg, S., et al. (2004) His tag effect on solubility of human proteins produced in Escherichia coli: A comparison between four expression vectors. **Journal of Structural and Functional Genomics**, 5: 217–229

Wooh, J.W., Kidd, R.D., Martin, J.L., et al. (2003) Comparison of three commercial sparse-matrix crystallization screens. **Acta Crystallographica - Section D Biological Crystallography**, 59: 769–772

Xiao, Z., Chen, Z., Gunasekera, A.H., et al. (2003) Chk1 mediates S and G2 arrests through Cdc25A degradation in response to DNA-damaging agents. **The Journal of biological chemistry**, 278: 21767–21773

Yang, Z.R., Thomson, R., McNeil, P., et al. (2005) RONN: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. **Bioinformatics**, 21: 3369–3376

Yasukawa, T., Kanei-Ishii, C., Maekawa, T., et al. (1995) Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin. **The Journal of biological chemistry**, 270: 25328–25331

Young, C.F. and P.G. (2012) **Protein Phosphorylation in Human Health**. Huang, C. (ed.) [online]. InTech. Available from: http://www.intechopen.com/books/protein-phosphorylation-in-human-health/phosphorylation-mediated-regulation-of-cdc25-activity-localization-and-stability [Accessed 26 May 2014]

Zhou, P., Lugovskoy, A.A., McCarty, J.S., et al. (2001) Solution structure of DFF40 and DFF45 N-terminal domain complex and mutual chaperone activity of DFF40 and DFF45. **Proceedings of the National Academy of Sciences of the United States of America**, 98: 6051–6055

Zhou, P. and Wagner, G. (2010) Overcoming the solubility limit with solubility-enhancement tags: Successful applications in biomolecular NMR studies. **Journal of Biomolecular NMR**. 46 pp. 23–31

Zou, Q., Bennion, B.J., Daggett, V., et al. (2002) The molecular mechanism of stabilization of proteins by TMAO and its ability to counteract the effects of urea. **Journal of the American Chemical Society**, 124: 1192–1202
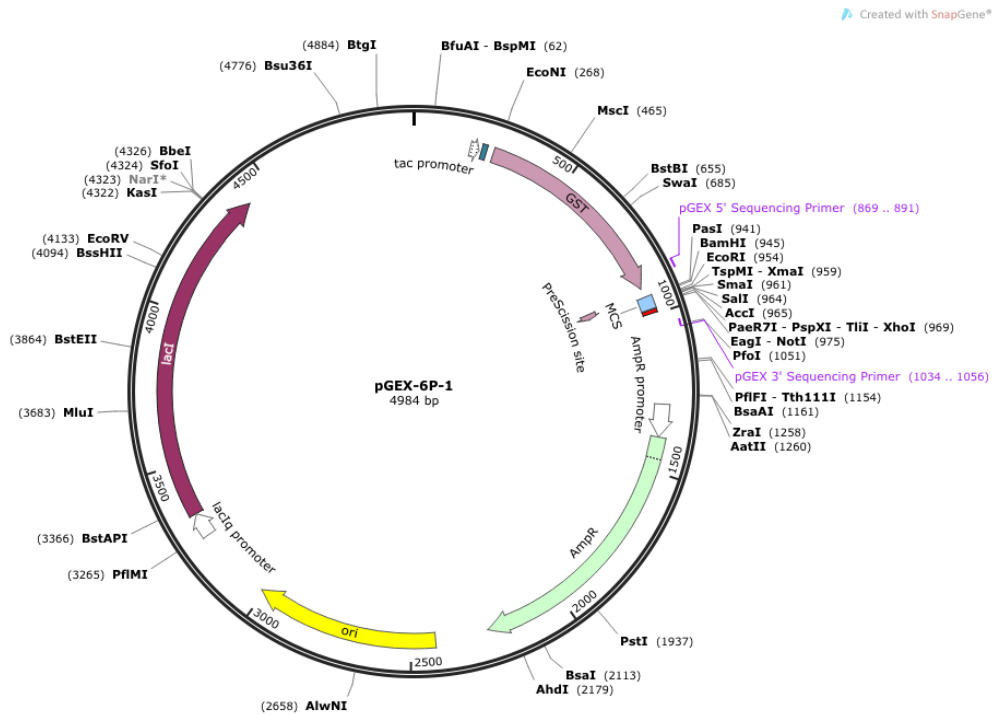
Appendix (A)

