Structure and inhibition of novel cyclin-dependent kinases

Sarah Elisabeth Dixon-Clarke

Nuffield Department of Medicine & University College

University of Oxford

Michaelmas term 2015

Thesis submitted for the degree of Doctor of Philosophy in Clinical Medicine

Structure and Inhibition of novel cyclin-dependent kinase.

Sarah Elisabeth Dixon-Clarke. University College.

Thesis submitted for the degree of Doctor of Philosophy in Clinical Medicine, Nuffield Department of Medicine. Michaelmas 2015.

Abstract

Protein phosphorylation by members of the cyclin-dependent kinase (CDK) family determines the cell cycle and regulates gene transcription. CDK12 and CDK16 are relatively poorly characterised family members containing atypical domain extensions and represent novel targets for structural studies, as well as cancer drug discovery. In this thesis, I developed protocols to express and purify the human CDK12 kinase domain in complex with its obligate partner, CycK. I solved three distinct crystal structures of the complex providing insights into the structural mechanisms determining CycK assembly and kinase activation. These structures revealed a C-terminal kinase extension that folded flexibly across the active site of CDK12 to potentially gate the binding of the substrate ATP. My structures also identified Cys1039 in the C-terminal extension as the binding site for the first selective covalent inhibitor of CDK12, which has enormous potential as a pharmacological probe to investigate the functions of CDK12 in the DNA damage response and cancer. I also identified rebastinib and dabrafenib as potent, clinically-relevant inhibitors of CDK16 and solved a co-crystal structure that defined the extended type II binding mode of rebastinib. Preliminary trials using these relatively non-selective compounds to inhibit CDK16 in melanoma and medulloblastoma cancer cell lines revealed rebastinib as the more efficacious drug causing loss of cell proliferation in the 1-2 micromolar range. Use of the co-crystal structure to design more selective derivatives would be advantageous to further explore the specific role of CDK16. Finally, I identified a D-type viral cyclin from Kaposi's sarcoma-associated herpesvirus that could bind to the CDK16 kinase domain and interfere with its functional complex with human CycY causing loss of CDK16 activity. These studies provide novel insights into the structural and regulatory mechanisms of two underexplored CDK family subgroups and establish new opportunities for cancer drug development.

Declaration

I declare that there are no parts of this thesis or its research herein have been reproduced or accepted for another award or degree or diploma at any other university or learning institution. This thesis contains no other person's work except where stated in text.

Word count: approximately 39,700 words (excluding references)

Acknowledgements

I would like to extend my sincere gratitude to my supervisor, Dr Alex Bullock, for giving me the opportunity to undertake my DPhil project and for his guidance, support and encouragement throughout.

I would also like to acknowledge the members of my group, past and present, for their help along the way to enable me to reach this stage and make my time here a pleasure; Dr Georgina Kerr, Dr Ellie Williams, Dr Caroline Sanvitale, Dr Peter Canning, Dr Sergio Galan-Bartual, Dr Fiona Sorrell, Vikram Ayinampudi, Dr Daniel Pinkas, Zhuoyao Chen and Joshua Bufton. In particular, I need to thank Dr Jon Elkins for his excellent tutelage and advice. Overall, thanks to a great many people at the SGC for their willingness to help, provide valuable advice, their sense of humour and for making it a fantastic environment to work in.

I would like to extend my thanks to our collaborators, Professor Nathanael Gray and his research group for enabling me to be part of the THZ531 project and the exciting results this brought. With thanks to Professor Kei Sakamoto and Saif Shehata for the collaboration on the CDK16 project, the sharing of knowledge and the results this has produced.

I would like to thank the Nuffield Department of Medicine for awarding me this prize studentship to undertake my DPhil. University College has made my time here particularly special, for all the opportunities it gives, the atmosphere it fosters and the friends I have made there.

I would like to acknowledge my family and in particular my Mum, Dad and brothers, Mat and Dan, for their unconditional love and faith in me. Lastly, I thank Stefano, for all his support.

Contents

Abbreviations

Å	Angstrom
A-loop	Activation loop
ADP	Adenosine 5’-diphosphate
AMP-PNP	Adenylyl imidodiphosphate
AML	Acute myeloid leukaemia
ATP	Adenosine-5'-triphosphate
BCA	Bicinchoninic acid *assay*
BCS	Basic ChemSpace
BsaI	*Bacillus stearothermophilus* restriction enzyme
BSA	Bovine serum albumin
CAK	Cdk-activating kinase
CCLE	Cancer Cell Line Encyclopedia
CCP4	Collaborative Computational Project Number 4
CDK	Cyclin-dependent kinase
cDNA	Complementary DNA
CML	chronic myeloid leukaemia
CO_2	Carbon dioxide
CS	Carcinosarcoma
CV	Column volume
CycA	Cyclin A
CycK	Cyclin K
CycL1	Cyclin L1
CycT1	Cyclin T1
Da	Daltons
DSIF	DRB sensitivity inducing factor
DFG	Asp-Phe-Gly
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
DSF	Differential scanning fluorimetry
DTT	Dithiothreitol

FBS	Fetal bovine serum
FDA	United States Food and Drug Administration
FL	Full length
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GBM	Glioblastoma multiforme
G-loop	Glycine-rich loop
GST	Glutathione S-transferase
HEPEs	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
HCS	Hampton Crystal Screen
HCl	Hydrochloric acid
HEK293	Human embryonic kidney cells
HIN	Hampton screen index
HIV	Human immunodeficiency virus
HPLC	High-performance liquid chromatography
IC_{50}	Half maximal inhibitory concentration
IDA	Iminodiacetic acid
ITC	Isothermal titration calorimetry
JCSG	Joint Centre for Structural Genomics
K_d	Dissociation constant
kDa	Kilodaltons
KD	Kinase domain
KSHV	Kaposi's sarcoma-associated herpesvirus
LB	Lysogeny broth
LC-MS	Liquid chromatography- mass spectrometry
LFS	Ligand friendly screen
LIC	Ligation-independent cloning
MBL	Medulloblastoma
MBP	Myelin basic protein
MGC	Mammalian Gene Collection
MES	2-(N-Morpholino)ethanesulfonic acid, 4-Morpholineethanesulfonic acid
MME	Monomethyl ether 550

$MgCl_2$	Magnesium chloride
MIDAS	Molecular dimensions crystallization screen
mM	Millimolar
$MnCl_2$	Manganese chloride
MM	Multiple myeloma
MS/MS	Tandem mass spectrometry
MW	Molecular weight
N/A	Not applicable
NaCl	Sodium chloride
NaF	Sodium fluoride
Na_3VO_4	Sodium orthovanadate
NCI60	US National Cancer Institute (NCI) 60 human tumour cell line anticancer drug screen
NELF	Negative elongation factor
nM	Nanomolar
P^{32}	Gamma phosphate group
PACT	pH-, anion- and cation-testing
PCR	Polymerase chain reaction
PBS	Phosphate buffered saline
PDB	Protein Data Bank
PEG	Polyethylene glycol
pI	Isoelectric point (pI)
PKA	Protein kinase A
P-loop	Phosphate binding loop
pRb	retinoblastoma tumour-suppressor gene-encoded protein
p-TEFb	Positive transcription elongation factor b
PVDF	Polyvinylidene fluoride
qPCR	Quantitative Polymerase chain reaction
RNA	Ribonucleic acid
RNAPII	RNA polymerase II
RPM	Revolutions per minute
RPMI	Roswell Park Memorial Institute medium
SC	Serous cystadenocarcinoma

S.D.	Standard deviation
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sf9	Spodoptera frugiperda
siRNA	short interfering Ribonucleic acid
T-ALL	T-cell Acute lymphoblastic leukaemia
TCEP	tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TFIID	Transcription factor II D
TFIIH	Transcription factor II H
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
TWEEN	Polyethylene glycol sorbitan monolaurate
Tm	Melting temperature
WT	Wild type
xg	Times gravity
µM	micromolar

Standard one or three letter abbreviations are used to denote nucleotides and amino acids.

Chapter 1 Introduction

1.1 Role of protein phosphorylation in cell signalling

The most widespread protein post-translational modification used in intracellular signal transduction is protein phosphorylation [1]. Protein phosphorylation classically involves the transfer of the γ-phosphate from adenosine triphosphate (ATP) to the hydroxyl moiety of specific amino acids within a recipient protein, including serine, threonine or tyrosine. However, other amino acids can also be phosphorylated such as histidine residues [2]. Phosphorylation is a reversible modification catalysed by protein kinases and removed by phosphatases [3]. Phosphorylation can modify the function of the target protein. For example, it can alter its enzymatic activity or its subcellular localization. It can also help to create a binding site to initiate or disrupt protein–protein interactions[4]. Finally, it can also help to protect or mark a protein for destruction by the ubiquitin-proteasome system [4, 5].

Protein phosphorylation affects every basic cellular process, including metabolism, growth, cell division, differentiation, motility, organelle trafficking, membrane transport, muscle contraction, neurobiology, immunity and morphogenesis [1, 6]. Indeed, it is estimated that 30% of all cellular proteins are phosphorylated on at least one residue [7]. Deregulated phosphorylation is also associated with multiple diseases including cancer, vascular disorders, and diabetes, as well as neurological and inflammatory diseases. As a consequence, protein kinases have become attractive therapeutic targets. Currently, 28 small molecule protein kinase inhibitors are approved by the United States Food and Drug Administration (FDA), and many hundreds of clinical studies targeting kinases are ongoing [12]. This thesis focuses on two kinases that have been linked to cancer.

1.2 The human kinome

Protein kinases are one of the largest gene families in eukaryotes making up approximately 2% of the human genome. Sequencing of the human genome has enabled the updating of early

protein kinase classifications and identified the full protein kinase complement, the so-called "kinome" [1, 8]. The generation of a phylogenetic tree for the human kinase catalytic domains has also enabled rapid comparison of related kinases for structural and functional insight in both normal and pathobiology (Fig. 1.1) [1, 8]. To date 538 protein kinases have been identified in the human kinome (http://kinase.com/web/current/), grouped within a hierarchy consisting of eight phylogenetic branches, consisting of CMGC, GYC, TK, TKL, STE, CK1, AGC, CAMK, other and atypical kinases (see Fig. 1.1 for a description of these abbreviations).

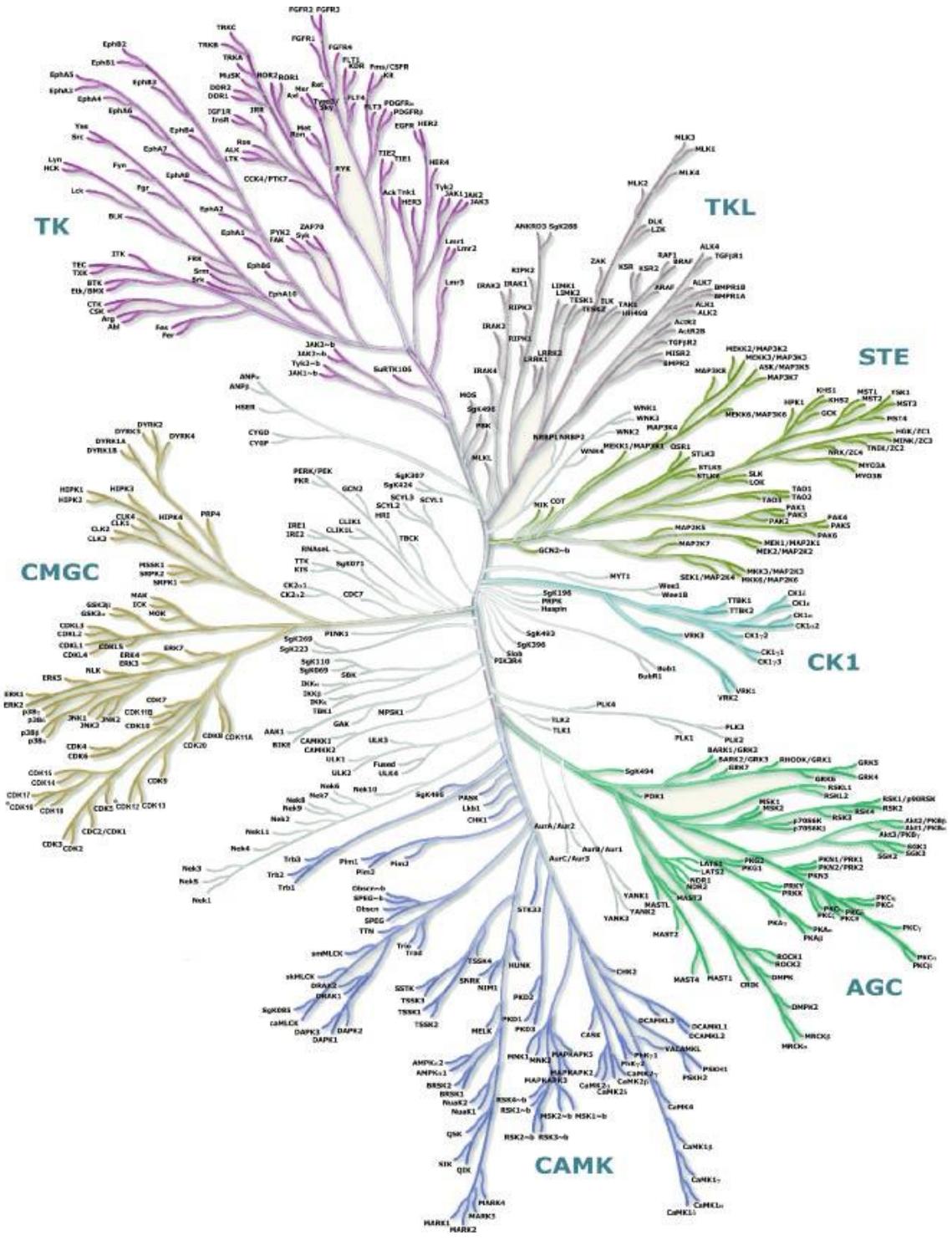


Figure 1.1. Schematic phylogenetic tree of the human kinome. Adapted from Manning *et al*., 2002 [1]. Abbreviations: AGC, named after the protein kinase A, G, and C families (PKA, PKC, PKG); CAMK, calmodulin/calcium-regulated kinases; CK1, named after casein kinase 1; CMGC, named after the CDK, MAPK, GSK3 and CLK families, STE, homologues of the yeast STE7, STE11 and STE20 genes; TK, tyrosine kinase; TKL, tyrosine kinase-like. Note, atypical kinases are not shown. CDK12 and CDK16 (*) are the focus of this thesis.

1.3 The kinase domain

Protein kinases must distinguish a specific phosphorylation site against a background of ~700,000 potentially phosphorylatable residues. Specificity is achieved in part by the structure of each kinase's catalytic domain. However, many kinases are structurally similar. Thus, additional mechanisms have evolved to ensure specificity. For instance, both local and distal (docking) interactions may occur between the kinase and substrate [9]). There may also be spatial and temporal regulation, particularly when kinases assemble into larger complexes, which may include other regulatory scaffolds and adaptor proteins [7].

The kinase domain was initially classified into 12 subdomains by Hanks and Hunter based on conserved sequence motifs in the primary structure (reviewed in [10]). These descriptions have been largely superceded by knowledge of the kinase domain structural fold derived from X-ray crystallography. The first crystal structure of a protein kinase was that of the cAMP-dependent protein kinase, also known as protein kinase A (PKA) [11]. Subsequently, additional structural work revealed the highly conserved features that determine the active catalytic domain [12-14].

Protein kinases are highly dynamic molecular switches that show a high degree of conformational plasticity in the way in which they may transition between an inactive and active state [12]. Kinases are often maintained in an "off" state that has minimal activity and a fully active state is necessarily kept under multiple layers of control. As nearly all protein kinases catalyze the same phosphotransfer reaction, unsurprisingly upon activation they adopt structurally similar conformations, but can have diverse inactive and intermediate conformations [14].

1.4 Structural and functional paradigms from the cyclin-dependent kinase family

1.4.1 The CDK family

Cyclin-dependent kinases (CDKs) belong to the CMGC (named after the CDK, MAPK, GSK3 and CLK families) branch of the kinome, which has 61 members in total (Fig. 1.1). CDKs are recognised as a diverse family of proline-directed serine/threonine kinases. In other words, they

show selectivity to phosphorylate serine and threonine residues (P0 position) that are followed by a proline (the so-called P+1 position). The best characterised CDK members act as key regulators of the cell cycle (CDK 1-4, 6), neuronal function (CDK5) or gene transcription (CDK 7-9, 11) [15, 16]. More recently, a number of novel CDKs have been described for which the biological roles largely remain to be elucidated (CDK10, CDK12-20) (Figure 1.2) [17]. The CDKs are so named due to the discovery of their dependence on cyclin binding partners for their activity [17]. These specific, regulatory 'cyclin' subunits assemble with their cognate CDK to generate the heterodimeric, catalytically active kinase. The first cyclin proteins (CycA and CycB) were first identified by their fluctuating levels during the cell cycle and their co-purification with CDK1 (also known as CDC2, cell division control protein 2) [18].

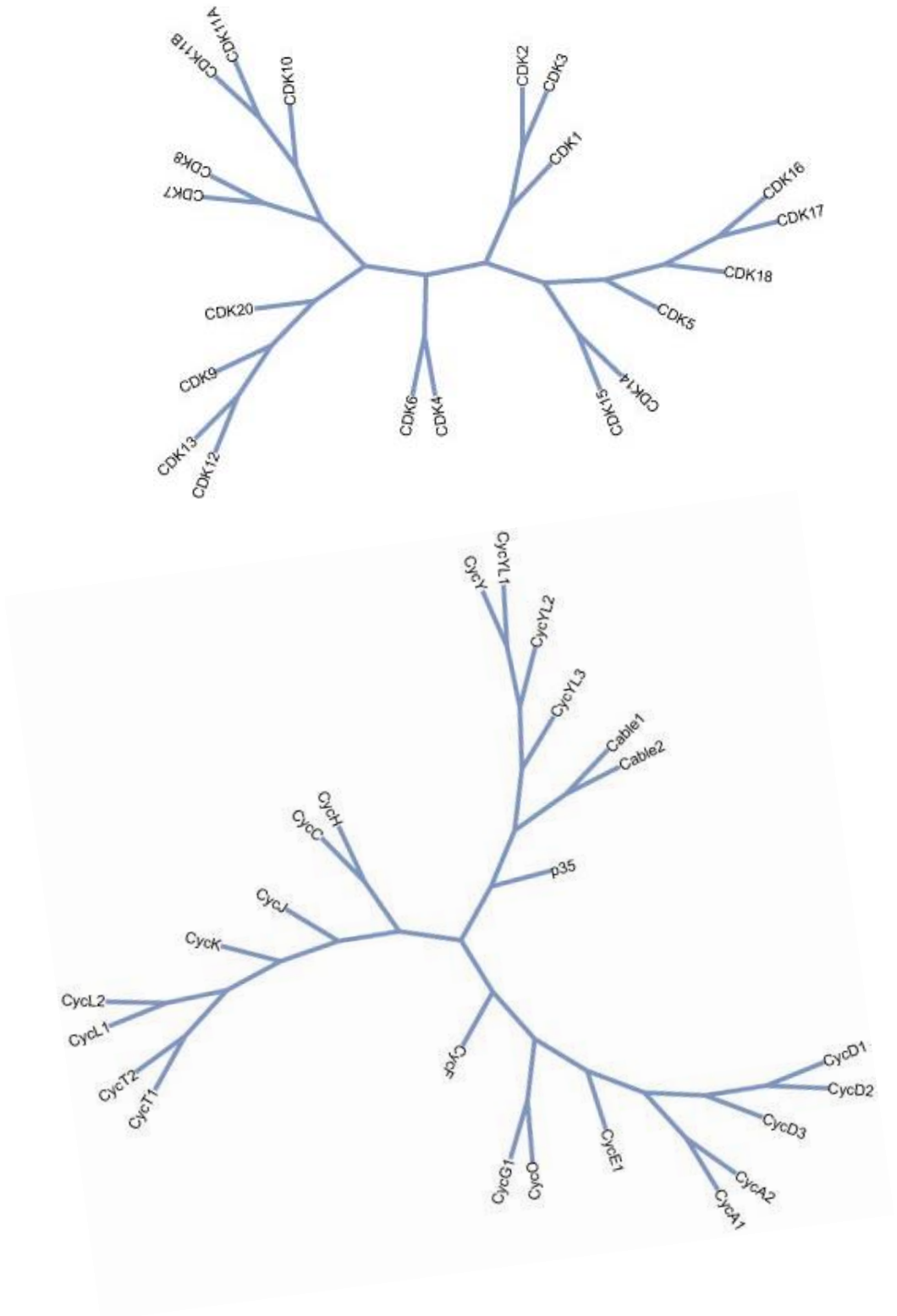


Figure 1.2. Phylogenetic classification of the CDK and cyclin protein families. All full-length CDK and cyclin sequences were obtained from Uniprot [19] (http://www.uniprot.org/) and analysed using the SGC SATurn program (http://www.thesgc.org/saturn).

1.4.2 Overview of the CDK and cyclin domain folds

The conserved kinase domain fold, as exemplified by CDK2 (Fig. 1.3a), consists of two structurally distinct subdomains or lobes. The amino-terminal lobe ("N-lobe") consists of a five-stranded β-sheet and one prominent α-helix, the αC helix (Fig. 1.3a). The carboxyl-terminal-lobe ("C-lobe") is larger and is predominantly α-helical. The loop between the β5 and αD secondary structural elements is the only part that connects the N-lobe to the C-lobe [10, 12, 14]. The kinase domain fold of CDKs contains two specific inserts that help the formation of protein-protein interactions. In the N-lobe, the β3-αC loop is atypically large to support the binding of the cyclin partner (described further below and shown in Fig. 1.4). A CDK/MAPK-specific insert is also present in the αG-αH loop with propensity to form two 3_{10} helices (Fig. 1.3a). This C-terminal insert in CDK2 mediates an interaction with Cks1 that targets CDK2 to different phosphoproteins, as well as with the phosphatase KAP [20, 21].

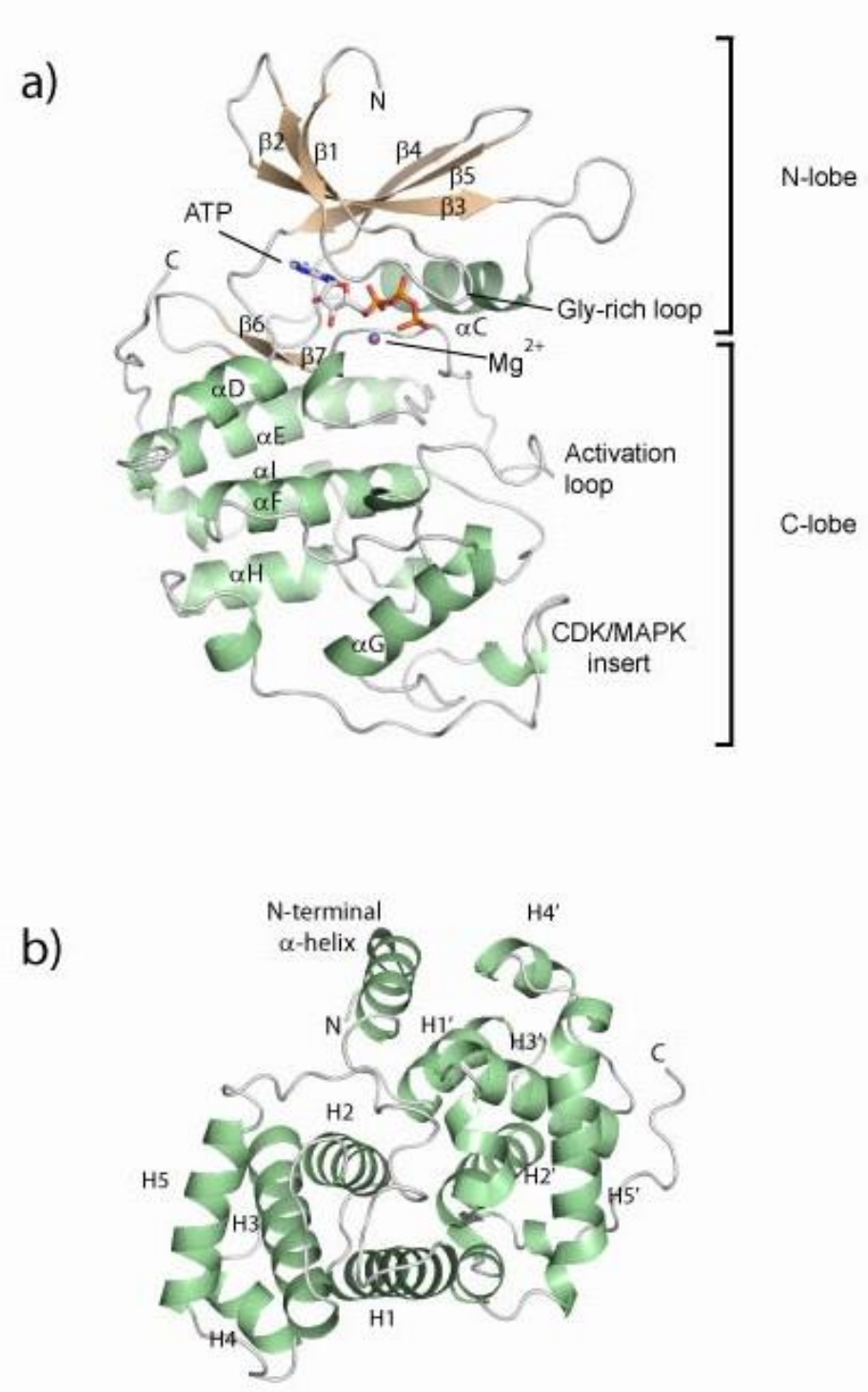


Figure 1.3. Overview of the CDK2 and cyclin A structures. a) Ribbon representation of the CDK2 crystal structure (PDB ID: 1FIN). The kinase fold contains two subdomains known as the N-lobe and C-lobe. ATP sits in a cleft between the two lobes. Secondary structure elements are labelled, including the unphosphorylated activation loop. Bound CycA is not shown for clarity. **b)** Overview of the cyclin fold in CycA (PDB ID: 1FIN). CycA contains two cyclin box domains each comprising 5 helices as well as a long N-terminal helix αN.

Between the two kinase lobes a deep cleft is formed that serves as the ATP-binding site [10, 12, 14]. ATP and Mg^{2+} interaction is enabled by multiple residues across both lobes (Fig. 1.3). ATP sits beneath a highly conserved loop in the N-lobe between the β1 and β2 strand, which contains a conserved glycine-rich motif (GXGXφG) where φ is usually tyrosine or phenylalanine. This loop is variably known as the glycine-rich loop ("G-loop") or phosphate-binding loop ("P-loop") [13, 14] (Fig. 1.4). The two β-strands typically sit above the adenine ring, whereas the glycine-rich loop mediates hydrogen bond interactions with the phosphate moieties and positions the ATP γ-phosphate for phosphoryl transfer [12]. The activation loop harbours the site of activating threonine phosphorylation, Thr160 in CDK2, and is therefore also sometimes described as the T-loop. Further description of important residues for phosphotransfer is given below.

The cyclin box is a compact domain of approximately 100 amino acids that folds into a bundle of five α-helices. Most cyclin proteins contain two cyclin box domains (Fig. 1.3b). The more conserved N-terminal domain (cyclin box A) mediates the interaction with the CDK, whereas the C-terminal domain (cyclin box B) is generally more divergent. CycA contains an additional helix at its N-terminus that contributes to CDK2 binding as well as a short C-terminal helix (Fig. 1.3b). Whilst CDKs are highly dynamic, the cyclin fold appears relatively rigid between apo- and CDK-bound conformations [22].

1.4.3 Structural basis for CDK/cyclin assembly and activation

The CDK subunit is completely inactive without its cyclin binding partner. Cyclins are best known as temporal regulators, but may also facilitate substrate recruitment and the subcellular localisation of CDKs [23]. Cyclin binding is centred on a unique "PSTAIRE" sequence motif within the αC helix of classical CDKs, such as CDK2, resulting in rearrangement of key catalytic residues [24]. This motif varies across the CDK family contributing to the cyclin-binding specificity; for example, it is PITAIRE in CDK12 and PCTAIRE in CDK16 [25, 26]. The αC helix is a highly dynamic regulatory element. In inactive kinases, it is frequently rotated outward from the active

site breaking side chain interactions that are critical for catalysis [10] (Fig. 1.4). Cyclin binding induces the αC to rotate inwards imparting partial activation (Fig. 1.4). Further structural rearrangements are required for a fully active CDK which is orchestrated following activation loop phosphorylation (discussed further below).

The binding surface on the cyclin subunit is largely comprised of residues from the α3 and α5 helices in the first cyclin box. These cyclin residues make extensive contact with the αC, as well as the surrounding β-structure in the kinase N-lobe (Fig. 1.4). In the CDK2/CycA complex, the N-terminal CycA helix αN also makes additional contacts with the kinase C-lobe (Fig. 1.4).

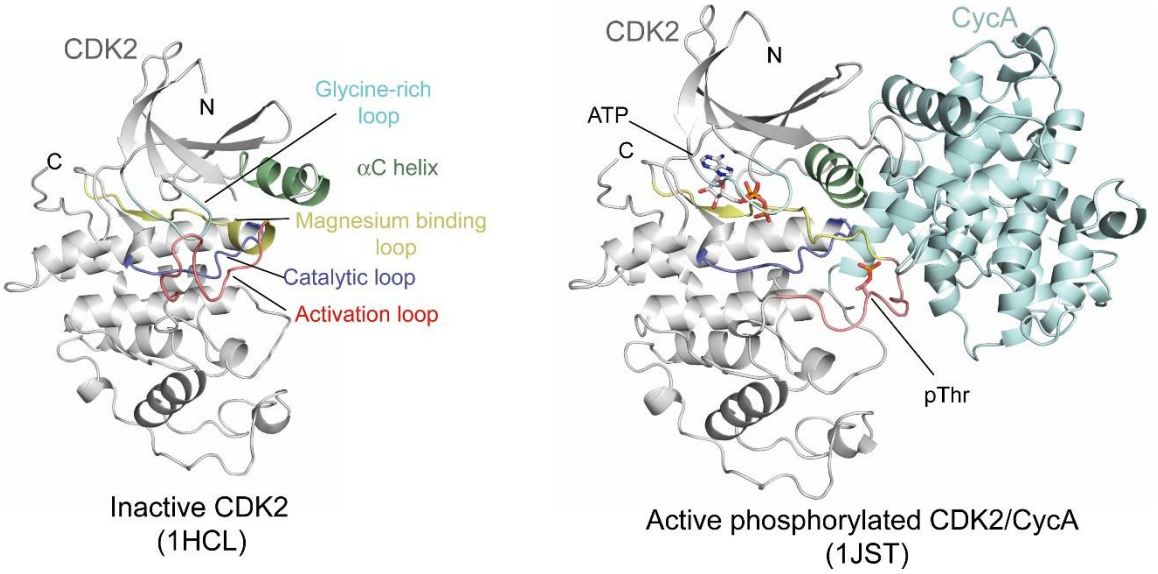


Figure 1.4. Structural basis for assembly of active CDK2/CycA. CDK2 is shown in a monomeric, inactive state. The αC helix (green) is rotated outwards from the ATP pocket disrupting critical interactions for catalytic activity. The activation loop (red) is positioned so that it blocks the binding pockets for ATP and substrate. Upon cyclin binding, there is an inward repositioning of the αC helix, enabling a critical ion pair to form between the Glu51 (αC) and the Lys33 (β3). Cyclin binding displaces the activation loop onto the C-lobe, enabling substrates to gain access to the active site. The active phosphorylated CDK2/CycA complex is shown. CAK phosphorylation of a specific threonine residue effectively locks the activation loop in an open conformation. Overall, these steric changes correctly position critical residues for the phosphotransfer reaction, including the DFG motif that coordinates Mg^{2+} ions and ATP in the binding cleft [6, 14].

Cyclin binding displaces the activation loop from an inhibitory position that may occlude the ATP pocket and onto the C-lobe, enabling substrates to gain access to the active site. In CDK1

this steric alteration is thought to make a specific threonine residue in the activation loop available for phosphorylation by a CDK-activating kinase (CAK), effectively locking the loop in an open conformation. However, some studies suggest that CDK2 is phosphorylated before its cyclin assembly [27]. The activation loop phosphorylation is required for full kinase activation. In particular, the phospho-threonine in CDK2 acts as an organising centre forming hydrogen bond interactions with the αC, the bound CycA subunit and the substrate (Fig.1.4). Overall, these steric changes correctly position key catalytic residues in the CDK to orientate bound ATP and substrate for the phosphotransfer reaction [6, 14]. However, several CDK family members appear not to require activation loop phosphorylation, including CDK5 and proteins of the PCTAIRE family (CDK16-18) [15, 22].

In mammals, the CDK-activating kinase (CAK) is a complex of CDK7, CycH and the RING finger protein, MAT1 [28]. CDK7/CycH is also part of the general transcription factor TFIIH, where it is involved in promoter clearance and progression from the pre-initiation to the initiation stage [28, 29],

1.4.4 Critical motifs for kinase activation

The inward movement of the αC helix induces a closed conformation of the N- and C-lobes that is coupled to nucleotide binding ability [14]. This establishes a critical salt bridge between a glutamate within the "PSTAIRE" sequence of the αC helix and a conserved catalytic lysine in the "VALK" motif in the β3 strand, a hallmark of the active state in protein kinases (Fig. 1.5) [24, 30].This rearrangement correctly positions the lysine to coordinate the α- and β-phosphates of the ATP for catalysis [10, 12, 14] (Fig. 1.5).

Upon phosphorylation, the activation loop becomes ordered in the C-lobe creating a short antiparallel β-sheet between β9 in the activation loop and β6 in the protein core [10]. The catalytic loop lies between β6 and β7 and is so named as it contains the catalytic "HRD" motif [10]. The histidine packs below the activation loop to stabilise its "DFG" motif. The arginine forms an ion-

pair with the phospho-threonine, which helps further to rotate the DFG motif into the required orientation for catalysis [10, 14]. The catalytic aspartate contributes to the phospho-transfer reaction. The magnesium binding regions follow the catalytic loop and include a conserved lysine as well as the DFG motif (Fig.1.5). The aspartate residue is an invariant catalytic residue that is responsible for chelating the magnesium ion(s) that position the phosphate for phospho-transfer [31]. In active kinases the phenylalanine adopts a buried "in" position packing in a hydrophobic pocket by the αC helix [13]. The glycine residue is highly conserved and likely contributes to the loop plasticity [12].

Active kinases have a highly conserved spatial motif, which comprises four hydrophobic residues that align to link the two kinase lobes in a so-called hydrophobic spine [10]. In CDK2, this is comprised of Leu66 (β4) and Leu55 (αC) from the N-lobe and Phe146 (activation loop) and His125 (catalytic loop) from the C-lobe (Fig. 1.5)

The active enzymes can be turned off in various ways. For example, the complex can be subject to inhibitory phosphorylations, such as Thr14 and Tyr15 in the glycine-rich loop of CDK2 [22]. The complex might also associate with a tight-binding inhibitor protein or the cyclin subunits may be degraded by highly specific ubiquitin-mediated proteolysis [32].

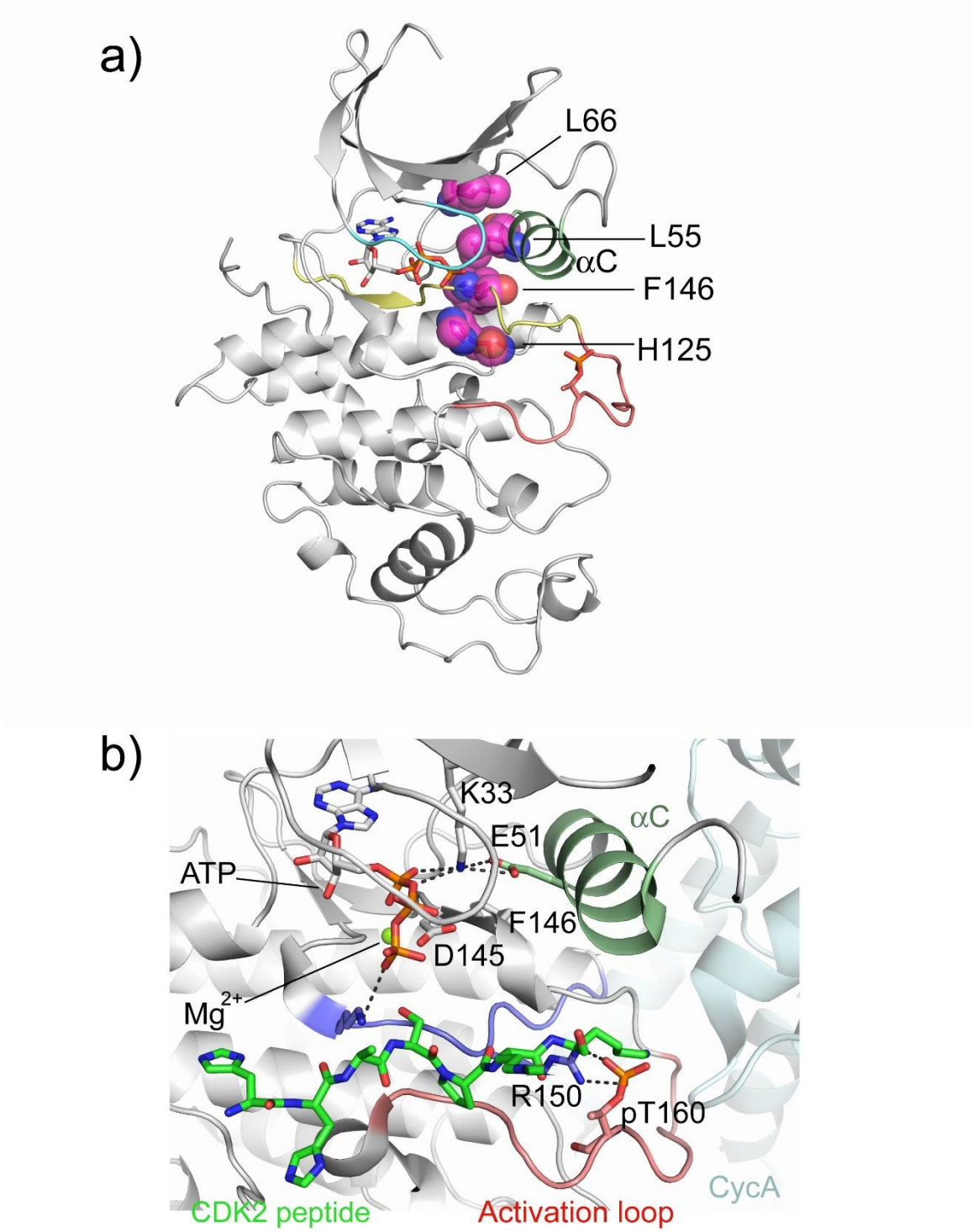


Figure 1.5. Key motifs for kinase activation. a) Activation of CDK2 aligns a hydrophobic spine comprising Leu55, Leu66, His125 and Phe146 that stabilizes the correct orientation of the N and C-lobes. b) The correct inward position of the αC enables a conserved salt bridge between Lys33 and Glu51. Additional DFG and HRD motifs contribute to ATP-Mg^{2+} and phospho-transfer. This structure shows the CDK2 optimal consensus substrate peptide, HHASPRK, bound to CDK2. (PDB IDs 1JST and 1QMZ)

1.4.5 Substrate recruitment

Kinase substrates bind across the C-lobe at the front of the ATP-binding pocket (Fig. 1.5b). CDK kinases have specificity for substrates containing a consensus SPxR/K sequence, where x is any amino acid [33].

CDKs have an unusual conformation of the activation loop that rotates Val164 (CDK2 numbering) away from the bound substrate. This gives the CDK kinases specificity for proline at the P+1 position as all other amino acids would be disfavoured by having an unsatisfied hydrogen bond donor in the kinase-substrate complex. A basic residue is preferred at the position P+3 as it can form an ion-pair with the phospho-threonine [22, 33].

Some cyclin subunits can also contribute to substrate binding and allow diversification of the consensus phosphorylation motif. For example, cell-cycle cyclins such as CycA have a hydrophobic pocket for docking of a distal recruitment motif (R/KxL) on the substrate [33]. Structural studies suggests the linker between the substrate's phosphorylatable residue and their recruitment motif is flexible and there does not appear to be a preferred path on the cyclin surface, thus diverse substrates may traverse this path as long as the linker is sufficient to span the distance [22].

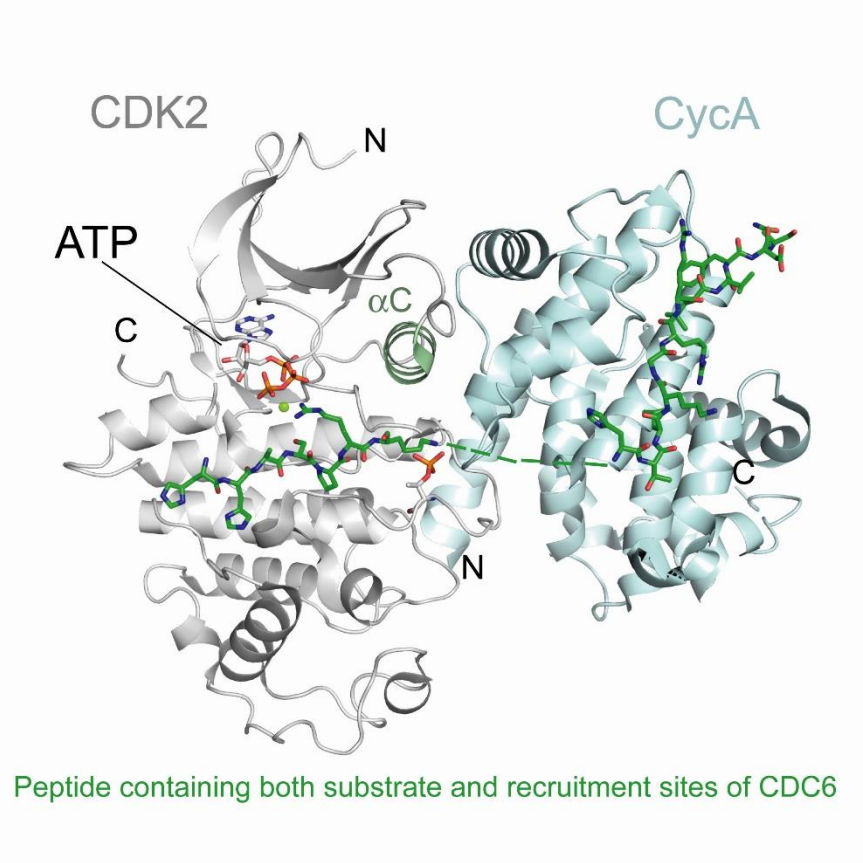


Figure 1.6. Substrate recruitment. A CDC6 derived peptide containing substrate peptide and the RXL-recruitment peptide on the CycA recruitment site (green). The linker region could not be built and thus a dashed line has been added to connect the two parts (PDB ID 2CCI) [33].

1.5 CDK family members roles

1.5.1 Cell cycle CDKs

The eukaryotic cell cycle is divided into four distinct phases: G_1, S, G_2, and M. CDK/cyclin complexes have been described as the engines and clocks that drive the switch like transitions [23, 34]. CDK4 and CDK6 bind D-type cyclins (CycD1-3) and act in early G1 phase; CDK2/CycE triggers S phase while CDK2/CycA along with CDK1/CycA regulate S-phase completion [34]. CDK1/CycB directs the cells into mitosis [35]. While these CDK concentrations are relatively constant, their cyclins subunit levels oscillate [34]. In addition, two classes of Cdk-inhibitory subunits (CKIs), the Cip/Kip and Ink4 families, control these CDK/cyclin complexes [23]. The Cip/Kip family (p21, p27, and p57) inhibit CDK2 and CDK4/6 complexes. The Ink4 family members (p15, p16, p18, and p19) exclusively bind to and inhibit CDK4 and CDK6 [32].

1.5.2 Transcriptional CDKs

CDKs partake in transcriptional regulation by phosphorylation of the carboxy-terminal domain (CTD) of the largest subunit of RNA polymerase II, Rbp1. The transcriptional cycle of initiation, elongation and termination is coordinated by the phosphorylation of the intrinsically unstructured conserved heptad ($Y_1S_2P_3T_4S_5P_6S_7$) repeats in the CTD [36, 37]. Within the consensus Ser2, Ser5 and Ser7 can all be phosphorylated [38]. The CTD acts as a scaffold for factors and is integral to the coupling of transcription to co-transcriptional events, including mRNA capping, splicing, editing, 3' end formation and histone modifications [39-41]. The CTD phosphorylation status changes dynamically during transcription, with the CTD kinases and phosphatases producing distinct phosphorylation patterns to give a complex regulatory code, known as the "CTD code", to direct recruitment of specific factors [41, 42].

Transcription factors bind directly to the DNA to modulate the transcription initiation step. Initiation requires the assembly of a pre-initiation complex (PIC) at which stage the CTD is hypo-phosphorylated [41, 43]. The PIC complex also includes the Mediator complex and general transcription factors, including TFIIH [43]. The CDK7/CycH/MAT1 which forms the mammalian CDK-activating kinase (CAK), is also part of the general transcription factor TFIIH. After PIC assembly, CDK7 phosphorylates the CTD at Ser5 during initiation and early transcription and is capable of phosphorylating Ser7 [40, 43]. CDK8 associates with CycC to form part of the Mediator complex in transcription and phosphorylates the CTD at Ser2 and Ser7, *in vitro* [41, 44]. The contribution of CDK8 to transcription *in vivo* is unclear. The Ser5 phosphorylation signal decreases gradually towards the poly-adenylation site, whereas the Ser7 phosphorylated levels remain high throughout the transcription cycle [43, 45].

RNAPII is blocked soon after initiation by DRB sensitivity–inducing factor (DSIF) and negative elongation factor (NELF). CDK9/CycT1 constitutes the active form of the positive elongation factor (P-TEFb) which coordinates the elongation phase of transcription by phosphorylating Ser2 positions in the CTD, as well as other elongation factors [46]. CDK9/CycT1

phosphorylates both DSIF and NELF, to release NELF and convert DSIF into a processivity factor [47]. Phosphorylation at Ser2 releases the CTD from pausing into an elongation mode [43]. Ser2 phosphorylation signals are highest downstream of the poly-A site, which modulates the recruitment of 3'-RNA processing factors [43, 45]. CDK9/CycT1 was the first transcriptional CDK/cyclin complex to have its structure determined. It revealed relative plasticity in the position of the cyclin relative to the CDK [46]. The interface was also sparse compared to cell cycle CDKs due to a 25° rotation in the relative position of the cyclin resulting in its interaction with the kinase N-lobe only [22]. CDK9 has been shown to autophosphorylate, in contrast to cell-cycle CDKs whose activation loop phosphorylation is strictly undertaken by CAK [48]. More recently, CDK12 and CDK13 have also been shown to phosphorylate the CTD at Ser2.

1.5.3 Novel CDK family members

1.5.3.1 CDK12-13

Human CDK12 (CrkRS) and its close homologue CDK13 (CDC2L5) are unusually large CDK family members (1,490 and 1,512 amino acids, respectively) containing a central kinase domain with flanking arginine-serine-rich (RS) domains [49-51]. CDK12 is conserved throughout evolution from yeast (Ctk1) to human [52].

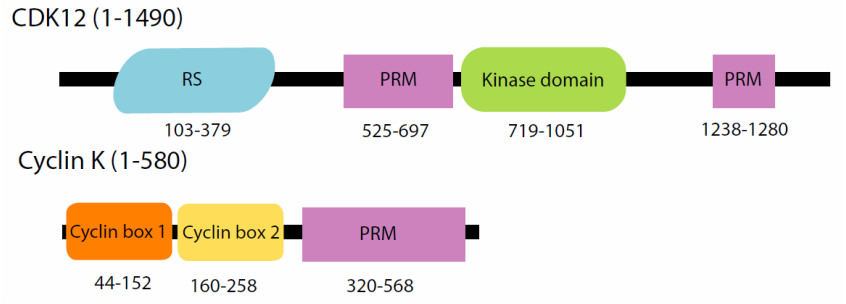


Figure 1.7. CDK12 and Cyclin K domain organization. Arginine/serine-rich (RS), proline-rich motif (PRM).