A1 - Vectors

<u>A1 (a) pNIC28-Bsa4</u>
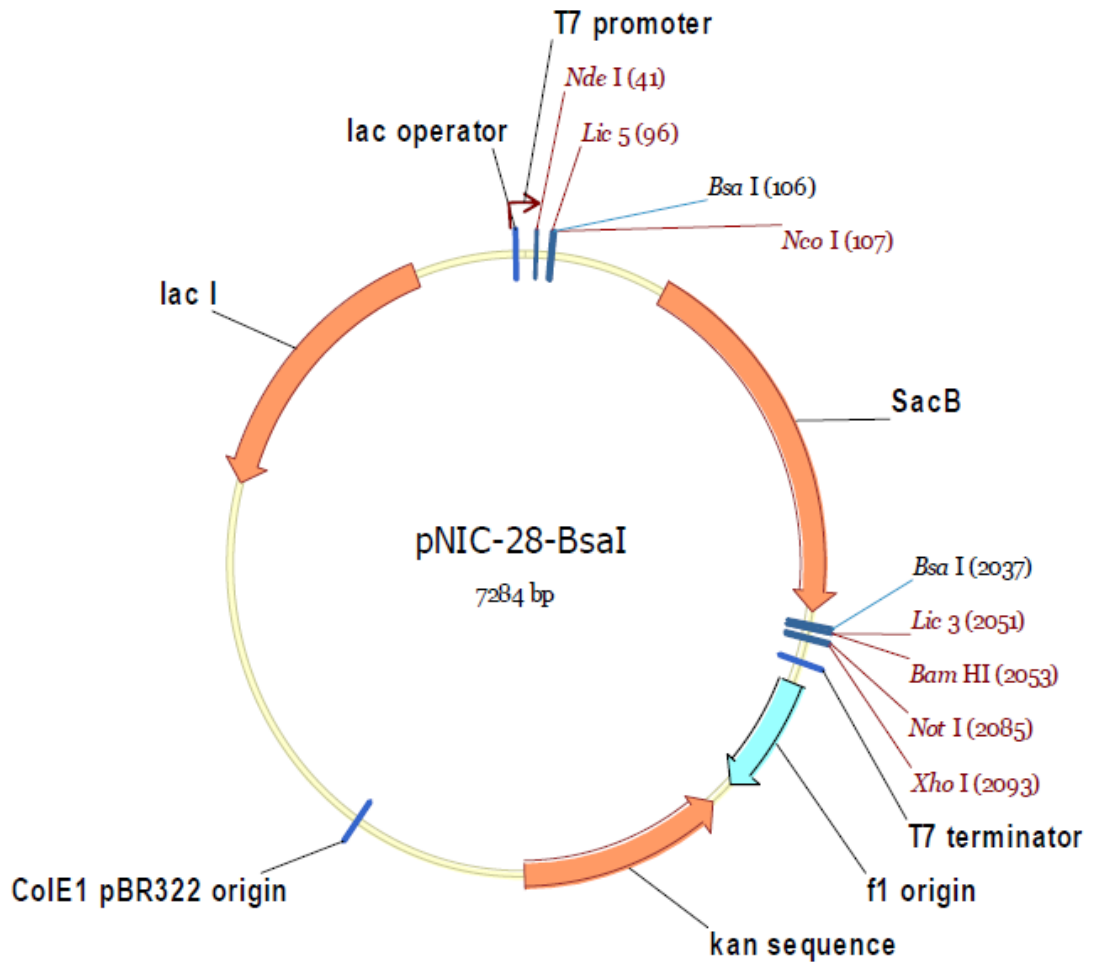
(**http://www.sgc.utoronto.ca/SGC-WebPages/Vector_PDF/pNIC28-Bsa4.pdf**)

pNIC28-Bsa4 vector (SGC, Oxford) contains a 22 N-terminal tag which includes a His_6 tag and TEV cleavage site. It has a kanamycin resistance marker and a T7 promoter to promote high-level protein expression.

A1 (b) pGEX-6P-1

pGEX-6P-1 vector (GE Healthcare, Buckinghamshire, UK) contains a GST N-terminal tag followed by a PresScission enzyme cleavage site. It is resistant to Ampicillin and has a tac promoter for increased protein expression.

A2 - Agar Plate Recipes

<u>LB Agar Recipe (make up 500 ml with water)</u>

  12.5 g   LB

   7 g   Agar

<u>M9 Agar Recipe (make up 500 ml with M9)</u>

  500 ml   M9 Salts

   7 g   Agar

  25 ml   Nutrient Mix

A3 - M9 Media Recipe

M9 Salts pH 7.4 (make up 2 l with water)

12g	Na_2HPO_4
6g	KH_2PO_4
1g	NaCl

M9 Nutrient Mix (make up 2 l with water)

20 ul	1 M $CaCl_2$
4 ml	1 M $MgSO_4$
800 ul	1 mM $FeCl_3$
2 ml	20 mg/ml Thiamine
4 g	D-glucose
2 g	$^{15}NH_4Cl$
1 ml	Metal mix

Metal Mix (make up 100 ml with water)

0.0169 g	$MnSO_4$
0.0175 g	$CuSO_4$
0.115 g	$ZnSO_4$
0.029 g	H_3BO_3

A4 – Purification Buffers

Hepes Buffers used in the purification of His-Cdc25C$_{270-462}$

Lysis/Equilibration Buffer: 50 mM Hepes, pH 7.4, 500 mM NaCl, 10 mM imidazole, and 0.5 mM TCEP

Wash Buffer: 50 mM Hepes, pH 7.4, 500 mM NaCl, 50 mM imidazole, and 0.5 mM TCEP

Elution Buffer: 50 mM Hepes, pH 7.4, 500 mM NaCl, 250 mM imidazole, and 0.5 mM TCEP

Superdex S75 Buffer: 50 mM Hepes, pH 7.4, 300 mM NaCl, and 0.5 mM TCEP

Histrap Purification Buffers

Lysis and Equilibration: 50 mM sodium phosphate (pH 7.4), 500 mM NaCl, 10 mM Imidazole and 1 mM TCEP

Wash Buffer: 50 mM sodium phosphate (pH 7.4), 500 mM NaCl, 70 mM Imidazole, and 1 mM TCEP

Elution Buffer: 50 mM sodium phosphate (pH 7.4), 500 mM NaCl, 200 mM Imidazole, and 1 mM TCEP

TEV Cleavage Buffer: 50 mM Sodium phosphate (pH 7.4), 250 mM NaCl, and 1 mM TCEP

GSTrap Purification Buffers

Lysis/Equilibration/Wash Buffer: 50 mM sodium phosphate (pH 7.5), 250 mM NaCl, 1 mM EDTA, and 1 mM TCEP