CDK12 has a threonine in the activation loop at an analogous position to the activatory threonine in CDKs that are phosphorylated by the Cdk-activating kinase (CAK) [50]. There are 20 arginine/serine (RS) motifs in the N-terminus of CDK12, which are characteristic structural features of members of serine/arginine (SR)-rich proteins that are involved in pre-mRNA processing [26] (Fig. 1.7). CDK12 and CDK13 contain a PITAIRE motif at the conserved position of the PSTAIRE α-helix cyclin binding motif found in CDK2.

CycK was first identified as a protein that could substitute for G1-phase cyclin genes in the budding yeast, *Saccharomyces cerevisiae* [53]. CycK is most closely related to human cyclins C and H and was dubbed a as a "transcriptional" cyclin, which was subsequently shown to be another cyclin binding partner for CDK9 [53, 54]. The CDK9/CycK complex could functionally substitute for CDK9/CycT *in vitro*. However, more recent studies showed that human CycK associates with CDK12 and CDK13 in two separate complexes, but not with CDK9 [55, 56]. Isoform 1 of human CycK is a 65 kDa, 580 amino acid, protein containing two canonical cyclin boxes that show similarity with cyclin T1, as well as a divergent C-terminal proline-rich region [55, 57] (Fig. 1.7). CDK12 and CDK13 may also regulate transcription via assembly with CycK [26, 55, 56, 58]. Ser2 phosphorylation is performed by both CDK12/CycK and CDK9/CycT, which are identified as orthologues of *S. cerevisiae* Ctk1/Ctk2 and Bur1/Bur2, respectively [58, 59]. *In vitro* similarly to other CTD kinases, CDK12 was shown to be promiscuous and able to phosphorylate all three CTD serines [38]. However, CDK12 showed the highest activity for CTD when pre-phosphorylated at Ser7 [45]. Knockdown of CDK12 had modest effect on global transcription, in contrast to CDK9 knockdown, however CDK12 preferentially regulates the expression of DNA damage response genes, such as BRCA1, ATR, ATM and Fanconi anemia proteins [55]. CDK12 also assembles with multiple RNA processing factors and couples the transcription of pre-mRNA to 3' end formation [40, 50, 52, 60, 61]. Alternative splicing of pre-mRNA is tightly coupled to transcription [43]. CDK12 complexes with cyclin L1 or L2 have also been reported to localise to nuclear speckles and to regulate RNA splicing [50, 62].

1.5.3.2 CDK14-18

An extended CDK subfamily has been identified, designated CDK14-18, with similar primary structures in their kinase domains, but diverged cyclin binding sequence motifs of "PFTAIRE" (CDK14-15) or "PCTAIRE" (CDK16-18) (Fig. 1.2). Little is known about their structure, function or regulation, revealing a significant gap in our understanding of potentially important signal transduction pathways.

In contrast to classical CDKs, which are composed of little more than the conserved central protein kinase catalytic core, CDK14-18 are characterized by large N- and C-terminal extensions. In addition, CDK16-18 contain a conserved sequence immediately N-terminal to the kinase domain. Phylogenetic analyses suggest that these CDKs are most closely related to the atypical protein CDK5, which has its own atypical cyclin p35 [17]. CDK5 is ubiquitously expressed, but is most abundant in the nervous system and has been ascribed post-mitotic, neuronal functionality [63].

The first crystal structure from this atypical family, comprising the CDK16 kinase domain in complex with indirubin E804 inhibitor, was solved by the SGC at 2.4 Å resolution (PDB ID: 3MTL) (Fig. 1.8). In the absence of a cyclin, CDK16 shows an inactive conformation with conformational plasticity in the N-lobe, as evidenced by the largely unstructured αC helix and an inactive ‘DFG-out’ conformation of the activation loop.

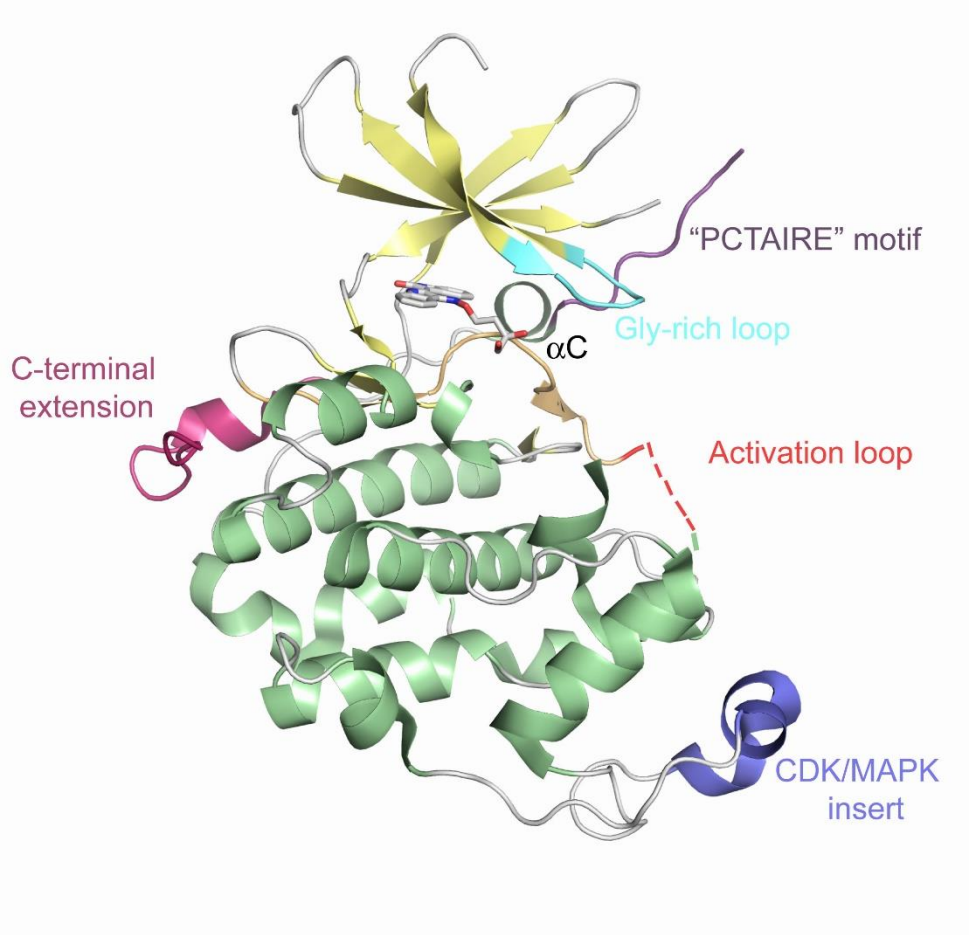


Figure 1.8. CDK16 kinase domain in complex with Indirubin E804, in an inactive conformation 2.4Å (PDB 3MTL). The CDK16 inactive catalytic state is apparent in the largely unstructured αC helix, reduced to a single turn, in which the PCTAIRE motif adopts an extended and somewhat flexible conformation. CDK16 kinase domain contains the typical CDK insert, which packs against the αG helix and has a PCTAIRE-family specific C-terminal extension. At the end of the PCTAIRE the catalytic αC-β3 salt bridge is formed by E393. Despite binding an ATP-competitive inhibitor, Indirubin E804, which forms three hydrogen bonds with the hinge region, the activation segment shows an inactive 'DFG-out' conformation.

Table 1.1. Cyclin binding partners elucidated for CDK14-18.

CDK	Cyclin binding partners
CDK14	CycD3 [64] CycY [65, 66]
CDK15	-
CDK16	CycY [67, 68] p35[69] CycI[68] CycK [68] CycYL1[68]
CDK17	-
CDK18	CycK [70]

The precise cyclin complement for CDK14-18 is unknown, although a number of potential cyclin binding partners have been suggested (Table 1.1). These include known partners of classical CDKs, such as CycD3 and CycK, which bind CDK4/6 [35] and CDK9 [71] [72], respectively. Recently, a novel cyclin CycY, has emerged as a potential binding partner for both CDK14 and CDK16 [65-68]. CycY has only poor sequence conservation with classical cyclins and displays several unusual properties. First, it is predicted to contain only a single cyclin box as observed in p35 [73], whereas other cyclins contain two. Second, CycY can be recruited to the plasma membrane by N-terminal myristoylation [65, 66], in contrast to other cyclins, which localize to the nucleus. Tethering of CycY to the plasma membrane enables the CDK14/CycY complex to phosphorylate the Wnt co-receptor LRP6, apparently linking growth factor signals to the cell cycle [65].

CDK16 is ubiquitously expressed, but is particularly high in terminally differentiated cells and in the brain where it is suggested to control neurite outgrowth [74, 75]. In addition, CDK16 is highly expressed in the testis and plays a unique role in the terminal steps of spermatogenesis [68]. Genetic knockdown in *C. elegans* suggested interdependent phenotypes for CYY-1, the CycY

ortholog, and PCT-1, a CDK16-18 ortholog [67], implicating parallel roles for CDK16/CycY and CDK5/p35 in the inhibition of retrograde axonal trafficking [67]. Co-immunoprecipitation of human CDK16 and CycY was recently confirmed in HEK293 cells by Mikocelvic *et al.* [68]. In contrast to classical CDKs, a conserved N-terminal region amongst the "PCTAIRE" kinases, was necessary for interaction in addition to the kinase domain demonstrating a novel regulatory mechanism. CDK16 has been implicated in other diverse biological and physiological processes, including vesicle trafficking [76, 77], glucose homeostasis [78, 79] and muscle differentiation [80].

1.6 Protein kinases as drug targets

Due to their important roles in cellular signaling and various human diseases, protein kinases are an area of intensive research. Protein kinase inhibitors can be divided into four classes. Currently, the majority of clinical and pre-clinical kinase inhibitors are reversible ATP-competitive type I inhibitors that take advantage of any unique features of the ATP-pocket to enable selectivity [81]. Given the high sequence conservation and the size of the kinome, achieving this selectivity is a significant challenge with most approved kinase inhibitors showing limited selectivity and targeting multiple kinases [82]. Comprehensive pre-validation of the kinase target prior to inhibitor development is an ongoing challenge. Type I inhibitors must also out-compete ATP. This class of inhibitors take advantage of interactions in the ATP pocket, often using conserved hinge region interactions, for example.

Type II inhibitors also target the ATP pocket, but bind to the inactive "DFG-out" conformation of the kinase, which can be structurally more diverse. Type II inhibitors occupy a hydrophobic pocket besides the αC helix created by an inversion of DFG phenylalanine [83]. Type III inhibitors are allosteric inhibitors that act outside the ATP-pocket. They do not compete with ATP binding, but can still lead to ATP displacement [83]. Type IV inhibitors target kinases outside their catalytic domain to interfere with a regulatory binding partner, for example. In this manner, these inhibitors can be very selective as the protein-protein interaction it targets is often unique

[83].

Selective covalent inhibitors are conceptually attractive as they may offer sustained target engagement as non-equilibrium binding limits the competition even at high endogenous substrate concentrations. Thus, they are more efficient for complete target inactivation [81, 84, 85]. Covalent inhibitors are also less sensitive to pharmacokinetic parameters, so provide long residence time and thus require less frequent dosing [85, 86]. The activity of the target can only be recovered by synthesis of new protein[81]. It is a particularly useful strategy then complete inactivation of the target is require. Conversely, covalent inhibitors may not be suitable for biological targets which have a rapid turnover or in instance where for mechanisms requiring short residence time, transient or partial inhibition [85, 86].

There has been some reluctance to pursue irreversible inhibitors due to the risks of haptenization and concerns about off-target effects [81]. Risk of toxic events can be lessened through optimization of non-covalent interactions to improve target receptor recognition and modulation of the electrophilic warhead, as hyper-reactive warheads may lead to other drug-induced toxicity [86, 87].

Selective covalent inhibitors would be extremely useful as probes for protein kinases whose function is poorly understood. Targeting cysteine residues offers a selectively filter, as these residues aren't uniformly conserved across the kinome [84]. Therefore, non-covalent recognition only needs to enable discrimination between kinases that possess an equivalently placed cysteine residue. Targeting cysteines has also been used as an entry point to inhibitor design by starting with a weakly potent covalent interaction that can then be optimized by the growth of inhibitor fragments often in combination with structural work [88].

There is evidence that covalent inhibitors may lessen the development of drug resistance from mutation of the binding site, by increasing biochemical efficient and non-equilibrium binding [86]. However, in some targeted therapies a resistance mutation at the cysteine residue is a common resistance mechanism for covalent inhibitors, representing an ongoing challenge in drug

development and the need for second and third generation covalent inhibitors [89].

1.7 Target discovery and validation in drug discovery

One of the most critical stages during drug-discovery is target identification and validation. Genetic knockouts or knockdowns can be time consuming and challenging to apply to cell cultures, but the advent of CRISPR/Cas9 editing and RNA-mediated silencing technologies, such as short interfering RNA (siRNA), has provided key strategies to examine the loss-of-function phenotypic effect of silencing a gene of interest. This allows investigators to interrogate the functional impact of protein loss. However, gene knockout or knockdown does not allow for individual protein domains to be assessed as therapeutic targets, or for the effects of specific post-translational modifications. CRISPR/Cas9 technologies have enabled site-directed mutagenesis in cells that partly addresses this, but the effects are permanent rather than transient. Readouts may be further complicated due to the level of redundancy in biological systems as silencing of a gene may promote alternative signaling pathway activation. Cancer cells in particular typically adapt to such genetic insults and gene re-expression can be insufficient for rescue limiting the confidence of the target validation.

Alternatively, biological systems may be chemically probed using small-molecule inhibitors of a target. Chemical probes are advantageous as they can be easily applied in high-throughput screening efforts to identify important therapeutic targets. In cells, the effect of a chemical probe may be rapid and useful for precise temporal control [90, 91]. Chemical probes are tunable and it is possible to titrate small-molecule inhibitors across a range of concentrations, which may reveal a dose-dependent spectrum of cellular phenotypes and provide quantitative control in a signaling pathway. If the binding is transient and reversible, then the biological effect induced on the target can be reversible [90, 91]. Inhibitors of an enzyme may also still allow it to perform protein-protein interactions, thus enzymatic and structural roles can be dissected [91]. Multiple chemical probes can also be easily tested combinatorially to screen for additive effects and synthetic lethal interactions [91]. This is also possible using siRNA, but this can introduce further non-specific effects and sometimes toxicity. Indeed, some genes may be essential such

that cell lines may not be viable with a gene knockout or severe knockdown. In these instances, chemical probes are particularly useful due to the transient and tunable nature.

Establishing target specificity is crucial to interpreting the readout and off-target effects. The main disadvantage of a chemical probe is potentially the lack of specificity. Off -target effects of small molecules may affect proteins of similar sequence and structure [90]. Consequently, strict guidelines have been suggested as criteria for development of a probe.

Ultimately, a combination of both genetic and chemical probe strategies are typically applied to aid target validation. Divergence between RNA interference and chemical probe results can often be attributed to protein-protein interactions and may further help dissect a signaling pathway [91]. Chemical inhibitors are also likely to better represent the true benefits of developing a drug.

1.8 The untargeted kinome

Kinases are among the most frequently mutated proteins in cancers, yet there are large gaps in our understanding. Analyses suggests that 25% of the kinome has unknown function while 50% of kinases are largely uncharacterized [82]. Moreover, challenges in target validation have narrowed the focus of kinase drug-discovery to a few well-validated targets. Novel members of the CDK family, including CDK12-18, are, therefore, interesting targets for the development of specific tool compounds that can be used to probe function and therapeutic potential.

1.9 CDK12 and CDK16 as targets for drug discovery

1.9.1 CDK12

CDK12 has been consistently been reported as having a role in the transcription of a subset of genes, the damage response genes [40, 55, 92, 93]. Dysregulation of CDK12 has been identified in several cancers, including ovarian, breast and gastric cancer as well as in lung and large intestine tumours [93-96]. CDK12 was amongst 127 significantly mutated genes in 12 cancer forms [97]. CDK12 mutations are also associated with homologous recombination (HR) defects in ovarian

cancers suggesting a tumour suppressor role [95]. Disabling CDK12 function reduced BRCA1 levels and sensitized these cells to PARP1/2 inhibition. In ERalpha-positive breast cancer CDK12 silencing caused resistance to tamoxifen and additional estrogen signaling inhibitors, by activation of the mitogen-activated protein kinase (MAPK)-signaling pathway [98]. Tumor suppressor function is further suggested by the presence of disabling CDK12 mutations in ovarian cancer [93-95]. These data suggest that CDK12/13 inhibitors alone, or in combination treatment, may represent an effective strategy for targeting cancers demonstrating extreme dependencies for the expression of CDK12/13–regulated genes.

However, several investigations have shown oncogenic roles for CDK12 in cancer progression. CDK12 is located in the q12-q21 region of chromosome 17 (17q12-q21), a region which is amplified in a large number of cancers, including ovarian and breast cancers [94, 99]. In breast cancer, Cdk12 was found to be co-amplified with ERBB2A the tyrosine kinase receptor gene, a protein which is amplified and overexpressed in about 20% of breast tumours [26, 100, 101]. There is a correlation between amplification and overexpression of genes from this region, which play functional roles in the pathogenesis of breast cancer [99]. Overexpression of *CDK12* correlated with aggressive disease in breast cancer with high tumour grade and proliferative index, based on tissue microarrays of invasive breast carcinomas [102]. Thus, CDK12 may also play an oncogenic role in cancer progression but the details underlying this remain to be elucidated.

Recently, CRISPR/Cas9 –mediated gene editing has been developed to generate cells expressing a selectively inhibitable CDK12 [103]. Whilst repression of CDK12 in this way enables investigation of the functional impact of CDK12 loss, its utility is limited to this particular cell line. Identification of inhibitors of CDK12/13 function is paramount to understanding the effects of pharmacologically inhibiting CDK12/13 in normal and disease states.

1.9.2 CDK16

Whilst the precise function of CDK16 is still being elucidated, recent investigations using siRNA experiments targeting CDK16 in prostate, breast, cervical cancer, and melanoma cell lines, showed CDK16 knockdown diminished cancer cell proliferation [104-106]. Interestingly, CDK16 knockdown in non-transformed cells did not affect proliferation [104]. An evaluation of CDK16 expression levels in primary tumors showed that its expression was significantly elevated in prostate and breast cancer [106]. In addition, RNAi showed that CDK16 played a role in the proliferation of medulloblastoma cells lines [107]. Overexpression of CycY has been linked to human cancer and knockdown of CycY has been shown to inhibit proliferation of glioma cells [108, 109].

The precise role of CDK16 in normal and pathobiology remains to be elucidated and chemical probes would be useful to aid in the elucidation of this. CDK16 forms an interesting target for drug discovery

1.10 Aims of this thesis

The aim of this thesis was to characterise the extended CDK family beyond the classical members CDK1 to CDK9. The focus will be largely upon two representative members CDK12 and CDK16 for which initial clones are available. Both proteins have atypical extensions beyond their kinase domain and their structure, regulation and functions are only poorly understood. I will aim to utilize bacterial and *Sf9* insect cell expression systems to obtain recombinant proteins and to attempt to determine their mode of regulation by identifying their cyclin binding partners and requirements for phosphorylation. Once a CDK/cyclin complex has been identified, they will be entered into crystallisation trials. Structural studies by X-ray crystallography will allow in depth examination of the mode of CDK regulation as well as their different structural features compared to other CDK family members. Specific kinase inhibitors have not been described for either CDK. Therefore, further efforts will be made to screen for inhibitors of CDK12 and CDK16 using

differential scanning fluorimetry (DSF) and other kinase activity assays. Where inhibitors are identified, X-ray crystallography may help to define the precise mode of binding. Should inhibitors be identified, I aim to explore their use in cancer cell lines to gain a better insight into their kinase function and potential for cancer drug discovery.

1.11 References

1. Manning, G., et al., *Evolution of protein kinase signaling from yeast to man.* Trends Biochem Sci, 2002. **27**(10): p. 514-20.
2. Attwood, P.V., et al., *Focus on phosphohistidine.* Amino Acids, 2007. **32**(1): p. 145-56.
3. Hunter, T., *Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signaling.* Cell, 1995. **80**(2): p. 225-36.
4. Cohen, P., *Protein kinases--the major drug targets of the twenty-first century?* Nat Rev Drug Discov, 2002. **1**(4): p. 309-15.
5. Cohen, P., *The regulation of protein function by multisite phosphorylation--a 25 year update.* Trends Biochem Sci, 2000. **25**(12): p. 596-601.
6. McClendon, C.L., et al., *Dynamic architecture of a protein kinase.* Proc Natl Acad Sci U S A, 2014. **111**(43): p. E4623-31.
7. Ubersax, J.A. and J.E. Ferrell, Jr., *Mechanisms of specificity in protein phosphorylation.* Nat Rev Mol Cell Biol, 2007. **8**(7): p. 530-41.
8. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
9. Schulman, B.A., D.L. Lindstrom, and E. Harlow, *Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(18): p. 10453-10458.
10. Taylor, S.S. and A.P. Kornev, *Protein kinases: evolution of dynamic regulatory proteins.* Trends Biochem Sci, 2011. **36**(2): p. 65-77.
11. Knighton, D.R., et al., *Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase.* Science, 1991. **253**(5018): p. 407-14.
12. Kornev, A.P., et al., *Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism.* Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17783-8.
13. Kornev, A.P. and S.S. Taylor, *Dynamics-Driven Allostery in Protein Kinases.* Trends Biochem Sci, 2015. **40**(11): p. 628-47.
14. Huse, M. and J. Kuriyan, *The conformational plasticity of protein kinases.* Cell, 2002. **109**(3): p. 275-82.
15. Endicott, J.A. and M.E. Noble, *Structural characterization of the cyclin-dependent protein kinase family.* Biochem Soc Trans, 2013. **41**(4): p. 1008-16.
16. Malumbres, M., *Cyclin-dependent kinases.* Genome Biol, 2014. **15**(6): p. 122.

17. Malumbres, M., et al., *Cyclin-dependent kinases: a family portrait.* Nat Cell Biol, 2009. **11**(11): p. 1275-6.

18. Evans, T., et al., *Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division.* Cell, 1983. **33**(2): p. 389-96.

19. *UniProt: a hub for protein information.* Nucleic Acids Res, 2015. **43**(Database issue): p. D204-12.

20. Bourne, Y., et al., *Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle-regulatory protein CksHs1.* Cell, 1996. **84**(6): p. 863-74.

21. Song, H., et al., *Phosphoprotein-protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2.* Mol Cell, 2001. **7**(3): p. 615-26.

22. Echalier, A., J.A. Endicott, and M.E. Noble, *Recent developments in cyclin-dependent kinase biochemical and structural studies.* Biochim Biophys Acta, 2010. **1804**(3): p. 511-9.

23. Morgan, D.O., *Cyclin-dependent kinases: engines, clocks, and microprocessors.* Annu Rev Cell Dev Biol, 1997. **13**: p. 261-91.

24. Jeffrey, P.D., et al., *Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex.* Nature, 1995. **376**(6538): p. 313-20.

25. Mikolcevic, P., J. Rainer, and S. Geley, *Orphan kinases turn eccentric: a new class of cyclin Y-activated, membrane-targeted CDKs.* Cell Cycle, 2012. **11**(20): p. 3758-68.

26. Kohoutek, J. and D. Blazek, *Cyclin K goes with Cdk12 and Cdk13.* Cell Div, 2012. **7**: p. 12.

27. Merrick, K.A. and R.P. Fisher, *Putting one step before the other: distinct activation pathways for Cdk1 and Cdk2 bring order to the mammalian cell cycle.* Cell cycle (Georgetown, Tex.), 2010. **9**(4): p. 706-714.

28. Lolli, G., et al., *The crystal structure of human CDK7 and its protein recognition properties.* Structure, 2004. **12**(11): p. 2067-79.

29. Lolli, G. and L.N. Johnson, *CAK-Cyclin-dependent Activating Kinase: a key kinase in cell cycle control and a target for drugs?* Cell Cycle, 2005. **4**(4): p. 572-7.

30. Cole, A.R., *PCTK proteins: the forgotten brain kinases?* Neurosignals, 2009. **17**(4): p. 288-97.

31. Nolen, B., S. Taylor, and G. Ghosh, *Regulation of protein kinases; controlling activity through activation segment conformation.* Mol Cell, 2004. **15**(5): p. 661-75.

32. Roussel, M.F., *The INK4 family of cell cycle inhibitors in cancer.* Oncogene, 1999. **18**(38): p. 5311-7.

33. Brown, N.R., et al., *The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases.* Nat Cell Biol, 1999. **1**(7): p. 438-43.

34. Hochegger, H., S. Takeda, and T. Hunt, *Cyclin-dependent kinases and cell-cycle transitions: does one fit all?* Nat Rev Mol Cell Biol, 2008. **9**(11): p. 910-6.

35. Malumbres, M. and M. Barbacid, *Mammalian cyclin-dependent kinases.* Trends Biochem Sci, 2005. **30**(11): p. 630-41.

36. Buratowski, S., *Progression through the RNA polymerase II CTD cycle.* Mol Cell, 2009. **36**(4): p. 541-6.

37. Fuda, N.J., M.B. Ardehali, and J.T. Lis, *Defining mechanisms that regulate RNA polymerase II transcription in vivo.* Nature, 2009. **461**(7261): p. 186-92.

38. Jeronimo, C., P. Collin, and F. Robert, *The RNA Polymerase II CTD: The Increasing Complexity of a Low-Complexity Protein Domain.* J Mol Biol, 2016.

39. Martinez-Rucobo, F.W., et al., *Molecular Basis of Transcription-Coupled Pre-mRNA Capping.* Mol Cell, 2015. **58**(6): p. 1079-89.

40. Liang, K., et al., *Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing.* Mol Cell Biol, 2015. **35**(6): p. 928-38.

41. Rzymski, T., et al., *CDK8 kinase--An emerging target in targeted cancer therapy.* Biochim Biophys Acta, 2015. **1854**(10 Pt B): p. 1617-29.

42. Corden, J.L., *Pol II CTD Code Light.* Mol Cell, 2016. **61**(2): p. 183-4.

43. Eick, D. and M. Geyer, *The RNA polymerase II carboxy-terminal domain (CTD) code.* Chem Rev, 2013. **113**(11): p. 8456-90.

44. Schneider, E.V., et al., *The structure of CDK8/CycC implicates specificity in the CDK/cyclin family and reveals interaction with a deep pocket binder.* J Mol Biol, 2011. **412**(2): p. 251-66.

45. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K.* Nat Commun, 2014. **5**: p. 3505.

46. Baumli, S., et al., *The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation.* EMBO J, 2008. **27**(13): p. 1907-18.

47. Larochelle, S., et al., *Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II.* Nat Struct Mol Biol, 2012. **19**(11): p. 1108-15.

48. Garber, M.E., et al., *CDK9 autophosphorylation regulates high-affinity binding of the human immunodeficiency virus type 1 tat-P-TEFb complex to TAR RNA.* Mol Cell Biol, 2000. **20**(18): p. 6958-69.

49. Marques, F., et al., *A new subfamily of high molecular mass CDC2-related kinases with PITAI/VRE motifs.* Biochem Biophys Res Commun, 2000. **279**(3): p. 832-7.

50. Ko, T.K., E. Kelly, and J. Pines, *CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles.* J Cell Sci, 2001. **114**(Pt 14): p. 2591-603.

51. Even, Y., et al., *CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo.* J Cell Biochem, 2006. **99**(3): p. 890-904.

52. Bartkowiak, B. and A.L. Greenleaf, *Expression, purification, and identification of associated proteins of the full-length hCDK12/CyclinK complex.* J Biol Chem, 2015. **290**(3): p. 1786-95.

53. Edwards, M.C., C. Wong, and S.J. Elledge, *Human cyclin K, a novel RNA polymerase II-associated cyclin possessing both carboxy-terminal domain kinase and Cdk-activating kinase activity.* Mol Cell Biol, 1998. **18**(7): p. 4291-300.

54. Fu, T.J., et al., *Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription.* J Biol Chem, 1999. **274**(49): p. 34527-30.

55. Blazek, D., et al., *The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes.* Genes Dev, 2011. **25**(20): p. 2158-72.

56. Cheng, S.W., et al., *Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II.* Mol Cell Biol, 2012. **32**(22): p. 4691-704.

57. Baek, K., et al., *Crystal structure of human cyclin K, a positive regulator of cyclin-dependent kinase 9.* J Mol Biol, 2007. **366**(2): p. 563-73.

58. Bartkowiak, B., et al., *CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1.* Genes Dev, 2010. **24**(20): p. 2303-16.

59. Bowman, E.A. and W.G. Kelly, *RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases.* Nucleus, 2014. **5**(3): p. 224-36.

60. Davidson, L., L. Muniz, and S. West, *3' end formation of pre-mRNA and phosphorylation of Ser2 on the RNA polymerase II CTD are reciprocally coupled in human cells.* Genes Dev, 2014. **28**(4): p. 342-56.

61. Eifler, T.T., et al., *Cyclin-dependent kinase 12 increases 3' end processing of growth factor-induced c-FOS transcripts.* Mol Cell Biol, 2015. **35**(2): p. 468-78.

62. Chen, H.H., Y.C. Wang, and M.J. Fann, *Identification and characterization of the CDK12/cyclin L1 complex involved in alternative splicing regulation.* Mol Cell Biol, 2006. **26**(7): p. 2736-45.

63. Lalioti, V., D. Pulido, and I.V. Sandoval, *Cdk5, the multifunctional surveyor.* Cell Cycle, 2010. **9**(2): p. 284-311.

64. Shu, F., et al., *Functional characterization of human PFTK1 as a cyclin-dependent kinase.* Proc Natl Acad Sci U S A, 2007. **104**(22): p. 9248-53.

65. Davidson, G., et al., *Cell cycle control of wnt receptor activation.* Dev Cell, 2009. **17**(6): p. 788-99.

66. Jiang, M., et al., *Cyclin Y, a novel membrane-associated cyclin, interacts with PFTK1.* FEBS Lett, 2009. **583**(13): p. 2171-8.

67. Ou, C.Y., et al., *Two cyclin-dependent kinase pathways are essential for polarized trafficking of presynaptic components.* Cell, 2010. **141**(5): p. 846-58.

68. Mikolcevic, P., et al., *Cyclin-dependent kinase 16/PCTAIRE kinase 1 is activated by cyclin Y and is essential for spermatogenesis.* Mol Cell Biol, 2012. **32**(4): p. 868-79.

69. Cheng, K., et al., *Pctaire1 interacts with p35 and is a novel substrate for Cdk5/p35.* J Biol Chem, 2002. **277**(35): p. 31988-93.

70. Rual, J.F., et al., *Towards a proteome-scale map of the human protein-protein interaction network.* Nature, 2005. **437**(7062): p. 1173-8.

71. Lin, X., et al., *P-TEFb containing cyclin K and Cdk9 can activate transcription via RNA.* J Biol Chem, 2002. **277**(19): p. 16873-8.

72. Yu, D.S., et al., *Cyclin-dependent kinase 9-cyclin K functions in the replication stress response.* EMBO Rep, 2010. **11**(11): p. 876-82.

73. Liu, D., S. Guest, and R.L. Finley, Jr., *Why cyclin Y? A highly conserved cyclin with essential functions.* Fly (Austin), 2010. **4**(4): p. 278-82.

74. Charrasse, S., et al., *PCTAIRE-1: characterization, subcellular distribution, and cell cycle-dependent kinase activity.* Cell Growth Differ, 1999. **10**(9): p. 611-20.

75. Graeser, R., et al., *Regulation of the CDK-related protein kinase PCTAIRE-1 and its possible role in neurite outgrowth in Neuro-2A cells.* J Cell Sci, 2002. **115**(Pt 17): p. 3479-90.

76. Liu, Y., et al., *Pctaire1 phosphorylates N-ethylmaleimide-sensitive fusion protein: implications in the regulation of its hexamerization and exocytosis.* J Biol Chem, 2006. **281**(15): p. 9852-8.

77. Palmer, K.J., J.E. Konkel, and D.J. Stephens, *PCTAIRE protein kinases interact directly with the COPII complex and modulate secretory cargo transport.* J Cell Sci, 2005. **118**(Pt 17): p. 3839-47.

78. Tang, X., et al., *An RNA interference-based screen identifies MAP4K4/NIK as a negative regulator of PPARgamma, adipogenesis, and insulin-responsive hexose transport.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2087-92.

79. Chen, X.Y., et al., *Brain-selective kinase 2 (BRSK2) phosphorylation on PCTAIRE1 negatively regulates glucose-stimulated insulin secretion in pancreatic beta-cells.* J Biol Chem, 2012. **287**(36): p. 30368-75.

80. Shimizu, K., et al., *Pctaire1/Cdk16 promotes skeletal myogenesis by inducing myoblast migration and fusion.* FEBS Lett, 2014. **588**(17): p. 3030-7.

81. Liu, Q., et al., *Developing irreversible inhibitors of the protein kinase cysteinome.* Chem Biol, 2013. **20**(2): p. 146-59.

82. Fedorov, O., S. Muller, and S. Knapp, *The (un)targeted cancer kinome.* Nat Chem Biol, 2010. **6**(3): p. 166-169.

83. Muller, S., et al., *The ins and outs of selective kinase inhibitor development.* Nat Chem Biol, 2015. **11**(11): p. 818-21.

84. Lanning, B.R., et al., *A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors.* Nat Chem Biol, 2014. **10**(9): p. 760-7.

85. Johnson, D.S., E. Weerapana, and B.F. Cravatt, *Strategies for discovering and derisking covalent, irreversible enzyme inhibitors.* Future Med Chem, 2010. **2**(6): p. 949-64.

86. Bauer, R.A., *Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies.* Drug Discov Today, 2015. **20**(9): p. 1061-73.

87. Mah, R., J.R. Thomas, and C.M. Shafer, *Drug discovery considerations in the development of covalent inhibitors.* Bioorg Med Chem Lett, 2014. **24**(1): p. 33-9.

88. Erlanson, D.A., et al., *In situ assembly of enzyme inhibitors using extended tethering.* Nat Biotechnol, 2003. **21**(3): p. 308-14.

89. Engel, J., J. Lategahn, and D. Rauh, *Hope and Disappointment: Covalent Inhibitors to Overcome Drug Resistance in Non-Small Cell Lung Cancer.* ACS Med Chem Lett, 2016. **7**(1): p. 2-5.

90. Spring, D.R., *Chemical genetics to chemical genomics: small molecules offer big insights.* Chem Soc Rev, 2005. **34**(6): p. 472-82.

91. Weiss, W.A., S.S. Taylor, and K.M. Shokat, *Recognizing and exploiting differences between RNAi and small-molecule inhibitors.* Nat Chem Biol, 2007. **3**(12): p. 739-44.

92. Chen, H.R., et al., *Cdk12 Regulates Neurogenesis and Late-Arising Neuronal Migration in the Developing Cerebral Cortex.* Cereb Cortex, 2016.

93. Ekumi, K.M., et al., *Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex.* Nucleic Acids Res, 2015. **43**(5): p. 2575-89.

94. Cancer Genome Atlas Research, N., *Integrated genomic analyses of ovarian carcinoma.* Nature, 2011. **474**(7353): p. 609-15.

95. Joshi, P.M., et al., *Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors.* J Biol Chem, 2014. **289**(13): p. 9247-53.

96. Kauraniemi, P., et al., *New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays.* Cancer Res, 2001. **61**(22): p. 8235-40.

97. Kandoth, C., et al., *Mutational landscape and significance across 12 major cancer types.* Nature, 2013. **502**(7471): p. 333-9.

98. Iorns, E., et al., *CRK7 modifies the MAPK pathway and influences the response to endocrine therapy.* Carcinogenesis, 2009. **30**(10): p. 1696-701.

99. Luoh, S.-W., *Amplification and expression of genes from the 17q11∼q12 amplicon in breast cancer cells.* Cancer Genetics and Cytogenetics, 2002. **136**(1): p. 43-47.

100. Sircoulomb, F., et al., *Genome profiling of ERBB2-amplified breast cancers.* BMC Cancer, 2010. **10**: p. 539.

101. Benusiglio, P.R., et al., *HapMap-based study of the 17q21 ERBB2 amplicon in susceptibility to breast cancer.* Br J Cancer, 2006. **95**(12): p. 1689-95.

102. Capra, M., et al., *Frequent alterations in the expression of serine/threonine kinases in human cancers.* Cancer Res, 2006. **66**(16): p. 8147-54.

103. Bartkowiak, B., C. Yan, and A.L. Greenleaf, *Engineering an analog-sensitive CDK12 cell line using CRISPR/Cas.* Biochim Biophys Acta, 2015. **1849**(9): p. 1179-1187.

104. Yanagi, T. and S. Matsuzawa, *PCTAIRE1/PCTK1/CDK16: a new oncotarget?* Cell Cycle, 2015. **14**(4): p. 463-4.

105. Yanagi, T., J.C. Reed, and S. Matsuzawa, *PCTAIRE1 regulates p27 stability, apoptosis and tumor growth in malignant melanoma.* Oncoscience, 2014. **1**(10): p. 624-33.

106. Yanagi, T., et al., *PCTAIRE1 phosphorylates p27 and regulates mitosis in cancer cells.* Cancer Res, 2014. **74**(20): p. 5795-807.

107. Cwiek, P., et al., *RNA interference screening identifies a novel role for PCTK1/CDK16 in medulloblastoma with c-Myc amplification.* Oncotarget, 2015. **6**(1): p. 116-29.

108. Yue, W., et al., *Overexpression of cyclin Y in non-small cell lung cancer is associated with cancer cell proliferation.* Sci China Life Sci, 2010. **53**(4): p. 511-6.

109. Xu, Y., et al., *Lentivirus-mediated knockdown of cyclin Y (CCNY) inhibits glioma cell proliferation.* Oncol Res, 2010. **18**(8): p. 359-64.

Chapter 2 Materials and Methods

2.1 Protein expression

2.1.1 Designing constructs

A number of the proteins used were designed prior to my arrival in the SGC. Decisions on new constructs design were informed by the solubility and stability of these and in light of new findings, as well as new proteins of interest. If cDNA was not already available in the SGC this was requested from the Mammalian Gene Collection (MGC) [1]. The full-length protein and a series of truncated fragments were designed based on bioinformatics predictions of structured domains [2].Secondary structure predictions using Psipred [3] and predictions of disorder using Globplot (Expasy, EMBL) [4] were used to inform decisions on domains boundaries. In some instances for proteins, crystal structures were available in the PDB, such as CycK and viral K-cyclin, so these boundaries were selected.

2.1.2 Ligation independent cloning (LIC)

The LIC method of cloning was employed. The LIC method is based on annealing short complementary single-stranded 5'-tails generated by the 3′-exonuclease activity of T4 DNA polymerase which has both polymerase (5' to 3') and exonuclease (3' to 5') activity. A 12–16 base pair overhang was added to the polymerase chain reaction (PCR) primers, which had incorporated appropriate extensions dependent upon the vector of choice [5].

This protocol is for viral K-cyclin cloning into a vector pNIC-zB.

Table 2.1. Primers for viral K-cyclin cloning.

	Forward primer	Reverse primer
pNIC-zB	TACTTCCAATCCATGGCAAC TGCCAATAACCCGC	TATCCACCTTTACTGTTAAT AGCTGTCCAGAATGCGCAG
Sequencing primers for PCR screen	TGTGAGCGGATAACAATTCC	AGCAGCCAACTCAGCTTCC

The appropriate viral K-cyclin primers were designed with vector specific 5' extensions (Table 2.1). The forward and reverse primers were mixed to a concentration of 5 μM of each (1.5 μl). The template DNA was diluted to 2.5 ng/μl. A Mastermix containing Platinum Pfx polymerase was added to a PCR-tube (100 μl tube) at 21 μl per tube and subsequently the primers (1.5 μl) then template (2.5 μl) were added.

PCR was run as follows;

94°C, 5 min;

(94°C, 30 sec; 52°C, 30 sec; 68°C, 2min) x30 cycles;

72°C, 10 min;

15°C hold

The PCR product was analysed on a 1 % agarose gel then purified using the Qiagen PCR purification and eluted in 30 μl. The purified PCR product was then treated

with T4 DNA polymerase (Novagen) in the presence dCTPs, for 30 minutes at 22 °C followed by 20 minutes at 75 °C, in a thermocycler (Techne, TC-512).

The pNIC-Zb vector was prepared by digestion with BsaI at 50 °C in a water bath, for 2.5 hours. The products were analysed on a 1 % agarose gel then purified using the Qiagen PCR purification kit and eluted in 50 µl. The digested vector was then T4 treated in the same way as the PCR product, except with dGTP. Annealing was then performed by mixing the treated plasmid (4 µl) and insert (8 µl) alongside appropriate controls of vector only and insert only. The mixture was incubated or 30 minutes at room temperature then transferred to ice, to which 50 µl of Mach1 competent cells were added for a further 30 minute incubation. The sample was then heat shocked at 42 °C for 45 seconds then chilled briefly before the addition of 1 mL of Lysogeny broth (LB) was added. After 90 minutes incubation at 37 °C, 100 µl of the transformation was plated onto LB-agar plates containing 50 µg/mL Kanamycin and 5% sucrose and colonies were left to grow overnight at 37 °C. A single colony was then screened by PCR with a Bioline Taq Red DNA polymerase to confirm annealing. The same loop was also used to inoculate LB with 50 µg/mL Kanamycin for overnight incubation at 37 °C, with shaking at 700 rpm. The next day 330 µl of 60 % sterile glycerol was added to the sample and flash frozen for storage at -80 ^{o}C.

2.1.3 Expression vectors

Viral K-cyclin was cloned into the transfer vector pNIC-Zb. Viral K-cyclin used in earlier work was a gift from Professor Ernest Laue (University of Cambridge) and had a N-terminal histidine tag with an enterokinase cleavage site.

Further expression constructs were prepared by the SGC using LIC. Schematics of the vectors used are shown in Figure 2.1. CDK16 was cloned into the vector pNIC-CH, which encodes a C-terminal hexahistidine tag which is non-cleavable.

Baculoviral expression constructs containing Tn7 sequences were transformed into the *E. coli* strain DH10Bac for *in vivo* recombination into bacmid DNA using InVitrogen's Bac-to-bac system.

Human CDK12, cyclin K, cyclin L1, Cyclin D1, Cyclin D3 and Cyclin I constructs were cloned into the baculoviral transfer vector pFB-LIC-Bse (Fig.2.1c) by LIC. The vector encodes an N-terminal hexahistidine tag and a Tobacco Etch Virus Protease A (TEV) cleavage site.

Human CDK16, CDK17, CDK18, Cyclin Y, Cyclin Y-like-1, Cyclin Y-like-2 and Cyclin Y-like-3 constructs were cloned into the baculoviral transfer vector pFB-CT10HF-LIC by ligation-independent cloning (Fig.2.1d). The vector encodes a C-terminal hexahistidine and FLAG tag with a Tobacco Etch Virus Protease A (TEV) cleavage site.

A pFBHTb expression vector containing Cdk-activating kinase (CAK1) from *Saccharomyces* cerevisiae (Uniprot P43568) was purchased from the MRC Protein and Phosphorylation Unit (University of Dundee, UK).

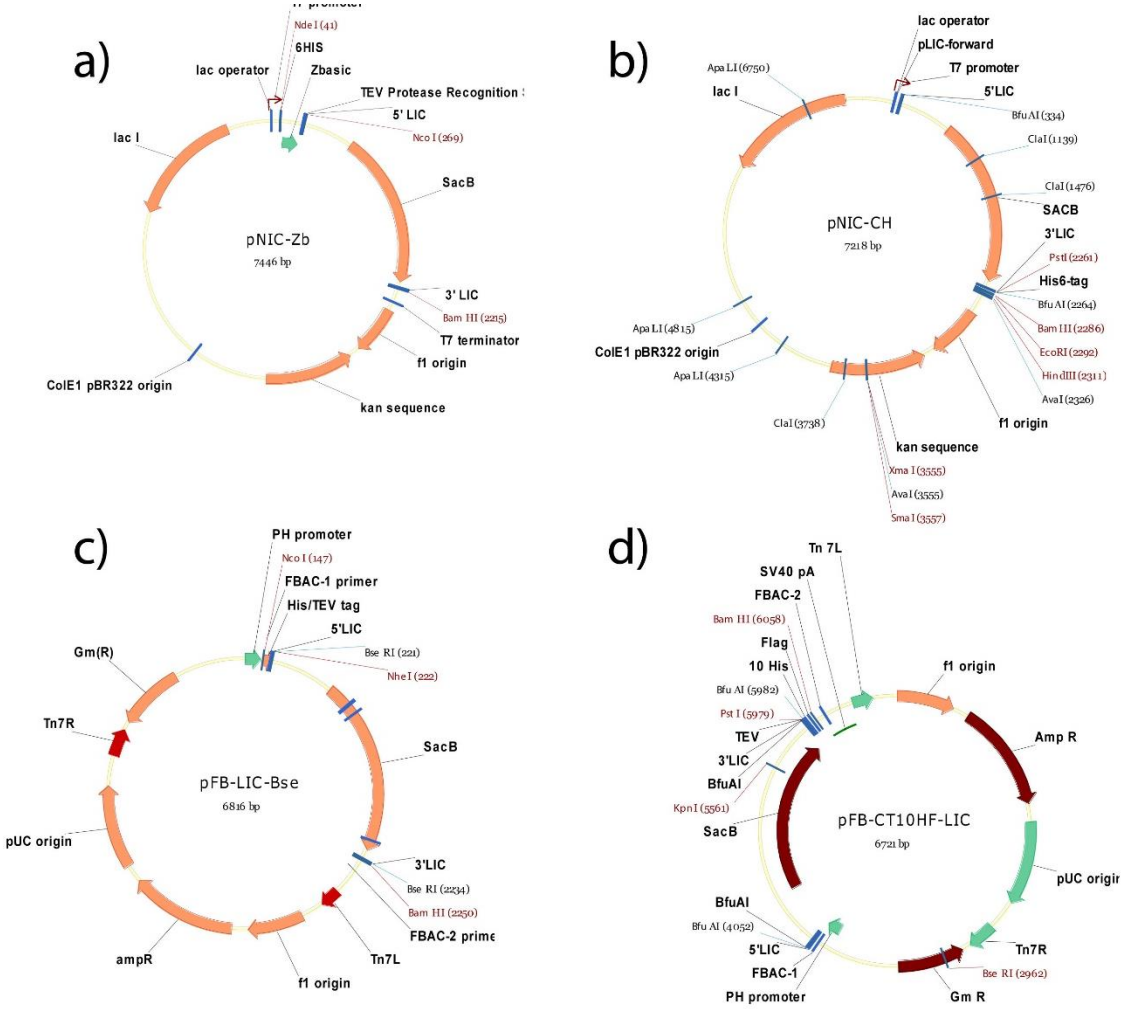


Figure 2.1. LIC-modified pFastBac vectors for bacterial and baculovirus expression (Taken from http://www.thesgc.org/reagents/vectors. Diagrams created by Dr P. Savitsky, Dr O.Gileadi and Dr G. Kochan[5]).

2.1.4 Escherichia coli expression

2.1.4.1 Transformation

For transformation, 1 µL of the plasmid was incubated on ice for 30 minutes with 50 µL of thawed BL21 Rosetta competent cells, which contain the bacteriophage T7 RNA polymerase gene. The cells were heat shocked for 45 seconds at 42°C in a water bath and then returned to ice for 2 minutes. To this, 300 µL of LB was added and it incubated at

37°C with shaking, at 180 *Revolutions per minute (*rpm), for 1 hour. Then 50 µL of the suspension was spread on a kanamycin and chloramphenicol LB-agar plate, which was incubated at 37°C overnight. A single colony was picked and grown in LB medium containing 50 µg/mL kanamycin, overnight at 37°C with shaking at 180 rpm. This was then used to make a glycerol stock, mixing with an equal volume of sterile glycerol and flash frozen in liquid nitrogen and stored at -80°C.