Elution Buffer: 50 mM sodium phosphate (pH 7.5), 250 mM NaCl, 1 mM EDTA, 20 mM Glutathione, and 1 mM TCEP
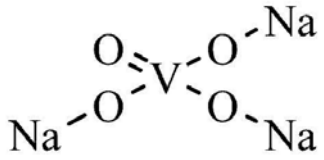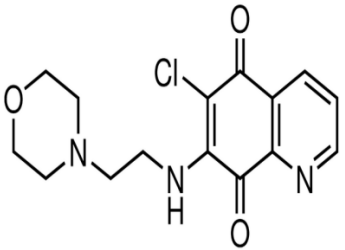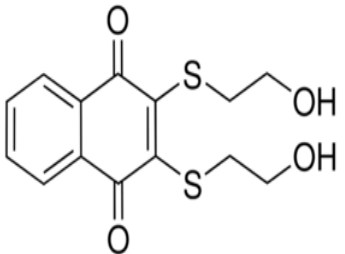
A5 – Table of Compounds (I)
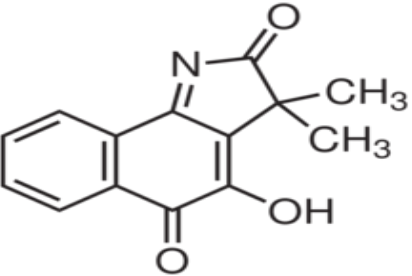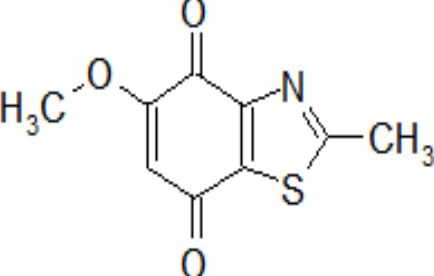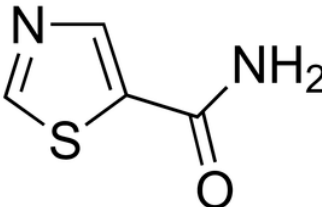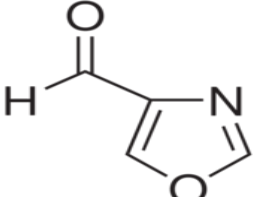
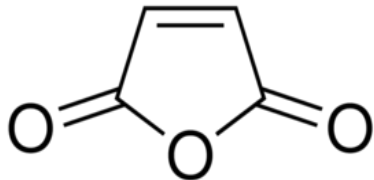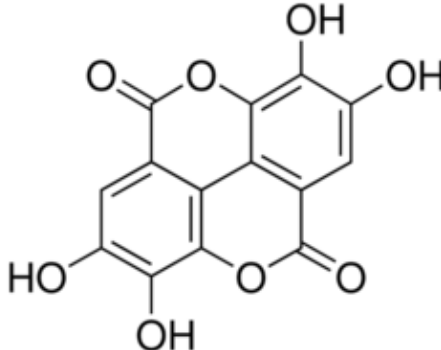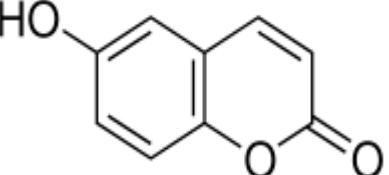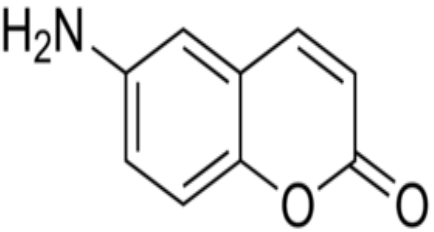| Compound	Source	Product Code
Sodium Orthovanadate	Sigma-Aldrich	S6508
NSC 663284 (Lazo et al., 2001) [6-Chloro-7-(2-morpholin-4-ylethylamino)quinoline-5,8-dione]	Sigma-Aldrich	N7537
NSC 95397 (Lazo et al., 2001) [2,3-Bis[(2-hydroxyethyl)thio]-1,4-naphthoquinone]	Sigma-Aldrich	N1786
BVT.948 (Liljebris et al., 2004) [4-Hydroxy-3,3-dimethyl-2H-benzo[g]indole-2,5(3H)-dione]	Sigma-Aldrich	B6060
5-Methoxy-2-methyl-1,3-benzothiazole-4,7-dione	Atlantic Research Chemicals	XS02005
Thiazole-5-carboxamide	Sigma-Aldrich	CDS020516
4-Oxazolecarboxaldehyde	SIGMA	697915
cis-Aconitic anhydride [*cis*-Propene-1,2,3-tricarboxylic anhydride]	Sigma-Aldrich	217808
Maleic anhydride [2,5-Furandione]	Sigma-Aldrich	M625
Ellagic acid [4,4',5,5',6,6'-Hexahydroxydiphenic acid 2,6,2',6'-dilactone]	Sigma-Aldrich	E2250
6-Hydroxycoumarin	Sigma-Aldrich	642665
6-Amino-chromen-2-one	Sigma-Aldrich	CDS019990
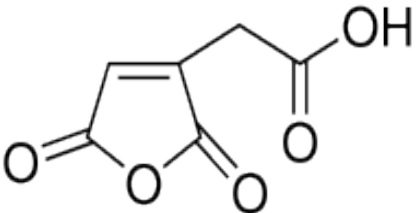
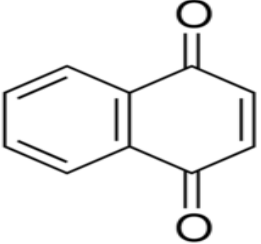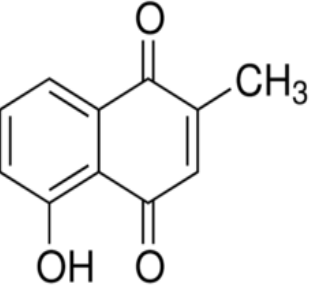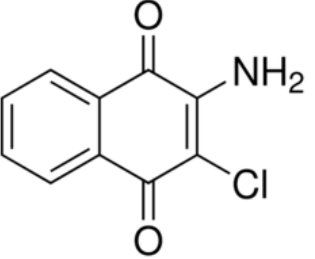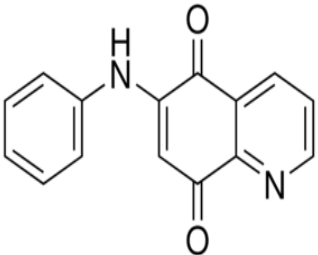| 1,4-Naphthoquinone	Sigma-Aldrich	152757
Plumbagin [5-Hydroxy-2-methyl-1,4-naphthoquinone]	Sigma-Aldrich	P7262
Quinoclamine [2-Amino-3-chloro-1,4-naphthoquinone]	Sigma-Aldrich	32719
6-Anilinoquinoline-5,8-quinone	Sigma-Aldrich	A6563
4-(2-Aminoethyl)morpholine	Sigma-Aldrich	A55004
7,8-dihydro-5(6H)-quinolinone	Sigma-Aldrich	PH009807
(2E)-2-(aminomethylidene)indene-1,3-dione	ChemDiv	4533-0056
3-(1,3-dioxo-3a,4,7,7a-tetrahydroisoindol-2-yl)propanoic acid	ChemDiv	4554-6814

A6 – Table of Compounds (II)
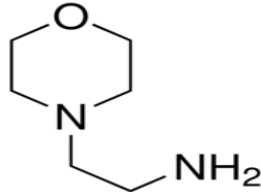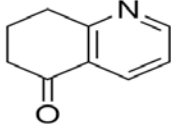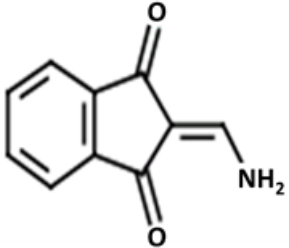
A6 – Table of Compounds (II)

Compound No.	Compound	Structure	Formula	MW	LogP	Interaction by NMR
1	Sodium Orthovanadate		Na_3VO_4	183.91	-4.64	Yes
2	NSC 663284 (Lazo et al., 2001) [6-Chloro-7-(2-morpholin-4-ylethylamino)quinoline-5,8-dione]		$C_{15}H_{16}ClN_3O_3$	321.76	0.19	Yes
3	NSC 95397 (Lazo et al., 2001) [2,3-Bis[(2-hydroxyethyl)thio]-1,4-naphthoquinone]		$C_{14}H_{14}O_4S_2$	310.29	1.13	Yes

4	BVT.948 (Liljebris et al., 2004) [4-Hydroxy-3,3-dimethyl-2H-benzo[g]indole-2,5(3H)-dione]		$C_{14}H_{11}NO_3$	241.24	1.71	Yes
5	5-Methoxy-2-methyl-1,3-benzothiazole-4,7-dione		$C_9H_7NO_3S$	209.22	0.59	No
6	Thiazole-5-carboxamide		$C_4H_4N_2OS$	128.15	-0.074	No
7	4-Oxazolecarboxaldehyde		$C_4H_3NO_2$	97.07	0.30	No