2.1.4.2 Bacterial expression

A glycerol stock (*E. coli* strain BL21 (Rosetta) or BL21(DE3)-R3-pRARE2 for other proteins) transformed with the expression construct was used to inoculate a 50 mL starter culture containing LB media with 50µg/ml kanamycin. The starter culture was grown overnight at 37°C with shaking at 180 rpm. The following morning, four flasks containing 1 L LB/kanamycin were each inoculated with 10 mL of the starter culture. Cultures were incubated at 37°C with shaking at 180 rpm until an OD 600nm $\geq$ 0.7 was reached. The flasks were then cooled to 18°C and 1 mM IPTG added to induce protein expression overnight. Cells were harvested by centrifugation at 5000 rpm at 4°C for 15 min. Cell pellets from each flask were frozen and stored at -20°C.

2.1.5 Baculoviral expression

P0 and P1 viruses were prepared by members of the SGC. I amplified P2 viruses from P1 in a 50-100 mL scale expression and the supernatant was stored at 4°C. *Spodoptera frugiperda (Sf9)* insect cells were infected with P2 virus at a density of 2,000,000 cells/mL for 48 hours or 72 hours, dependent on the stability and yield of the protein as determined by small-scale test expression and purification. In some instances, proteins were co-expressed and thus the level of P2 virus volume used to infect the cells was varied to optimize the co-expressed protein ratio. Cells were harvested at 900xg, at 4°C for 25 min. Cell pellets from each flask were frozen, sometimes in 40 mL of 50 mM HEPES, pH 7.5;

500 mM NaCl; 5% Glycerol; 5 mM Imidazole, 0.5 mM TCEP supplemented with 1µg/mL protease inhibitor cocktail set III (Calbiochem), and stored at -20°C. Some proteins were found to be more stable when frozen as pellet alone or in some instances, immediate purification following expression was necessary to limit degradation.

2.1.6 Mammalian expression

P2 virus was provided by Grazyna Kochan, SGC. Proteins were expressed using baculovirus infected HEK293 cells. Cells were infected with 5 mL P2 virus at around 90% confluence for 72 hours. Cells were harvested at 1000rpm, 4°C for 5 min. Cell pellets from each flask were frozen and stored at -20°C.

2.2 Purification

2.2.1 Extraction method

Frozen cells were thawed at room temperature for 20 minutes then placed on ice. Samples were made up to an appropriate volume with binding buffer (25 mL per 1 L equivalent culture) with 1 µg/mL protease inhibitor cocktail set III (Calbiochem). Cells were lysed by sonication (Sonics, Vibra Cell) over 12 minutes at 35% amplitude (sonicator pulsing 'ON' for 5 seconds and 'OFF' for 10 seconds). For mammalian cells, 1% triton was incubated with cells for 30 minutes prior to sonication. The cell lysate was clarified by centrifugation at 21000 rpm at 4°C for 1 h in the presence of 0.15% PEI to precipitate unwanted nucleic acid. The supernatant was recovered for purification.

2.2.2 Small scale test expression and purification

For small scale testing, 3 mL cultures were grown using a 24-well block and harvested at 48 or 72 hours post-infection. Cells were lysed by sonication, using a 24-probe sonicator, in binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM Imidazole, 0.5 mM TCEP) supplemented with protease inhibitor cocktail set III

(Calbiochem) with sonication for 4 minutes (sonicator pulsing 'ON' for 5 seconds and 'OFF' for 10 seconds). The lysate was clarified by centrifugation. The expressed proteins were purified using nickel-sepharose resin (GE Healthcare) and analysed by SDS PAGE.

2.2.3 Nickel-affinity Chromatography

Some 5 mL of a 50 % nickel-sepharose slurry (Amersham) was applied onto a 1.5 x 10 cm column. The column was first washed with 5 column volume (CV) of deionised distilled H_2O, and then equilibrated with 5 CV binding buffer. Lysate sample was filtered and loaded onto the column under gravity flow. The column as then washed with 10 CV of binding buffer followed by 6 CV of wash buffer. Bound samples were eluted by an imidazole step gradient using 2 CV elution buffers containing 50 mM, 100 mM, 150 mM and 200 mM imidazole. Elution fraction samples were analysed by SDS-PAGE, as were subsequent purification steps in combination with electrospray ionisation mass spectrometry (ESI-MS).

Nickel-Affinity Chromatography Buffers:

Binding buffer: 50 mM HEPES, pH 7.5; 500 mM NaCl; 5% Glycerol; 5 mM Imidazole, 0.5 mM TCEP

Wash buffer: 50 mM HEPES, pH 7.5; 500 mM NaCl; 5% Glycerol; 30 mM Imidazole, 0.5 mM TCEP

Elution buffers: 50 mM HEPES, pH 7.5; 500 mM NaCl; 5% Glycerol; 50 to 250 mM Imidazole, 0.5 mM TCEP

2.2.5 Enzymatic Treatment

To eluted protein 0.5 mg TEV endoprotease (which cleaves the recognition sequence ENLYFQ*(G/S)) was added) to cleave the affinity tag. Eluates were incubated with the protease overnight at 4°C. Cleavage was assessed my ESI-MS.

2.2.6 Addition of reducing reagent during purification stages

Between each purification stage, 10 mM DTT was added to maintain an adequate reduced state to prevent protein oxidation and in some instances, 10 mM arginine/glutamine mixture was added. Simultaneous addition of L-arginine and L-glutamine has been shown to be effective in preventing protein aggregation and precipitation, which can occur during protein concentrating [6]. It is thought these charged amino acids interact to mask oppositely charged groups on the protein surface, whilst the aliphatic hydrophobic side chains cover exposed hydrophobic parts, thus reducing the surface "stickiness" which can cause aggregation [6].

2.2.7 Size-exclusion chromatography

Size-exclusion chromatography was employed primarily as a purification step but also during characterization of protein complex formation. Protein samples were concentrated using an Amicon Ultra centrifugal filter with a 10kDa cut-off (<3mL) and loaded onto a pre-equilibrated size-exclusion chromatography column (sephadex 200 HiLoad 16/60 column or sephadex 75 HiLoad 16/60 column) and run at a rate of 1 mL/min on the AKTA express at 4ºC (GE Healthcare). Fractions were collected and protein was detected by the peak in UV absorbance and accompanied by analysis by SDS-PAGE, prior to pooling of appropriate fractions.

Size exclusion chromatography buffer:

300 mM NaCl, 50 mM HEPES pH 7.5, 0.5 mM TCEP

2.2.8 Ion-Exchange Chromatography

Ion-exchange chromatography separates proteins based on their charge. Depending on the pI of the protein and protein complex, a buffer was chosen with a pH 1-1.5 difference to the pI. Anion exchange, using a HiTrap Q HP column (5 mL) (GE Healthcare) and cation exchange, using a HiTrap SP column (5 mL) (GE Healthcare) were used, as detailed in results sections. The column was pre-washed with high salt buffer and then equilibrated with the no salt buffer. The sample was loaded onto a pre-equilibrated column (flow rate 3 mL/min) and a linear elution gradient was run from 0 to 1 M NaCl.

Anion exchange buffers:

High salt buffer: 50 mM HEPES pH 7.5, 1M NaCl, 0.5 mM TCEP

No salt buffer: 50 mM HEPES pH 7.5, 0.5 mM TCEP

Cation exchange buffers:

High salt buffer: 50 mM MES pH 6.5, 1M NaCl, 0.5 mM TCEP

No salt buffer: 50 mM MES pH 6.5, 0.5 mM TCEP

Any deviation from this is specified in the results sections.

2.3 Protein characterization

2.3.1 Electrospray ionisation mass spectrometry (ESI-MS)

Protein masses were determined by ESI-MS, using an Agilent LC/MSD TOF system with reversed-phase HPLC coupled to electrospray ionization and an orthogonal time-of-flight mass analyzer. Proteins were desalted prior to mass spectrometry by rapid elution off a C3 column with a gradient of 5-95% isopropanol in water with 0.1% formic

acid. Spectra were analyzed and exported using the Agilent Masshunter Qualitative Analysis software.

2.3.2 Determination of protein concentration

Protein concentrations were measured at A280nm on a Nanodrop spectrophotometer (Thermo Scientific, Waltham, MA) against a buffer blank, using the molar extinction coefficient and molecular weight of the protein as derived from the primary sequence using Expasy Protparam tool. Molar extinction coefficients were estimated by using standard values for each tyrosine (ε = 1490 $M^{-1}cm^{-1}$) and tryptophan (ε = 5500 $M^{-1}cm^{-1}$) residue in the protein. Protein concentration was determined using the Beer-Lambert's Law (equation 1), where A is the absorbance of the protein, ε is the molar extinction coefficient (M^{-1} cm^{-1}), c is the protein concentration (M) and l is the path length (cm).

Equation 1: $A = \varepsilon c l$

2.3.3 Phosphorylation mapping (LC-MSMS)

Phosphorylation mapping was performed by Rod Chalk (SGC). Following sample cleavage with trypsin, potential phosphopeptides were analysed by LC-MSMS using a Dionex U3000 nano HPLC coupled to a Bruker Esquire HCT ion trap mass spectrometer.

2.3.4 Western blotting- HEK293 cells

The protein sample was run on SDS-PAGE and then transferred onto PVDF membrane (GE Healthcare) and incubated with blocking solution (TBST + 5% (w/v) skimmed milk powder) for 1 hour at room temperature. The membrane was then washed with PBS + 0.05% TWEEN 3 times, for 10 minutes each and probed with the monoclonal anti-flag M2-peroxidase (HRP) antibody (Sigma) (1:2333) dilution at 4 °C

overnight. The membrane was washed with PBS + 0.05% TWEEN and dried briefly. Protein bands were detected using ECL (Pierce) and an LAS4000 image reader.

2.4 Assessing protein-protein interactions

2.4.1 Pull-down assays

Pull-down assays were undertaken with purified proteins. 'Bait' protein was applied on to a gravity flow 30 mL column containing either Ni-sepharose or anti-FLAG M2 agarose equilibrated in binding buffer. Samples were kept agitated on the column without flow for 30 min to facilitate batch binding before the flow-through was collected. The column was washed two times to eliminate non-specific binding, prior to addition of the purified 'prey' protein at equimolar concentration and a further 30 minute incubation with agitation. The flow through was collected before an additional wash step and a final elution with either 250 mM imidazole elution buffer (for His-tagged bait) or binding buffer containing 100 µM FLAG peptide (for FLAG-tagged bait). Each fraction was analysed by SDS-PAGE. Control experiments were performed similarly by addition of 'prey' protein to beads in the absence of 'bait'.

2.4.2 Isothermal titration calorimetry (ITC)

ITC was used to measure protein binding. In ITC one binding partner is titrated into a solution containing the potential interacting partner. The direct thermodynamic observed in binding is the heat associated with this event and ITC provides a direct measurement of this heat change [7].

Calorimetric measurements were performed at 10°C using a VP-ITC calorimeter (MicroCal). Samples were dialysed against 50 mM HEPES pH 7.5, 150 mM NaCl and 1 mM DTT buffer overnight at 4°C in D-tube dialyzer midi tubes, with a 3.5 kDa cut-off (Millipore) and degassed before use. Each titration experiment consisted of an initial (2µl)

injection followed by a further 19 sequential injections of 15 μL. Intervals between each injection allowed equilibrium to complete. Heats of dilution were measured in blank titrations by injecting the test protein into a cell containing dialysis buffer and were subtracted from the binding heats. Data were fit to a 1:1 binding model using the Origin software package provided with the instrument (MicroCal).

2.5 Activity assays

2.5.1 Peptides

CDK2 consensus substrate peptide, HHASPRK, was synthesized by GenScript."PCTAIRE-tide" peptide with the consensus sequence (PKSPKARKKL) was ordered from Pepceuticals [8].CDK12 RNAPII CTD heptapeptide repeat peptides, with and without potential priming phosphorylation sites, were synthesized by Severn Biotech. CDK12 modified peptide identified in the Bosken *et al.*, paper "YSPTSPS(P)YSPTSPSSPS(P)YSPTSPS(P)Y-PEG" [9] was synthesized by Severn Biotech.

2.5.2 *In vitro* kinase assay

The ADP-Glo Kinase assay (Promega) was used to determine kinase activity. All reactions were performed at room temperature.

Assays were performed in 96-well plates with final volumes of 25 μL. For CDK16 with K-cyclin activity assays, the two-step assay first involves a 30 min kinase reaction buffered in 1 x kinase reaction buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mg/mL BSA) containing the protein kinase (0.01mg/mL), optional cyclin (0.01mg/mL), ultrapure ATP (100 μM) (Promega), with or without a substrate (CDK2 peptide, HHASPRK (200 μM), MBP (5 μM), casein (20 μM) or PCTAIRE-tide (20 μM)) and with or without an specific inhibitor (5 μM). Reactions were terminated with an equal volume

of ‘ADP-glo Reagent’ to deplete the remaining ATP. Following a 40 minute incubation, ‘Kinase Detection Reagent’ was added which both converts the ADP product back to ATP and induces a luciferase/luciferin reaction for quantification using a luminometer (POLARstar Omega, BMG Labtech). CDK2/CycA complex ability to phosphorylate the CDK2 peptide, was used alongside CDK16 activity assays as a positive control. For CDK16 with CycY activity assays, assays were performed in 384-well plates with final volumes of 5 µL. The two-step assay first involves a 30 min kinase reaction buffered in 1 x kinase reaction buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 20 mM $MgCl_2$, 0.1 mg/mL BSA) containing the protein kinase (2.5 µM), optional cyclin (2.5 µM), ultrapure ATP (100 µM) (Promega), with or without a PCTAIRE-tide (10 µM)) and with or without an specific inhibitor (6 µM). Reactions were terminated with an equal volume of ‘ADP-glo Reagent’ to deplete the remaining ATP. Following a 40 minute incubation, ‘Kinase Detection Reagent’ was added which both converts the ADP product back to ATP and induces a luciferase/luciferin reaction for quantification using a luminometer (POLARstar Omega, BMG Labtech). CDK2/CycA complex ability to phosphorylate the CDK2 peptide, was used alongside CDK16 activity assays as a positive control.

Assays were performed in 384-well plates with final volumes of 5 µL. For CDK12/CycK activity assays, the two-step assay first involves a 30 min kinase reaction buffered in 1 x kinase reaction buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mg/mL BSA) containing the protein kinase (2.5 µM), ultrapure ATP (1 mM) (Promega), with or without a substrate (0.5 mM). Reactions were terminated with an equal volume of ‘ADP-glo Reagent’ to deplete the remaining ATP. Following a 60 minute incubation, ‘Kinase Detection Reagent’ was added which both converts the ADP product back to ATP and induces a luciferase/luciferin reaction for quantification using a luminometer (POLARstar Omega, BMG Labtech).

2.5.3 Radioactivity assay

2.5.3.1. CDK12/CycK radiometric activity assay

Kinase assays on CDK12/CycK performed by Dr S. –W. Grace Cheng at the Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. Purified CDK complexes were incubated in kinase assay buffer (50 mM Tris pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 0.1% v/v NP-40, 1 mM DTT, 20 µM Na β-glycerophosphate, EDTA-free complete protease inhibitor cocktail) with varying amounts of cold ATP/ [γ-32P]ATP and GST-CTD substrate which contains all 52 heptad repeats of human RNA Pol II C-terminal domain [10]. Reactions were incubated in a circulating 30°C water bath for 15, 30 or 60 minutes. Kinase reactions were stopped with the addition of 6x SDS-PAGE loading dye. Samples were heated at 85°C for 5 minutes and resolved by 4-12% SDS-PAGE. Gels were subsequently dried on 3MM Whatmann paper and imaged with a FujiFilm FLA-7000 scanner. Phosphorylated bands were quantified using FujiFilm MultiGauge™ software [11].

2.5.3.2 CDK16 radiometric kinase activity assay

Kinase assays were performed by Saif Shehata, École Polytechnique Fédérale de Lausanne Nestlé Institute of Health Sciences S.A., Lausanne, Switzerland. The purified CDK16/CycY and inhibitor mixtures, were each assayed for phosphotransferase activity in a final assay volume of 50 µL containing 50 mM HEPES, pH 7.5, 0.1 mM EGTA, 10 mM magnesium acetate, 0.1 mM [γ-^{32}P]ATP (300 c.p.m./pmol), 0.03% Brij 35, 1 mM DTT and 50 µM PCTAIRE-tide (PKSPKARKKL) [8]. CDK16 and the compound were incubated for 5 minutes prior to the addition of the CycY. Reactions were incubated at 30^{0}C room-temperature for 30 minutes and terminated by spotting onto P81 paper and immersion in 75 mM H_3PO_4. P81 filters were washed three times for 10 min with H_3PO_4, rinsed with acetone, air-dried and incorporation of γ-^{32}P determined by Cherenkov counting.

2.6 Assessing ligand binding

2.6.1 Differential scanning fluorimetry (DSF)

DSF monitors thermal unfolding of a protein in solution in the presence of a fluorescent probe with affinity for hydrophobic parts of the protein, in this instance SYPRO Orange was used, which is favourable due to its high signal-to-noise ratio [12] (Fig. 2.2). The dye binds poorly to a folded, globular protein giving only low fluorescence but as the protein is heated and unfolds, more hydrophobic regions are exposed to which the dye binds and this can be measured as a change in fluorescence. The midpoint of the unfolding defines the apparent the melting temperature (Tm). Ligand binding to a native protein causes an increase in the protein's melting temperature that is proportional to the ligand's binding affinity and concentration [12]. Thus, thermal shift screening experiments are performed to determine the ΔTm calculated between the Tm of the protein alone and the Tm of the protein with ligand.

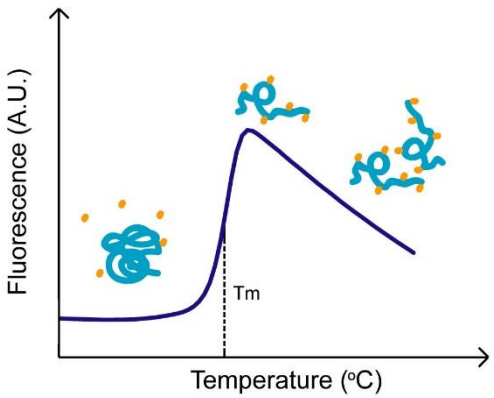


Figure 2.2. Schematic representation of the thermal unfolding of a protein. The dye binds poorly to folded protein giving only a low fluorescence. As the temperature increases, the protein will unfold revealing more hydrophobic regions to which the fluorescent dye may bind. The Tm is the midpoint of the unfolding. Following a peak in intensity, a gradual decrease in fluorescence will be observed as the protein comes out of solution as it precipitates and aggregates [12].

Experiments were performed in 96-well plate format, on an Mx3005p RT–PCR machine (Stratagene). Protein (final concentration 2 µM) was buffered in 10 mM HEPES pH7.5, 500 mM NaCl and mixed with SYPRO Orange. Compounds were added to 10 µM final concentration and an optical seal applied to the plate. Samples were heated at 1°C per min from 25°C to 95°C in a series of cycles. Fluorescence intensity was measured every 1°C. The data were then exported and analysed in Excel and Graphpad Prism using the Boltzmann equation (equation 2). The fluorescence intensity is plotted as a function of temperature to generate a sigmoidal curve, that can be described as a two-state transition [12]. The inflection point of the transition curve (Tm) can be calculated using the Boltzmann equation (Equation 2), where LL and UL are the values of minimum and maximum intensities, respectively, and a denotes the slope of the curve within the Tm.

[Equation 2] $$y = \mathrm{LL} + \frac{(\mathrm{UL} - \mathrm{LL})}{1 + \exp\left(\frac{Tm - x}{a}\right)}$$

2.7 X-ray crystallography

X-ray crystallography is a powerful tool to elucidate protein three-dimensional structure. The first challenge of crystallography is engineering a soluble protein construct that will transition from a disordered, soluble state and self-assemble into an ordered, crystal lattice. The protein must be purified to homogeneity then at this stage entered into crystallization trials [13]. Once a soluble, concentrated protein sample has been obtained, crystallography can be broadly divided into stages of protein crystal growth, data collection, phasing and modelling.

For crystals to form, the protein solution must enter supersaturated, thermodynamically metastable state [13, 14]. The protein sample must thus be concentrated to a high concentration but avoiding aggregation. At this stage nucleation may occur and a crystal can form whilst the solution equilibrates (Fig. 2.3). The most common technique is vapour diffusion and the method utilized here, is the sitting drop method. A protein drop is mixed with precipitant at various ratios in a small volume and sits in a well beside a larger reservoir of solution containing the same precipitant, which is then tightly sealed. As the droplet is more concentrated, the water vaporizes from the droplet into the reservoir, which serves to the increase the concentrations of both the protein and precipitant until the protein solubility limit is reached and at this stage supersaturation, nucleation and phase separation may occur [13].

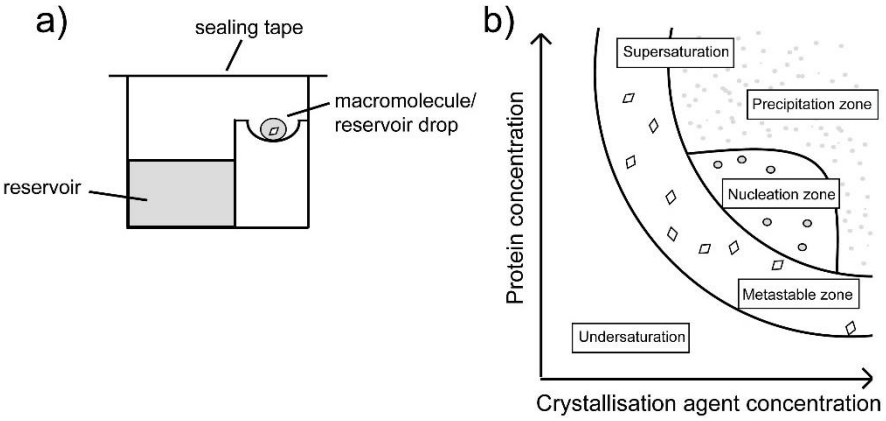


Figure 2.3. Protein crystal growth. a) A sitting drop in a vapour diffusion experiment. **b)** A phase diagram. The various zones of precipitation, nucleation and supersaturations are represented in a phase diagram which shows protein concentration against precipitant concentration (Adapted from Chayen & Saridakis [14]).

2.7.1 Crystallization trials

Finding the condition in which your protein will crystallise, may be achieved from screening a large number of variable parameters in parallel, such as precipitant choice and

concentration, buffer type, pH and temperature [13, 14]. Sparse-matrix sampling is employed in the SGC (2.7.2).

Proteins were buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM DTT or with buffers as specified in the results sections. All crystallization experiments were performed in 96-well, three-subwell SWISSCI plates (SWISSCI AG) using the sitting-drop vapour-diffusion method and crystals were grown at either 4°C or 20°C. Reservoirs of 20 µL were used for equilibration against 150 nL fixed-volume crystallization drops, which were prepared at 2:1, 1:1 and 1:2 protein:reservoir volume ratios, in most instances. Plates were set up using the TTP Labtech's mosquito Crystal liquid handling robot, dispensing protein drops and transfer the mother liquor. Crystallization drops were imaged regularly at set intervals over a period of two months using an automatic inspection system (Rigaku).

Two different sparse matrices were employed in most instances, offering trial conditions selected from known crystallization conditions for macromolecules. Parallel experiments under identical experimental parameters such as protein concentrations, protein buffers, drop ratios and temperatures were set up with four primary, coarse screens, namely the commercial Hampton Research Crystal Screen 1 and 2 (HCS) and the Hampton Research Index (HIN) screen, as well as the sparse-matrix screen "JCSG" which is a screen developed by a collection of academics at the Joint Center for Structural Genomics, and the SGC equivalent to a PACT (a pH-, anion- and cation-testing) screen, "LFS" (ligand friendly screen) [16, 17].

JCSG is made up of a combination of conditions from the Joint Center for Structural Genomics (JCSG), which was designed to extend conditions from other screens to maximize the coverage of crystallization parameter space with no redundancy that using

multiple vendor screens would bring [16]. The LFS screen does not cover a wide crystallization space but rather aims to decouple the components of each condition which can provide information about the protein, even in the absence of crystals [16].

The BCS screen, a further coarse screen, was used in some instances which samples widely used polyethylene glycol (PEG) precipitants through the use of PEG smears, which are "mixtures of different PEGs with a requirement of either neutral or cooperatively positive effects of each component on crystal growth" [17].

In some instances, further fine screens were utilised based around conditions in which crystals or crystalline precipitate was observed, which could be designed and made in-house using the Multiprobe II plus liquid handler (PerkinElmer).

2.7.1.1 Crystallization trials with inhibitors or nucleotides

For crystallisation trials of protein with an inhibitor or nucleotide, the inhibitor or nucleotide was added to the protein when the protein sample was more dilute and then subsequently concentrated in a 10 kDa MWCO Amicon Ultra concentrator to a desired amount in the presence of the inhibitor or nucleotide.

2.7.1.2 Crystallization trials with covalent inhibitors

The purified CDK12/cyclin K complex was incubated overnight at 4°C, with covalent inhibitor (THZ1, THZ-4-90-1 or THZ531) at 1.5-times the determined protein concentration. Near complete covalent binding was shown by a mass shift in the CDK12 of 566 Da or 559 Da depending on the compound, as determined using mass spectrometry. Excess free compound was removed by size-exclusion chromatography using a HiLoad Superdex S75 26/60 column (GE Healthcare). The protein-inhibitor complex was concentrated in a 10 kDa MWCO Amicon Ultra concentrator and the final concentration

was determined on the Nanodrop, using an estimated extinction coefficient of 35870 $M^{-1}cm^{-1}$ and a molecular weight of 37252 Da. to 5.5 mg/mL buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP. Crystals were grown at 20°C in 195 nL sitting drops mixing 75 nL protein solution and 20 nL crystal seed stock with 100 nL of a reservoir solution comprising 0.1 M HEPEs pH 7.0, 10% PEG8000, 0.2 M magnesium chloride. Crystals were cryo-protected with mother liquor supplemented with an additional 15% ethylene glycol and vitrified in liquid nitrogen after mounting.

2.7.2 Microseeding

A microseed stock was prepared on a method adapted from D'Arcy et al.[15]. Microseeding takes crystalline material generated in an experiment to add to a new experiment. This crystalline material can act as a nucleation site for crystal growth, removing the need for *de novo* nucleation [15]. Crystals of the same protein were used. 6µl of the reservoir solution was added to the well containing crystals and this was resuspended several times and transferred to an Eppendorf tube on ice containing a "seed bead". This process was repeated three times to ensure all the crystals were transferred. The Eppendorf tube was vortexed for 2 minutes, with stopping every 30 seconds to cool the tube on ice. The microseed stock was used immediately, adding 20 nL to each subwell.

2.7.3 Mounting

The solution which surrounds the formed crystals in which they grew is described as the "mother liquor". Crystals when removed for mounting are kept in mother liquor to prevent dehydration. Mother liquor was supplemented with the chosen cryo-protectant, generally ethylene glycol, and added to the crystal drop. The crystal was cryo-cooled in liquid nitrogen after mounting. Mounting was done at the same temperature at which the plate had been incubated.

2.7.4 Data collection

Diffraction data were collected at Diamond Light Source on beamlines I02, I03, I04, I04-1 and I24.

2.7.5 Data processing

Data were indexed and integrated using XDS [18] or MOSFLM [19] and scaled using AIMLESS [20]. The likely content of the asymmetric unit was determined using Matthews coefficient in the CCP4 suite of programs [21]. Phases were found using molecular replacement in PHASER [22] with various input PDB models as stated in the results sections. A PDB file for AMP-PNP and the inhibitor THZ-5-31-1, was created using PRODRG in the CCP4 suite [21]. A PDB model for rebastinib was created using ACEDRG of the CCP4 suite [21]. Models were built initially using BUCCANEER [21] or COOT [23] and then refined and modified using alternate rounds of REFMAC5 [24] and COOT. In case some cases the final rounds of refinement were done in PHENIX [25].

The refined structures were validated with MolProbity [26] and the atomic coordinate files deposited in the Protein Data Bank. Structure figures were prepared with PyMOL [27].

The protein structure analysis was undertaken using the 'Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute (ePISA) (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html), to determine the CDK/Cyclin complex interfaces (A^2) [28].

Omit maps were calculated using PHENIX[25] for structures by removing the region of interest, protein or ligand, from the PDB file and running PHENIX with the structure factor file. The omit map was contoured at 2.5σ.

2.8 Cell culture and assays

2.8.1 Maintenance of cancer cell cultures

Melanoma cancer cell lines SK-Mel-28 (ATCC HTB-72), IGR-37, IG7-39 and 501-Mel were obtained from Professor Colin Goding's laboratory (Ludwig Institute, University of Oxford) [29]. Medulloblastoma cancer cell line, DAOY (ATCC HTB-186), was obtained from Professor Nicola Sibson's laboratory (Department of Oncology, University of Oxford).

The five cancer cell lines were grown in plastic flasks as adherent cultures (T25 or T75 cm^2 flasks) in culture media conditions as detailed below.

Cell lines were maintained at 37°C, at 5% CO_2 and passaged every two days for the DAOY and 501-Mel cell lines, whereas SK-Mel-28, IGR-37 and IG7-39 were passaged every three days.

Melanoma cell line maintenance media:

Growing Medium: RPMI + 10% FBS + 1% pent/strep

Freezing medium: 80% RPMI + 10% FBS + 1% pent/strep + 10% DMSO

DAOY cell line maintenance media:

Growing Medium: DMEM + 10% FBS + 1% Glutamine

Freezing medium: 70% DMEM + 10% Serum + 20% DMSO

2.8.2 Western blotting

Protein expression was analysed by Western blotting. Cells were plated at $4x10^5$cells/well in 6-well plates and incubated at 37°C, at 5% CO_2 for 72 hours. Cells were then harvested and lysed in 20 mM Tris–HCl pH 7.5, 150 mM NaCl, 1 % Triton X-100, 25 mM NaF, 25 mM Na_3VO_4, 2 mM Na β-glycerophosphate and protease inhibitors (Roche) on ice for 30 min. Protein concentration was determined by BCA assay (Pierce)

and 20 µg run on 4–12 % Bis–Tris gel (Life Technologies). The protein was transferred onto PVDF membrane (GE Healthcare) pre-activated by methanol for 1 minute and incubated with blocking solution (TBST + 5% (w/v) skimmed milk powder) for 1 hour at room temperature. The membrane was then washed with PBS + 0.05% TWEEN 3 times for 10 minutes each and probed with the relevant primary antibody at 4 °C overnight. After membrane washing with PTBS + 10% TWEEN the appropriate secondary antibodies were added and incubated for 1 hour at room temperature. The membrane was washed with PBS + 0.05% TWEEN and dried briefly. Protein bands were detected using ECL (Pierce) and an LAS4000 image reader.

Antibodies used were: CDK16 polyclonal antibody was from Sigma-Aldrich (rabbit, 1:250 dilution). Cyclin Y polyclonal antibody was from Proteintech (rabbit, 1:500 dilution). p27 (C-19): SC-528 polyclonal antibody was from Santa Cruz Biotechnologies (rabbit, 1:100 dilution). β-tubulin antibody was from Sigma-Aldrich (mouse, 1:500 dilution). Secondary antibodies anti-rabbit IgG (Sigma, #A6667) and Anti-mouse IgG (Dako) were diluted at 1:2000.

2.8.3 RNA prep and quantitative PCR

Cells were plated at $3.8x10^5$cells/well in 6-well plates in 3ml. After 72 hours, cells were harvested and total RNA was prepared using the Qiagen RNeasy RNA extraction kit. A total of 0.5 µg total RNA was converted to cDNA using Superscript III reverse transcriptase (Invitrogen) using oligo (dT) primers and following the manufacturer's protocol. Triplicate wells were subjected to comparative quantitative PCR using SensiMix SYBR Low-ROX (Bioline) and gene-specific primers for CDK16, cyclin Y and p27 and GAPDH. The plate was then briefly spun down and placed into a Real-Time PCR machine (Stratagene Mx3005p). The data were then exported and analysis was performed in MxPro QPCR Software (Agilent Technologies). Expression levels were normalised to GAPDH

and the relative quantity (dRn) of untreated samples was calculated. The data is shown as relative levels to a one of the cancer cell lines. Primer sequences for qPCR are provided in Table 2.2.

Table 2.2. Primer sequences for qPCR.

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
$p27^{Kip1}$	GGTTAGCGGAGCAATGCG	TCCACAGAACCGGCATTTG
CDK16	GCAGTGACCCTGGAGAGG	TCAAGTCCTCGTGCACAATC
Cyclin Y	CCAGCACAATATTCCTCAGT	CTAAGAGCATCCTTCCATCT
GAPDH	CGACCACTTTGTCAAGCTCA	AGGGGAGATTCAGTGTGGTG

2.8.4 Cell Viability Measurements

The cancer cell lines were seeded at $3x10^4$ cells/well in a 96-well plate, at 50 µL volume in each specified media per cell line. After 24 hours, an additional 50 uL media was added containing either the inhibitors, 0.5 % DMSO or media only to achieve various inhibitor concentrations whilst ensuring the same volume of DMSO in all test wells. The plates were incubated at 37°C, at 5% CO_2 for 72 hours. Cell viability was measured using the CellTiter-Glo® Luminescent cell-viability assay (Promega) on a luminometer (GloMax-Multi Microplate Multimode Reader, Promega) after 72 hours. The CellTiter-Glo® Luminescent cell-viability assay Promega protocol provided was followed, with some deviations following advice within the laboratory of other users. Clear, flat-bottom plates, were used rather than white plates, as they are also compatible with the luminometer as cross-talk is deemed negligible. This enabled images to be taken of the wells on the

(2.9.4). The CellTiter-Glo reagent was diluted in PBS, 1 in 4, as this has been shown to be as effective.

2.8.5 Imaging of cancer cell lines

At 72 hours, prior to the cell viability assay measurements, images were taken of the cells using the Floid Cell Imaging Station light microscopy (Thermofischer).

2.9 References

1. The, M.G.C.P.T., et al., *The completion of the Mammalian Gene Collection (MGC).* Genome Research, 2009. 19(12): p. 2324-2333.
2. Gileadi, O., et al., *High throughput production of recombinant human proteins for crystallography.* Methods Mol Biol, 2008. 426: p. 221-46.
3. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* J Mol Biol, 1999. 292(2): p. 195-202.
4. Linding, R., et al., *GlobPlot: Exploring protein sequences for globularity and disorder.* Nucleic Acids Res, 2003. 31(13): p. 3701-8.
5. Savitsky, P., et al., *High-throughput production of human proteins for crystallization: the SGC experience.* J Struct Biol, 2010. 172(1): p. 3-13.
6. Golovanov, A.P., et al., *A simple method for improving protein solubility and long-term stability.* J Am Chem Soc, 2004. 126(29): p. 8933-9.
7. Perozzo, R., G. Folkers, and L. Scapozza, *Thermodynamics of protein-ligand interactions: history, presence, and future aspects.* J Recept Signal Transduct Res, 2004. 24(1-2): p. 1-52.

8. Shehata, S.N., et al., *Analysis of substrate specificity and cyclin Y binding of PCTAIRE-1 kinase.* Cell Signal, 2012. 24(11): p. 2085-94.

9. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K.* Nat Commun, 2014. 5: p. 3505.

10. Cheng, S.W., et al., *Interaction of cyclin-dependent kinase 12/CrkRS with cyclin K1 is required for the phosphorylation of the C-terminal domain of RNA polymerase II.* Mol Cell Biol, 2012. 32(22): p. 4691-704.

11. Dixon-Clarke, S.E., et al., *Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding.* Sci Rep, 2015. 5.

12. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. 2(9): p. 2212-21.

13. Rupp, B., *Biomolecular Crystallography. Principles, Practice, and Application to Structural Biology*. 2010, Abingdon, UK: Garland Science. 800.

14. Chayen, N.E. and E. Saridakis, *Protein crystallization: from purified protein to diffraction-quality crystal.* Nat Methods, 2008. 5(2): p. 147-53.

15. D'Arcy, A., et al., *Microseed matrix screening for optimization in protein crystallization: what have we learned?* Acta Crystallogr F Struct Biol Commun, 2014. 70(Pt 9): p. 1117-26.

16. Newman, J., et al., *Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy.* Acta Crystallogr D Biol Crystallogr, 2005. 61(Pt 10): p. 1426-31.

17. Chaikuad, A., S. Knapp, and F. von Delft, *Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens.* Acta Crystallogr D Biol Crystallogr, 2015. 71(Pt 8): p. 1627-39.

18. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. 66(Pt 2): p. 125-32.

19. Leslie, A.G.W. and H.R. Powell, *Processing diffraction data with mosflm*

Evolving Methods for Macromolecular Crystallography, R.J. Read and J.L. Sussman, Editors. 2007, Springer Netherlands. p. 41-51.

20. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. 69(Pt 7): p. 1204-14.

21. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. 67(Pt 4): p. 235-42.

22. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. 40(Pt 4): p. 658-674.

23. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. 66(Pt 4): p. 486-501.

24. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallogr D Biol Crystallogr, 2011. 67(Pt 4): p. 355-67.

25. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. 66(Pt 2): p. 213-21.

26. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. 66(Pt 1): p. 12-21.

27. Schrödinger LLC., *The PyMOL Molecular Graphics System*. Version 1.2r3pre.

28. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. 372(3): p. 774-97.

29. Halling-Brown, M.D., et al., *canSAR: an integrated cancer public translational research and drug discovery resource.* Nucleic Acids Research, 2012. 40(Database issue): p. D947-D956.

Chapter 3 Structures of the CDK12/CycK complex

3.1 Introduction

The CDK12/CycK complex has been characterised as a transcriptional CDK complex showing comparable activity to the CDK9/CycT complex. Both CDKs have been shown to phosphorylate Ser2 in the heptad repeat of the RNA Pol II C-terminal domain [1, 2] . The CycK subunit was also initially reported as an alternative cyclin partner of CDK9, but more recent evidence suggests that it associates specifically with CDK12 and the close homologue CDK13 [3] . At the start of this project there were no available structures of CDK12 or CDK13 to explain their activity or cyclin binding preferences. The aim of this chapter was therefore to elucidate the first structure of the kinase domain of CDK12 and its proposed complex with CycK. This knowledge may shed light upon the unique features of an active CDK12/CycK complex and how these compare to the structure of the CDK9/CycT1 complex [4, 5].

Staff at the SGC had previously tried to express human CDK12 protein without success. A number of constructs exploring truncations between CDK12 residues Arg661 to Gly1099 were cloned into a modified pFastBac vector and expressed in *Spodoptera frugiperda (Sf9)* insect cells. These constructs were designed to include the predicted kinase domain and variable extensions, including some incorporating the proline-rich motif (Fig. 3.1). However, no constructs were identified that gave soluble expression. Similar poor results were obtained when expressing a range of CDK12 truncations between Arg701 to Glu1059 in *E. coli* cells.

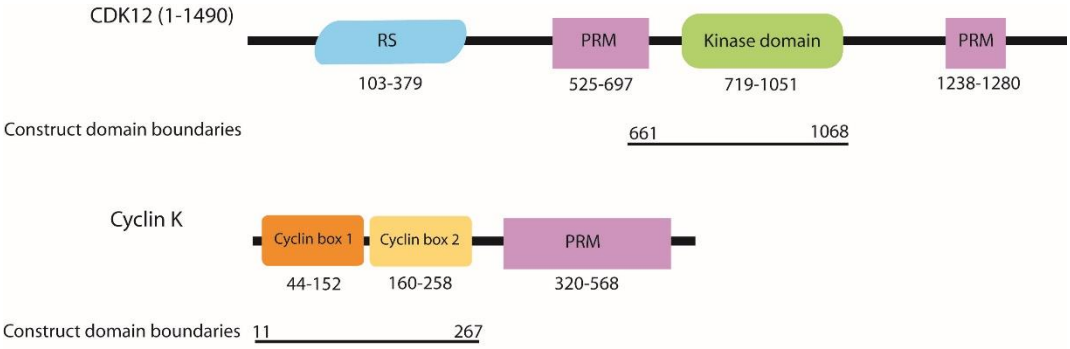


Figure 3.1. Schematic showing the domain organisation of human CDK12 and CycK and construct domain boundaries. The longest construct domain boundaries are shown below as a black line (RS, arginine-serine-rich; PRM, proline-rich motifs).

In this chapter, I establish co-expression and purification protocols for a variety of human CDK12/CycK complexes and report the crystal structures of two CDK12/CycK complexes solved in the presence of AMP-PNP and containing variable truncations of a C-terminal kinase extension. The structures exhibit traits of an active CDK12/CycK complex, consistent with their activation loop threonine phosphorylation performed by CAK.

3.2 Results

3.2.1 Soluble expression of CDK12 requires CycK

Some 14 hexahistidine-tagged CDK12 constructs exploring truncations between residues Arg661 to Gly1099 were retested for their soluble expression in *Sf*9 insect cells. Small scale (3 mL) nickel-affinity purifications indicated that none of the 14 were solubly expressed as observed previously (Fig. 3.2). Cyclin-dependent kinases display inherent plasticity that may be stabilized by binding of their cognate cyclin. Therefore, co-expression of CDK12 constructs was performed with the potential cyclin partners, including cyclin K1 (CycK) and cyclin L1 (CycL1). Neither of the CycL1 constructs gave soluble expression and CDK12 protein was again not recovered (Fig. 3.2). In contrast, the majority of CDK12 constructs were solubly expressed when co-expressed with CycK$^{11\text{-}267}$

(cyclin domain corresponding to PDB ID: 2I53) allowing purification of the various CDK12/CycK complexes (Fig. 3.2). Of note, it was observed that CDK12 constructs with the shortest N-terminus at residue Val724 were insoluble and not rescued by co-expression with CycK. The shortest truncation of the C-terminus at residue Asp1032 was also found to be detrimental.

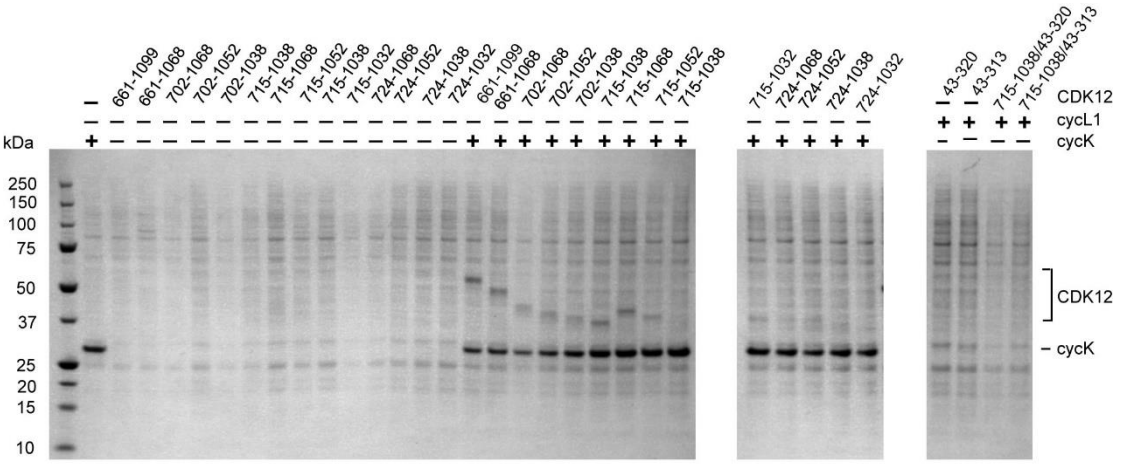


Figure 3.2. Soluble expression of CDK12 requires CycK. Small scale nickel-affinity purifications from 3 mL baculoviral expression of hexahistidine-tagged proteins. CDK12 constructs are indicated as well as $CycK^{11-267}$, $CycL1^{43-320}$ and $CycL1^{43-313}$.

3.2.2 Different constructs of CDK12 show variable phosphorylation

Mass spectrometry revealed that many of the expressed CDK12 proteins were mono-phosphorylated to a greater or lesser extent (Fig. 3.3), whereas constructs with shorter C-termini were essentially unphosphorylated (Fig. 3.3).

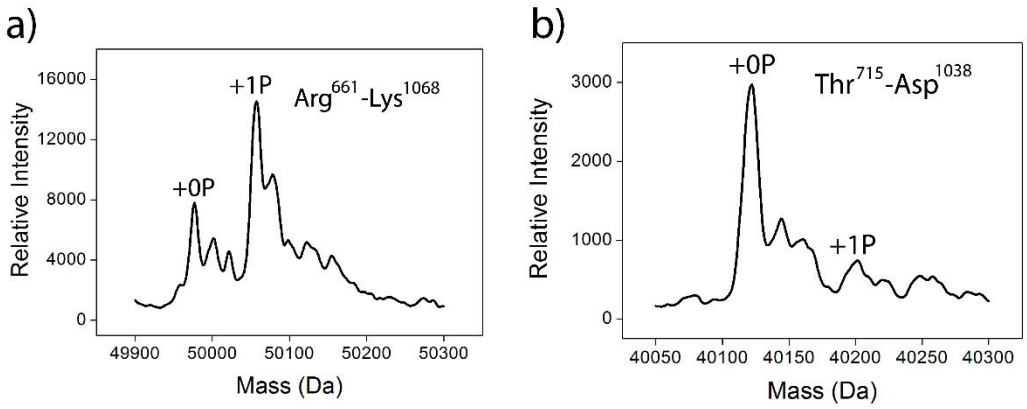


Figure 3.3. Phosphorylation status of hexahistidine-tagged CDK12 protein constructs following co-expression with CycK in *Sf9* insect cells. Following small-scale co-expression and nickel-affinity purification, the CDK12 constructs exhibited variable levels of phosphorylation, shown here by representative long and short CDK12 constructs. a) CDK12$^{661\text{-}1068}$ was a longer construct and the mono-phosphorylated form was the predominant species. b) CDK12$^{715\text{-}1038}$ was one of the shortest constructs and was predominantly unphosphorylated. The two mass peaks represent the native unphosphorylated protein (0P) as well as a species containing a single phosphorylation on CDK12 (shown as +1P).

3.2.3 Optimisation of CDK12/CycK complexes for crystallisation

The largest complex comprising CDK12$^{661\text{-}1099}$ and CycK$^{11\text{-}267}$ had appeared promising in the small scale test expressions. However, this complex had a low yield upon scale up of its expression and subsequent purification steps resulted in protein precipitation, despite addition of a reducing agent, 10 mM dithiothreitol (DTT), and a mixture of 10 mM L-arginine and 10 mM L-glutamine, which is reported to help reduce protein aggregation and precipitation [6] (Data not shown). The final yields were too low for this protein complex to enter into crystallisation trials. Subsequent effort was therefore focused on other more truncated complexes that were more stable and were purified with higher yields.

Small scale (3 mL) trials indicated that co-expression for 72 hours gave higher yields than 48 hours (see Fig. 3.4 for example data using the CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex). However, the relative expression levels of CDK12 and CycK remained unequal with a large excess of CycK protein observed (Fig. 3.4). The CDK12 and CycK constructs

all had a hexahistidine tag as they were originally cloned for individual expression. For complexes, it is usually preferable if each protein has a different tag or if one is untagged so that affinity chromatography may be used to capture homogeneous complex and to exclude any excess of one monomer. In this case in the interest of time it was decided to test whether further optimisation of the expression and purification steps could be undertaken, instead of re-cloning, to remove excess CycK to achieve a homogeneous 1:1 protein complex.