8	cis-Aconitic anhydride [cis-Propene-1,2,3-tricarboxylic anhydride]		$C_6H_4O_5$	156.09	-0.83	No
9	Maleic anhydride [2,5-Furandione]		$C_4H_2O_3$	98.06	-0.18	No
10	Ellagic acid [4,4',5,5',6,6'-Hexahydroxydiphenic acid 2,6,2',6'-dilactone]		$C_{14}H_6O_8$	302.19	0.94	No
11	6-Hydroxycoumarin		$C_9H_6O_3$	162.14	1.51	No

12	6-Amino-chromen-2-one		$C_9H_7NO_2$	161.16	1.07	No
13	1,4-Naphthoquinone		$C_{10}H_6O_2$	158.15	1.67	No
14	Plumbagin (Hafeez et al., 2012) [5-Hydroxy-2-methyl-1,4-naphthoquinone]		$C_{11}H_8O_3$	188.18	1.78	Yes
15	Quinoclamine [2-Amino-3-chloro-1,4-naphthoquinone]		$C_{10}H_6ClNO_2$	207.61	1.69	Yes

16	6-Anilinoquinoline-5,8-quinone		$C_{15}H_{10}N_2O_2$	250.25	1.84	Yes
17	4-(2-Aminoethyl)morpholine		$C_6H_{14}N_2O$	130.19	-1.00	No
18	7,8-dihydro-5(6H)-quinolinone		C_9H_9NO	147.2	1.21	No
19	(2E)-2-(aminomethylidene)indene-1,3-dione		$C_{10}H_7NO_2$	173.05	0.69	Yes
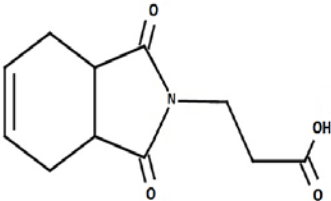
20	3-(1,3-dioxo-3a,4,7,7a-tetrahydroisoindol-2-yl)propanoic acid		$C_{11}H_{13}NO_4$	223.08	0.025	No

Table A6: A list of compounds tested by NMR spectroscopy

The table includes the structures of the compounds, their molecular formula, molecular weight, and predicted LogP values. The structures of the compounds were obtained from the suppliers' websites and the Molinspiration, cheminformatics website was used to predict the LogP values for the compounds.

A7 – Table listing the Software used

Program/Database	Use	Reference	Web Source
ATSAS 2.5.1 Suite of programs (programs used: PRIMUS, GNOM, DAMMIF, DAMAVER, SUPCOMB, and CRYSOL)	Analysis of SAXS data	(Konarev et al., 2003; Franke and Svergun, 2009; Volkov and Svergun, 2003; Kozin and Svergun, 2001; Svergun et al., 1995)	http://www.embl-hamburg.de/biosaxs/software.html
Agilent Technologies Primer Design tool	Primer Design	-	http://www.genomics.agilent.com/primerDesignProgram.jsp
CDSSTR	Analysis of CD data	(Sreerama and Woody, 2000; Whitmore and Wallace, 2004)	http://dichroweb.cryst.bbk.ac.uk/html/home.shtml
Chimera	Visualisation of SAXS ensembles	(Pettersen et al., 2004)	https://www.cgl.ucsf.edu/chimera/
DisMeta server	Protein secondary structure predictions	(Huang et al., 2014)	http://www-nmr.cabm.rutgers.edu/bioinformatics/disorder/
EOM	Analysis of SAXS data	(Bernadó et al., 2007)	http://www.embl-hamburg.de/biosaxs/eom.html
PDB	Database of protein 3D structures	-	http://www.rcsb.org/pdb/home/home.do
ProParam	Prediction of protein properties	Gasteiger et al., 2005)	http://web.expasy.org/protparam/
SEDFIT	Analysis of AUC data	(Brown and Schuck, 2006)	-
I-TASSER	Generation of Cdc25C 3D structure models	(Zhang, 2008)	http://zhanglab.ccmb.med.umich.edu/I-TASSER/