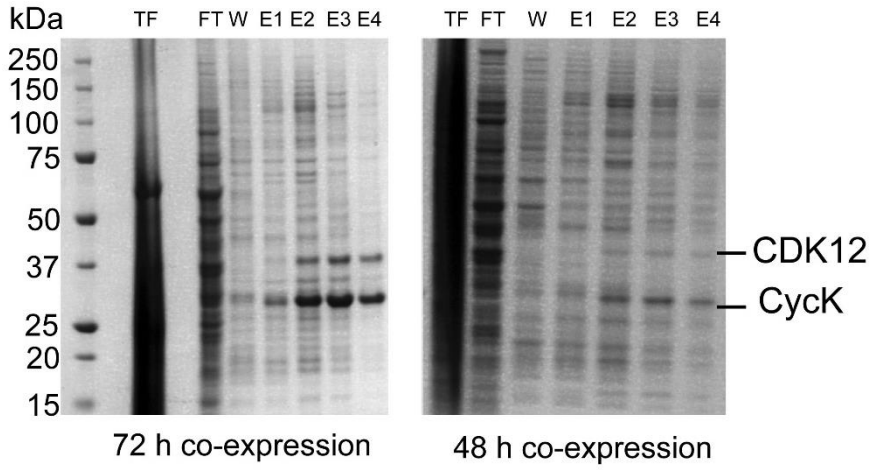


Figure 3.4. Small-scale 50 mL test co-expression of CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ for 48 hours or 72 hours in *Sf9* insect cells. A higher level of protein expression was seen after 72 hours expression (left-hand side) as opposed to only 48 hours expression (right-hand side) when using a 1:1 ratio of P2 virus volume. (TF=total fraction; FT=flow-through fraction; W=wash fraction, E1-E4=elution fractions. Please refer to materials and methods for more details including buffer composition.)

Scale up trials using 1L cultures and co-expression for 72 hours were conducted with several different CDK12 constructs. It was found that more equal expression levels of CDK12 and CycK could be achieved by adjusting the ratio of the volume of the two P2 viruses used for infection of the *Sf9* cells. Twice the volume of CDK12 P2 virus compared to CycK was particularly successful in co-expression (Fig.3.5). Subsequent size-exclusion chromatography could then be used to separate the remaining excess of CycK protein.

Example purifications of the CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ and CDK12$^{715\text{-}1051}$/CycK$^{11\text{-}267}$ complexes are shown in figure 3.5.

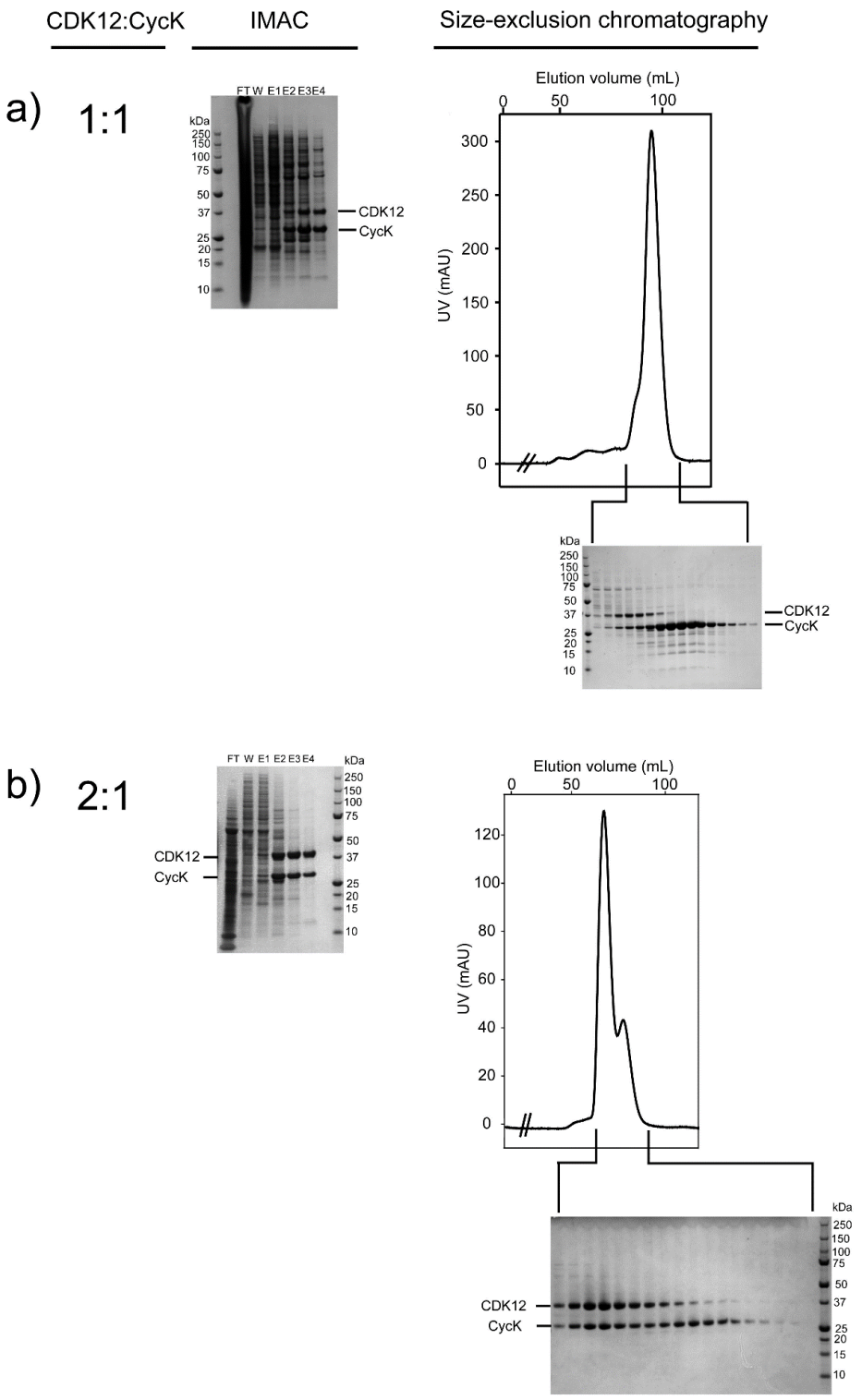


Figure 3.5. Expression and purification steps undertaken to achieve a 1:1 complex of CDK12 and CycK. a) Example purification of CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ from co-expression using a 1:1 P2 virus ratio for infection of *Sf9* insect cells. CycK was in excess of CDK12 after IMAC, but was removed by subsequent size-exclusion chromatography. In this example, fractions 1-5 were pooled for further characterisation. **b)** Example purification of CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ from co-expression using a 2:1 P2 virus ratio for infection of *Sf9* insect cells. This gave more equal expression of CDK12 and CycK and subsequent size-exclusion chromatography again separated excess CycK. Elution fractions 2-7 were then pooled for further purification steps (not shown). (FT=flow-through fraction; W=wash fraction, E1-E4=elution fractions. Please refer to materials and methods for more details.)

Co-expression of CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ in *Sf9* insect cells yielded CDK12 protein that was essentially unphosphorylated. The experimental mass of 37,686 Da was 89 Da smaller than the theoretical mass as expected after cleavage of the N-terminal initiator methionine and acetylation of the new N-terminal glycine which precedes the hexahistidine tag. (Fig. 3.6). CycK had an experimental mass of 30,403 Da following tag cleavage, which matched the theoretical mass (Data not shown). Given CDK12 has a threonine in the activation 'T-loop' that is typically phosphorylated by CAK in other CDKs, I attempted to phosphorylate CDK12 *in vitro* by incubation of the purified CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex with recombinant CAK from *Candida albicans* (a gift from my supervisor Dr Alex Bullock). A small scale trial yielded near complete mono-phosphorylation of CDK12, but no modification on CycK. Upon scale up, 1.2 mg of CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex was treated with 80 µg of CAK in the presence of 1 mM ATP, 5 mM $MgCl_2$ and 0.1 mM $MnCl_2$ at room temperature, which resulted again in mono-phosphorylated CDK12, as detected by intact mass spectrometry (Fig. 3.6). For final clean up, size exclusion chromatography using a HiLoad Superdex S75 26/60 column (GE Healthcare) buffered in 50 mM HEPES pH 7.5, 300 mM NaCl and 1 mM TCEP. Lastly, the proteins underwent reverse nickel-affinity purification to remove remaining CAK, which was expressed with a non-cleavable hexahistidine tag. Interestingly, the larger construct CDK12$^{715\text{-}1052}$ was mono-phosphorylated after co-expression without further CAK treatment (Fig. 3.6). Again, this complex could be purified to homogeneity using size-exclusion chromatography (Fig 3.5).

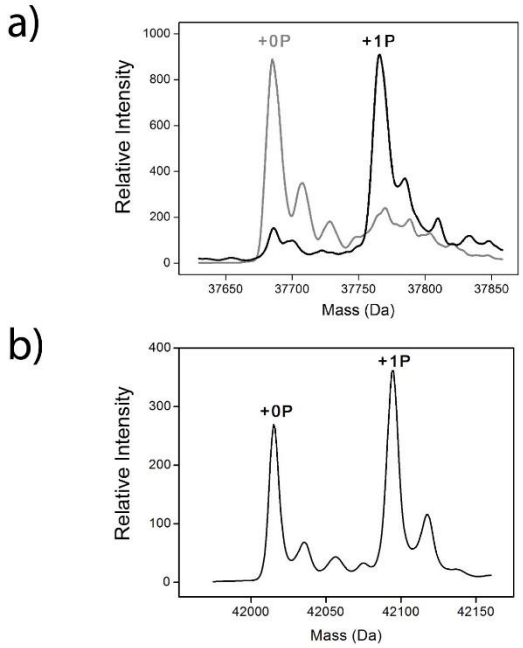


Figure 3.6. Deconvoluted intact mass spectra for two CDK12 constructs. a) Deconvoluted intact mass spectrum for CDK12$^{715\text{-}1038}$ purified following co-expression with CycK$^{11\text{-}267}$ and obtained before and after treatment with recombinant CAK from *Candida albicans*. This short CDK12 construct was essentially unphosphorylated until CAK treatment. **b)** Deconvoluted intact mass spectrum for CDK12$^{715\text{-}1052}$ purified following co-expression with CycK$^{11\text{-}267}$. The two mass peaks represent the native protein (0P) as well as a larger species containing a single phosphorylation on CDK12 Thr893 (shown as +1P).

3.2.4 Autophosphorylation trials of the CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ and CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complexes

The phosphorylation status of the different CDK12 constructs following co-expression in insect cells was varied, although essentially all showed at least some proportion of mono-phosphorylated species. In insect cell expression systems, post-translational modifications may be performed by endogenous host kinases similarly to mammalian expression. Thus, the recombinant CDK12 proteins may be modified either by an endogenous CAK, or by autophosphorylation. Indeed, CDK9 can autophosphorylate on its activation loop in *cis* [4]. To address this question, *in vitro* autophosphorylation trials were performed with two different purified CDK12/CycK complexes. Neither the CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex (Fig. 3.7) nor the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complex (Fig. 3.8) showed any evidence of autophosphorylation. This would suggest that the mono-

phosphorylation of CDK12 upon expression resulted from an endogenous CAK, although it cannot be ruled out that other host factors (such as chaperones) could enable CDK12 autophosphorylation in cells that is not observed after purification. This could be tested in future using a kinase-dead mutant of CDK12. Notably, Liang *et al*., have shown that Flag-purified full-length CDK12, CDK13 and CDK9 can all autophosphorylate in an *in vitro* kinase assay, although the sites were not determined. Based on my data, it is unlikely that the CDK12 site was within the kinase domain [7].

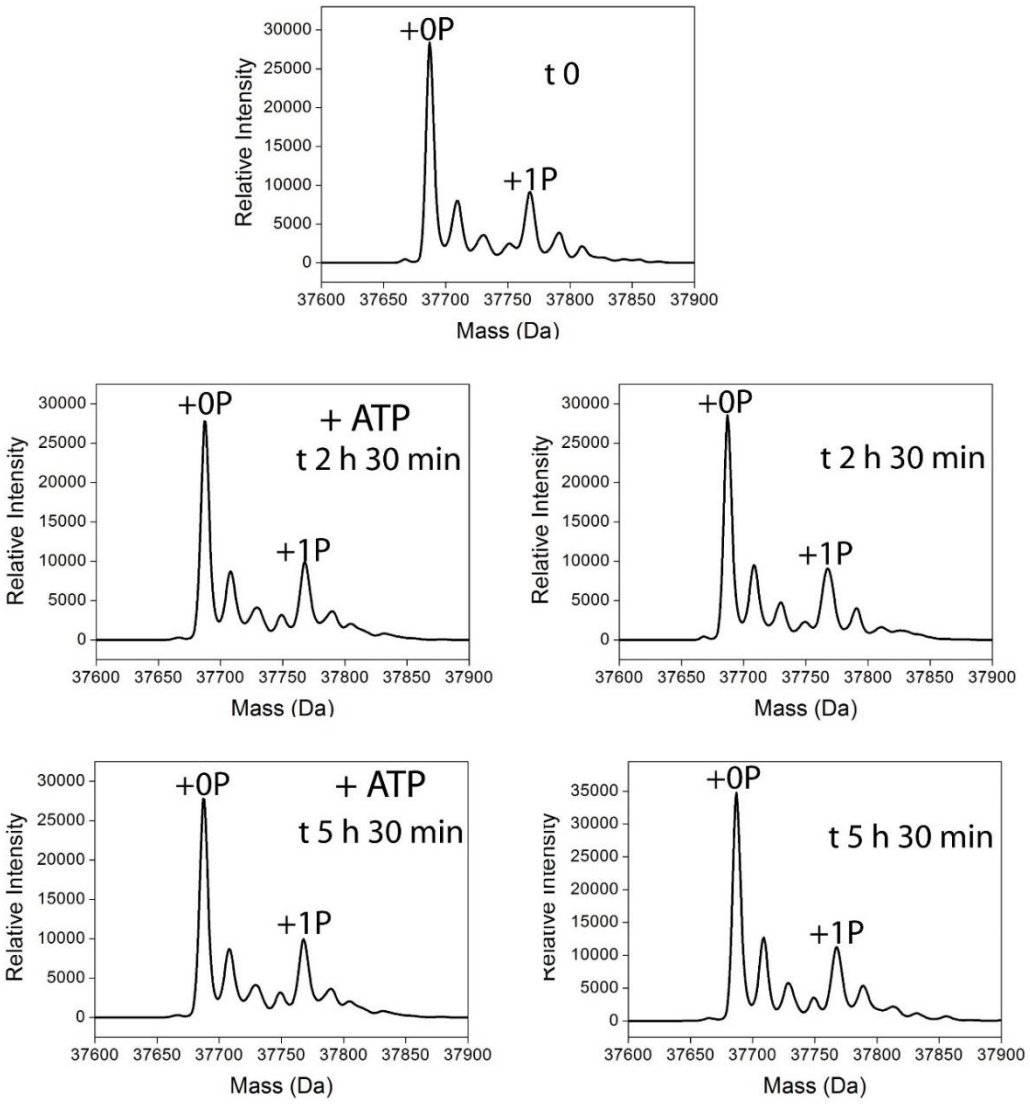


Figure 3.7. CDK12$^{715\text{-}1038}$ did not autophosphorylate *in vitro*. CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ was incubated at room temperature in the presence of 1 mM ATP, 5 mM $MgCl_2$ and 0.1 mM $MnCl_2$. No change in the level of mono-phosphorylated CDK12 was apparent.

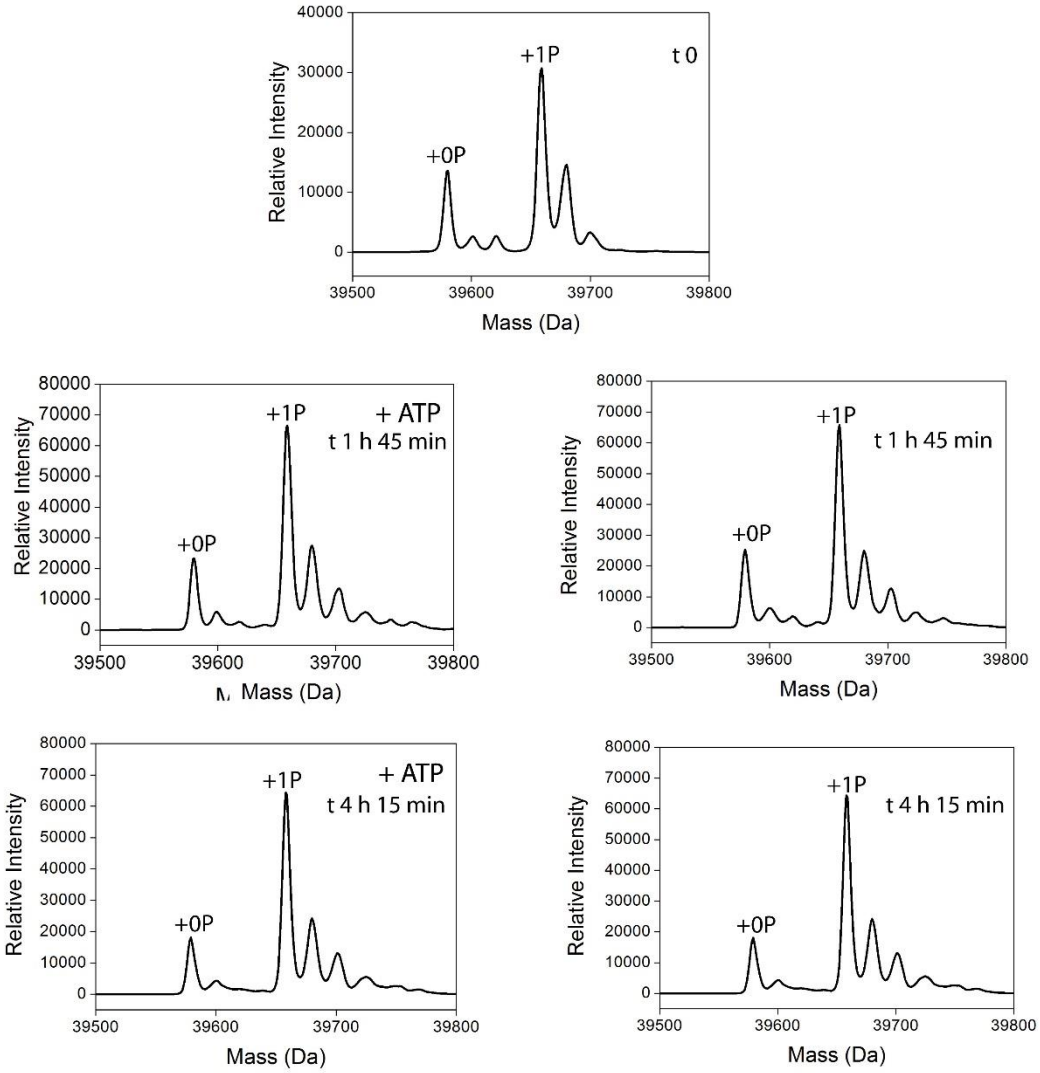


Figure 3.8. CDK12$^{715\text{-}1052}$ did not autophosphorylate *in vitro*. CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ was incubated at room temperature in the presence of 1 mM ATP, 5 mM $MgCl_2$ and 0.1 mM $MnCl_2$. No change in the level of mono-phosphorylated CDK12 was apparent.

3.2.5 Crystallization trials of the CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ and CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complexes

Crystal trials were performed using the vapour diffusion method with sitting drops. The first screens were set up using the smaller, predominantly unphosphorylated CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex. This protein complex buffered in 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM imidazole, 0.5 mM TCEP, 10 mM DTT was

concentrated to 7.2 mg/mL and mixed with the non-hydrolysable ATP analogue, adenylyl imidodiphosphate (AMP-PNP) at 1 mM concentration. A variety of crystallisation screens were tested in which the reservoir solutions were supplemented with an additional 0.1 mM $MgCl_2$ (Table 3.1). Initial crystals were obtained at 4 °C after 20 days in a condition containing 0.20 M potassium thiocyanate; 0.1 M bis-tris propane pH 6.5; 20.0% PEG 3350;10.0% ethylene glycol. While these crystals diffracted to 2.3 Å resolution they exhibited multiple lattices that could not be indexed. The mounted crystal was subsequently analysed by intact mass spectrometry, which revealed only the mass for CycK (Fig 3.9). SDS-PAGE of this purification had also suggested an excess of CycK in the complex (not shown), indicating again that free CycK may have crystallized alone. Other crystals from the same protein preparation gave only poor diffraction.

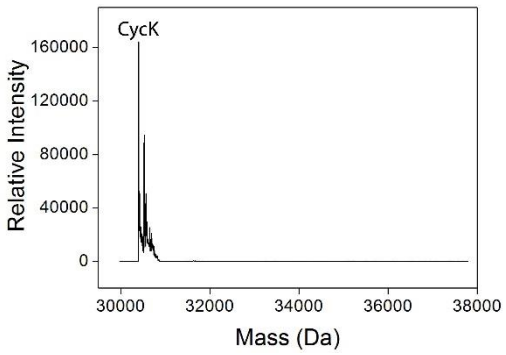


Figure 3.9. Deconvoluted intact mass spectrum obtained from a crystal of the unphosphorylated CDK12$^{715-1038}$/CycK$^{11-267}$ complex. A peak at 30402 Da corresponds to the mass of CycK$^{11-267}$. A peak at 30,526 Da corresponds to a shift of +124 Da of unknown identity. No peak is present for CDK12$^{715-1038}$, which has an expected mass of 37686 Da.

Subsequently, crystal trials were performed using CDK12$^{715-1038}$/CycK$^{11-267}$ complex that had been mono-phosphorylated by CAK treatment. The presence of the CAK meant further purification steps were undertaken including a reverse purification by IMAC

to remove the CAK and a final size exclusion chromatography step to remove any slight excess of CycK. The final protein sample was buffered in 50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM imidazole, 0.5 mM TCEP, 10 mM DTT and concentrated to 4.3 mg/mL. AMP-PNP was added to a final concentration of 1 mM. A supplement of 1 mM $MgCl_2$ was added to the reservoir solutions in each crystal screen in an effort to increase AMP-PNP binding (Table 3.1). Viable crystals grew at 4 °C after 4 days in 150 nL sitting drops mixing 100 nL protein solution with 50 nL of a reservoir solution comprising 20% PEG3350, 0.1 M bis-tris propane pH 6.5, 0.2 M sodium nitrate, 10% ethylene glycol, 1 mM $MgCl_2$ (Fig. 3.10). After 24 days the crystals were cryo-protected and mounted.

Crystal screens were also performed with the longer $CDK12^{715-1052}$/$CycK^{11-267}$ complex, which was largely mono-phosphorylated without CAK treatment. This protein complex was buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP, 10 mM DTT, 5 mM L-arginine, and 5 mM L-glutamate and concentrated to 6 mg/mL. The ATP analogue AMP-PNP was added to a final concentration of 1 mM together with 5 mM $MgCl_2$ (Table 3.1). Crystals appeared after 2 days, grown at 20 °C in 150 nL sitting drops mixing 50 nL protein solution with 100 nL of a reservoir solution comprising 20% PEG3350, 150 mM DL-malic acid (Fig. 3.10). At 16 days the crystal was mounted and cryo-protected with mother liquor supplemented with an additional 15% ethylene glycol, 3 mM AMP-PNP, 5 mM $MgCl_2$ and vitrified in liquid nitrogen.

Table 3.1-Summary of the crystallization trials performed with selected CDK12/CycK complexes.

Proteins	**CDK12$^{715-1038}$/ CycK$^{11-267}$**	**CDK12$^{715-1038}$/ CycK$^{11-267}$**	**CDK12$^{715-1052}$/ CycK$^{11-267}$**
Protein Concentration (mg/mL)	7.2	4.3	6
Phosphorylation Status	+0P	+1P	+1P
Coarse screens	LFS, JCSG, HIN, HCS	LFS, JCSG, HIN, HCS	LFS, JCSG, HIN, HCS
Fine screen	-	LFS-H1, LFS-H5, LFS-F4-G4-H4	HIN-F1, HIN-H10_LFS-E11
Additions	1 mM AMP-PNP to protein sample + 0.1 mM $MgCl_2$ to reservoir	1 mM AMP-PNP to protein sample + 1 mM $MgCl_2$ to reservoir	1 mM AMP-PNP + 5 mM $MgCl_2$ to protein sample
Crystal Plate Temperature (°C)	4 or 20	4 or 20	4 or 20

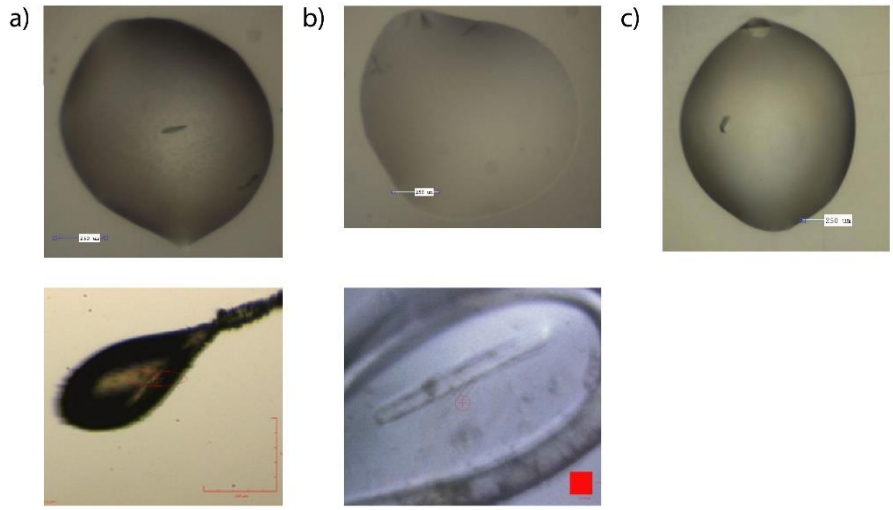


Figure 3.10. Representative crystals obtained for different CDK12/CycK complex preparations.
a) Crystals of the unphosphorylated CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex shown after 20 days in the presence of AMP-PNP and a LFS coarse screen condition comprising 0.20 M potassium thiocyanate, 0.1 M bis-tris propane pH 6.5; 20.0% PEG 3350; 10.0% ethylene glycol, 0.1 mM $MgCl_2$. Before mounting, crystals were cryo-protected with mother liquor supplemented with an additional 15% ethylene glycol and vitrified in liquid nitrogen. **b)** Crystals of mono-phosphorylated CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ shown after 19 days in the presence of AMP-PNP and a LFS coarse screen condition, in a sitting drop of 150 nL (mixing 100 nL protein solution with 50 nL of a reservoir solution) comprising 0.2 M sodium nitrate, 0.1 M bis-tris propane pH 6.5, 20% PEG3350, 10% ethylene glycol, 1 mM $MgCl_2$. Before mounting, crystals were cryo-protected with mother liquor supplemented with an additional 15% ethylene glycol and vitrified in liquid nitrogen. **c)** Crystals of phosphorylated CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ shown after 12 days in the presence of AMP-PNP, 5 mM $MgCl_2$ and a HIN coarse screen condition comprising 0.15 M DL-malic acid, 20% PEG3350, pH 7. Mother liquor was supplemented with 3 mM AMP-PNP, 5 mM $MgCl_2$, and 15% ethylene glycol prior to mounting and cryo-cooling in liquid nitrogen.

3.2.6 Structure determination of the CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex

Diffraction quality crystals in the monoclinic space group $P2_1$ were obtained for the phosphorylated CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complex. Diffraction data were collected at 100 K on Diamond Light Source beamline I24. The resulting structure of the CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ was solved by molecular replacement using Phaser [8] and CDK9 (PDB ID: 4BCG) [9] and CycK (PDB ID: 2I53) [10] as search models and refined at 3.15 Å resolution

(see chapter 2 for full methods). Data collection and refinement statistics are shown in Table 3.2.

The electron density maps for the structure were generally of high quality. There were two CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complexes in the asymmetric unit. The structure confirmed that the activation loop residue Thr893 was the site phosphorylation (Fig. 3.11). The electron density was most complete for the complex comprising chains A and C. CycK (chain A) was modelled between residues Pro22-Gln260 and the bound CDK12 subunit (chain C) between residues Trp719-Pro1031, respectively, with the exception of two regions of poor electron density between CDK12 Gln797-Lys803 and Lys975-Lys976.

For the CDK12/CycK complex comprising chains D and B complex, the CDK12 (chain D) was traced between residues Cys723-Glu1024, with poorer electron density between Asp761-Glu763, Lys796-Gly807, Glu825-Val829, and the activation loop Ser886-Arg890. CycK (chain B) was traced between Pro22-Gln260 with the exception of residues Met230-Tyr231.

3.2.7 Structure determination of the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complex

Diffraction quality crystals in space group $P2_1$ (but with different cell dimensions) were also obtained for the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complex. Diffraction data were collected at 100 K on Diamond Light Source beamline I02. The structure was solved by molecular replacement using my earlier CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ structure as a search model (PDB ID: 4UN0). This structure was refined at 3.15 Å resolution and similarly contained two protein complexes in the asymmetric unit (see Table 3.2 for data collection and refinement statistics for both structures).

In the complex of the longer CDK12 construct, the CDK12 chains were traced between residues Asp718-Gln1050, with the exception of residues Asp798-Ala799 and

Ser889 in chain A and residues Arg1049-Gln1050 in chain C, which were excluded from the model. CycK chains B and D were traced between residues Thr20-Met265 and Pro22-Lys262, respectively.

Table 3.2. Diffraction data collection and refinement statistics

Complex	**CDK12$^{715-1038}$/CycK$^{11-267}$**	**CDK12$^{715-1052}$/CycK$^{11-267}$**
PDB accession code	4UN0 (supersedes 4CJY)	4CXA
Chain IDs	Chains A/B (CDK12/CycK)	Chains C/A (CDK12/CycK)
	Chains C/D (CDK12/CycK)	Chains D/B (CDK12/CycK)
Data Collection		
Beamline	Diamond Light Source, I24	Diamond Light Source, I02
Wavelength (Å)	0.9686	0.97868
Resolution[a] (Å)	57.82-3.15 (3.37-3.15)	30.53-3.15 (3.37-3.15)
Spacegroup	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	49.6, 148.8, 92.2	69.8, 138.8, 71.9
α, β, γ (°)	90.0, 94.2, 90.0	90.0, 105.0, 90.0
No. unique reflections[a]	22,936 (4145)	22,908 (4132)
Completeness[a] (%)	99.4 (99.3)	99.8 (99.9)
I/σ(I)[a]	8.7 (2.0)	6.4 (1.5)
R_{merge}[a]	0.19 (0.87)	0.19 (0.98)
Redundancy[a]	4.3 (4.2)	5.0 (4.9)
Refinement		
Ligands	-	AMP-PNP
No. atoms in refinement	8190	8694
R_{factor} / R_{free} (%)	21.7/27.1	22.4/27.9
Average B factor (Å^2)	55	73
R.m.s. deviations		
Bond lengths (Å)	0.002	0.004
Bond angles (°)	0.58	0.87
Molprobity		
Ramachandran favoured	94%	94%
Ramachandran allowed	5%	5%
Crystallization conditions	20% PEG3350, 10% Ethylene glycol, 0.1M bis-tris propane pH6.5, 0.2M sodium nitrate	0.15M DL-malic acid, 20% PEG3350, pH 7

[a] Values in brackets show the statistics for the highest resolution shells. R.m.s. indicates root-mean-square.

3.2.8 Overview of the CDK12$^{715-1052}$/CycK$^{11-267}$ complex

Overall, the structures of the two CDK12/CycK complexes are similar to the previously determined structures of other CDK complexes, although as predicted they most closely resemble those of the transcriptional CDK class. The structures also confirmed that the single phosphorylation site suggested by mass spectrometry was at the activation loop residue Thr893 where electron density for the phosphate moiety was clearly observable.

However, there are also notable features within the CDK12/CycK complexes that are novel. The most striking example is the C-terminal kinase extension that is revealed in the longer CDK12$^{715-1052}$ structure (Fig 3.11). This feature shows an additional helix outside the canonical kinase fold, the αK helix. Significantly, the CDK12 C-terminal extension makes multiple contacts with both the N- and C-lobes of the kinase domain as well as the bound nucleotide (Fig. 3.11).

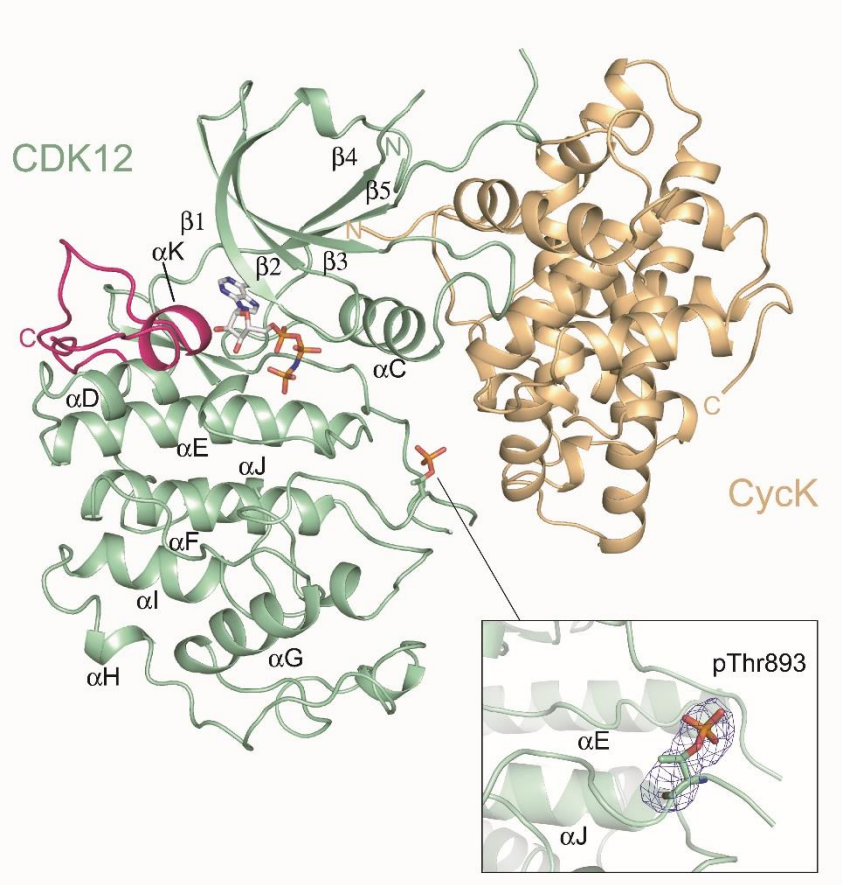


Figure 3.11. Overview of the CDK12$^{715-1052}$/CycK$^{11-267}$ complex structure with AMP-PNP bound. The boxed omit map shows the density for the phosphorylated threonine residue (pThr893) in the activation loop. Omit map contoured at 2.5σ. (PDB: 4CXA, chains A/B)

3.2.9 Re-refinement of the C-terminus of CDK12 in the CDK12$^{715-1038}$/CycK$^{11-267}$ complex

The structure determination of the larger complex comprising CDK12$^{715-1052}$ provided greater confidence in the conformation of the C-terminal kinase extension. Using these extra data, I re-examined my refinement of the CDK12$^{715-1038}$/CycK$^{11-267}$ structure, which had been deposited previously in the PDB as 4CJY. This remodelling allowed CDK12$^{715-1038}$/CycK$^{11-267}$ chain C to be extended from Pro1031 to Trp1036 and CDK12 chain D to be extended from Leu1020 to Glu1024, to align with structure

CDK12$^{715-1052}$/CycK$^{11-267}$, chain A. The newly refined structure was redeposited and assigned the new PDB ID 4UN0, which supersedes 4CJY.

3.2.10 The structures show three distinct protein complexes

The different chains in the CDK12/CycK structures share a common heterodimeric assembly of their folded domains, but exhibit significant differences in their protein C-termini yielding three distinct protein complexes (Fig. 3.12). Most notably, the structural comparisons reveal significant conformational plasticity in the C-terminal kinase extension of CDK12$^{715-1052}$, which packs alongside the ATP pocket in chain A, but at the back of the kinase domain in chain C (PDB ID: 4CXA, Fig. 3.12a). The ligand AMP-PNP, included in all crystallizations, is seen only in chain A where its interactions are stabilized by the presence of the αK helix (Fig. 3.12a). Similarly, there was very poor or no electron density for AMP-PNP in the structures of the shorter CDK12$^{715-1038}$ construct, which highlights the importance of the αK extension for ATP interaction.

Perhaps as a result of these varied features, other subtle differences are also observed across the different CDK12 structures. Overall, the nucleotide-bound chain (4CXA chain A) shows a more closed conformation of the ATP pocket presumably due to the tighter packing of the glycine-rich loop (Fig. 3.12b). Closer packing of the αC helix is also observed in the CDK12$^{715-1052}$ chains resulting in a subtle twist in their N-terminal lobe (Fig. 3.12b) and a small shift in the position of the bound cyclin (Fig. 3.12a). The structure of the CycK subunit appears rigid and is essentially unchanged from the unbound protein [10]. Differences in the CycK structure are observed only for the N- and C-termini as well as a flexible region centred on the H4' helix in the second cyclin box (Fig. 3.12c).

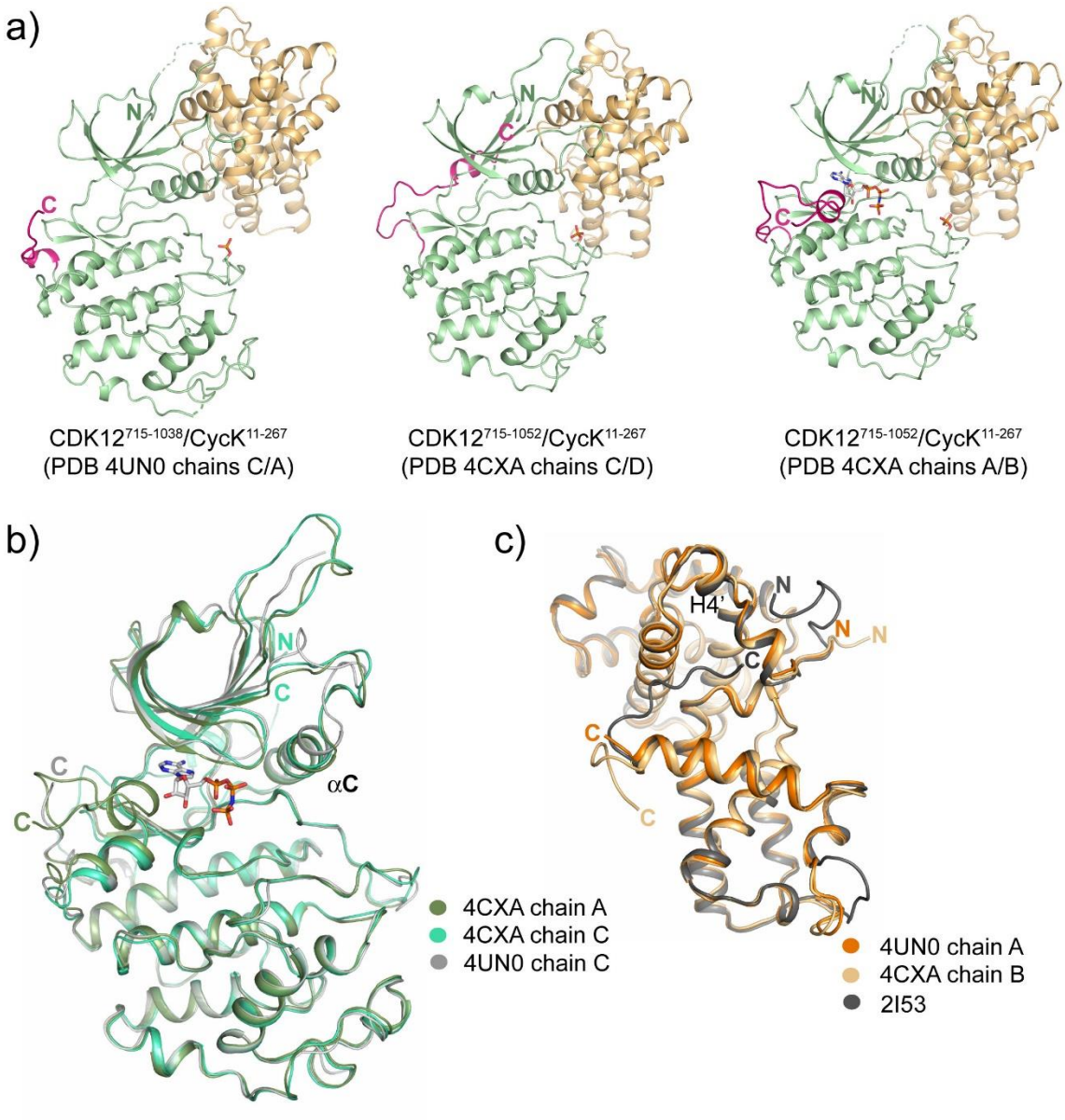


Figure 3.12. Three distinct structural states of the CDK12/CycK complex. a) Ribbon representation overview of the complex structures of CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ and CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$. CDK12 is coloured green, except for the C-terminal kinase extension shown in pink. CycK is coloured light brown. The C-terminal kinase extension, including αK, adopts distinct packing conformations in the different chains in the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ structure. AMP-PNP, included in all crystallisations, is observed only in chain A of this structure where its binding is stabilized by the C-terminal extension. All CDK12 subunits were phosphorylated on Thr893 (shown by sticks). **b)** Superposition of the CDK12 chains from available crystal structures. In addition to changes in the C-terminal region, there are subtle differences in the packing of the glycine-rich loop, the β4-β5 loop and the tilt of the αC helix. **c)** Superposition of the CycK chains from available crystal structures. The apo-structure of CycK (PDB ID: 2I53) shows differences to the complex structures in the region of the H4' helix in the second cyclin box as well as in the N- and C-termini.

3.2.11 Structural features of the active CDK12 kinase domain

The CDK12 chains display an active conformation of the kinase domain fold consistent with their cyclin interaction and their phosphorylation at the activation loop residue Thr893 (Fig. 3.13a). The kinase fold of CDK12 retains the conserved CDK/MAPK insert, which is located between the αG and αH helices (Fig 3.13a). The correct orientation of the αC helix in the AMP-PNP-bound structure is confirmed by the canonical salt bridge formed between the catalytic residues Lys756 (β3) and Glu774 (αC) (Fig. 3.13b). While AMP-PNP was clearly observed, no co-ordinated magnesium ions could be reliably built at this resolution.

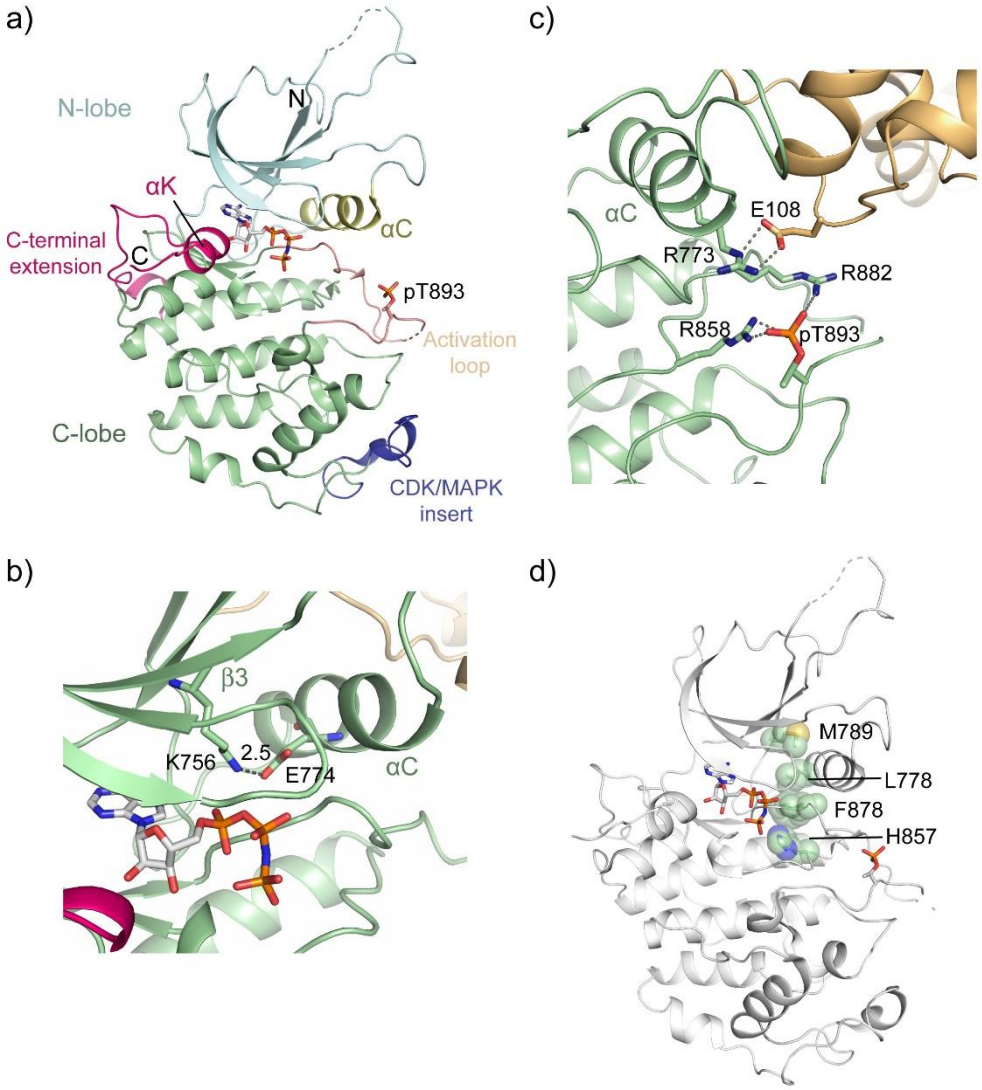


Figure 3.13. Structural features of the active CDK12 subunit a) Ribbon representation highlighting selected structural features of the CDK12$^{715\text{-}1052}$ subunit bound to AMP-PNP (PDB ID 4CXA, chain A). **b)** Salt bridge formed between Lys756 in the β3 strand and Glu774 in the PITAIRE motif of the αC helix. **c)** Phosphorylation at CDK12 Thr893 helps to stabilize the active kinase conformation. Shown are the hydrogen bond interactions of pThr893 with Arg882 (activation loop) and Arg858 (catalytic loop) as well as the nearby interaction of Arg773 (PITAIRE motif, αC helix) with CycK Glu108. d) CDK12 adopts an active conformation as shown by the correct alignment of the hydrophobic spine (indicated residues shown in spacefill representation).

Phosphorylation of Thr893 stabilizes both the activation and catalytic loops through hydrogen bond interactions with Arg882 and Arg858, respectively (Fig. 3.13c). In contrast to CDK2, there are no direct interactions between the phosphate and the cyclin. However, a proximal residue Arg773 from the PITAIRE motif does form electrostatic contacts with Glu108 of CycK (helix H3) (Fig. 3.13c). At the start of the activation segment, the 'DFG'

motif residues Asp877, Phe878 and Gly879, adopt the 'DFG-in' conformation characteristic of active kinases. Arg882 is one of eight deleterious CDK12 mutations identified in ovarian cancer (Fig 3.14) [11]. Its mutation to leucine significantly impairs kinase activity [12] and is predicted to break critical interactions between phospho-Thr893 and the activation loop.

Finally, a hydrophobic regulatory spine, noted as a conserved spatial motif in active kinases, is also established across the N- and C-terminal lobes of CDK12 and comprises Leu778 within the αC helix, Met789 from the β4 strand, His857 from the catalytic loop 'HRD' motif and Phe878 from the activation loop 'DFG' motif (Fig. 3.13d).

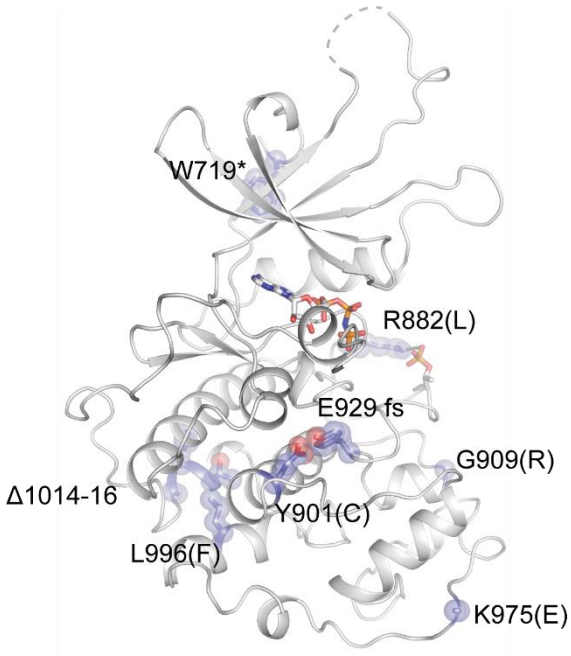


Figure 3.14. Ovarian cancer-associated mutations [11] that impair CDK12 activity are mapped onto the structure. These are predicted to destabilize the protein fold.

3.2.12 Interactions in the CDK12/CycK interface

Overall, the binding of CDK12 and CycK fits the model of the transcriptional CDKs first established by the structure of the CDK9/CycT1 complex [4]. As expected for this transcriptional CDK class, interactions in the protein-protein interface are limited to the kinase N-lobe and the first cyclin box motif, unlike the cell cycle CDK2/ CycA complex in which the cyclin spans both kinase lobes (Fig. 3.15a).

The new CDK12/CycK structures reveal a large β4-β5 loop, which forms a notable insertion in CDK12 and contributes additionally to the overall binding surface area (Fig. 3.15a). The β4-β5 loop sits atop CycK where it inserts CDK12 Phe802 into a nest of hydrophobic residues, including Val142, Val143 and Ile146 from CycK H5. This packing is further stabilized by the intervening CycK residue Arg145 in helix H5, which adopts alternative rotamers in the two subunits in the asymmetric unit to hydrogen bond with CDK12 Thr794 or Gly807, respectively (Fig. 3.15a). Interestingly, Arg145 is conserved only in CycL1, which is another suggested CDK12 interaction partner (Fig. 3.15a). A structure-based sequence alignment also reveals the conservation of Phe802 within CDK13.

Below the β4-β5 loop the CDK12 'PITAIRE' helix (αC) packs between CycK helices H3 and H5 (Fig. 3.15b). Here, the core of the protein-protein interface is hydrophobic, which is characteristic of CDK/cyclin packing with contributions from the kinase N-terminus (Trp719) as well as the αC and surrounding β-sheet (Fig. 3.15b) [13]. At the core of the hydrophobic interface is CycK Phe153, which is shielded from solvent by other aromatic residues. Beyond the hydrophobic core there are a number of electrostatic interactions that make up the periphery of the binding interface, including a salt bridge between CDK12 Arg773 (αC) and CycK Glu108 (H3) (Fig. 3.15c).

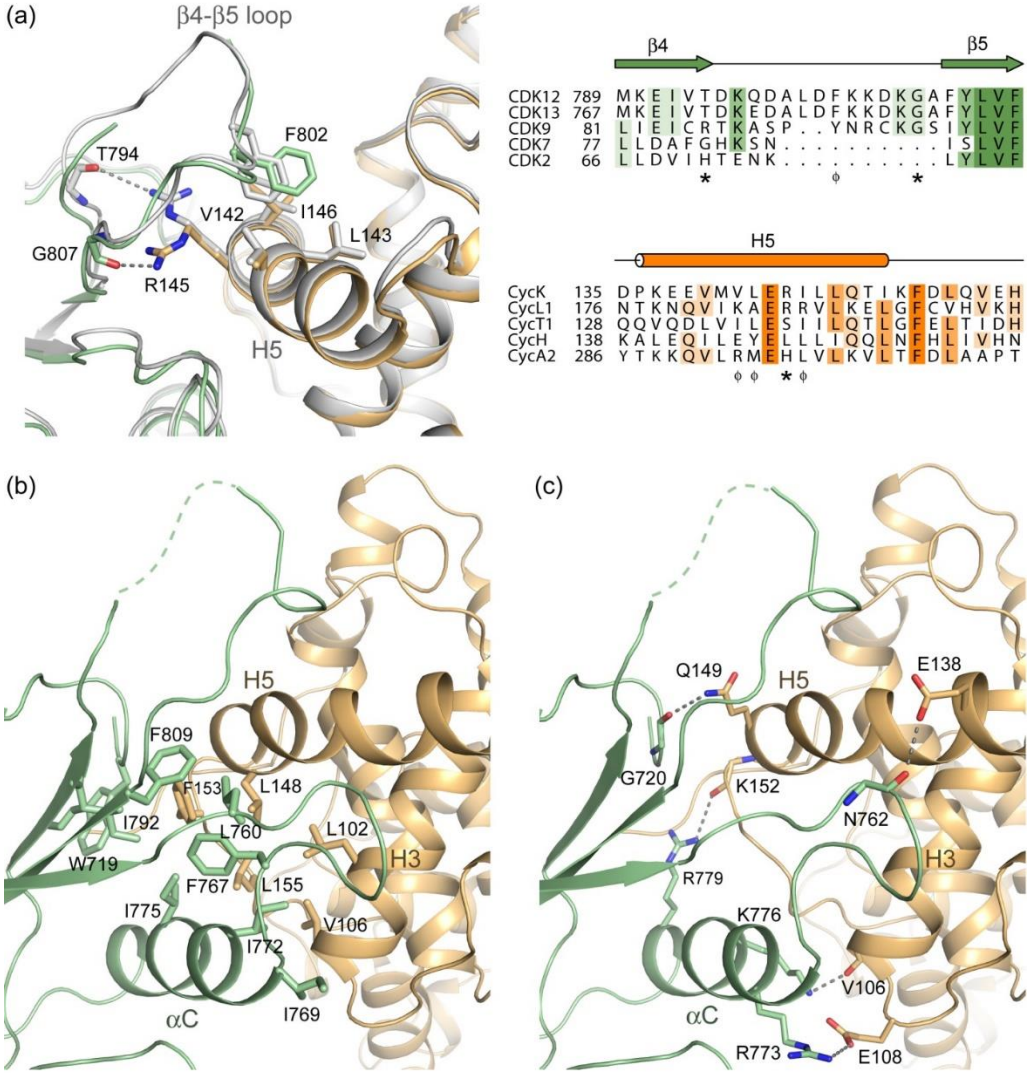


Figure 3.15. Side chain interactions in the CDK12/CycK interface. a) Binding of the CDK12 β4-β5 loop to CycK. Shown is a superposition of the two complexes present in the asymmetric unit of the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ structure (chains A and B are coloured green and light brown, respectively, whereas chains C and D are coloured grey). A structure-based sequence alignment (right panel) reveals an insertion in the β4-β5 loop of CDK12 and CDK13 as well as the sequence divergence across this region and the interacting H5 helix of CycK. The positions of displayed residues participating in hydrophobic (ϕ) and hydrogen bond (*) interactions are marked under the alignment. **b)** Hydrophobic interactions define the core of the protein-protein interface. **c)** Electrostatic interactions cluster outside the core interface.

As observed in the CDK9/CycT1 complex, the CDK12 interface with CycK shows a ~25° rigid body rotation compared to the equivalent CDK2/CycA complex. As a consequence the interface is limited to the kinase N-lobe and is therefore much reduced in size. Indeed, the interface between CDK12 and CycK has a contact area that is only 69% of the size of the of the equivalent CDK2/CycA complex, although this is notably larger than the CDK9/CycT1 interface which is just 59% of the size of CDK2/CycA (Fig. 3.16). The difference between CDK12 and CDK9 in part arises from the large β4-β5 loop insertion in CDK12, which sits above the cyclin as well as the interaction between the PITAIRE residue Arg776 and CycK Glu108. In addition, the CycK N-terminus (Thr20-Pro22) forms some contact with the kinase N-lobe of CDK12. The size of the interface also varies between the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ and CDK12$^{715\text{-}1038}$/CycK$^{11\text{-}267}$ complexes as well as within the different complexes contained within the asymmetric unit of each crystal structure (Fig. 3.16).

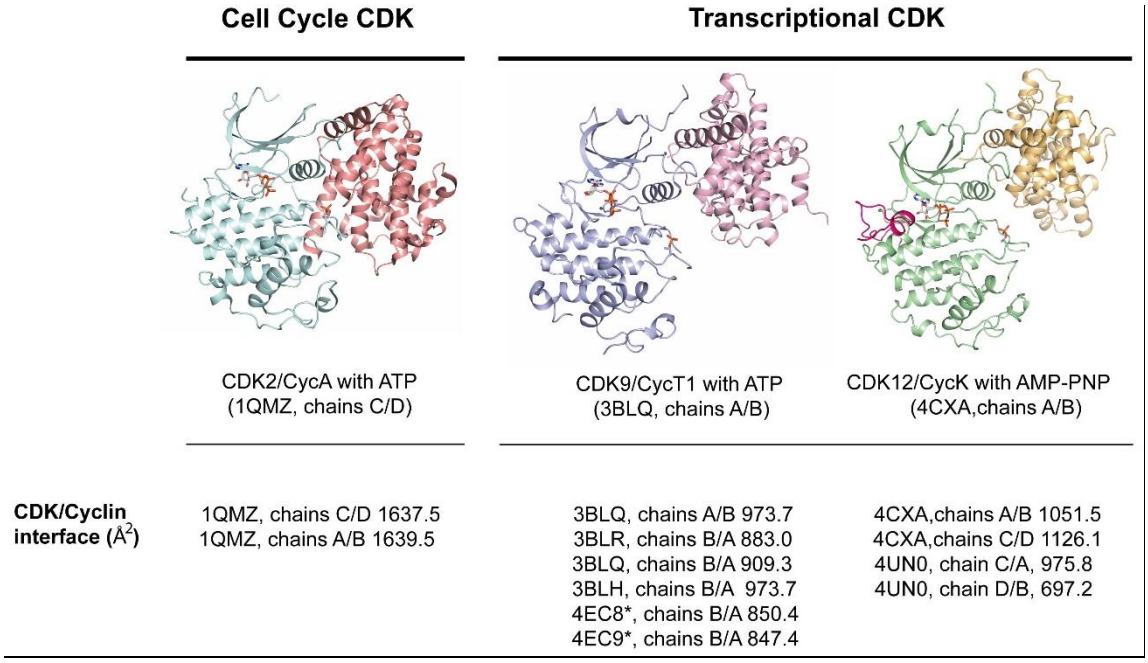


Figure 3.16. Comparison of the binding interfaces in different cell cycle and transcriptional CDK/cyclin complexes. Representative structures are shown together with the sizes of their protein interfaces in determined with the program ePISA [14] (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html). The asterisk denotes complexes containing full-length CDK9.

3.2.13 Conformational plasticity of the C-terminal kinase extension

An unexpected feature of the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complex structure is the conformational plasticity of the C-terminal kinase extension. The conformation of the two CDK12 molecules in the asymmetric unit diverges following Leu1025, which packs against the αE helix at the back of the kinase domain (Fig. 3.17a). In these structures, the C-terminal kinase extension wraps around the front of the ATP pocket where it runs parallel to the β6 strand before turning away in a perpendicular direction (Fig. 3.17a). By contrast, the C-terminus in the second CDK12 complex continues a path across the back of the kinase domain to pack within 4.3 Å of the bound CycK subunit (Fig. 3.17a). Interestingly, the αK helix, encompassing residues 1040-HELWS-1044, is formed in both CDK12 conformations suggesting that this secondary structural element is stably folded (Fig. 3.17b).

The C-terminal kinase extension has been shown to be critical for CDK12 activity [15]. When engaged by the ATP pocket this element acts to enclose the bound AMP-PNP molecule adding significantly to pocket surface area (Fig. 3.17c). As well as contacts by His1040 and Glu1041 to the AMP-PNP moiety, there are stabilizing interactions with the kinase hinge region, including a hydrogen bond between Asp1038 and the hinge residue Tyr815 (Fig. 3.17d). Van der Waals interactions with the N-lobe β1 strand are additionally established by both His1040 and Leu1042. Finally, there are hydrophobic and electrostatic interactions with the C-lobe, including a hydrogen bond between the main chain amide of Glu1041 and side chain of Asp819. As noted earlier, at this resolution no waters or magnesium ions could be modelled. The visible structure in the current work extends to the polybasic tail region comprising 1045-KKRRRQ-1050, which may form a putative interaction site for capture of the CTD substrate.

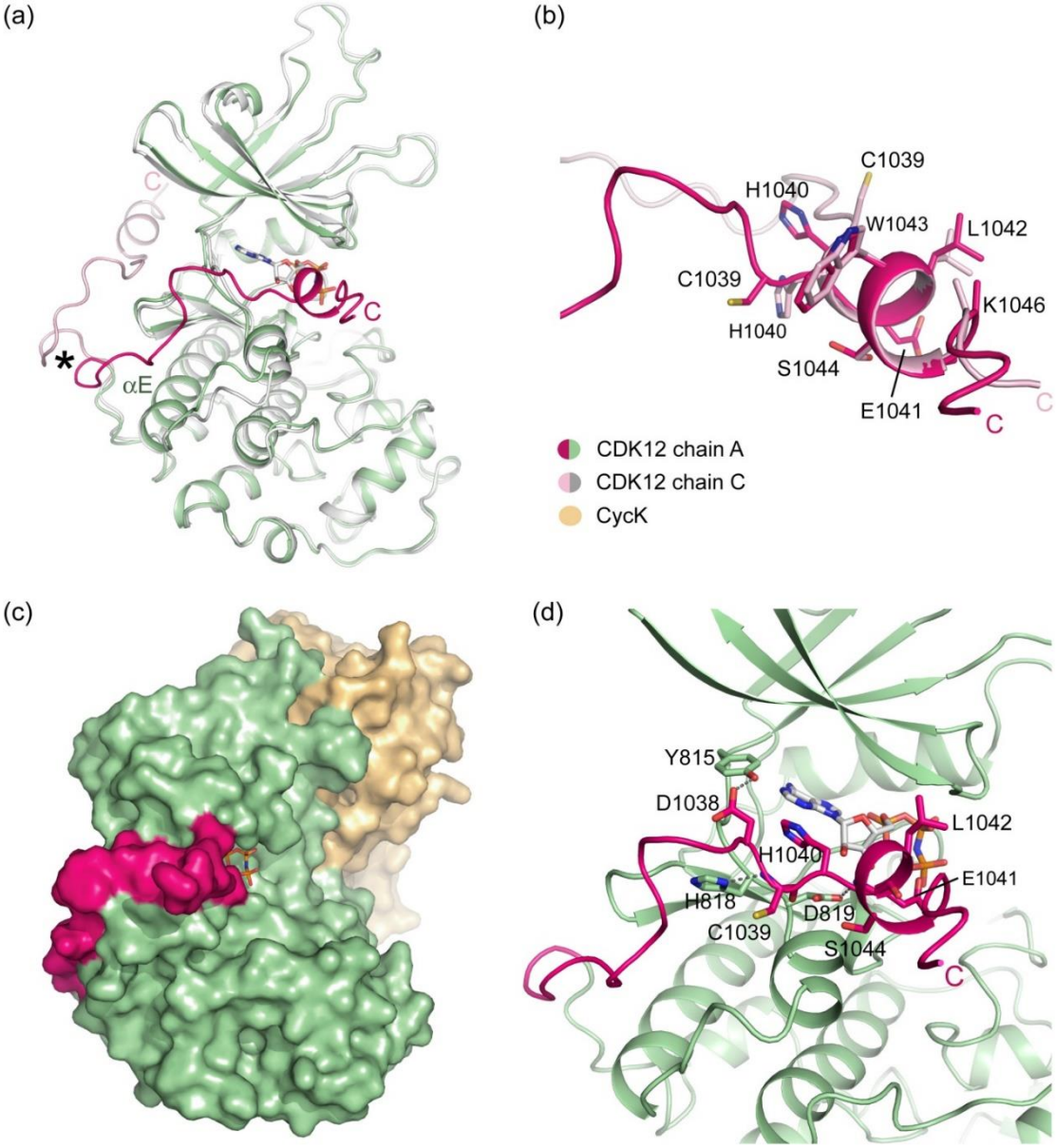


Figure 3.17. Alternative conformations of the CDK12 C-terminal extension. a) Superposition of the two CDK12$^{715\text{-}1052}$ subunits in the asymmetric unit (PDB ID: 4CXA). An asterisk marks the point at which the two chains diverge following Leu1025. In chain A (green and bright pink), the C-terminal extension folds in front of the ATP pocket, whereas in chain C (grey and pale pink) the C-terminus extends across the back of the kinase domain. **b)** Superposition reveals that the αK helix is stably formed in both chains despite their alternative packing arrangements. **c)** Molecular surface representation of chains A and B highlighting the enclosure of the bound AMP-PNP molecule by the C-terminal kinase extension. Coloured as above. d) Selected interactions of the C-terminal kinase extension packing at the front of the ATP pocket (CDK12 chain A).

3.2.14 Full length CDK12/CycK is required for maximal activity

Attempts were made to characterize the activity of the $CDK12^{715-1052}$/$CycK^{11-267}$ complex using the Promega ADP-Glo kinase assay and a range of peptides based on the CTD consensus repeat, $Y_1S_2P_3T_4S_5P_6S_7$. In this assay format, any ADP product resulting from kinase activity is modified to induce a luminescent signal. At this time Bosken *et al.* reported that recombinant CDK12 had a distinct preference for CTD substrates with prior phosphorylation of Ser7 and that it failed to target heptad repeats containing the common Lys7 variant [15]. Therefore, a number of variably phosphorylated peptides were purchased containing two heptad repeats (Fig. 3.18a) to allow for potential priming phosphorylation as well as a peptide containing three repeats which was found to be optimal by Bosken *et al* (Fig. 3.18b).

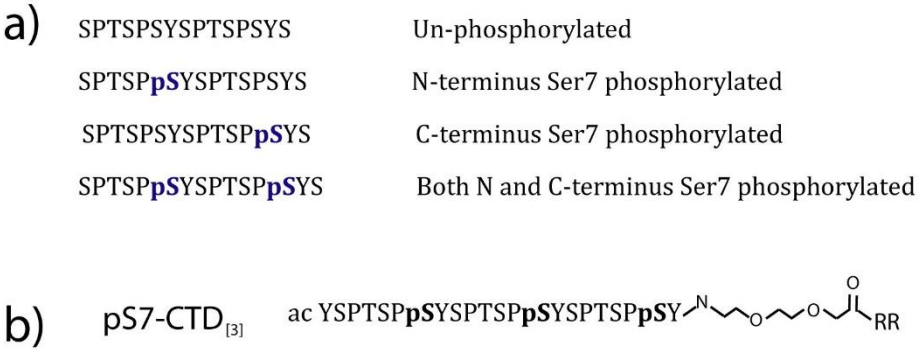


Figure 3.18. Sequences of selected CTD peptides. a) A range of peptides containing two heptad repeats were purchased with variable Ser7 phosphorylation. **b)** Sequence of an optimal CTD peptide determined by Bosken *et al*. that contained three consecutive heptad repeats with Ser7 phosphorylation (pS7-$CTD_{[3]}$) [15].

Preliminary studies using the ADP-Glo kinase assay failed to show any detectable kinase activity against the shorter CTD peptides irrespective of whether the two consensus heptad repeats contained phosphorylated serine 7 (Fig. 3.19a). Similar negative results

were subsequently obtained using the optimised "pS7-$CTD_{[3]}$" peptide, which has three phosphorylated heptad repeats (Fig. 3.19b). Again, no activity was observed above background levels.

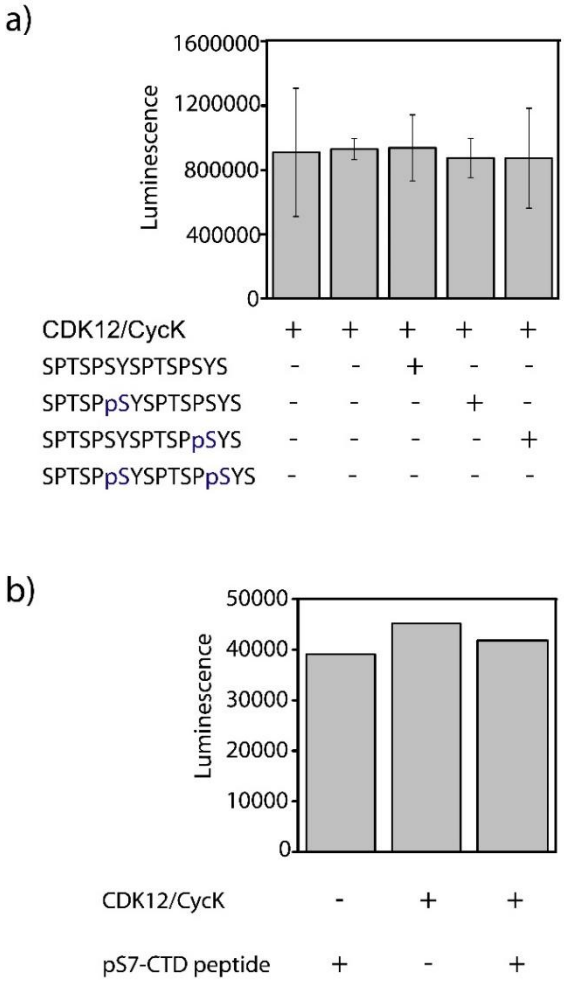


Figure 3.19. Preliminary investigation of $CDK12^{715-1052}$/$CycK^{11-267}$ activity against selected CTD peptides. Kinase activity was measured using the ADP-Glo assay according to the manufacturer's protocols (Promega). In this assay, ADP resulting from kinase activity is modified to induce a luminescent signal. **a)** Trial experiments using a range of CTD peptides containing two heptad repeats with and without potential priming phosphorylation sites. No peptide-directed kinase activity was detected. Error bars show the standard deviation from three replicates. **b)** Trial experiments using the optimised "pS7-$CTD_{[3]}$" peptide again showed no detectable activity.

In light of these negative results, it was decided to ship the recombinant CDK12 complexes to our collaborators, Dr Grace Cheng and Professor Gregg Morin, at the University of British Columbia who had established a radiometric assay for CDK12 using

a GST-CTD fusion as a substrate with 52 native heptad repeats (facilities for work with radioisotopes were not available at the SGC). *In vitro* kinase assays were performed by Dr Grace Cheng to compare the activities of the two CDK12/CycK complexes used in the structural studies against the full-length proteins available in Professor Gregg Morin's laboratory. The activity of the full length CDK12/CycK1 complex was comparable to that of the equivalent CDK9/CycT1 complex (Fig. 3.20). By contrast, the complex comprising CDK12$^{715\text{-}1052}$ exhibited a 10-fold reduction in activity, while the activity of the CDK12$^{715\text{-}1038}$ truncation was severely diminished (Fig. 3.20). These differences suggest that other domains within the full length proteins make important contributions to the substrate interactions. In addition, it suggests that the level of activity of the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complex may have been below the limit of detection in the ADP-Glo kinase assay.

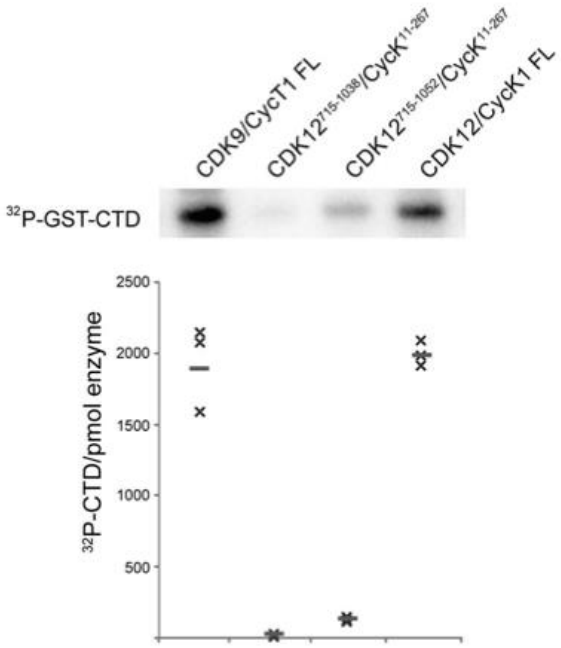


Figure 3.20. CDK12 truncations show diminished activity against the RNA Pol II CTD. Comparison of *in vitro* kinase activity against a GST-CTD substrate by various CDK12 complexes and CDK9/CycT1 (FL denotes full length proteins). A representative autoradiograph detecting ^{32}P-GST-CTD product is shown together with a graphic representation of 3 biological replicates of each phosphorylation reaction. This work was performed by Dr Grace Cheng at the University of British Columbia.

3.3 Discussion

In an effort to further characterize human CDK12/CycK complexes I was able to achieve soluble expression of various CDK12 constructs in *Sf9* insect cells by co-expression with a cognate cyclin partner, CycK. This strategy of co-expression may be useful as a general approach to rescue the soluble expression of other CDKs in future work, such as CDK13. It was also demonstrated that phosphorylation of the CDK12 activation loop may be performed by CAK, either via co-expression or *in vitro* in the presence of ATP and $MgCl_2$. Thus, a reliable expression and purification strategy has been established to obtain various crystallisable CDK12 kinase domain constructs in complex with CycK and with activation loop phosphorylation. Elucidation of the structure will be useful to provide insight into potential druggable sites as discussed in chapter 4.

The structures provide different snapshots of the CDK12/CycK complex in its phosphorylated, active configuration. Overall, they elucidate the layers of control that are necessary to achieve an active complex, including the requirement of a CycK binding partner as well as activation loop phosphorylation by CAK, which stabilises the structural arrangement of the activation loop that is likely to facilitate the engagement of substrate.

The most significant conformational differences were observed in the C-terminal kinase extension, which contains the unusual αK helix. Plasticity in this region is consistent with earlier observations for CDK9 [5]. This CTD-directed kinase also contains a C-terminal tail that folds across the ATP-binding site, although it displays a distinct αK position (Fig. 3.21). Conformational cycling of the CDK9 C-terminal tail is essential for an ordered kinetic pathway during the catalytic cycle [5]. For example, an active CDK9 conformation for ATP-binding is stabilised by the ordered C-terminal tail structure that

partially covers the ATP-binding site similarly to the “closed” conformation observed in the AMP-PNP-bound CDK12 structure. These structures could prevent free exchange of ADP for ATP. In support of these models, the closed conformation of CDK12 was also observed in an independent structure of the CDK12/CycK complex determined in parallel by Bosken *et al.* [15].

A model proposed for the catalytic activity of CDK9/cycT1 suggests ordered recruitment of substrates, with a subsequent ternary complex formation of CDK9-ATP which then binds to the CTD substrate enabling phosphotransfer, followed by CTD substrate release and lastly, ADP release. This ordered pathway has implications for phosphorylation within the CTD multiple consensus sequence repeats [5]. Further experiments will be required to see if CDK12 utilises a similar mechanism.

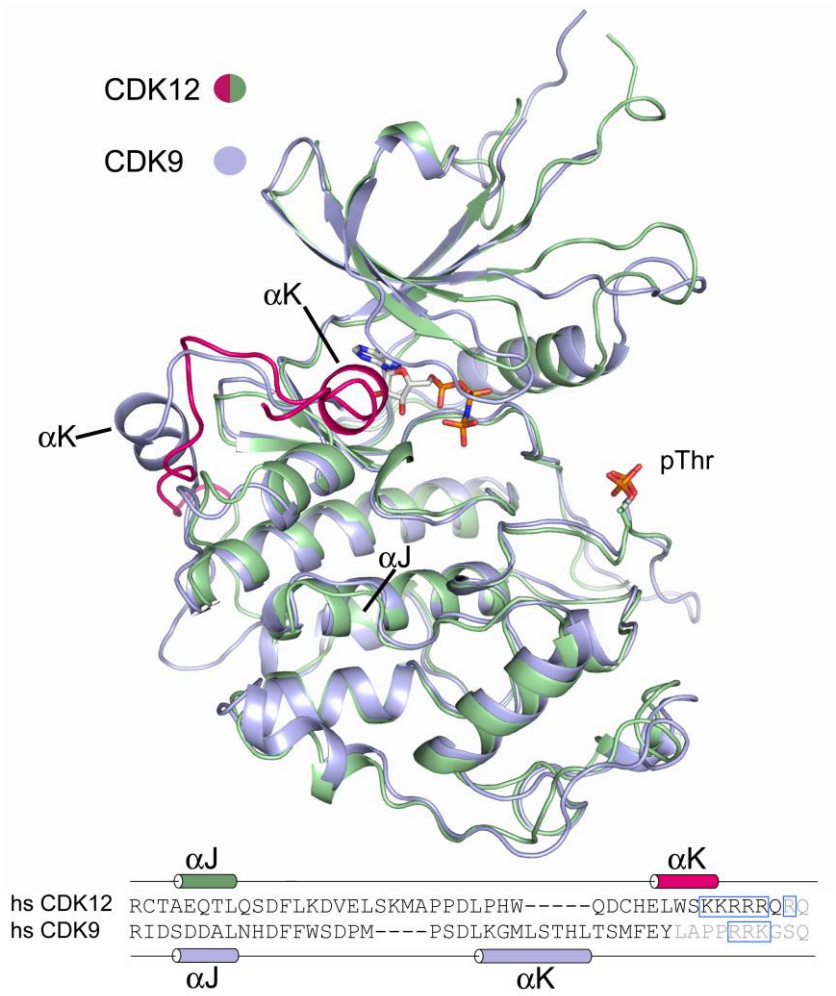


Figure 3.21. Structural comparison of CDK12 and CDK9. The overall kinase fold is similar between the two CDKs. However, the αK helix is located at distinct positions in CDK12 and CDK9 (shown are CDK12 PDB 4CXA and CDK9 PDB 4EC8). A sequence alignment of *Homo sapiens* (hs) CDK12 and CDK9 is also shown covering the C-terminal kinase extensions. Secondary structure elements are highlighted. Boxed in blue is the polybasic region conserved in these CTD-directed kinases.

Consistent with elements of the CDK9 model, AMP-PNP was observed to be stably bound in our CDK12 structures only when trapped in the ATP pocket by the αK helix. Interestingly, we also observed that the αK helix was stably folded when alternatively positioned at the back of the kinase domain. This path for the C-terminus is somewhat reminiscent of the folding in the closely related MAPK family [16]. However, the limited number of specific contacts suggests that this region may be free in solution to explore

multiple conformations. Thus, it remains unproven whether this reflects a true representation of the apo-CDK12 kinase structure or merely one αK position sampled by the flexible C-terminus that was trapped within the crystal lattice. It would be of interest to investigate these motions in future by molecular dynamics simulations as well as by site-directed mutagenesis.

The CDK12/CycK structures also provide new details of this protein-protein interaction. Similarly to other CDK-bound cyclins, the CycK subunit appears relatively rigid, with few differences from the determined apo-structure [10]. The cyclin binding interface of the transcriptional CDKs is notably smaller than those regulating the cell cycle. It was found that this loss is offset slightly in CDK12 by an unusual β4-β5 loop insertion that packs atop the H5 helix of the first cyclin box motif. Moreover, this loop is conserved in the CDK13 kinase domain, which also binds to the CycK subunit. However, as with other transcriptional CDKs, the cyclin interaction of CDK12 is restricted to the kinase N-lobe. Structural studies of the CDK9/CycT1 complex have revealed that C-lobe interactions may be mediated instead by other accessory proteins that direct kinase activity, such as HIV-1 Tat [17]. In this quaternary complex, the Tat protein makes extensive contacts with the CycT1 subunit as well as the CDK9 activation loop to modulate their protein-protein interaction. The host cellular binding partners that Tat may mimic remain to be elucidated. The kinase activation segment in CDK12 is similarly exposed and available to bind other partners to fine-tune the function and activity. It will be interesting to discover in future if any such factors can be identified that modulate CDK12 similarly to the positive and negative regulators of P-TEFb, including Brd4, HEXIM1 and the noncoding 7SK small nuclear RNA, respectively [18]. Interestingly, CycK has also been found to interact with the HIV multifunctional accessory protein Nef and to inhibit *HIV*-1 gene expression and replication [19].

By comparison other CDKs, the CDK12 and CDK13 kinases are significantly extended with large regions N- and C-terminal to the kinase domain. In addition, human cyclin K1 has a large C-terminal extension, which includes proline-rich motifs (PRM) and multiple arginine-serine-rich (RS) motifs (Fig. 3.1). It is conceivable these extensions could also mimic accessory proteins to modulate CDK12 activity. Indeed, my collaborator Dr Grace Cheng found that the full length CDK12/CycK1 complex was significantly more active than the truncated complexes used in this study.

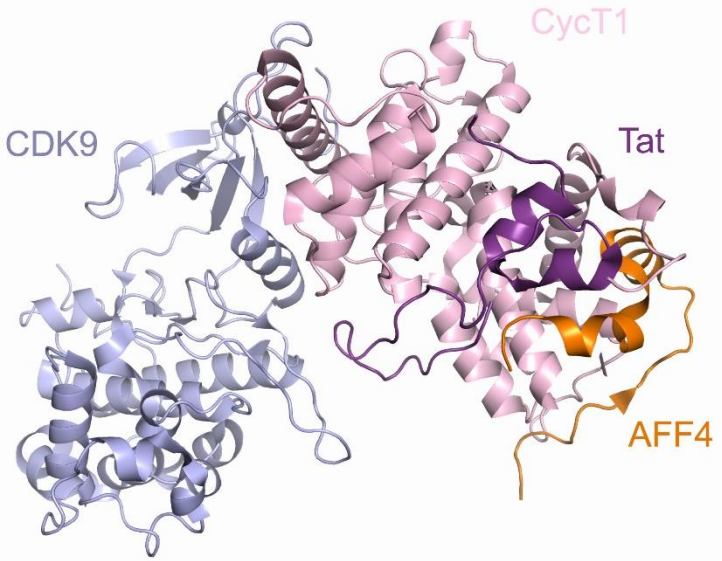


Figure 3.22. Crystal structure of HIV-1 Tat complexed with human P-TEFb and AFF4. (PDB ID 4OR5).

The CDK12 kinase acts late in the transcriptional cycle and is expected to engage a negatively charged CTD substrate [15]. The final C-terminal positions in the CDK12$^{715\text{-}1052}$ structure form a polybasic cluster that is also loosely conserved in CDK9 (Fig. 3.21). The folding of the C-terminal extension across the ATP pocket is therefore also hypothesized as an adaptation to support the recruitment of the CTD substrate by provision of a complementary charged surface [5, 15]. Dynamic movement of the C-terminus may also

promote recruitment by increasing the potential search space for protein-protein interactions. Further biochemical and structural analyses are required to validate these hypotheses.

Overall, this work extends our understanding of the structural mechanisms that determine the activity of the CDK12/CycK complex. Further, it emphasizes the importance of protein flexibility as well as the contribution of regulatory elements outside the core catalytic domains. Finally, the packing of the αK helix offers a novel ATP pocket environment for the design of specific inhibitors that will be explored in chapter 4.

3.4 References

1. Bartkowiak, B., et al., *CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1.* Genes Dev, 2010. 24(20): p. 2303-16.
2. Bowman, E.A. and W.G. Kelly, *RNA polymerase II transcription elongation and Pol II CTD Ser2 phosphorylation: A tail of two kinases.* Nucleus, 2014. 5(3): p. 224-36.
3. Kohoutek, J. and D. Blazek, *Cyclin K goes with Cdk12 and Cdk13.* Cell Div, 2012. 7: p. 12.
4. Baumli, S., et al., *The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation.* EMBO J, 2008. 27(13): p. 1907-18.
5. Baumli, S., et al., *The CDK9 tail determines the reaction pathway of positive transcription elongation factor b.* Structure, 2012. 20(10): p. 1788-95.
6. Golovanov, A.P., et al., *A simple method for improving protein solubility and long-term stability.* J Am Chem Soc, 2004. 126(29): p. 8933-9.

7. Liang, K., et al., *Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing.* Mol Cell Biol, 2015. 35(6): p. 928-38.

8. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. 40(Pt 4): p. 658-674.

9. Shao, H., et al., *Substituted 4-(thiazol-5-yl)-2-(phenylamino)pyrimidines are highly active CDK9 inhibitors: synthesis, X-ray crystal structures, structure-activity relationship, and anticancer activities.* J Med Chem, 2013. 56(3): p. 640-59.

10. Baek, K., et al., *Crystal structure of human cyclin K, a positive regulator of cyclin-dependent kinase 9.* J Mol Biol, 2007. 366(2): p. 563-73.

11. Cancer Genome Atlas Research, N., *Integrated genomic analyses of ovarian carcinoma.* Nature, 2011. 474(7353): p. 609-15.

12. Joshi, P.M., et al., *Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors.* J Biol Chem, 2014. 289(13): p. 9247-53.

13. Echalier, A., J.A. Endicott, and M.E. Noble, *Recent developments in cyclin-dependent kinase biochemical and structural studies.* Biochim Biophys Acta, 2010. 1804(3): p. 511-9.

14. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. 372(3): p. 774-97.

15. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K.* Nat Commun, 2014. 5: p. 3505.

16. Zhang, F., et al., *Atomic structure of the MAP kinase ERK2 at 2.3 A resolution.* Nature, 1994. 367(6465): p. 704-11.

17. Tahirov, T.H., et al., *Crystal structure of HIV-1 Tat complexed with human P-TEFb.* Nature, 2010. 465(7299): p. 747-51.

18. Zhou, Q. and J.H. Yik, *The Yin and Yang of P-TEFb regulation: implications for human immunodeficiency virus gene expression and global control of cell growth and differentiation.* Microbiol Mol Biol Rev, 2006. 70(3): p. 646-59.

19. Khan, S.Z. and D. Mitra, *Cyclin K inhibits HIV-1 gene expression and replication by interfering with cyclin-dependent kinase 9 (CDK9)-cyclin T1 interaction in Nef-dependent manner.* J Biol Chem, 2011. 286(26): p. 22943-54.

20. Network, T.C.G.A.R., *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. 513(7517): p. 202-9.

Chapter 4 A covalent inhibitor of CDK12

4.1 Introduction

Specific, potent inhibitors of CDK12 would be highly valuable reagents to further investigate the regulation and functions of CDK12. However, only a few CDK12 inhibitors have been reported and these lack selectivity. Bosken *et al*, recently established a radiometric assay to test a series of small molecule protein kinase inhibitors for their ability to inhibit CDK12/CycK as well as P-TEFb [1]. Flavopiridol, a pan-CDK ATP-competitive inhibitor currently in clinical trials was the most potent against CDK12 with an IC_{50} value of 1.4 μM [1]. This was one-order of magnitude weaker than against CDK9 and in concurrence, Liang *et al*. subsequently showed flavopiridol gave only modest decreases in both CDK12 and CDK13 activity in comparison to CDK9 [1, 2]. Of the other tested inhibitors, only Purvalanol A and B, and CR8 reduced CDK12 activity nearly 10-fold but at 100-fold excess concentration. Thus, the absence of selective CDK12/13 small molecule inhibitors as tool compounds impedes the possibility to investigate the consequences of acute and prolonged inhibition of CDK12/13 under normal and pathological conditions.

During screening for new inhibitors of kinases involved in gene transcription, Kwiatkowski *et al.* identified and characterised a covalent CDK7 inhibitor, THZ1 [3]. THZ1 is a phenylaminopyrimidine bearing a potentially cysteine-reactive acrylamide moiety. THZ1 binds the ATP-pocket and has the ability to covalently modify CDK7 Cys312 located on a C-terminal extension of the kinase domain, which traverses the C-lobe close to the ATP-pocket (Fig. 4.2). Sequence alignment within the CDK family showed that a cysteine at this position was unique to CDK7, although CDK12 and CDK13 had a cysteine within four amino acids of Cys312 (Fig. 4.1a). At higher concentrations, THZ1 could also inhibit CDK12 kinase activity [3]. Comparison of the CDK7 model and CDK12 crystal structures, confirmed that both kinases have similarly positioned, yet spatially

distinct, cysteines in their C-terminal domains (Figure 4.2b) [3, 4]. Importantly, CDK9 lacks a C-terminal cysteine. THZ1 thus provided a starting point for the rational design of a more specific CDK12/13 inhibitor.

Having developed protocols to purify and crystallise CDK12/CycK complexes (Chapter 3), I established a collaboration with Professor Nathanael Gray (Dana Farber Cancer Institute, Boston) to apply these protocols to help the development of THZ1 analogues with selectivity for CDK12. In particular, the key aim was to use X-ray crystallography to determine the binding mode of these inhibitors. My earlier CDK12/CycK crystal structure with AMP-PNP showed that Cys1039 sits close to the ATP pocket but is directed into solvent (Fig. 4.1b). A co-structure would therefore be important to understand how to target this cysteine more selectively.

In this chapter, I test a number of THZ1 analogues for their covalent interaction with CDK12. I then establish protocols to obtain a co-crystal structure of the first selective, covalent CDK12 inhibitor "THZ531". This compound will provide a useful probe for the elucidation of the role of CDK12 in both normal and pathobiology.

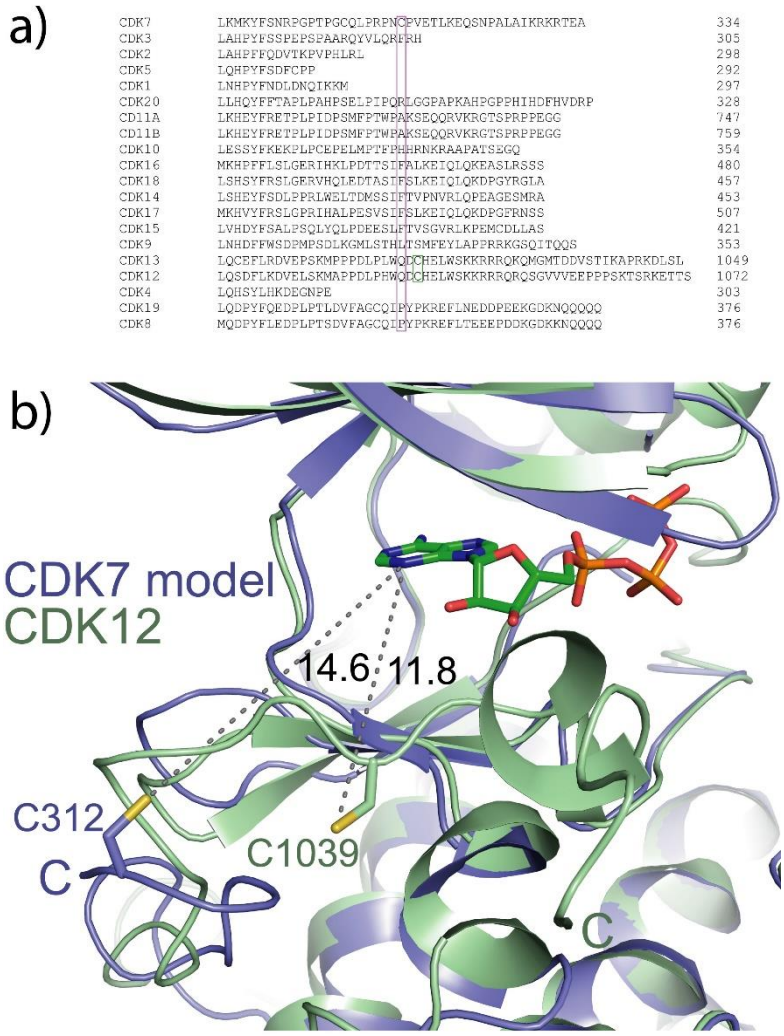


Figure 4.1 Cysteine residues present in CDK7 and CDK12 kinase domains. a) Sequence alignment of CDKs. Sequences of human CDKs were taken from Uniprot and aligned using Blast [5] against CDK7. The position of CDK7 Cys312, at which THZ1 covalently binds, is highlighted across the CDK structures in a purple box. CDK12 and CDK13 kinase domain C-terminal cysteine residues are highlighted in a green box. **b)** Superposition of a published CDK7 model [3] and my structure of CDK12$^{715\text{-}1052}$ (PDB: 4CXA).

4.2 Results

4.2.1 DSF screening for inhibitors of the CDK12/CycK complex

Prior to work with the THZ1 analogues, a differential scanning fluorimetry (DSF) assay was tested as a screen to identify inhibitors of CDK12/CycK from the library of

kinase inhibitors held at the Structural Genomics Consortium. For these experiments I used the $CDK12^{715-1052}$/$CycK^{11-267}$ protein complex as CDK12 could only be purified by co-expression with CycK (I do not yet have a method to purify soluble CDK12 alone). The DSF assay monitors the thermal unfolding of a protein in the presence of a fluorescent dye (sypro orange), which binds to exposed hydrophobic regions of the protein [6]. The protein is heated in solution and upon unfolding, more hydrophobic regions are exposed and this can be measured as a change in fluorescence. The midpoint of the unfolding defines the apparent the melting temperature (Tm). Ligand binding to a native protein causes an increase in the protein's melting temperature that is proportional to the ligand's binding affinity and concentration [6]. Thus, thermal shift screening experiments are performed to determine the ΔTm calculated between the Tm of the protein alone and the Tm of the protein with ligand.

The results suggested that the thermal shift assay is unsuitable for CDK12/CycK complexes. In DSF, the fluorescence intensity is plotted as a function of temperature to generate a sigmoidal unfolding curve that can be described by a two-state transition [6]. However, for the CDK12/CycK complex two apparent unfolding transitions were observed both in controls containing only the protein-dye solution and in samples containing an additional 2% DMSO as a matched control for wells containing inhibitor compounds (which were diluted from stocks in 100% DMSO) (Fig. 4.2a).

Further DSF experiments showed that the CycK monomer had an apparent Tm value of 58.5 °C (Fig. 4.2b), which correlated well with the second unfolding transition observed for the CDK12/CycK complex (Fig. 4.2a). It is therefore possible that the first transition represented the dissociation and unfolding of CDK12 (apparent Tm of approximately 50°C, while the subsequent transition was CycK unfolding (Fig. 4.2a). Due to this unfortunate behaviour it was difficult to reliably use the DSF assay for inhibitor

screening against the CDK12/CycK complex. However, there was evidence for the binding of flavopiridol. In the presence of this compound the complex showed a single unfolding transition with an apparent Tm of 58°C (Fig. 4.2a). While difficult to interpret, this may indicate a thermal shift (ΔTm) against the complex of 8°C.

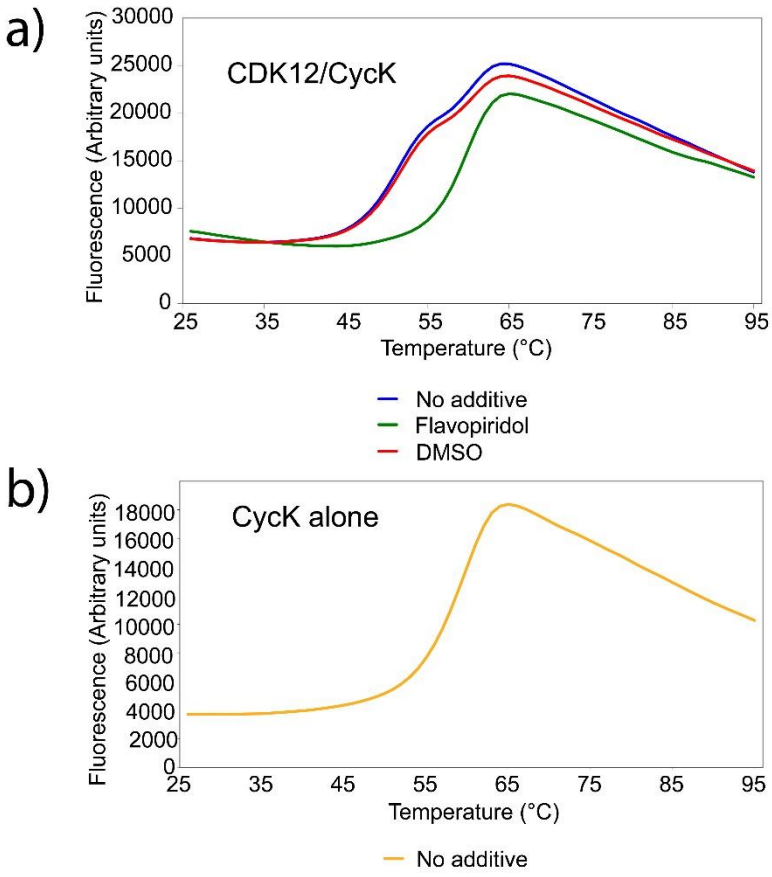
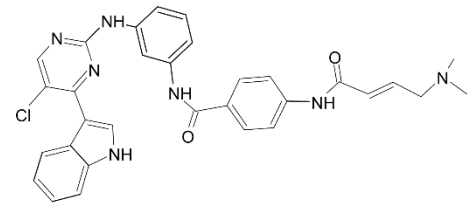
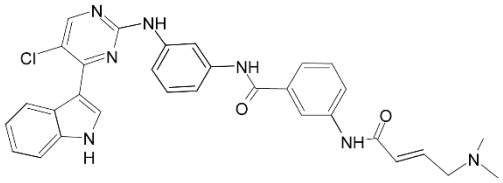


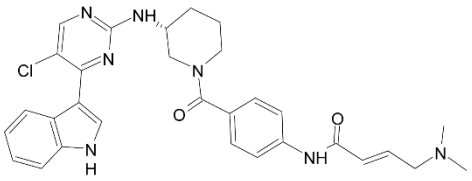
Figure 4.2. DSF screening for CDK12 inhibitors. a) Thermal unfolding curves for the $CDK12^{715-1052}$/$CycK^{11-267}$ complex mixed with either 10 μM flavopiridol (green curve), 2% DMSO (red curve) or no additive (blue curve). **b)** Thermal unfolding curve for the CycK alone with no additive (orange curve).

4.2.2 THZ1 and THZ-3-95-1 covalently bind to CDK12

The first compounds received from Professor Nathanael Gray were THZ1 and an analogue THZ-3-95-1 (all inhibitors used in this study are shown in Fig. 4.3 and described in Table 4.1). It was first necessary to test their ability to covalently bind to the purified recombinant human CDK12/CycK complexes and to determine the regions of CDK12 required for binding.







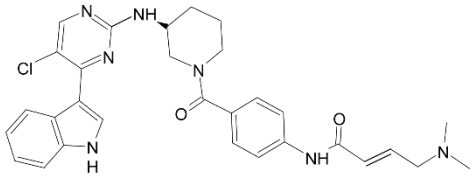


Figure 4.3-Structure of THZ compounds

Table 4.1-THZ compounds.

Compound	Molecular weight (Da)	Modification of THZ1
THZ1	566	-
THZ-3-95-1	566	The di-amide phenyl ring is para-substituted in THZ1 and meta-substituted in THZ-3-95-1
THZ-4-80-1	559	Di-amide phenyl ring is still para-substituted, however hydrogenation of the meta-substituted aniline ring in THZ1 generates a piperidine ring, which alters the orientation of the fixed acrylamide motif to selectively target CDK12 C1039
THZ531	559	R-enantiomer of THZ-4-80-1
THZ532	559	S-enantiomer of THZ-4-80-1

By comparison with the CDK7-THZ1 complex, the predicted binding site was Cys1039 within the CDK12 C-terminal kinase extension. This region has been shown to be relatively flexible and can enclose bound nucleotide in the ATP-binding pocket [1, 7]. However, my structures reveal another cysteine, Cys862, which is also in close proximity to the ATP-binding pocket (Fig. 4.4). Cys862 is located in the loop following the β7 sheet, and sits somewhat buried in the C-lobe some 14 Å from the adenine of AMP-PNP, as seen in the crystal structure (Fig. 4.4).

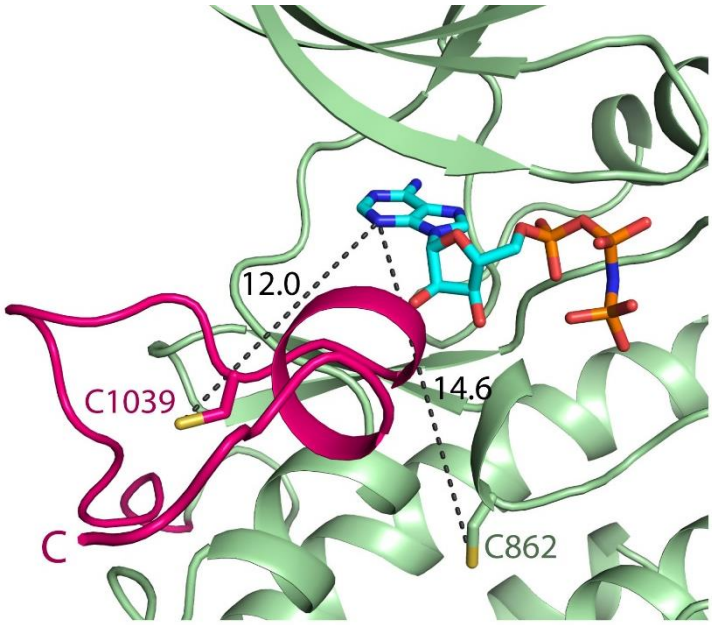


Figure 4.4. The position of Cys862 and Cys1039 in CDK12. AMP-PNP in the crystal structure is used as a reference point to measure the distance of the two potential cysteine residues in CDK12 which THZ compounds may bind to (PDB: 4CXA).

Using a range of CDK12 constructs spanning between residues 661-1068, a small-scale assay was performed with mass spectrometry to determine if Cys1039 was the binding site of the THZ series compounds. Most valuable for these experiments was the shorter CDK12 construct ending at Asp1038. Various CDK12/CycK complexes were incubated with either THZ1 or THZ-3-95-1 on ice for 40 minutes before any covalent interaction was assessed by mass spectrometry (Table 4.2, Fig. 4.5). All CDK12 constructs

with C-terminal residues up to or beyond Gln1052 were able to bind to both THZ1 and THZ-3-95-1, whereas CDK12$^{702\text{-}1038}$ and CDK12$^{715\text{-}1038}$ did not show a shift in mass representative of the covalent inhibitor binding (Table 4.2, Fig.4.5). CycK showed no change in mass, confirming the covalent inhibitor did not bind any of the seven cysteine residues present in the CycK construct (Fig. 4.6). From this it was possible to deduce that THZ1 and THZ-3-95-1 bind CDK12 at residue Cys1039 as this is the only cysteine between Asp1038 and Gln1052.

Table 4.2 Binding of covalent inhibitors to various CDK12 constructs.

N-Terminal Residue	C-Terminal Residue	THZ1 Binding[1]	THZ-3-95-1 Binding[1]
Arg661	Lys1068	Yes	Yes
Pro702	Lys1068	Yes	Yes
Pro702	Gln1052	Yes	Yes
Pro702	Asp1038	No	No
Pro702	Asp1038	No	No
Thr715	Lys1068	Yes	Yes
Thr715	Gln1052	Yes	Yes
Val724	Gln1052	Yes	Yes

[1] Detected by a corresponding mass shift in the intact mass spectrum.

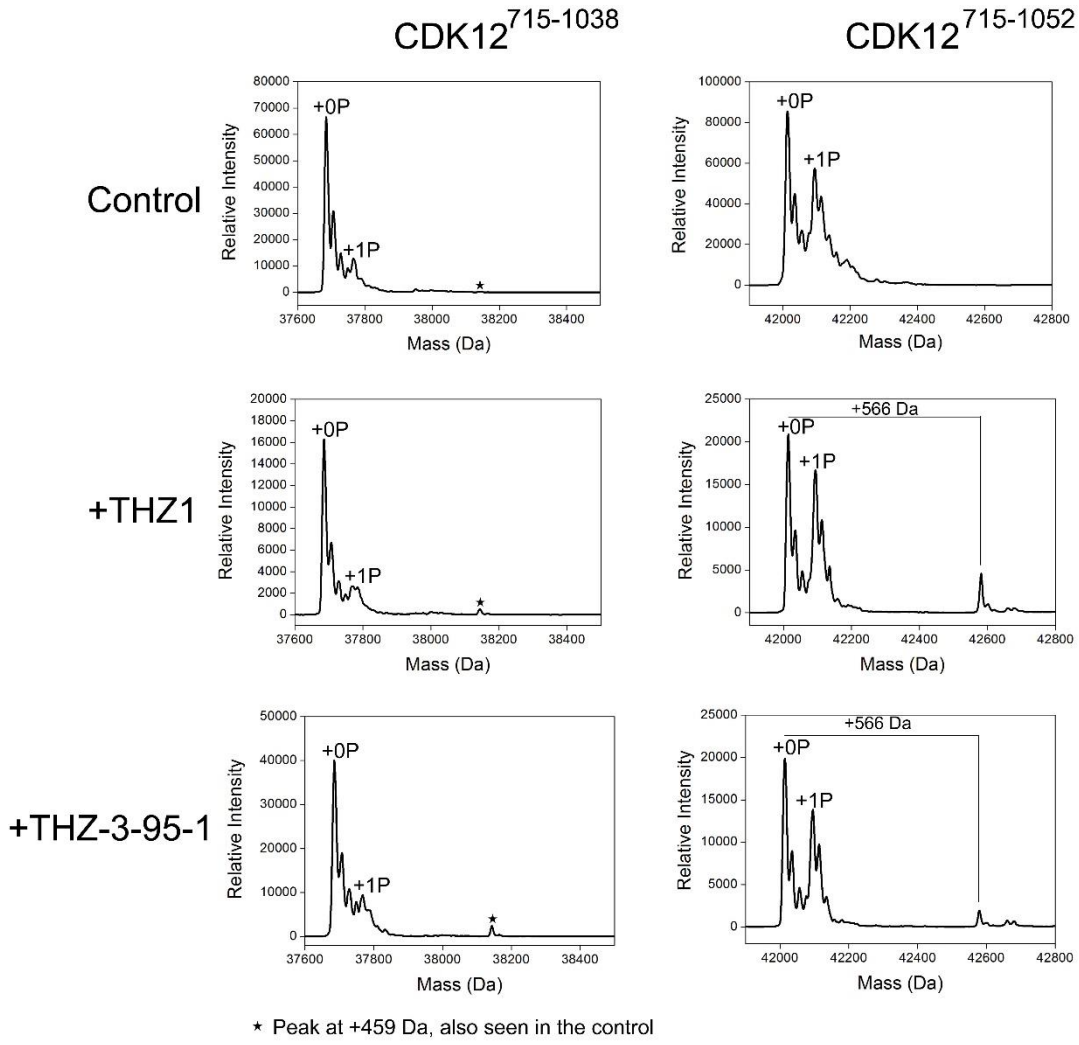


Figure 4.5. Mass spectrometry to assess covalent inhibitor binding to CDK12. Indicated CDK12 constructs were co-expressed with CycK$^{11-267}$ in 50 mL cultures of Sf9 cells and purified by IMAC. The hexahistidine tagged CDK12/CycK complexes were incubated with indicated inhibitors for 40 minutes and then analysed by mass spectrometry. The mass spectra show THZ1 and THZ-3-95-1 binding to CDK12$^{715-1052}$ (Mass of 42,102 Da), but not to CDK12$^{716-1038}$ (Mass of 40,208.5 Da), by a mass shift of 566 Da. The control corresponds to CDK12/CycK complexes incubated in the absence of inhibitors. Note, a small peak was also seen at +459 Da in the CDK12$^{716-1052}$ sample, but this was also present in the control mass spectra.

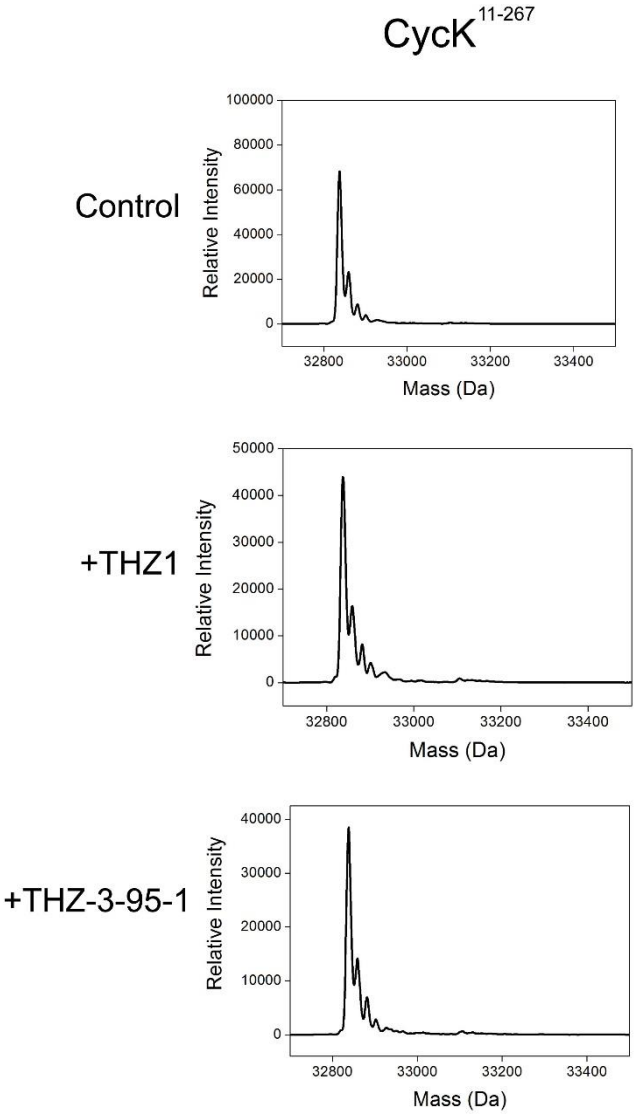


Figure 4.6. Mass spectra confirming the lack of covalent inhibitor binding to CycK. Hexahistidine tagged CDK12/CycK complexes were incubated with indicated inhibitors for 40 minutes and then analysed by mass spectrometry. Representative mass spectra are shown for the $CycK^{11-267}$ subunit in the absence (control) or presence of indicated inhibitors. Neither THZ1 nor THZ-3-95-1 bound to CycK (Mass of 32926 Da) as indicated by the lack of a mass shift of 566 Da.

4.2.3 Optimisation of a purification strategy for CDK12/CycK/inhibitor complexes

Following the initial small-scale trials, which established CDK12 residue Cys1039 as the likely site at which the covalent inhibitor binds, the next aim was to crystallise a CDK12/CycK complex with a THZ series inhibitor bound. To do so, I needed to ensure

that the THZ-bound CDK12 was the dominant species, which was not achieved in small-scale trials. The CDK12 construct spanning residues Thr715-Glu1052 was selected for crystallisation trials predominantly, as protocols had already been established to obtain a good yield of the $CDK12^{715-1052}$/$CycK^{11-267}$ complex. Previous crystal structures obtained with AMP-PNP bound (Chapter 3), had shown that the length of this C-terminal extension was sufficient to wrap around the front of the ATP pocket to form a number of stabling interactions with the bound nucleotide. Thus from hereon, CDK12/CycK refers to the $CDK12^{715-1052}$/$CycK^{11-267}$ complex.

I established in chapter 3 that the CDK12 activation loop can be phosphorylated *in vitro* at Thr819 by a CDK-activating kinase (CAK) from *Candida albicans*. To more efficiently and reliably obtain homogeneously phosphorylated CDK12, it was decided to try co-expression of a CAK. A pFastBac plasmid of CAK1 from *Saccharomyces cerevisiae* (referred to as CAK from here on) was purchased from MRC-PPU (Medical Research Council Protein Phosphorylation Unit), University of Dundee. Baculovirus for recombinant human CDK12, human cyclin K and the yeast CAK were then co-expressed in Sf9 insect cells. The level of each P2 virus was adjusted to optimise both the level of CDK12 phosphorylation and the desired 1:1 ratio of the CDK12 and CycK subunits.

The expressed protein complex was purified using nickel-sepharose resin. However, CAK, CDK12 and CycK each had a histidine tag, so further purification steps were necessary to achieve a homogeneous 1:1 CDK12/CycK complex and to remove the co-purified CAK. A cation-exchange chromatography step performed on a 5 mL HiTrap SP HP sepharose high-performance column (GE Healthcare), was introduced to remove CAK by taking advantage of the difference in calculated pI values of 2.4 between the CDK12/CycK complex and CAK (pI values for CDK12/CycK complex=8.0; CDK12=8.3;CycK=7.1;CAK=5.6, as determined by ProtParam [8]) (Fig. 4.7b). A high

salt buffer of 50 mM MES, pH 6.5, 0.5 mM TCEP, 1 M NaCl was selected in partner with a no salt buffer for the ion-exchange chromatography. SDS PAGE and mass spectrometric analysis confirmed that CAK was no longer present and that excess CycK had also been removed at this stage. The CDK12/CycK complex was then buffer exchanged into a new buffer in a 10 kDa MWCO Amicon Ultra spin concentrator, into a 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP buffer, to raise the pH of the buffer for the subsequent step and to remove any trace of the DTT added before ion-exchange chromatography, as this sulfhydryl-containing reducing agent may interfere with the covalent inhibitor binding to the CDK12 cysteine residue.

The CDK12/CycK complex was concentrated to at least 0.8 mg/mL and the compound was added at a molar ratio that was 1.5 times in excess of that of the protein complex. The mixture was incubated with a selected THZ series compound overnight at 4°C until near complete binding of CDK12 was achieved. This was determined by mass spectrometry using an Agilent LC/MSD TOF system. As before, binding of the compound was monitored by changes to the intact mass of phosphorylated CDK12 from 39,659 Da to a value corresponding to addition of the mass of the THZ series compound (Table 4.1). An example mass spectrum is shown in Fig. 4.8. For final clean-up, proteins underwent further purification by size-exclusion chromatography on a HiLoad S75 Superdex 16/60 column (GE Healthcare), to remove any excess compound or CycK (Fig.4.7). The protein-containing elutions were pooled and concentrated for subsequent crystal trials.

Crystals were cryo-protected with mother liquor supplemented with an additional 15 or 20 % ethylene glycol and vitrified in liquid nitrogen after mounting. Crystals were primarily screened using in-house diffractometers from Bruker. Further crystal screening and data collection was carried out at Diamond Light Source, on various beamlines.

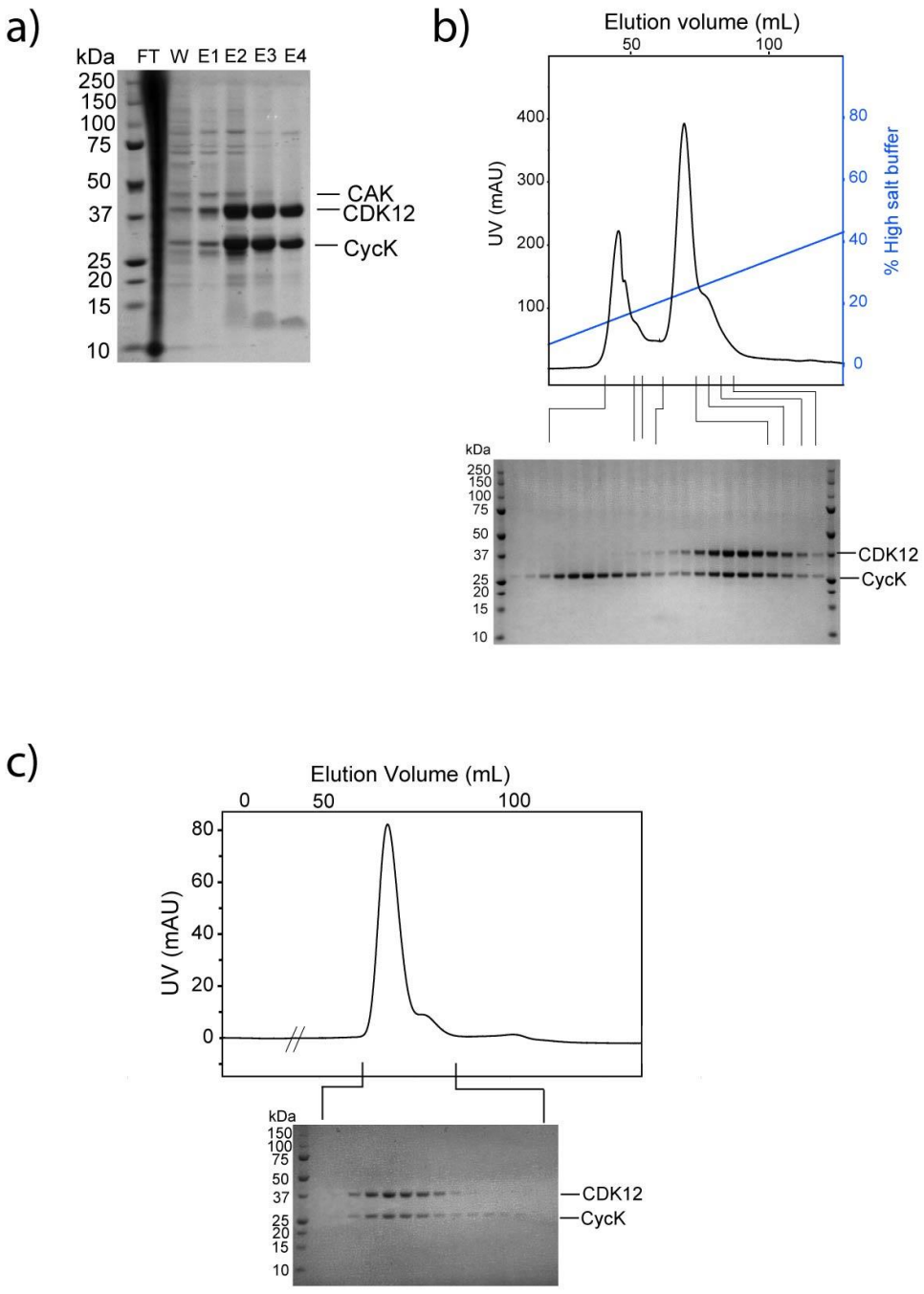


Figure 4.7. Purification of CDK12/CycK co-expressed with CAK. a) SDS PAGE gel showing fractions from nickel-affinity chromatography of co-expressed CDK12, CycK and CAK (Tagged CDK12 is 42,102 Da, tagged CycK is 32,926 Da and tagged CAK is 45,427 Da, correct sized bands are labelled). Abbreviations shown are FT, flow through fraction, W, wash fraction; E1-4, elution fractions 1-4. Elution fractions E2-E4 were pooled. **b)** Following the removal of the hexahistidine tags by TEV protease, the samples was further purified by cation exchange chromatography on a 5 mL HiTrap SP HP column (GE Healthcare), to remove any remaining CAK and excess CycK. Figures show the UV trace from the chromatography (top) and SDS PAGE of indicated elution fractions (below). Elution fractions from the main peak were pooled and buffered exchanged to remove DTT before addition of a covalent THZ series inhibitor. Buffers used were a high salt buffer of 50 mM MES, pH 6.5, 0.5 mM TCEP, 1M NaCl and a no salt buffer of 50 mM MES, pH 6.5, 0.5 mM TCEP. **c)** UV trace and SDS PAGE gel of elutions following size-exclusion chromatography purification of CDK12/CycK with bound THZ531, on a HiLoad S75 Superdex 16/60 column (GE Healthcare).

4.2.4 Crystallization and preliminary refinement of a CDK12/CycK/THZ1 complex structure

THZ1 was the first compound available for crystallisation trials. At this preparative scale of protein purification near complete binding of THZ1 to CDK12 was achieved (Fig.4.8.). Crystallisation trials were set up after a subsequent final purification step of size-exclusion chromatography. For crystallisation trials, the CDK12/CycK/THZ1 complex was concentrated to 6.3 mg/mL buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP and put into coarse screen trials; Hampton Crystal Screen (HCS), Hampton Crystal Screen (HCS), JCSG and LFS screens (Table 4.3). Diffraction quality crystals were grown at 4°C or 20°C in 150 nL sitting drops, mixing protein solution with reservoir solution at a ratio of 2:1, 1:1 or 1:2. Crystals were cryo-protected with mother liquor supplemented with an additional 15% ethylene glycol and vitrified in liquid nitrogen after mounting. Numerous CDK12/CycK/THZ1 complex crystals were obtained from coarse screens, but all of the 21 crystals mounted and tested yielded diffraction limits of between 3.4-5.5 Å resolution (Table 4.3., Fig. 4.9a). Fine screens based on precipitant conditions of the best diffracting crystals were then set up. These were based around an original HIN crystal screen condition of 25% PEG3350, 0.1 M bis-tris pH 5.5 and a HCS condition comprising 12% PEG20K, 0.1 M MES pH 6.5. New crystals were grown at 20°C in sitting drops and 6 crystals were mounted and screened for diffraction (Table 4.3). Viable diffraction quality crystals grew in the monoclinic space group *P* 31 2 1 in a final precipitant condition consisting of 21% PEG3.3K, 0.1 M bis-tris pH 5.5. The resulting CDK12/CycK/THZ1 structure was solved by molecular replacement using Phaser and separate CDK12 and CycK search models from my earlier structure (PDB ID: 4CXA). One CDK12/CycK complex was found in the asymmetric unit (Table 4.4; Fig. 4.9a). The structure was subsequently refined at 3.1 Å resolution (see chapter 2 for full methods). The

electron density for the CDK12 C-terminal kinase extension was not clearly defined beyond Pro1034. Moreover, there was only weak electron density inside the ATP-binding pocket that could not be increased upon refinement. Indeed, numerous rounds of refinement failed to provide any clear evidence for THZ1 to be bound in the CDK12 model or to elucidate the position of Cys1039 (data not shown).

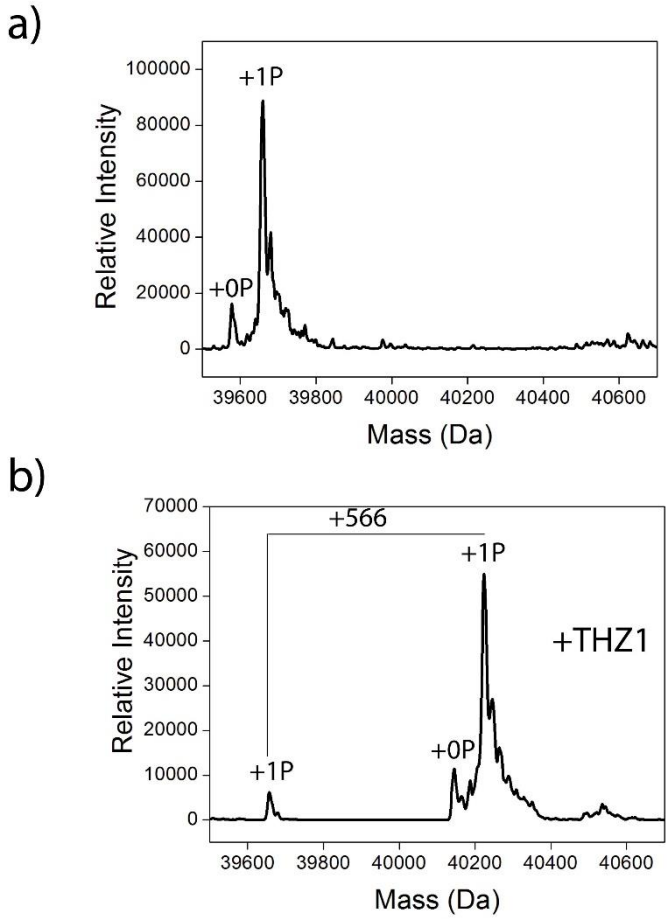


Figure 4.8. Deconvoluted intact mass spectra for CDK12, before and after covalent binding of the THZ1 compound. a) Intact mass spectrum of $CDK12^{715\text{-}1052}$. Mono-phosphorylated (+1P) $CDK12^{715\text{-}1052}$ is the predominant species (Mass of 39659 Da). **b)** Intact mass spectrum of $CDK12^{715\text{-}1052}$ with covalently bound THZ1. Phosphorylated (+1P) $CDK12^{715\text{-}1052}$ with THZ1 bound is the predominant species. A small amount of un-phosphorylated (0P) $CDK12^{715\text{-}1052}$ with THZ1 bound is also observed as well as some phosphorylated (+1P) $CDK12^{715\text{-}1052}$ without bound THZ1.

Table 4.3. Details of crystallisation trials for CDK12/CycK with THZ series compounds

Compound	THZ1	THZ-4-80-1	THZ531
CDK12/CycK complex concentration (mg/mL)	6.3	5.2-6.6	5.5-6.5
Coarse screens	HIN, HCS, JCSG, LFS	HIN, HCS	HIN, HCS, JCSG, LFS
Fine screens	HIN-D7, HCS-F10		HCS-F10-H12, HCS-D6
Additions	-	Crystal seeding	Crystal seeding
Temperature (°C)	4 or 20	4 or 20	20
Outcome	Crystals with diffraction to 3.1-5.5 Å. Dataset at 3.1 Å (unable to build ligand)	Crystals with diffraction to 3.6-7.8 Å	Crystals with diffraction to 2.7-3.6 Å Dataset at 2.7 Å (PDB: 5ACB)

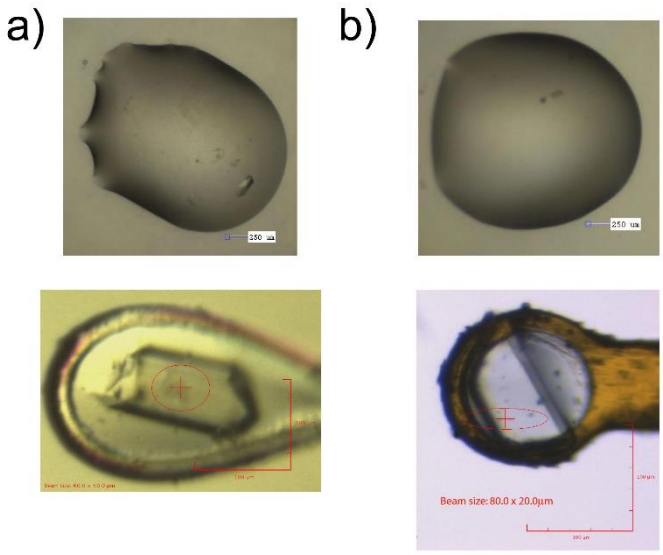


Figure 4.9. Crystals of CDK12/CycK with bound inhibitors. a) Crystals of CDK12/CycK with THZ1 were seen after 1 day at 20ºC in sitting drops with a reservoir containing a HIN fine screen condition of 21% PEG3.3K, 0.1 M bis-tris pH 5.5. After 6 days, the crystal was cryo-protected with an additional 25% ethylene glycol and mounted. Diffraction data were collected at the Diamond I04-1 beamline. **b)** Crystals of CDK12/CycK with THZ531 were seen after 2 days from microseeds in a JCSG coarse screen condition comprising 10% PEG8000, 0.2 M magnesium chloride, 0.1 M HEPES pH 7.0. After 22 days, the crystal was cryo-protected with an additional 15% ethylene glycol and mounted. Diffraction data were collected at the Diamond I03 beamline.

4.2.5 Crystallization trials of a CDK12/CycK/THZ-4-80-1 complex

To improve inhibitor selectivity for CDK12 my collaborators (Gray *et al.*), modified the substituent bearing the acrylamide to alter the orientation of the acrylamide warhead in an effort to better target Cys1039. Following iterative synthesis and testing of a variety of analogues, they developed the derivative THZ-4-80-1 (Table 4.1; Fig. 4.3) in which a piperidine ring was incorporated in the acrylamide ‘arm’. They observed that this compound resulted in more selective binding to CDK12 and CDK13 relative to CDK7 and CDK9 (Professor Nathanael Gray, personal communication). I was given this new compound for further structural studies.

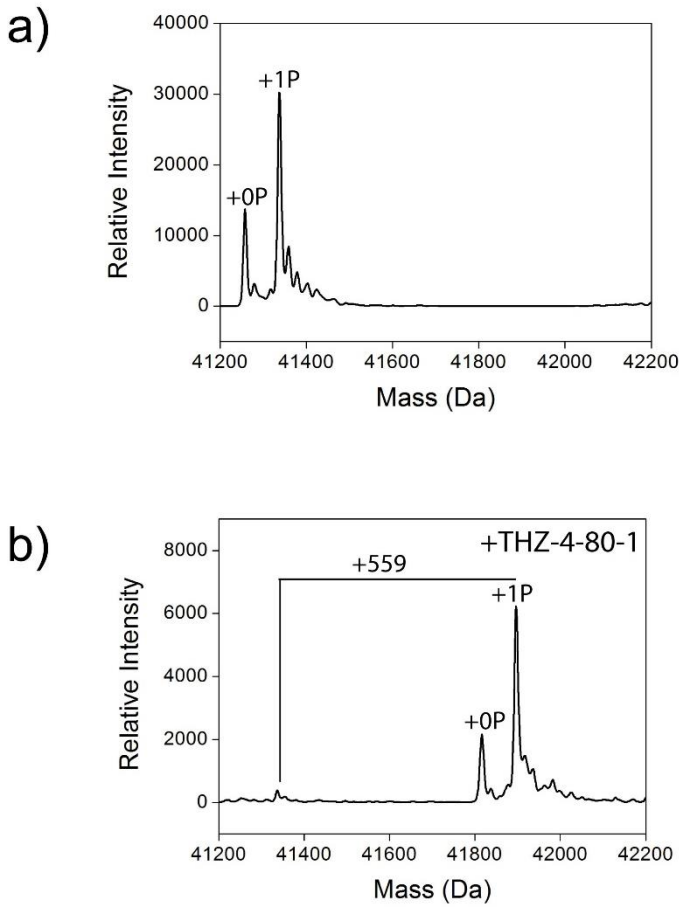


Figure 4.10. Deconvoluted intact mass spectra for CDK12, before and after covalent binding of the THZ-4-80-1 compound. a) Intact mass spectrum of $CDK12^{715-1068}$. Mono-phosphorylated (+1P) $CDK12^{715-1068}$ is the predominant species (Mass of 41,337 Da). b) Intact mass spectrum of $CDK12^{715-1052}$ with covalently bound THZ-4-80-1. Phosphorylated (+1P) $CDK12^{715-1068}$ with THZ-4-80-1 bound is the predominant species. A lesser amount of un-phosphorylated (0P) $CDK12^{715-1060}$ with THZ-4-80-1 bound is also observed as well as some phosphorylated (+1P) $CDK12^{715-1068}$ without bound THZ-4-80-1.

In initial crystallization trials, a longer CDK12 construct was trialled to see if this may improve the diffraction quality. The phosphorylated $CDK12^{715-1068}$/$CycK^{11-267}$ complex was prepared in the same way as described previously for the $CDK12^{715-1052}$/$CycK^{11-267}$ complex and incubated with THZ-4-80-1 at 4°C (Fig.4.10.). Near complete binding to CDK12 was achieved, as seen by an addition of 559 Da to the CDK12 mass of 41258 Da (Fig.4.10.). Coarse screens crystallization screens of HIN. HCS, JCSG and LFS were set up for co-crystallisation trials utilising THZ-4-80-1 at 5.9mg/mL, at 4ºC or 20 ºC,

in 150 nL sitting drops, mixing protein solution with reservoir solution at a ratio of 2:1, 1:1 or 1:2. These yielded only clusters of small needles (Fig. 4.11a).

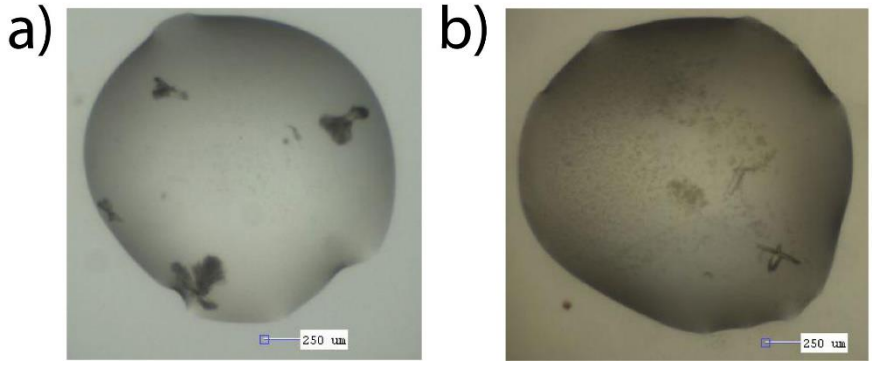


Figure 4.11. Crystallization trials for multiple CDK12 constructs with THZ-4-80-1. a) Representative needle-like crystals formed in clustered on the CDK12$^{715\text{-}1068}$/CycK$^{11\text{-}267}$ complex. This 150 nL sitting drop is from the JCSG coarse screen at 4ºC in a sitting drop mixing 100 nL protein solution with 50 nL of a reservoir solution comprising 0.1M Bis tris pH 5.5, 25% PEG3350, 0.2M NaCl. This was subsequently used in a microseed stock. **b)** A co-crystal of CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ with THZ-4-80-1. The crystal grew after 27 days at 20ºC in a sitting drop in the HIN coarse screen condition with microseeding (25% PEG1500). Diffraction data was collected at 3.6 Å.

Preliminary attempts were then made to crystallise the CDK12$^{715\text{-}1052}$/CycK$^{11\text{-}267}$ complex with THZ-4-80-1. Near complete binding of THZ-4-80-1 to CDK12$^{715\text{-}1052}$ was achieved, as seen by an addition of 559 Da to the CDK12 mass of 39,659 Da (Data not shown). Some of the needles in a JCSG coarse screen for the CDK12$^{715\text{-}1068}$/CycK$^{11\text{-}267}$ complex (0.1M Bis tris pH 5.5, 25% PEG3350, 0.2M NaCl), were used to make a microseed stock that was then used in subsequent crystallization trials (Fig. 4.10a). Coarse crystallization screens of HIN and HCS were set up for co-crystallisation trials at 6.6 mg/mL, at 20 ºC in 170 nL sitting drop, mixing protein solution with reservoir solution at a ratio of 2:1 and 1:1, with 20 nL microseed stock addition to each well (Fig. 4.11b). Crystals collected diffracted to between 3.6-7.8 Å (Table 4.3).

During further biochemical investigation using this compound, Gray *et al.* established that the enantiomers of THZ-4-80-1 had different affinity for CDK12.

Specifically, the R-form (THZ531) was found to have a higher affinity than the S-form (THZ532) (see Fig. 4.1 for compound structures). Therefore, further crystallisation trials with THZ-4-80-1 were halted and subsequently undertaken using the preferred R-enantiomer, THZ531 (Table 4.3).

4.2.6 Crystallization trials and structure determination of a CDK12/CycK/THZ531 complex

Co-crystallisation trials with THZ531 were undertaken with $CDK12^{715-1052}/CycK^{11-267}$ complex. Incubation of THZ531 with purified mono-phosphorylated $CDK12^{715-1052}/CycK^{11-267}$ complex resulted in near complete covalent binding as determined by mass spectrometry (Fig.4.12.). For crystallisation trials, the CDK12/CycK/THZ531 complex was concentrated to 5.5 mg/mL buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP. Viable diffraction quality crystals were grown at 20°C in 195 nL sitting drops mixing 75 nL protein solution and 20 nL crystal seed stock with 100 nL of a reservoir solution comprising 0.1 M HEPES pH 7.0, 10% PEG8000, 0.2 M magnesium chloride (Fig. 4.9b). Crystals were cryo-protected with mother liquor supplemented with an additional 15% ethylene glycol and vitrified in liquid nitrogen after mounting.

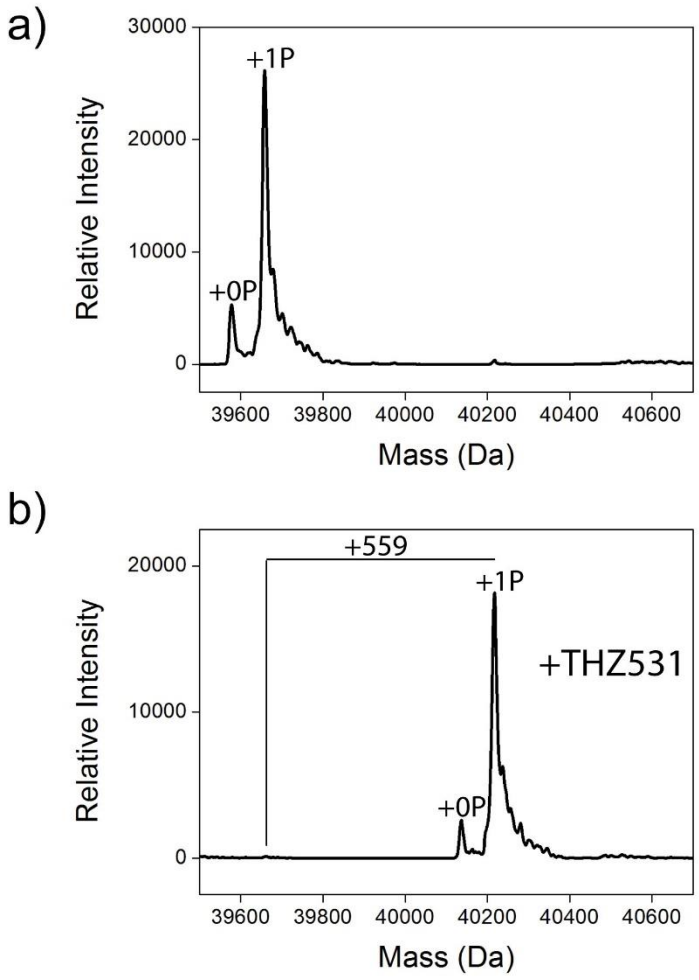


Figure 4.12. Deconvoluted intact mass spectra for CDK12, before and after covalent binding of the THZ531 compound. a) Intact mass spectrum of $CDK12^{715\text{-}1052}$. Phosphorylated (+1P) CDK12 is the predominant species prior to addition of THZ531 compound (Mass of 39659 Da). **b)** Intact mass spectrum of $CDK12^{715\text{-}1052}$ with THZ531 bound. Phosphorylated (+1P) CDK12 with THZ531 bound is the predominant species. Some small amounts of un-phosphorylated (0P) CDK12 with THZ531 bound and phosphorylated (+1P) CDK12 (unbound) are present.

Diffraction data were collected at 100 K on Diamond Light Source beamline I03. The resulting CDK12/CycK/THZ531 structure was solved by molecular replacement with Phaser using the separate CDK12 and CycK subunits from PDB ID: 4CJY (subsequently redeposited as 4UN0) as search models. The resulting structure was refined at 2.7 Å resolution. An initial file with PDB coordinates for the inhibitor, THZ531, was created using PRODRG in the CCP4 suite [9]. See Table 4.4 for full diffraction data collection and refinement statistics.

Table 4.4 Data collection and refinement statistics (molecular replacement)

	CDK12$^{715-1052}$/cyclin K$^{11-267}$/THZ531
PDB accession code	5ACB (To be released)
Chain IDs	Chains C/B (CDK12/CycK) Chains D/A (CDK12/CycK)
Data collection	
Beamline	Diamond Light Source, I03
Wavelength (Å)	0.97625
Space group	$P2_1$
Cell dimensions	
a, *b*, *c* (Å)	*49.8, 148.7, 91.6*
α, β, γ (°)	90.0, 93.8, 90.0
No. unique reflections	38,486
Resolution (Å)	41.3-2.70 (2.77-2.70)*
R_{merge}	0.071 (0.569)
CC(1/2)	0.997 (0.694)
$I / \sigma I$	12.5 (1.9)
Completeness (%)	98.7 (99.0)
Redundancy	3.0 (3.1)
Refinement	
No. reflections	34,240 (2581)
R_{work} / R_{free} (%)	22.14/26.19
Average B factor (Å^2)	60.8
No. atoms in refinement	
Protein	8993
Ligand/ion	80
R.m.s. deviations	
Bond lengths (Å)	0.0072
Bond angles (°)	1.1369
Molprobity	
Ramachandran favoured (%)	96
Ramachandran outliers (%)	4
Crystallization conditions	10% PEG8000, 0.2 M magnesium chloride, 0.1 M HEPES pH 7.0

*Single crystal. Values in parentheses are for highest-resolution shell. R.m.s. indicates the root-mean-square.

4.2.7 Overview of the CDK12/CycK complex with THZ531 bound

The co-crystal structure of the CDK12/CycK complex with bound inhibitor demonstrated that THZ531 covalently binds to Cys1039 (Fig. 4.11). There were two CDK12/CycK complexes in the asymmetric unit and each bound THZ531 in a different conformation (Fig. 4.13). The electron density maps for the structure were generally of high quality. The electron density was most complete for the CDK12/CycK complex comprising chains A and C. For this complex, the CDK12 subunit (chain C) was traced between residues Glu716-His1040, with the exception of residues Asp877-Tyr892. CycK (chain A) was traced between residues Thr20-Met260, with the exception of residues Lys228-Tyr231. In the second complex, CDK12 (chain D) was traced between residues Asp718-Ser1044, but poorer electron density prevented modelling between residues Arg779-Asp801 and Asn885-Ser889. CycK (chain B) was traced between residues Thr20-His267.

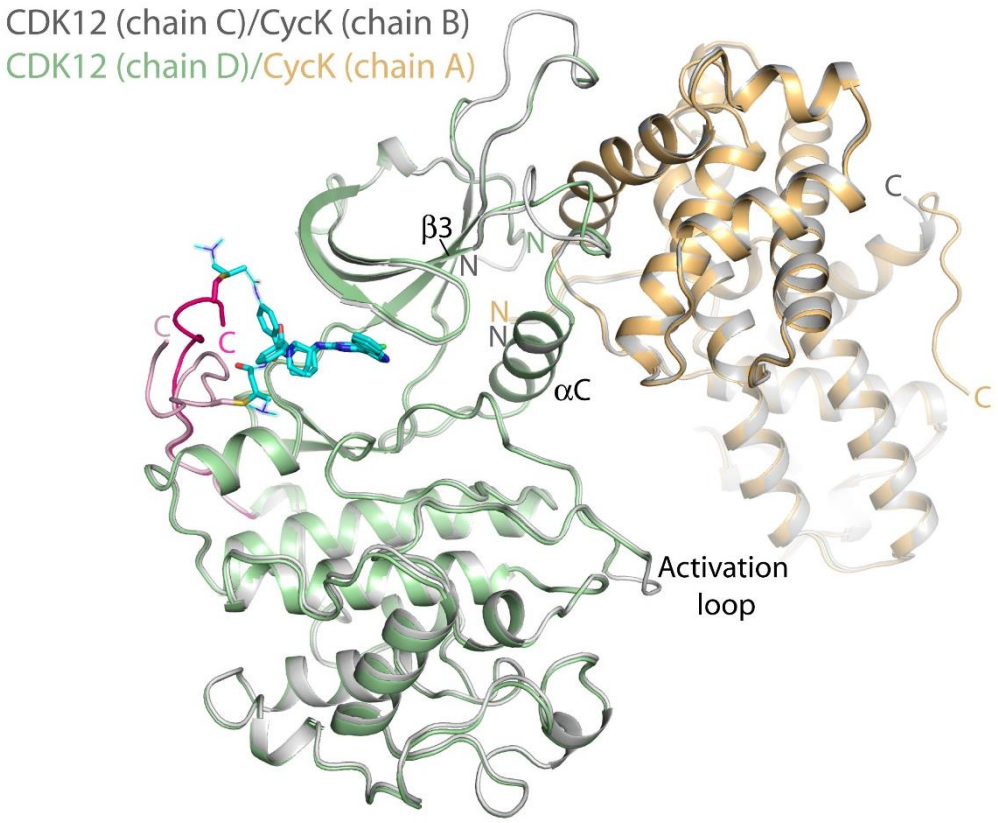


Figure 4.13. Overview of the crystal structure of CDK12/CycK showing two different binding conformations of THZ531. There were two CDK12/CycK complexes within the asymmetric unit and THZ531 bound to each complex in a different conformation. In particular, there was a significant difference in the structures of the two complexes within the C-terminal kinase extension (coloured pink). A different conformation is also observed in the CDK12 loop connecting αC and β3.

Overall, the structures of the two complexes in the asymmetric unit are well conserved, except for the C-terminal kinase extensions and their interactions with the THZ1 inhibitor which will be discussed in section 4.2.7 below. As expected, the CDK12 kinase domains exhibit an active conformation with a phosphorylated activation loop. The core interactions between the CDK and cyclin are also maintained in each complex. However, one region of divergence between the two complexes is the CDK12 loop between the N-lobe β3 strand and the αC helix. Here, residues in the loop between residues 760-DNEK-765 in chains C and D vary in position by as much as 8.1 Å (for example, Asn762, Fig. 4.14a). This solvent-exposed loop makes no significant interactions and further inspection

reveals that flexibility in this loop is apparent across the range of available CDK12 structures determined by x-ray crystallography (Fig. 4.14b).

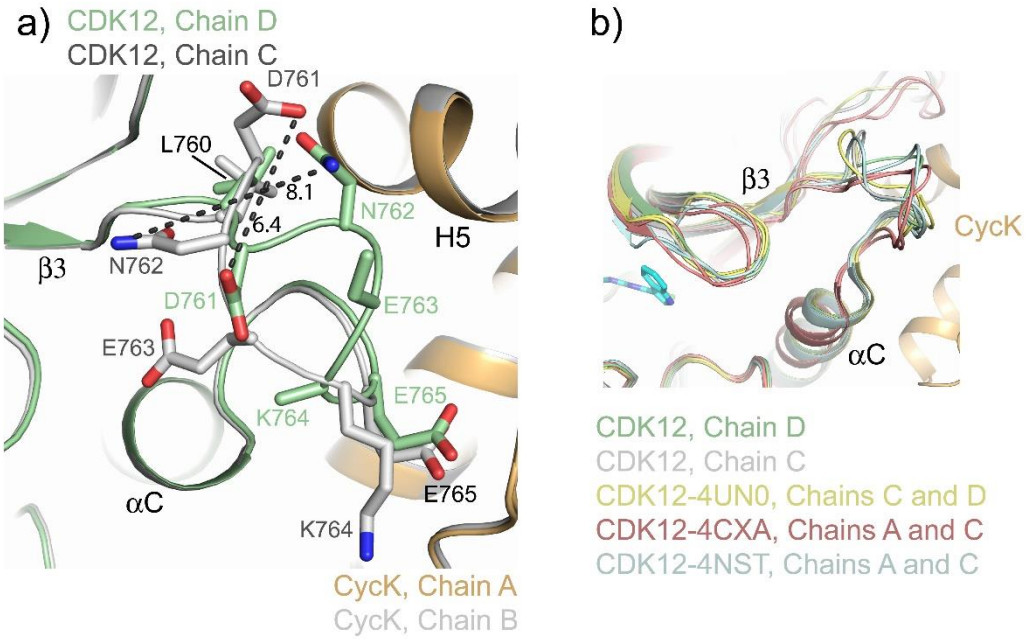


Figure 4.14. The CDK12 loop connecting β3 and αC is flexible. a) There is a shift in the position of the β3-αC loop in the two CDK12 subunits located in the asymmetric unit of the crystal structure. Distances between selected residues are indicated as a dashed line. There is 6.4 Å between the Asp761 carbonyls in each chain and 8.1 Å between the Asn762 amine in each chain. **b)** Superposition of all available CDK12 crystal structures reveals the common flexibility in the β3-αC loop.

4.2.8 THZ531 can bind to CDK12 in two different conformations

Most significantly, the structure revealed that the labile αK helix in the C-terminal kinase extension can be displaced from the CDK12 surface to allow Cys1039 to reorient towards the ATP–binding pocket (Fig. 4.15a). The 3-aminopyrimidine of THZ531 binds to the kinase hinge region forming two hydrogen bonds to Met816, while the pendant 3-indolyl and piperazine groups mediate further hydrophobic interactions in the ATP-binding pocket. Different conformations of the secondary amide were observed in the two THZ531 complexes causing the solvent exposed inhibitor groups to pack either against the N-lobe

β1 strand or C-lobe αD. Both sites were only poorly defined by the electron density, except for the cross-linked Cys1039 positions, which were clearly visible in the maps (Fig. 4.15b). The CDK12 C-terminal extension in both THZ531-bound conformations is shifted from the AMP-PNP-bound CDK12 structure, highlighting the flexibility of this region (Fig. 4.14c).

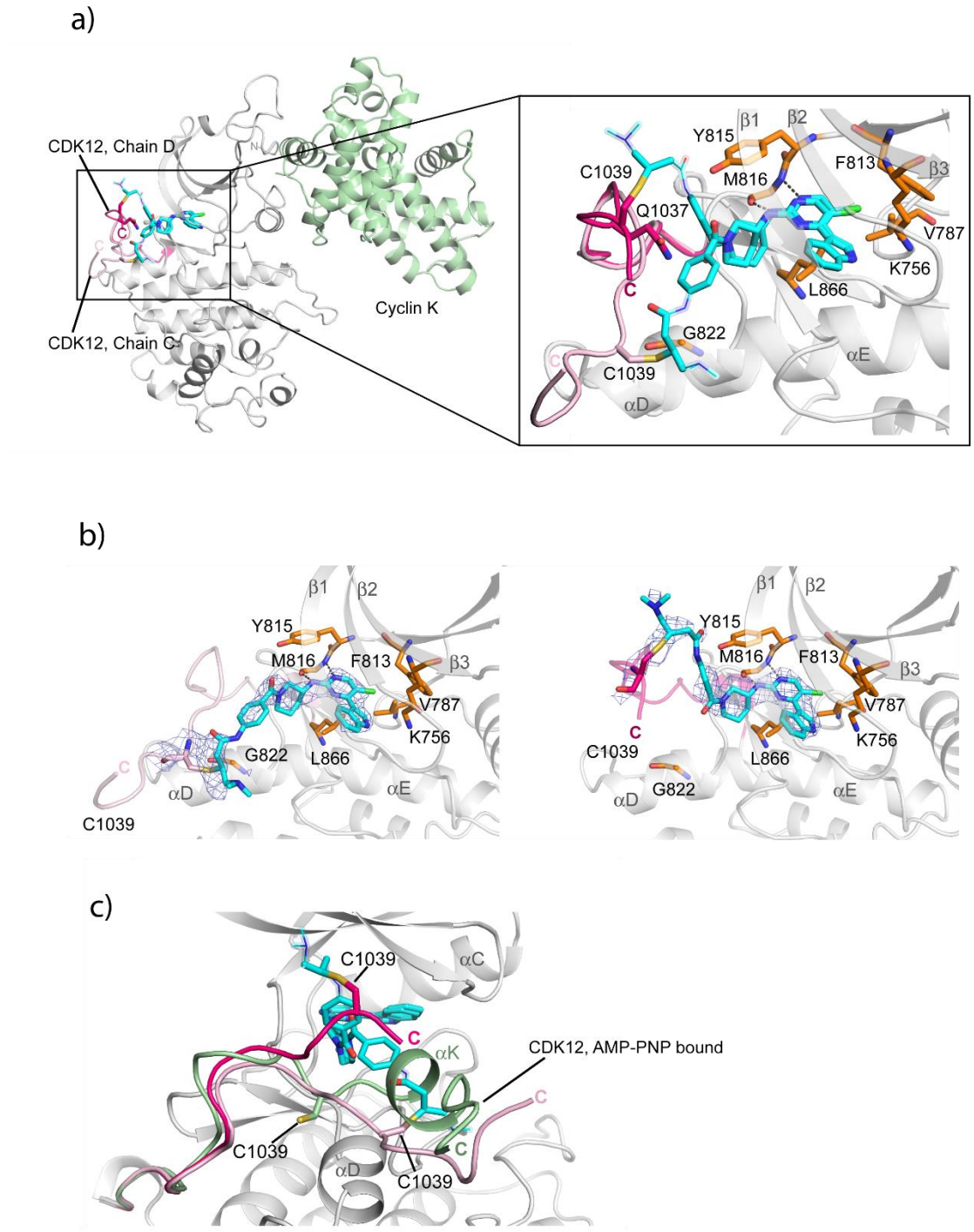


Figure 4.15. THZ531 co-crystal structure with CDK12/CycK. a) Overview of the CDK12/CycK structure showing THZ531 bound in different conformations in the two CDK12/CycK complexes present in the asymmetric unit. Hydrogen bonding between THZ531 and the CDK12 Met816 in the kinase hinge region is shown as dashed lines. The CDK12 C-terminal extensions in chains C and D are shown in pale and bright pink, respectively. Other CDK12 kinase domain secondary structure elements from chain D are shown in grey. The inhibitor THZ531 is shown in cyan**. b)** Omit maps contoured at 2.5σ for THZ531 and the C-terminal extension residue Cys1039 are shown for each THZ531 conformation. **c)** Superposition of the CDK12-THZ531 structures (grey with bright or pale pink C-terminal extensions) and the AMP-PNP bound structure of CDK12 (green, PDB: 4CXA).

4.2.9 Comparison of the CDK7-THZ1 and CDK12-THZ531 binding models

There are currently no available crystal structures for the CDK7-THZ1 complex. Nonetheless, the structure of the THZ531 inhibitor covalently bound to CDK12 allows the binding mode of THZ531 to be compared to a published docking model of the CDK7-THZ1 complex [3]. Notably, the equivalent sulfhydryl in CDK7, Cys312, is at a more distant site compared to either THZ531 conformation (11.7 Å and 14.6 Å from CDK12 Cys1039 in chain C and chain D, respectively), suggesting a basis for the respective selectivity of THZ1 and THZ531 (Fig.4.16). Seemingly, this interaction is the main specificity determinant with no further significant non-covalent interactions gained between the CDK12 and CDK7 binding models.

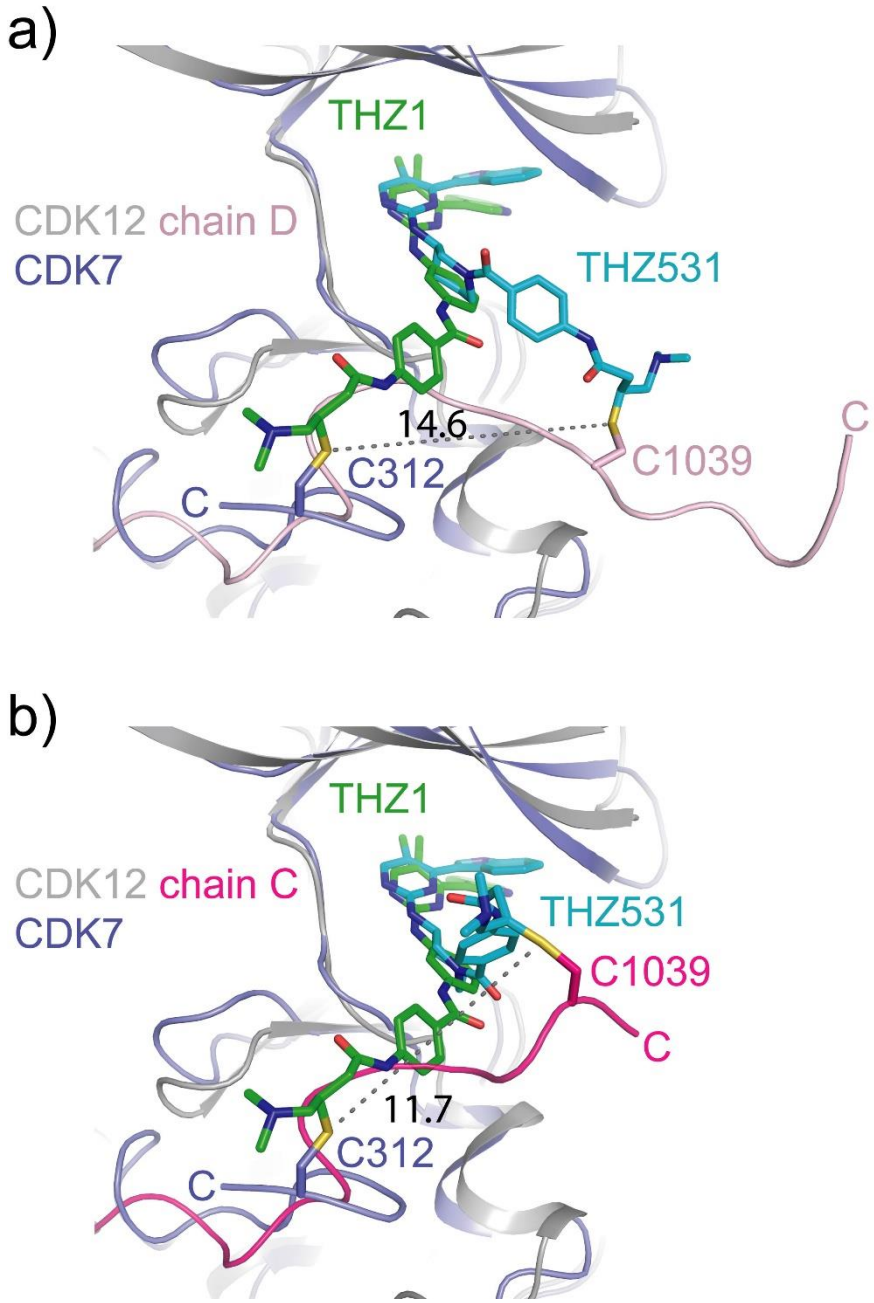


Figure 4.16. Superposition of the CDK7-THZ1 docking model and the different conformations of THZ531 bound to CDK12. a) The position of CDK12 Cys1039 in chain D with bound THZ531 is 14.6 Å away from the position of Cys312 in the CDK7 model with THZ1. **b)** The position of CDK12 Cys1039 in chain C with bound THZ531 is 11.7 Å away from the position of Cys312 in the CDK7 model with THZ1.

Further structural comparisons were made to assess the flexibility of the C-terminal kinase extensions in CDK12 and CDK7 as this has significant potential to determine the binding selectivity of any future compounds in the THZ series. Superposition of available

CDK12 crystal structures reveals that the position of Pro1034 is conserved in both CDK12-THZ531 chains, as well as in the CDK12$^{715\text{-}1038}$ structure (PDB 4UN0 chain C) and the AMP-PNP-bound chain of the CDK12$^{715\text{-}1052}$ structure (PDB 4CXA chain A). Further, the position and packing of Pro1034 are similarly conserved in the two CDK12 chains located in the asymmetric unit of the CDK12-ADP complex solved by Bosken *et al.* (PDB 4NST). The only exception is CDK12$^{715\text{-}1052}$ chain C (PDB 4CXA) in which the entire C-terminal extension is wrapped around the back face of the kinase domain. The preferred position of Pro1034 appears to be an anchor point for the C-terminal kinase extension that is stabilised by a number of hydrophobic interactions with CDK12 residues in the αD helix and the following loop (Fig. 4.17a). Beyond this point, the C-terminal extension is evidently flexible to accommodate the THZ531 inhibitor interaction through Cys1039.

In the crystal structure of CDK7 (PDB 1UA2) the C-terminal region is not defined beyond Asn311 and so the position of Cys312 has only been modelled [3, 4] . Interestingly, a proline residue in CDK7 (Pro310) adopts an equivalent position to CDK12 Pro1034 and appears to be similarly anchored by hydrophobic interactions (Fig. 4.16b). CDK7 Pro310 and CDK12 Pro1034 therefore appear as conserved positions for structure-based sequence alignment (Fig. 4.17b). Pro310 is notably in close proximity to Cys312 in CDK7, whereas CDK12 Pro1034 is five residues upstream from Cys1039 (Fig. 4.17b). Thus, if the anchoring of Pro310 is truly important, CDK7 Cys310 could not occupy a position around the ATP-binding site for THZ531 to form a covalent bond, consistent with the structural models in Fig. 4.16.

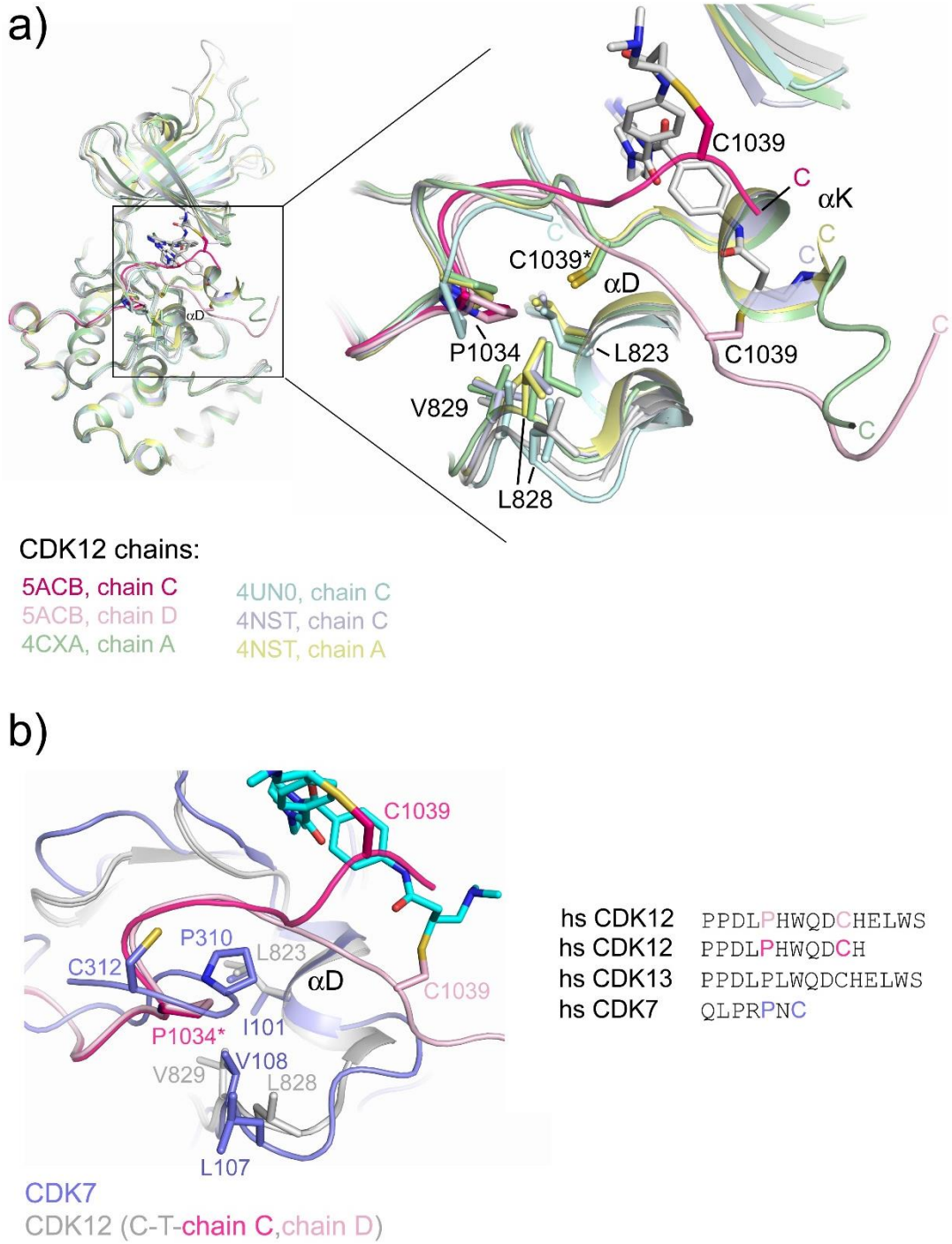


Figure 4.17. C-terminal kinase extensions in CDK7 and CDK12 show a structurally conserved proline. a) Superposition of available CDK12 crystal structures shows a conserved position of Pro1034, whereas further C-terminal regions are structurally diverged. Pro1034 is nestled in a hydrophobic pocket made up of Leu823 (αD helix) and Leu828 and Val 829 (loop following αD). CycK is omitted for clarity. C1039* labels Cys1039 from PDB 4NST and 4CXA chain A. For clarity only the THZ531 ligand is shown. **b)** Superposition of CDK12-THZ531 (PDB 5ACB) and the published CDK7 model reveals a similarly positioned proline in each structure anchored in position by a number of hydrophobic interactions. This proline forms the basis for a structure-based sequence alignment (right). 'C-T' refers to the CDK12 C-terminal extension.

4.3 Discussion

The efficacy of protein kinase inhibitors can be limited by their fast dissociation kinetics [10]. Covalent kinase inhibitors can bypass this problem and have potential for long lasting biological actions [11]. However, as a consequence, a major concern has been the selectivity of these compounds as reactive cysteine residues are generally found in a number of different protein kinases, as well as other ATP-binding proteins. To date, most covalent kinase inhibitors have been directed towards cysteines located within the kinase ATP-binding pocket [11]. My collaborator, Professor Nathanael Gray, recently introduced a new inhibitor design strategy by selectively targeting a distal cysteine residue present in the C-terminal kinase extension of CDK7 [3]. My first structures of the CDK12/CycK complex (chapter 3) confirmed that a cysteine was somewhat similarly located in the C-terminal extension of CDK12 and I have demonstrated here that the lead CDK7 inhibitor, THZ1, can also bind covalently to recombinant CDK12. In this chapter, I have solved the 2.7 Å co-crystal structure of CDK12/CycK bound to THZ531, which was developed as a highly selective CDK12 inhibitor by Professor Gray and colleagues (to be published). Data they have collected suggest that THZ531 inhibits CDK12 and CDK7 with IC_{50} values of 0.16 μM and 25 μM, respectively (Professor Gray, personal communication).

The new structure suggests the main determinant of protein kinase selectivity within this inhibitor series is the positioning of the cysteine in the C-terminal kinase extension. The two conformations observed for THZ351 binding to Cys1039 show the degree of flexibility inherent in the CDK12 C-terminal extension. The limited positions available to the bound inhibitor appear to dictate the final positions of Cys1039 in the different CDK12 chains, although conversely, the protein's flexibility allows THZ531 to adopt its two observed conformers. Regardless, the structure defines the mode of THZ531 binding to

CDK12.

Overall, both the non-covalent and covalent binding features of the THZ series may make contributions to the selectivity. The classical non-covalent hinge interaction is likely necessary to initially target the inhibitor to the kinase ATP-binding pocket and to orient it into a position favourable to enable the "warhead" to selectively target the nucleophilic cysteine residue. A large ATP-mimetic group with potent binding may be undesirable as it may increase off-target binding across the wider kinome. Instead, the Gray lab achieved CDK12 selectivity with THZ531 by modifying the linker region in THZ1 between the 3-aminopyrimidine and the reactive acrylamide motif. The co-crystal structure of THZ531 suggests that further linker designs should be tried to further define the optimal position for selectively targeting Cys1039. For example, shorter linkers might further disfavour binding to CDK7 Cys312.

Gray *et al.*, have also shown that THZ531 can potently inhibit CDK13, which is a close paralog of CDK12 and has a strictly conserved cysteine residue in its C-terminal kinase extension. Ideally, a chemical probe should differentiate between CDK12 and CDK13. At present, no crystal structure of CDK13 exists to guide the design of such a molecule, although the CDK12 structure provides a valuable template for homology modelling. CDK12 and CDK13 have 43.7% sequence similarity overall and the kinase domain has 91% sequence identity conservation [5, 12].

Fortunately, CDK9 does not have a C-terminal cysteine residue (Fig. 4.1a). Thus, THZ531 offers some potential as a tool to dissect the different transcriptional roles of CDK9 and CDK12 (work ongoing by Gray *et al.*). Moreover, THZ531 provides a level of specificity against the kinome that other pan-inhibitors of CDK kinases, such as flavopiridol, do not provide. The value of THZ531 as a chemical probe is further supported

by the availability of the poorly binding S-enantiomer, THZ532, which under the time constraints has not been characterised in this work. The unexpected divergent conformers of THZ531 in the CDK12 crystal structure make it somewhat challenging to fully decipher the structural basis for the enantiomer specificity. Presumably, the conformation of the THZ532 ligand when bound into the ATP pocket does not optimally place the "warhead" in the vicinity of Cys1039 for the reaction to occur efficiently prior to the ligand dissociating again.

The profiling of cancer cell lines has identified a subset of cell lines, including human T-cell acute lymphoblastic leukaemia (T-ALL), that have exceptional sensitivity to THZ1 [3]. Further studies with THZ1 have been undertaken in various forms of cancer both *in vitro* and *in vivo.* For example, small cell lung cancer was also particularly sensitive to THZ1, which is therefore of significant interest as a lead for potential clinical candidates [13]. A number of other studies have utilised THZ1 to further characterise the biological role of CDK7 and have shown an unexpected function in RNA polymerase II pausing and mRNA capping [14]. In future, it will be of interest to see the comparable activities of THZ531 and the insights its enables into CDK12 function.

4.4 References

1. Bosken, C.A., et al., *The structure and substrate specificity of human Cdk12/Cyclin K.* Nat Commun, 2014. 5: p. 3505.
2. Liang, K., et al., *Characterization of human cyclin-dependent kinase 12 (CDK12) and CDK13 complexes in C-terminal domain phosphorylation, gene transcription, and RNA processing.* Mol Cell Biol, 2015. 35(6): p. 928-38.
3. Kwiatkowski, N., et al., *Targeting transcription regulation in cancer with a covalent CDK7 inhibitor*. Nature, 2014. 511(7511): p. 616-20.

4. Lolli, G., et al., *The crystal structure of human CDK7 and its protein recognition properties.* Structure, 2004. 12(11): p. 2067-79.

5. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. 25(17): p. 3389-402.

6. Niesen, F.H., H. Berglund, and M. Vedadi, *The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability.* Nat Protoc, 2007. 2(9): p. 2212-21.

7. Dixon-Clarke, S.E., et al., *Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding.* Sci Rep, 2015. 5.

8. Wilkins, M.R., et al., *Protein identification and analysis tools in the ExPASy server.* Methods Mol Biol, 1999. 112: p. 531-52.

9. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. 67(Pt 4): p. 235-42.

10. Muller, S., et al., *The ins and outs of selective kinase inhibitor development.* Nat Chem Biol, 2015. 11(11): p. 818-21.

11. Liu, Q., et al., *Developing irreversible inhibitors of the protein kinase cysteinome.* Chem Biol, 2013. 20(2): p. 146-59.

12. McWilliam, H., et al., *Analysis Tool Web Services from the EMBL-EBI.* Nucleic Acids Res, 2013. 41(Web Server issue): p. W597-600.

13. Christensen, C.L., et al., *Targeting transcriptional addictions in small cell lung cancer with a covalent CDK7 inhibitor.* Cancer Cell, 2014. 26(6): p. 909-22.

14. Nilson, K.A., et al., *THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing.* Mol Cell, 2015. 59(4): p. 576-87.

Chapter 5 Identification of rebastinib and dabrafenib as inhibitors of CDK16

5.1 Introduction

The "PCTAIRE" kinase family remains a poorly characterised, but functionally important and intriguing group of CDK kinases. The three kinase members, PCTAIRE1-3 (also known as PCTK1-3), were recently renamed as CDK16-18, respectively [1]. Prior to my doctoral work, the cyclin requirement of these kinases was uncertain, but a breakthrough had been made at the SGC with the first crystal structure of a PCTAIRE family kinase domain. However, the 2.4 Å structure of human CDK16 (PDB ID: 3MTL, to be published) showed an inactive conformation of the catalytic domain with apparent conformational plasticity in the N-lobe, as evidenced by the largely disrupted αC helix and a "DFG-out" conformation of the activation loop.

CDK16 has high expression in the brain and testis and has been linked to X-linked intellectual disability [2, 3], as well as spermatogenesis [4]. In addition to these important biological roles in development, CDK16 has also been implicated in the growth of several cancers and is therefore a potential target for novel therapeutics.

Recent studies have shown that knockdown of CDK16 mRNA in prostate, breast, cervical and melanoma cell lines diminishes cancer cell proliferation. This is in contrast to non-transformed cells, in which CDK16 knockdown does not seem to have a role. A potential CDK16 target was also identified, the tumour suppressor p27 (Kip1, cyclin-dependent kinase inhibitor 1B). CDK16 is able to phosphorylate p27 at Ser10, which promotes its ubiquitination and degradation. CDK16 knockdown led to p27 accumulation and apoptosis in cancerous, but not in non-transformed cells. In prostate, breast and cervical cancer cell lines this up-regulation of p27 led to cell-cycle arrest in late G2-M phase and

apoptosis, but not in melanoma cell lines. This implies the effect of CDK16 knockdown is heterogeneous. Interestingly, non-malignant cells are not affected in the same way which is therapeutically favourable [5-8].

To date there are no selective CDK16 inhibitors that could be used as chemical probes of CDK16 activity and therapeutic potential. However, the methods determined for the recombinant expression of soluble CDK16 protein at the SGC allow for the screening of available small molecule libraries offering opportunities for inhibitor identification. Differential scanning fluorimetry (DSF) is a valuable technique for initial inhibitor screens as it does not require a knowledge of suitable substrates nor active protein. DSF can identify small molecule compounds that bind to the native protein and increase its stability, as described in chapter 2 and 4.Some initial DSF screens had been performed using CDK16 prior to my arrival and had identified indirubin E804 as the best ranked CDK16 inhibitor with a Tm shift of 8.3 °C. This compound had therefore been co-crystallised in the $CDK16^{163-478}$ kinase domain structure (PDB ID: 3MTL). Indirubin E804 is a rather promiscuous kinase inhibitor that inhibits various CDKs as well as the Src kinase [9, 10]. Based on this known co-structure, indirubin E804 has been used in some recent published work as a pharmacological probe for CDK16 alongside RNAi to demonstrate an apparent role for CDK16 in the proliferation of medulloblastoma cell lines [5].

In this chapter, I extended the previous work at SGC to include the screening of CDK16 against a collection of clinical kinase inhibitors. Two of these inhibitors, rebastinib and dabrafenib, were identified with higher Tm shifts than indirubin E804. I was also able to co-crystallise CDK16 with rebastinib to reveal further structural features of the CDK16 kinase domain in the "DFG-out" conformation. Both rebastinib and dabrafenib were then tested in enzymatic and preliminary cellular assays to explore potential CDK16 involvement in human cancer.

5.2 Results

5.2.1 Purification of CDK16 kinase domain

The cloning, expression and purification of a number of PCTAIRE family constructs had been established at the SGC prior to my arrival (discussed further in chapter 7). From these efforts, only one construct had been identified with a propensity to crystallise, namely a bacterial expression construct of human CDK16$^{163-478}$ containing a phosphomimetic S319D mutation within the kinase activation loop (although the physiological requirement for phosphorylation at this position has not been established). The host vector was pNIC-CH, which adds a non-cleavable hexahistidine tag to the protein C-terminus for nickel-affinity purification.

To further characterise the structural and inhibitor binding properties of CDK16 I transformed this kanamycin-resistant plasmid into BL21(DE3)-R3-pRARE2 cells. These cells have been selected to be resistant to bacteriophage infection (strain "R3") and carry the chloramphenicol-resistant plasmid pRARE2, which provides expression of extra tRNA synthetases to allow translation of rarely used codons in *E. coli*. As expected, expression of the mutant CDK16$^{163-478}$ construct (hereafter CDK16 unless specified) in these cells gave a high yield of soluble protein that was captured by nickel-affinity chromatography (Fig. 5.1a). After a wash step, the CDK16 protein was eluted using a step gradient of 50-250 mM imidazole in a running buffer of 50 mM HEPES, pH 7.5; 500 mM NaCl; 5% Glycerol; 0.5 mM TCEP. These elution fractions were subsequently pooled and concentrated in a 10 kDa MWCO Amicon Ultra concentrator to 46.1 mg/mL, which was determined spectrophotometrically on a NanoDrop ND-1000 (Thermo Scientific, Waltham, MA) using an estimated extinction coefficient of 35,870 $M^{-1}cm^{-1}$ and a molecular weight of 37,252 Da. The sample was aliquoted and flash frozen in liquid nitrogen for storage at -80°C until required. The sample was thawed on ice later for further purification by size-exclusion

chromatography using a HiLoad S200 Superdex 16/60 column (GE Healthcare) (Fig. 5.1b). The purified protein was homogeneous and had an experimental mass determined by LC/MS (Agilent LC/MSD TOF system) of 37252.2 Da in agreement with the expected molecular weight of 37251.8 Da. This protein was then used in subsequent inhibitor screening assays and crystallization trials.

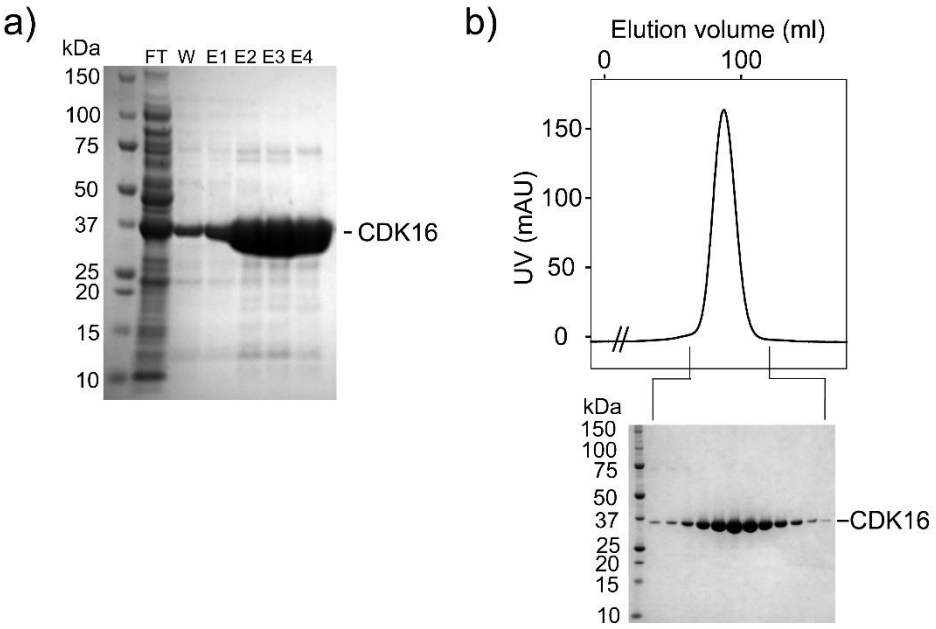


Figure 5.1. Purification of CDK16$^{163-478}$. a) SDS-PAGE gel showing fractions from nickel affinity purification of hexahistidine-tagged CDK16$^{163-478}$ (Molecular weight of 37251.8 Da), from expression in *E.coli* from 1 L culture volume. Fractions E1, E2, E3 and E4 were pooled (FT=flow-through fraction; W=wash fraction, E1-E4=elution fractions. Please refer to materials and methods for full buffer details.) **b)** An ultraviolet light absorbance trace from size-exclusion chromatography recording the elution of hexahistidine-tagged CDK16$^{163-478}$ from a HiLoad S200 Superdex 16/60 column (GE Healthcare). Indicated fractions were collected and analysed on an SDS-PAGE gel (lower panel).

5.2.2 Screening for inhibitors of CDK16

Recently, the SGC acquired a collection of ~150 clinically-relevant kinase inhibitors. I screened this panel of inhibitors against purified recombinant CDK16 kinase domain using DSF. Whilst there were a number of moderate binders, two compounds stood out with a Tm shift well above all previously screened inhibitors (Fig. 5.2). These were

rebastinib (DCC-2036) and dabrafenib (Tafinlar; GSK2118436), with Tm shifts of 12.6 ºC and 10.5 ºC, respectively (average values from two independent experiments). Rebastinib is a multi-targeted type II kinase inhibitor that was developed to inhibit BCR-ABL as well as the drug resistant gatekeeper mutant ABL^{T315I} [11]. It entered phase I clinical trials for chronic myeloid leukaemia [11]. Dabrafenib is an ATP-competitive, reversible inhibitor of mutant $BRAF^{V600E}$ that has been approved for clinical use in advanced melanoma [12].

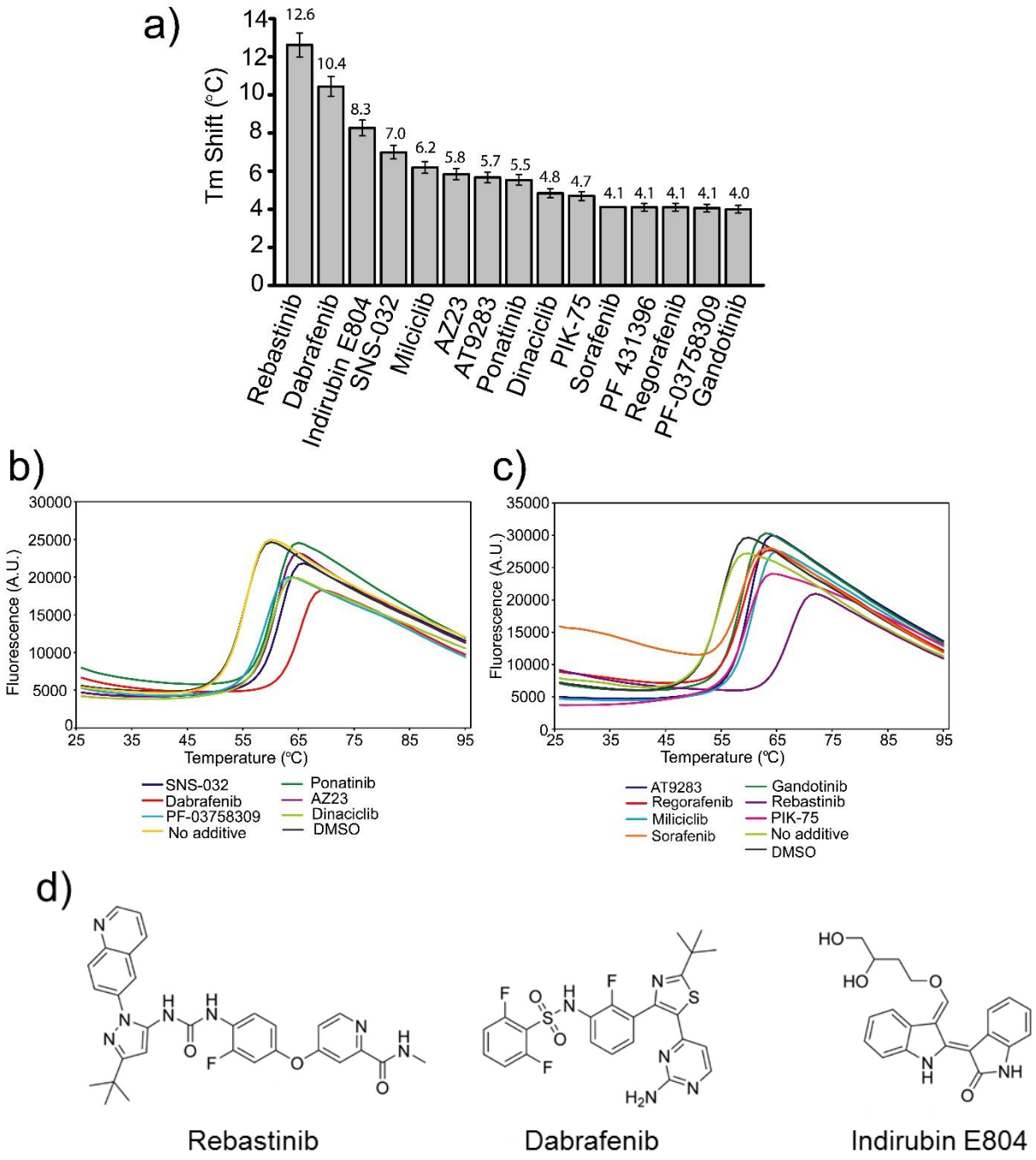


Figure 5.2. DSF screen for clinically-relevant inhibitors of CDK16. a) Bar chart showing the top 15 most potent binders identified for CDK16, as identified by a high Tm shift value. The error bars show the standard deviation of Tm shift values recorded from two separate experiments that used the same stock preparation of CDK16$^{163\text{-}478}$, but varied DSF experimental buffers consisting of 10 mM HEPES pH 7.5 and either 150 mM or 500 mM NaCl. Salt concentration was found to have no influence on these data. Mean values were therefore calculated and are shown above each bar. Data for Indirubin E804 were collected at the SGC prior to my arrival by Dr Oleg Fedorov. **b)** and **c)** are representative raw DSF data for each inhibitor identified in **a)**. **d)** Chemical scaffolds of most potent binders identified for CDK16.

Other inhibitors that gave a Tm shift above or close to 5 ºC included a number of CDK inhibitors, such as SNS-032 (7.0 ºC), milciclib (6.2 ºC) and dinaciclib (4.8 ºC). Further inhibitor descriptions are provided in Table 5.1. Later experiments showed that full length CDK16 had the same inhibitor binding profile (data not shown).

Table 5.1 Description of selected kinase inhibitors from the DSF screen.

Compound	Tm shift (ºC)	Known targets	Clinical stage
Rebastinib	12.6	BCR-ABL inhibitor for wild-type ABL and ABL(T315I) with IC50 values of 0.8 nM and 4 nM. It also inhibits SRC, LYN, FGR, HCK, KDR, FLT3, and Tie-2 and low activity is seen towards c-Kit, PDGFRα and PDGFRβ. CDK2/CycD has an IC50 >5 μM [11].	Phase I
Dabrafenib	10.5	Mutant $BRAF^{V600E}$ specific inhibitor with an IC50 of 0.8 nM in cell-free assays, with less potency against BRAF (wild type) (3.2 nM) and c-Raf (5 nM)[12].	Approved
Indirubin E804	8.3	ATP-competitive inhibitor that directly inhibits Src kinase activity IC50 of 0.43 μM [9].	Pre-clinical
SNS-032	7.0	ATP-competitive inhibitor of CDK9/CycT with IC50 of 4 nM and an IC50 for CDK2/CycA 38nM in cell-free assays. It is 10- and 20-fold selective over CDK1/CDK4 complexes and >1 μM for CDK6. It inhibits CDK7/CycH with IC50 of 62 nM [13, 14].	Phase I
Milciclib	6.2	ATP-competitive CDK inhibitor with an IC50 of 45 nM for CDK2/CycA. It is >3-fold more selective than for CDK1/CycB, CDK2/CycE, CDK4/CycD1, CDK5/p35 and CDK7/CycH. It also inhibits TrkA with an IC50 of 53 nM [15].	Phase 2

5.2.3 Crystallisation of CDK16 with rebastinib

Rebastinib is a type II kinase inhibitor that has been crystallised previously with the ABL tyrosine kinase [11]. To understand its binding to a CDK family kinase, crystallisation trials were undertaken with CDK16. The protein was buffer exchanged into a buffer that had previously proven successful for crystallisation consisting of 25 mM HEPES, pH 7.5, 150 mM NaCl, 5% glycerol, 10 mM DTT. This protein was then concentrated in a 10 kDa MWCO Amicon Ultra concentrator to 5 mg/mL. At this stage, rebastinib was added to 1.5 mM concentration, diluted from a 50 mM stock in DMSO, and the protein sample was then concentrated further to 16.4 mg/mL. Crystallisation trials at either 4 °C or 20 °C were set up in various coarse screens using 200 nL sitting drops that mixed protein and precipitant at ratios of 2:1, 1:1 or 1:2 (Table 5.2). A number of crystals grew as rods or as clusters of plates, but when screened no crystals yielded X-ray diffraction.

Further fine screens were set up based around the initial crystallisation conditions. CDK16 was prepared as described above with 1.5 mM rebastinib, but used at a final concentration of 13 mg/mL. Fine screens included "HIN-D10-E10", a matrix around Hampton Index screen conditions from wells D10 (20% PEG5000MME, 0.1 M bis-tris pH 6.5) and E10 (45% polypropylene glycol 400, 0.1 M bis-tris pH 6.5) (Appendix A2) and "JCSG-D1" (0.1 M HEPES pH 7, 25% PEG1000, 20% glycerol) (Appendix A3). In addition, a further coarse screen was set up to search more widely for diffracting crystals. This was the Basic ChemSpace (BCS) screen, which employs the use of PEG smears, which are mixtures of different PEGs with a requirement of either neutral or cooperatively positive effects of each component on crystal growth [16]. Crystals with diffraction between 2.5-3.2Å were obtained.

A cluster of crystals appeared in the BCS coarse screen plate after 1 day at 20°C in a sitting drop that mixed protein to precipitant at a ratio of 2:1 and a reservoir containing a 25 % PEG “medium” smear (PEG 2000,PEG 3350, PEG 4000, PEG 5000MME) and 0.1 M citrate pH 5.5 (Fig. 5.3). No crystals formed in the same condition using the alternative 1:1 or 1:2 ratios of protein to precipitant. After 9 days, the crystals had grown sufficiently for mounting. Some 25% ethylene glycol was added to the drop and several fragments of these plates were broken off and mounted. Screening at the Diamond Light Source I02 beamline showed strong diffraction.

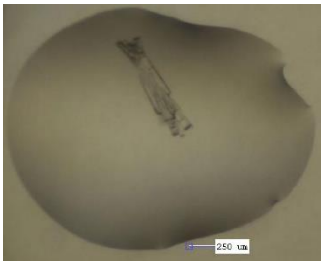


Figure 5.3. Crystals of CDK16$^{163\text{-}478}$ with rebastinib. The crystals grew at 20°C in a 200 nL sitting drop using a 2:1 ratio of protein to a precipitant consisting of a 25 % PEG “medium” smear (PEG 2000,PEG 3350, PEG 4000, PEG 5kmme) and 0.1 M citrate pH 5.5. This image shown was taken on day 7. Fragment of the crystal were subsequently mounted on day 9.

5.2.4 Structure determination of CDK16 with rebastinib

Diffraction data for CDK16 were collected on one crystal at 100 K on Diamond Light Source beamline I02. Data were indexed and integrated using XDS [17] then merged using AIMLESS [18] in the CCP4 suite of programs [19]. Indexing assigned the space group *P* 2 2 2_1. Initial molecular replacement using Phaser and a model of the previous CDK16 structure (PDB ID: 3MTL) generated only a poorly fitting model to the electron density. Further molecular replacement attempts using smaller CDK16 subdomains to allow for more structural changes were similarly unsuccessful and did

not yield a clear solution. Therefore, other space group possibilities from indexing were explored to find the correct space group assignment.

The diffraction data were finally integrated and indexed successfully using MOSFLM [20], which indicated a *P* 2 2 2 space group, and then scaled with AIMLESS [18]. Initial phases were calculated in PHASER using molecular replacement with apo-CDK16 as a template [21]. Program settings were adjusted to enable a search in all alternative space groups in order of general occurrence in the PDB database. A solution was found in the space group $P2_12\ 2_1$ and then forwarded for re-integration and further processing with AIMLESS [18].

After obtaining the initial electron density maps by rigid body refinement in REFMAC5 [22], a structural model was generated by automated building in Buccaneer [23]. Several rounds of NCS (non-crystallographic symmetry) restrained refinement were then performed in REFMAC5 [22] combined with manual modelling with COOT [24]. Processing statistics are provided in Table 5.3.

The electron density maps for the structure were generally of high quality. Two CDK16 molecules were found in the asymmetric unit. CDK16 chain A was modelled between residues Met162-Ala474, with the exception of regions of poor electron density between Ser312-Lys316. CDK16 chain B was modelled between residues Met162-Ser475, except for residues Lys311-Val322. Both protein chains showed clear density in the ATP-binding pocket for the inhibitor, rebastinib. A PDB file and cif dictionary for rebastinib (DCC-2036) were generated in PRODRG using the rebastinib chemical molecule found in the ABL kinase complex structure (PDB 3QRJ) [11]. An additional attempt to improve the rebastinib cif dictionary using ACEDRG showed no difference with the PROGDRG cif file. The refined structure was validated with MolProbity [25].

Table 5.2. Diffraction data collection and refinement statistics (molecular replacement)

	CDK16$^{163-478}$
PDB accession code	5G6V
Chain IDs	Chains A and B
Data collection	
Beamline	Diamond Light Source, I02
Wavelength (Å)	0.91742
Space group	*P* 2_1 2 2_1
Cell dimensions	
a, *b*, *c* (Å)	*57.19, 86.94, 146.25*
α, β, γ (°)	90.0°, 90.0°, 90.0°
No. unique reflections	37,175 (3157)
Resolution (Å)	28.98-2.20 (2.27-2.20)*
R_{merge}	0.093 (0.454)
CC(1/2)	0.996 (0.818)
$I / \sigma I$	9.4 (2.8)
Completeness (%)	98.3 (97.4)
Redundancy	5.2 (4.9)
Refinement	
Ligand	Rebastinib (DCC-2036; PDB ID is 919)
Resolution (Å)	86.94-2.20 (2.257-2.20)
No. reflections	35328
R_{work} / R_{free}	20.77/27.12
Average B factor (Å^2)	40.433
R.m.s. deviations	
Bond lengths (Å)	0.0089
Bond angles (°)	1.3439
Molprobity	
Ramachandran favoured (%)	96
Ramachandran allowed (%)	4
Crystallization conditions	25% PEG medium smear (PEG 2000,PEG 3350, PEG 4000, PEG 5000MME) and 0.1 M citrate pH 5.5

*Single crystal. Values in parentheses are for highest-resolution shell. R.m.s. indicates root-mean-square.

5.2.5 Overview of the co-crystal structure of CDK16 with rebastinib

The CDK16 structure with rebastinib retains the classical protein kinase fold with structural features of an inactive CDK (Fig. 5.5). The kinase N-lobe is composed of a five-stranded β-sheet and an ordered αC helix that contains the diverged cyclin binding sequence motif "PCTAIRE". In the largely alpha helical C-lobe the activation loop could not be fully modelled in either chain (Fig. 5.4), presumably owing to its flexibility in the absence of a cyclin. However, it is still clear that the activation loop is in an inactive DFG-out conformation that would occlude the ATP-pocket.

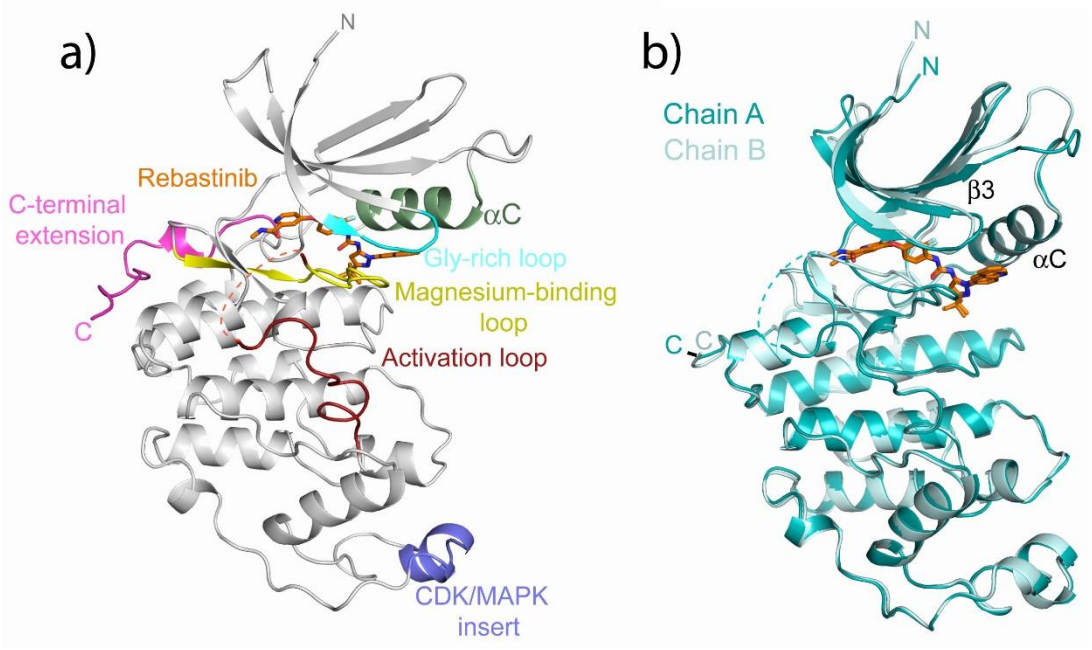


Figure 5.4. Overview of the crystal structure of CDK16 in complex with rebastinib. **a)** Ribbon representation of CDK16 chain A with bound rebastinib (orange). CDK16 is shown grey with additional coloured regions to highlight different structural features. **b)** Superposition of CDK16 chains A and B from the asymmetric unit. The β3 strand and αC helix are labelled for reference. A dashed line represents regions that were not modelled due to poor or missing electron density.

Despite the absence of a cyclin binding partner, the αC helix shows an "in" conformation, likely stabilised instead by the inhibitor binding (Fig. 5.5a). In contrast, the

helical structure was severely disrupted in the previous structure of CDK16 in complex with the type I inhibitor, indirubin E804 (PDB ID: 3MTL; Fig. 5.5a).

The two CDK16 monomers in the asymmetric unit pack head to tail with crystal contacts involving both the αC helix and the β3-αC loop. Perhaps as a result, the flexible loop between β3 and αC is seen in two different conformations in the two CDK16 chains (Fig.5.4b).

While the C-lobe conformation is well-conserved between the two CDK16 crystal structures, the N-lobe in the rebastinib complex is shifted slightly relative to the indirubin E804 complex (Fig. 5.5a). This change likely reflects accommodation of the extended type II inhibitor molecule, whereas the glycine-rich loop can better sit atop the indirubin E804.The final C-terminal region of CDK16 extends behind the αI helix, but then folds back towards the front side of the kinase domain where it finishes beside the αD helix.

The CDK16 construct that was crystallised in these two structures had a phosphomimetic substitution S319D in a potential activation loop phosphorylation site to mimic the classical CDK2 activation site (pThr160). Asp319 was not modeled in the indirubin E804 complex, but could be modelled in chain A of the rebastinib co-structure. In this DFG-out conformation of the activation loop, Asp319 is seemingly stabilised by a weak hydrogen bond with the Lys311 backbone (Fig. 5.5b). However, the density was weak for Lys311 and it was not possible to build the side chain.

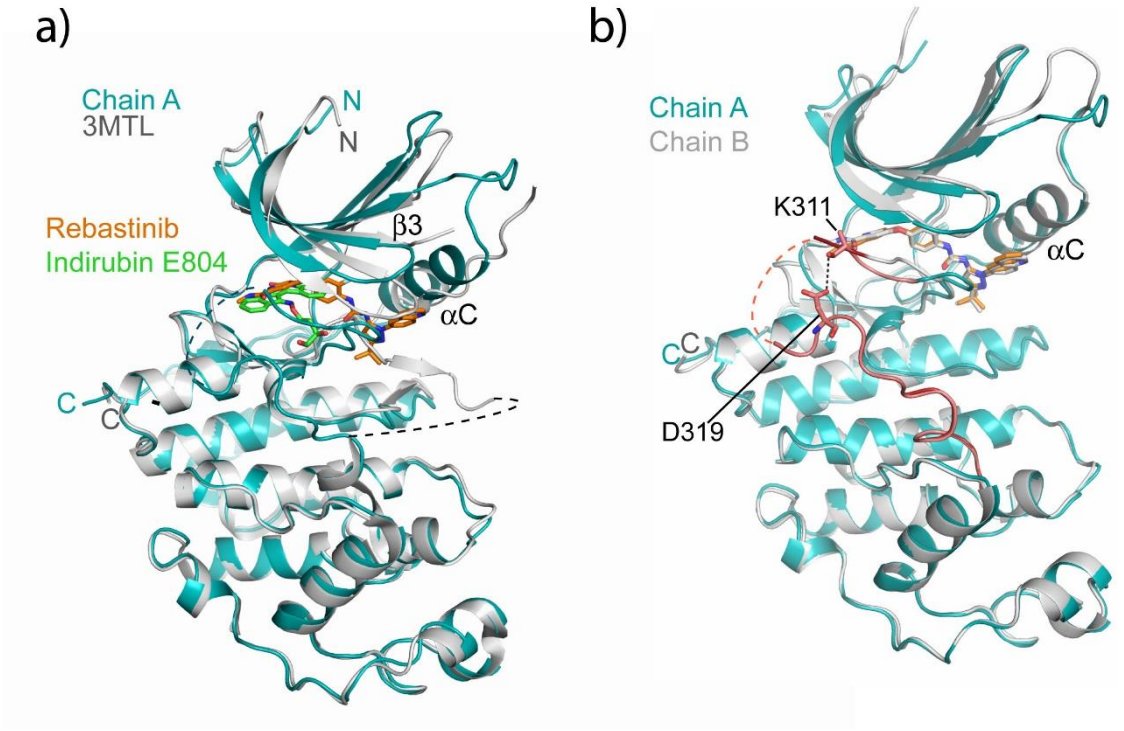


Figure 5.5. Structural comparison of CDK16 complexes with rebastinib and indirubin E804. a) Superposition of CDK16 chain A (teal) with rebastinib bound (orange) and CDK16 (grey) with indirubin E804 (green) bound (PDB ID 3MTL). **b)** Superposition of chains A (teal) and B (grey) of the CDK16 structure with the inhibitor, rebastinib (orange). In chain A, part of the activation loop (orange) could be built which was not visible in chain B, or the corresponding indirubin E804 complex. This included the site of the phosphomimetic S319D substitution. Gaps in the chains are modelled as dashed lines.

5.2.6 Rebastinib binding in the CDK16 active site

As anticipated, the binding of rebastinib is facilitated by an inverted conformation of the "DFG motif", with the Phe305 flipped out to create an extended binding pocket below the αC helix in which the quinoline moiety sits (Fig. 5.6a). Rebastinib is a type II inhibitor that was designed to target wild-type ABL and the gatekeeper mutant ABL^{T315I} [11] (Fig. 5.6b).

Rebastinib forms five hydrogen bonds in the ATP-binding pocket of CDK16 (Fig. 5.6a), as similarly observed in the equivalent ABL structure (Fig. 5.6b). The carboxamide-substituted pyridine ring forms two hydrogen bonds with the backbone amide and carbonyl

of the ATP-hinge residue, Leu243. The urea moiety's nitrogens form two hydrogen bonds with Glu211 from the αC "PCTAIRE" motif, which also participates in a salt bridge with catalytic lysine (Lys194). A further hydrogen bond is formed between the backbone amide of Asp304 and the urea carbonyl (Fig. 5.6a).

The t-butyl group occupies the hydrophobic pocket vacated by the inverted Phe305 from the "DFG" motif (Fig. 5.6a). Interestingly, Phe305 in chain B adopts a slightly different conformation, although this remains the "out" conformation, whereas other key residues are observed in the same orientation (Fig. 5.6c). The central phenyl ring is sandwiched between Phe305 and Phe240, which provides π-π packing interactions. This optimal arrangement is enabled by the phenyl's fluorine substituent, which is positioned *ortho*- to the urea moiety to provide a stereoelectronic bias for a planar conformation of the fluoro-phenyl ring and the urea function [11] (Fig. 5.6a).

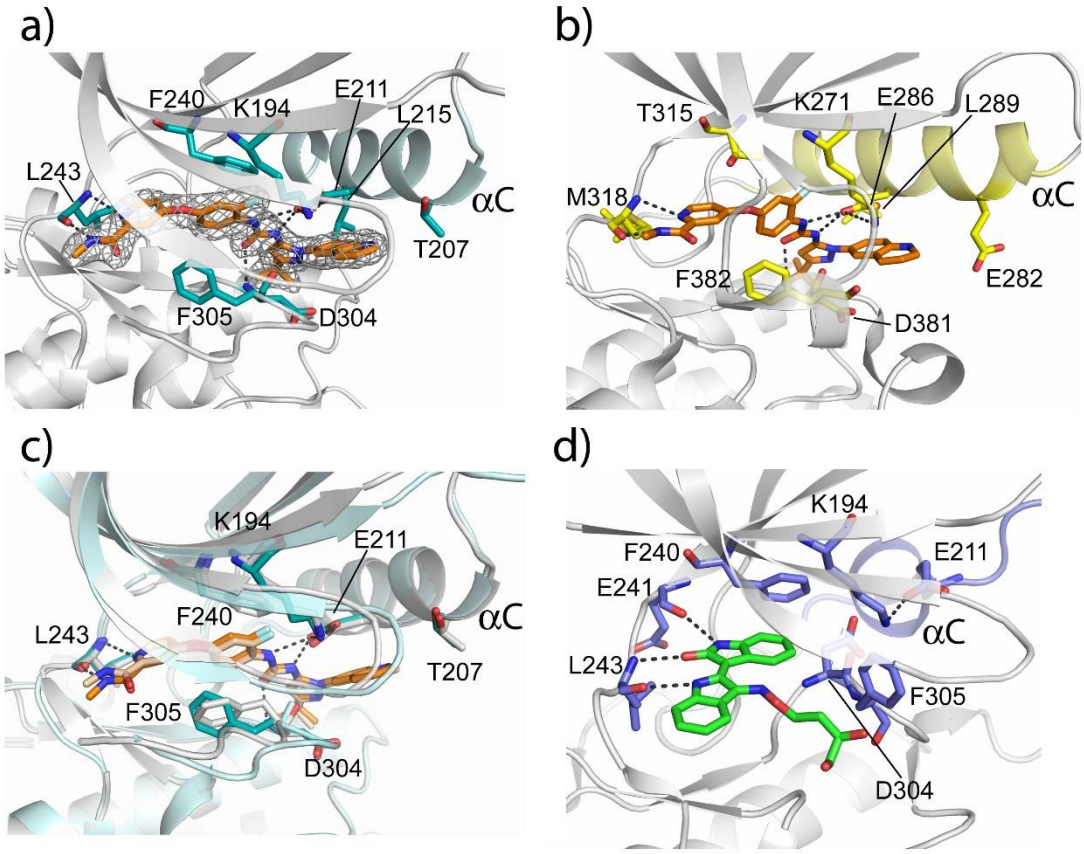


Figure 5.6. Comparison of inhibitor binding modes in CDK16 and ABL. a) Interactions of rebastinib (orange) with CDK16 (chain A shown grey; selected side chains are coloured cyan). A 2*Fo*-2*Fc* electron density map (grey mesh) is shown for the bound inhibitor contoured at 1 σ. **b)** Interactions of rebastinib with wild-type ABL showing key residues in yellow (PDB ID: 3QRI) **c)** Superposition of CDK16 chains A (teal) and B (grey). Phe305 in the two chains adopts slightly different conformations. However, rebastinib binds in the same position. Hydrogen bonding is shown as dashed black lines. **d)** Interactions of Indirubin E804 (green) with CDK16 showing key residues in blue (PDB ID: 3MTL).

Rebastinib retains binding to ABL gatekeeper mutants, such as ABL^{T315I}, by avoiding closer contact with this position [11]. As such, rebastinib is also able to bind to CDK16 in which the gatekeeper residue is Phe240. Indeed, as noted above, the inhibitor sits atop the central phenyl ring, provided favourable hydrophobic contacts.

Further comparison of the binding pockets of CDK16 and ABL reveals a number of other side chain differences that could be useful for the future modification of rebastinib to more selectively target CDK16. For example, the binding of the rebastinib quinoline

moiety in ABL is capped by Glu282 at the N-terminus of the αC (Fig. 5.6b). By comparison, CDK16 has a small Thr207 side chain at this position that would allow extension of the quinolone (Fig. 5.6b). In addition, the urea moiety in rebastinib is packed against Met290 in ABL, which corresponds to Leu215 in CDK16.

The extended binding mode of rebastinib is quite distinct from that of the type I inhibitor, indirubin E804 (Fig. 5.6d). While this inhibitor is expected to bind to a "DFG-in" conformation, its complex with CDK16 shows an atypical "DFG-out" conformation. However, the inhibitor does not form any interactions with the pocket vacated by the Phe305 side chain. This atypical binding may be explained by the rather poorly structured N-lobe in this complex, including a disrupted αC, which provides significant freedom for the activation loop to form a cage-like structure around the bound indirubin E804 molecule (Fig. 5.6d). The inhibitor does, however, make canonical interactions with the kinase hinge region including two hydrogen bonds with Leu243 and a third with the main-chain carbonyl of Glu241 (Fig. 5.6d).

As a consequence of the "DFG-out" conformation observed in these complexes, both CDK16 structures display a disrupted alignment of the hydrophobic spine, which is a conserved feature in active kinases (Fig. 5.7)[26]. As noted above, in the indirubin E804 complex, this likely reflects the absence of a bound cyclin, whereas in the rebastinib complex the disruption is forced by the type II inhibition mode.

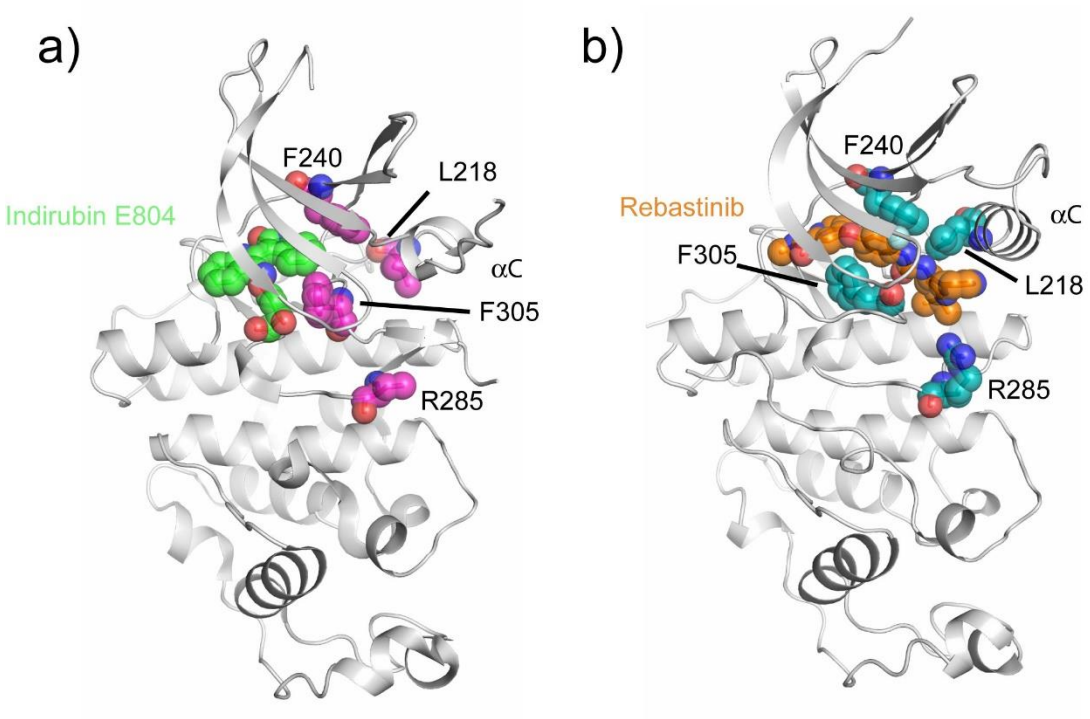


Figure 5.7. The activating hydrophobic spine is not aligned in the CDK16 crystal structures. a) CDK16 with bound indirubin E804 adopts an inactive conformation as shown by disruption of the hydrophobic spine. There is disorder around the αC helix and Leu218 is rotated away from the active site. The position of Arg285 in the catalytic loop "HRD" motif is shown for reference. **b)** CDK16 in complex with rebastinib similarly adopts an inactive conformation with disruption of the hydrophobic spine. In this complex, rebastinib binds across the expected position of the hydrophobic spine and sterically interferes with its formation. Indicated residues are shown in space-fill representation.

5.2.7 Determination of IC_{50} values for CDK16 inhibition by rebastinib and dabrafenib

During my investigation of CDK16, the group of Professor Kei Sakamoto at the Nestlé Institute of Health Sciences (Lausanne) established protocols to reconstitute an active CDK16 complex for use in an *in vitro* radiometric kinase assay [27]. In their system, bacterially-expressed full-length human CDK16 is mixed with CycY, a proposed cognate cyclin, which is expressed in COS-1 cells and co-purified with 14-3-3 (further CDK16 characterisation is detailed in Chapter 6). I was invited to visit the laboratory of Professor Kei Sakamoto in Lausanne to use this assay to establish IC_{50} values for rebastinib and dabrafenib. Once there, I discussed and designed the experiments with Saif Shehata. However, due to the constraints of the health and safety policy within the institute, I was

unable to undertake the practical work as a visiting scientist. Therefore, these radiometric kinase assays were performed instead by Saif Shehata.

The inhibitors rebastinib, dabrafenib and indirubin E804 were tested for their ability to inhibit CDK16/CycY activity at a range of inhibitor concentrations in duplicate, and the assays repeated three times. The purified CDK16/CycY and inhibitor mixtures, were each assayed for phosphotransferase activity in a final assay volume of 50 μL containing 50 mM HEPES, pH 7.5, 0.1 mM EGTA, 10 mM magnesium acetate, 0.1 mM [γ-^{32}P]ATP (300 c.p.m./pmol), 0.03% Brij 35, 1 mM DTT and 50 μM PCTAIRE-tide (PKSPKARKKL). PCTAIRE-tide is a peptide which satisfies the substrate-specificity requirements of CDK16, as determined by Shehata *et al*. [28]. CDK16 and the compound were incubated for 5 minutes prior to the addition of the CycY (no difference was observed using a longer 10 minute pre-incubation). Reactions were incubated at 30^{0}C for 30 minutes and terminated by spotting onto P81 paper and immersion in 75 mM H_3PO_4. P81 filters were washed three times for 10 min with H_3PO_4, rinsed with acetone, air-dried and incorporation of γ-^{32}P determined by Cherenkov counting.

Rebastinib and dabrafenib showed comparable IC_{50} values of 32 nM and 34 nM, respectively (Fig.5.8). By comparison, indirubin E804 gave weaker inhibition with an obtained an IC_{50} value of 84 nM. Overall these data are in good agreement with the DSF assay results, which yielded the same rank order of potencies (Fig. 5.3.)

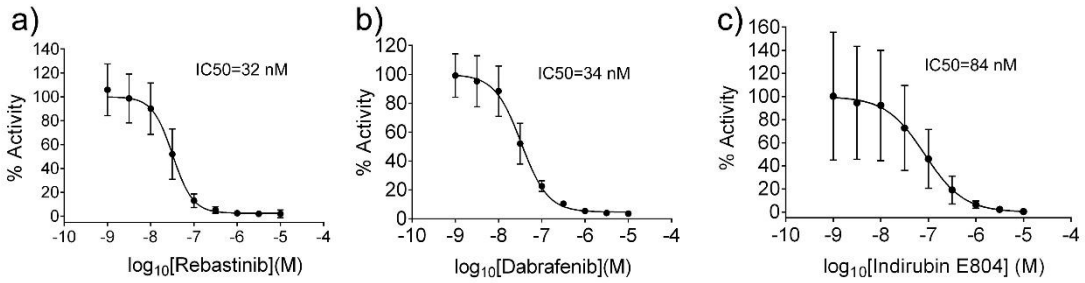


Figure 5.8. Radiometric *in vitro* kinase assays showing inhibition of full-length CDK16/CycY by rebastinib, dabrafenib and indirubin E804. Experiments were performed by Saif Shehata (Nestlé Institute of Health Sciences (Lausanne).

5.2.8 Preliminary investigation of the effect of rebastinib and dabrafenib on cancer cell lines

Recent studies have suggested a potential contribution of CDK16 to the growth of cancers cells, including prostate, breast, cervical cancer, and melanoma cell lines [6-8]. In particular, CDK16 can phosphorylate the cell cycle inhibitor p27 (encoded by the CDKN1B gene) on Ser10 and promote its degradation in cancer cells [6-8]. A high-throughput siRNA screen identified CDK16 amongst those genes for which silencing selectively impaired the proliferation of c-Myc-overexpressing medulloblastoma cell lines. Further, CDK16 was shown to be upregulated in a number of medulloblastoma cell lines and Indirubin E804 was used as a pharmacological inhibitor for CDK16, alongside siRNA [5]. CDK16 was similarly identified as being upregulated in melanoma cancer cells relative to normal skin tissue [6].

Several groups were identified in the University of Oxford that were working specifically on these cancers and were able to provide cell lines to me for inhibitor testing. Professor Nicola Sibson (Oncology Department) studies brain cancer metastasis and kindly provided some DAOY medulloblastoma cells, derived from desmoplastic cerebellar medulloblastoma.

Professor Colin Goding's group (Ludwig Institute for Cancer Research) specialises in melanoma research and kindly provided four malignant melanoma derived cell lines, SK-Mel-28, 501-Mel, IGR-37 and IGR-39, where IGR-37 is a metastasis from the same patient as IGR-39 [29]. IGR-37, IGR-39 and SK-Mel-28 have $BRAF^{V600E}$ mutations whereas 501-Mel does not [30-33].

I therefore began preliminary investigations into the effect of rebastinib and dabrafenib on each of these cancer cells lines. Dabrafenib is a potent inhibitor of the $BRAF^{V600E}$ mutation that is associated with melanoma and therefore formed a positive control for these cell lines. These preliminary studies where set up to complement the published siRNA results and to elucidate if CDK16 could be a relevant target for a larger investigations, including the development of a highly selective chemical probe for CDK16.

5.2.8.1 CDK16 is expressed in the various cancer cell lines

The five cancer cell lines were grown in plastic flasks as adherent cultures in culture media conditions as detailed in chapter 2. To confirm the presence of CDK16 and CycY in each of the available cancer cell lines, I analysed their protein expression by Western blotting using specific antibodies for these proteins as well as p27. Western blotting of β-tubulin was also used as an internal control for total protein expression levels. CDK16 appeared on the blots as a ladder of 2-3 bands, consistent with the vendor datasheet (HPA001366 antibody documentation, Sigma) which labels the lower MW band as CDK16 (Fig. 5.9). The higher MW bands were detected by the vendor with this antibody in a number of human cell lines including RT-4 and EFO-21. CDK16 was present in each cell line and was qualitatively most highly expressed in IGR-39 (Fig. 5.9). CycY was only very weakly detectable in SK-Mel-28, 501-Mel and DAOY cells, while p27 conversely was

only weakly detected in IGR-37 and IGR-39 (Fig.5.9). This is an interesting observation as it supports a potential role for the CDK16/CycY complex in the downregulation of p27.

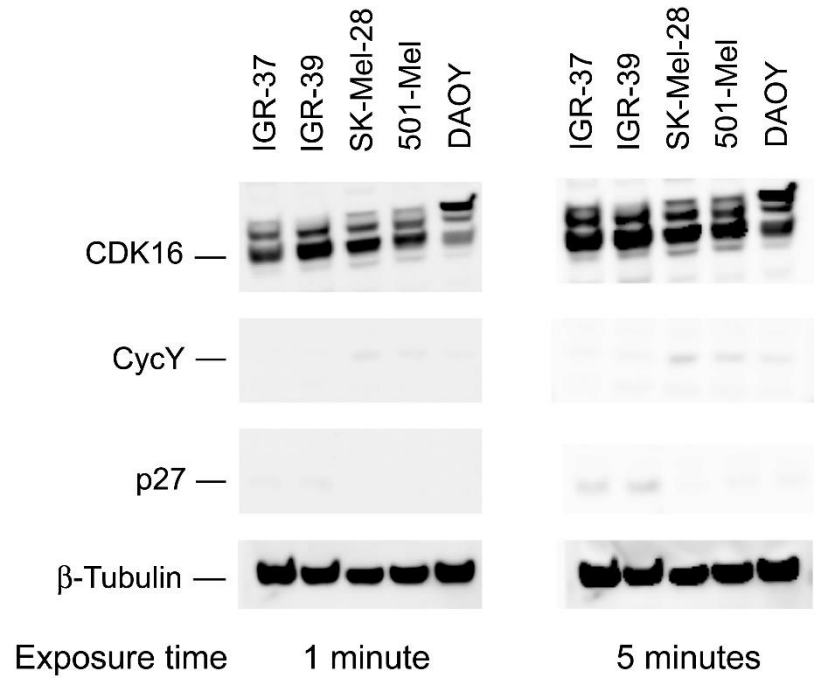


Figure 5.9. Expression levels of CDK16, CycY and p27 across cancer cell lines. Protein expression levels were determined by Western blotting after 72 hours. β-tubulin was used as a control. Images were taken after 1 minute and 5 minutes exposure.

Quantitative polymerase chain reaction (qPCR) was also employed to specifically analyse the mRNA levels of CDK16, CycY and p27 in the untreated cancer cell lines. The levels of CDK16, CycY and p27 in each cell line were normalised relative to their levels in the SK-Mel-28 (Fig. 5.10a) or DAOY (Fig. 5.10b) cell lines. The mRNA level of CDK16 varied across the different cell lines, but was generally higher in the melanoma cell lines than the DAOY cells. The highest expression was found in the IGR-39 cells, which correlates well with the protein level indicated by the Western blot. CycY mRNA expression was also highest in IGR-39, but lowest in the IGR-37 cell line. The 501-Mel cell line showed the highest level of p27 mRNA expression (Fig. 5.10), despite the lack of

detectable protein in this cell line (Fig. 5.9). The difference in mRNA and protein abundance further suggests the posttranslational regulation of p27.

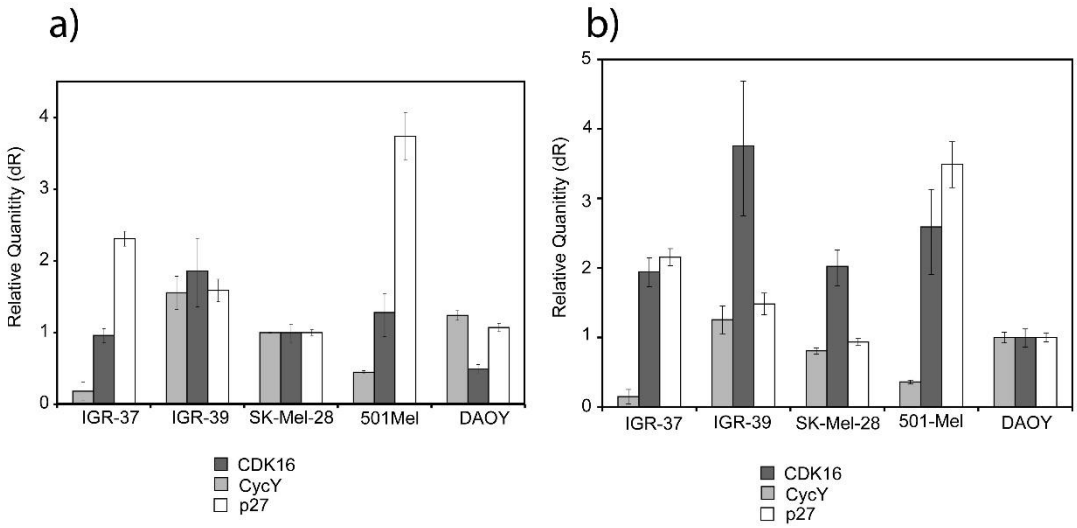


Fig 5.10. Comparison of CDK16, CycY and p27 mRNA expression levels across cancer cell lines a) mRNA expression levels for each gene were determined by qPCR and are shown relative to the corresponding mRNA levels in SK-Mel-28 cells, which were normalised to 1.0. **b)** mRNA expression levels for each gene are shown instead relative to the levels in DAOY, which were normalised to 1.0.

5.2.8.2 Cancer cell viability following treatment with rebastinib or dabrafenib

To test the effect of the compounds on each cell line, cell viability was measured following 72 hours in the presence of rebastinib or dabrafenib using the CellTiter-Glo® Luminescent cell-viability assay (Promega). It was predicted that the melanoma cell lines with the $BRAF^{V600E}$ mutation (SK-Mel-28, IGR-37 and IGR-39) would be sensitive to dabrafenib. While this compound forms a useful positive control, it is does not allow the relative effects of inhibiting CDK16 and $BRAF^{V600E}$ to be distinguished. A third compound, PLX-4720, was therefore selected as a further control treatment. PLX-4720 is

a similarly potent $BRAF^{V600E}$ inhibitor, but does not bind to CDK16 according to published binding assays by Ambit Biosciences [34]. Standard treatments for inhibition of medulloblastoma cell lines, such as DAOY, are not available.

Cell viability was determined by quantitation of the ATP present, as an indicator of metabolically active cells, by a coupled reaction in which mono-oxygenation of luciferin is catalysed. Luminescence readout data was represented as a percent of the untreated control set as 100%. EC50 values (μM) were calculated using GraphPad Prism (GraphPad Software, Inc) by the equation Y=Bottom + (Top-Bottom)/(1+10^((LogEC50-X))) for a sigmoidal dose-response curve of the transformed data.

Rebastinib showed good activity to block the proliferation of three of the four melanoma cell lines with EC_{50} values of 1.2-1.8 μM. Indeed, its effects against the SK-MEL-28 cell line (EC_{50} = 1.2 μM) were equipotent to the control compound PLX-4720 (EC_{50} = 1.1 μM) (Figure 5.11. and Table 5.3.). Only the 501-Mel cell line was more resistant, but was still inhibited by rebastinib with EC_{50} = 5.4 μM. By contrast, neither of the BRAF inhibitors showed useful inhibition of this cell line. Overall, these results with rebastinib are encouraging as it does not inhibit BRAF and was the most potent CDK16 inhibitor. Moreover, rebastinib also blocked the proliferation of DAOY cells with an EC_{50} = 1.8 μM.

Surprisingly, the cell viability response curves from the dabrafenib experiments were not suitable for curve fitting. Overall, this inhibitor appeared far less toxic than rebastinib. For example, inspection of the cells under a microscope revealed that all cells had entered apoptosis in the presence of 20 μM rebastinib, whereas significant numbers of cells were still viable in 20 μM dabrafenib (see appendix A4-8). Similarly, the response curves for PLX-4720 could only be fit for the SK-MEL-28 cells. It has been widely noted

that melanoma cells can readily acquire resistance to BRAF inhibition. For example, inhibitors can induce BRAF-RAF dimerization paradoxically inducing the unwanted downstream activation of ERK [35]. Thus, it is possible that these experiments may be limited by the use of laboratory cell lines, rather than primary patient-derived cells, although treatments are clearly desirable for drug-resistant cells too.

Table 5.3. Overview of cancer cell line viability following treatment with rebastinib, dabrafenib or PLX-4720.

Cancer cell line	EC_{50} (µM)		
	Rebastinib	**Dabrafenib**	**PLX-4720**
IGR-37	1.2	>20	>20
IGR-39	1.5	>20	>20
SK-Mel-28	1.2	>20	1.1
501-Mel	5.4	>20	>20
DAOY	1.8	>20	N/A

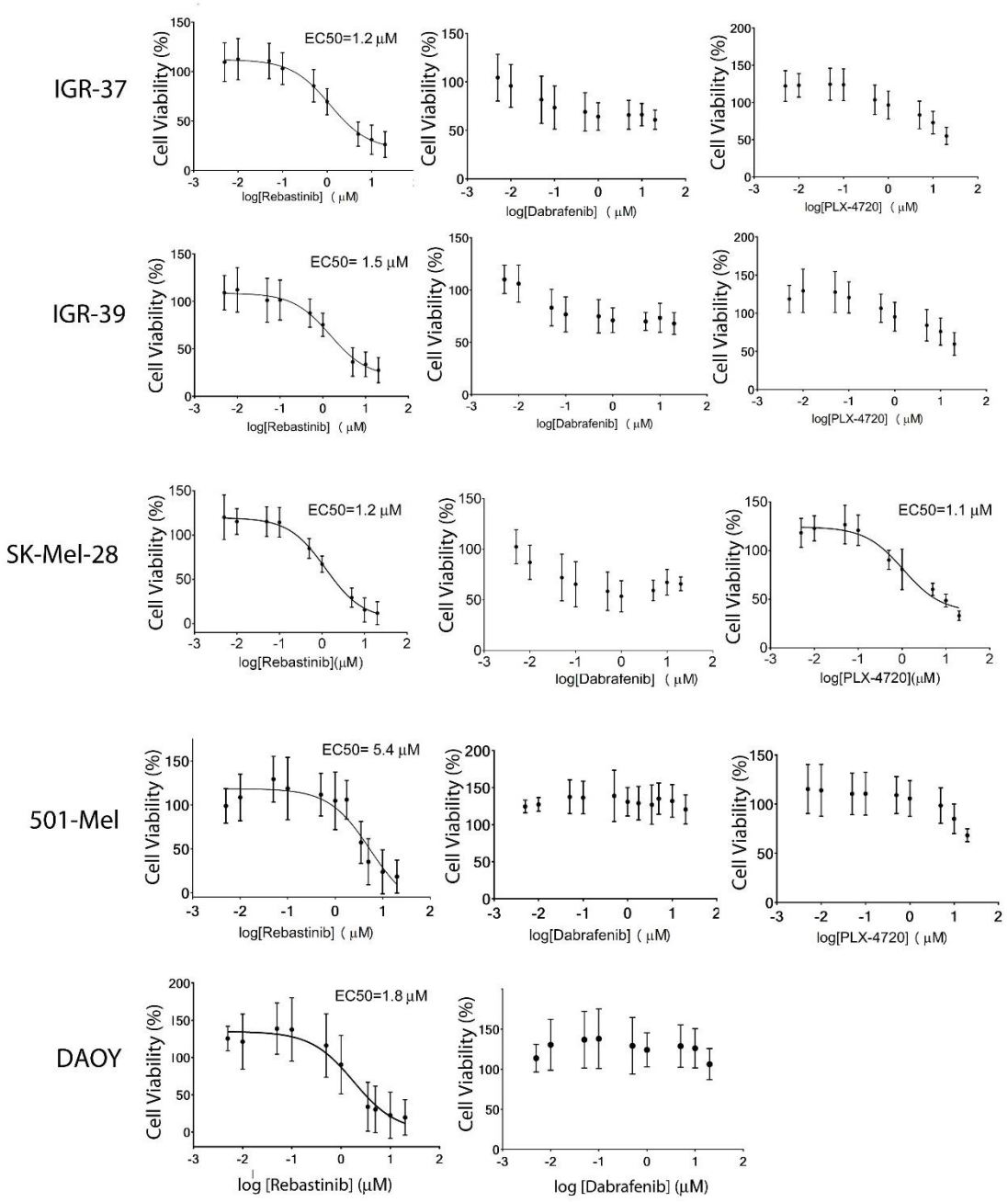


Figure 5.11. Cell viability of IGR-37, IGR-39. SK-Mel-28, 501-Mel and DAOY following treatment with rebastinib, dabrafenib or PLX-4720. Cell viability was determined for cells treated for 72 hours with rebastinib, dabrafenib or PLX-4720, as indicated. Data represent mean value±S.D. of three independent experiments performed in triplicates and normalized to the untreated control, except for SK-Mel-28 which is from one experiment in triplicate. Cell viability was assessed using a luminescence-based readout for ATP (CellTiter Glo) throughout. Data are represented as a percent of the untreated control set as 100%. Concentrations are displayed as molar (M). EC50 values (µM) were calculated using GraphPad Prism (GraphPad Software, Inc).

5.3 Discussion

In this chapter, I have identified rebastinib and dabrafenib as two clinically-relevant kinase inhibitors that can bind potently to CDK16. A co-crystal structure of the CDK16 kinase domain in complex with rebastinib identifies the structural features that determine its type II inhibitor binding mode. Dabrafenib is a type I inhibitor with a distinct chemical scaffold that shows superior potency to indirubin E804. Therefore, it would also be of interest to determine the co-crystal structure of dabrafenib to provide further valuable insights into how an ATP-competitive inhibitor may target CDK16 with such potency.

In the absence of selective chemical probes, and as the elucidation of the role of CDK16 in normal and pathobiology continues, these compounds may be valuable pharmacological tools for the future. Previous studies in cancer cell lines have used indirubin E804 and SNS-032 to explore CDK16 inhibitor potential [5, 36]. Both rebastinib and dabrafenib have been shown to be more potent inhibitors of CDK16 than these two compounds (Fig. 5.3). Ultimately, a more specific chemical probe for CDK16 is needed. The co-crystal structure of CDK16 and rebastinib will help to guide efforts in structure-based drug design, for example to exploit the ligand properties in favour of CDK16 selectivity over ABL.

In a study suggesting that CDK16 contributes to cancer cell line resistance to apoptosis in prostate cancer cell lines, the compound SNS-032 sensitised the cells to TRAIL-induced apoptosis, which was attributed to CDK16 inhibition [36]. In this situation, comparing results from different CDK16 inhibitors, such as rebastinib, dabrafenib and indirubin E804, may help to dissect whether an effect is common to CDK16 or an off-target

effect. Such approaches may be particularly necessary to discriminate between closely-related CDKs once their inhibitor-binding profiles are established [37].

Specificity is a significant potential challenge in developing a CDK16-specfiic chemical probe due to the conservation within the 'PCTAIRE' and 'PFTAIRE' CDKs, (CDK14-18). Type II inhibitors, such as rebastinib, target the monomeric inactive state, which is generally more structurally diverse, although both dabrafenib and rebastinib are multi-targeted kinase inhibitors with broad specificity. Kinase assay data for both compounds are collated in the Harvard Medical School LINCS database (library of integrated network-based cellular signatures) [38]. A KiNativ assay in this collection showed that 1 μM rebastinib gave 94% inhibition of CDK14, but none against wild-type BRAF. Thus, rebastinib may be a useful tool for crystallisation of CDK14 and for dissecting potential of CDK16 and BRAF. CDK16 was not included in this study. An alternative KinomeSCAN assay using an inhibitor concentration of 10 μM revealed that dabrafenib was a pan-CDK inhibitor, including CDK14-18.

An *in vitro* enzymatic kinase assay performed by collaborators at the Nestlé Institute of Health Sciences (Lausanne) established IC_{50} values for CDK16 inhibition by rebastinib and dabrafenib that were in the low nanomolar range. Based on existing gene knockdown results for CDK16, both compounds were subsequently tested in various cancer cell lines as a preliminary screen to establish if these cells were also sensitive to loss of the catalytic activity of CDK16. For instance, phenotypes from transient chemical inhibition can often differ to those of gene deletion or knockdown. Ideally, in future work with more available time these complementary studies should be performed in parallel. My preliminary cellular work confirmed that CDK16 was expressed in four tested melanoma cell lines as well as the DAOY medulloblastoma cell line. The EC50 values for rebastinib were in the low micromolar range across each cancer cell line and were notably more

efficacious than those using dabrafenib. The might suggest that CDK16 inhibition is not particularly synergistic with BRAF inhibitor, or simply reflect an acquired resistance mechanism. Notably, the effects of rebastinib in the melanoma cell line do support the targeting of alternative kinase signalling pathways as a strategy to combat potential resistance to BRAF inhibitors. Repurposing a clinically validated drug is an attractive, efficient prospect.

Whilst melanoma and medulloblastoma cancer cell lines were used in these preliminary investigations, there is available evidence to suggest that the CDK16 gene expression levels in these cells are only mid-range by comparison with other cancers (Fig. 6.12). Determining additional efficacy data to compare the effects of rebastinib in a low and a high expressing CDK16 cancer cell line may help further to highly the relevance of CDK16. Additionally, such studies could be guided by analyses of the genetic status of CDK16 (Fig. 6.13).

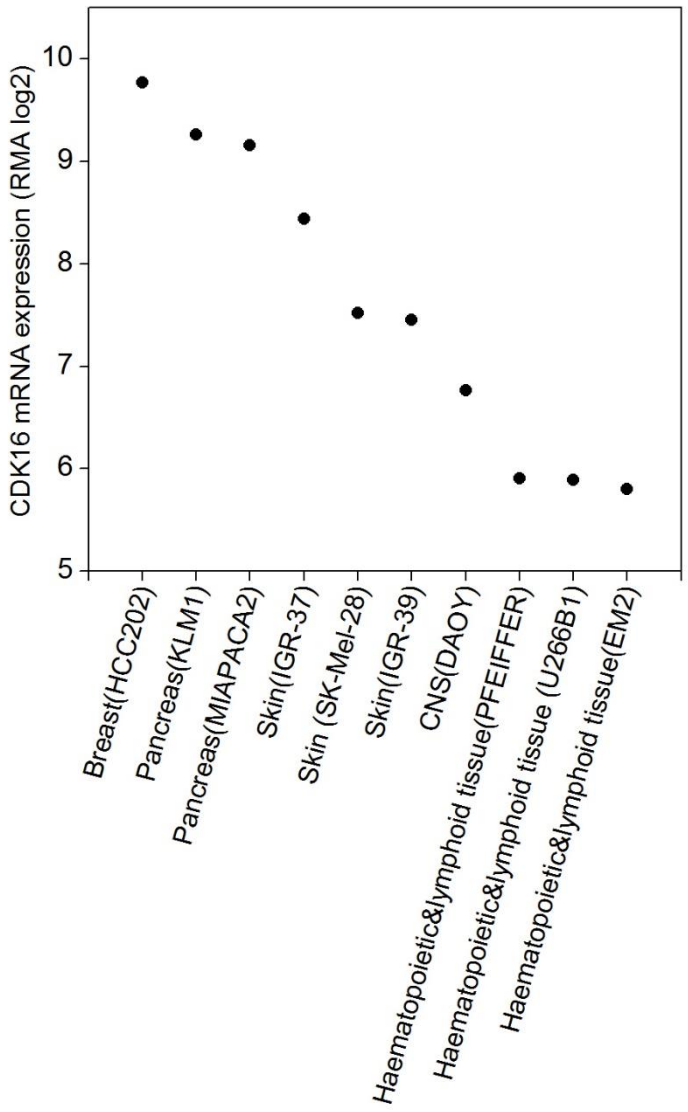


Figure 6.12. Gene expression data for CDK16 in various cancer cell lines. Data were extracted from the CCLE gene expression database. Gene Expression data for CDK16 extracted from CCLE_Expression_Entrez_2012-10-18.

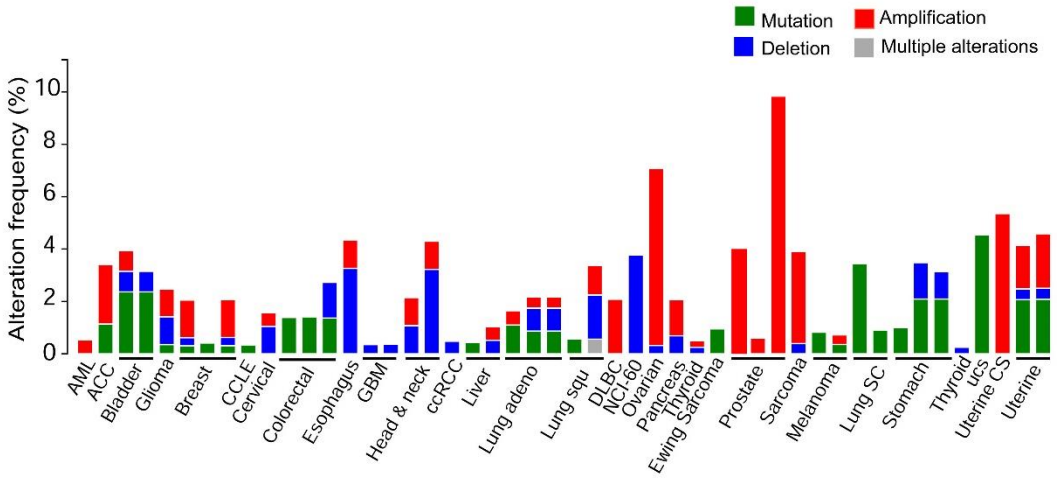


Figure 5.1. The frequencies of genetic changes of CDK16 are summarized across multiple cancer types. For each of the indicated cancer types, the frequencies of mutation (green), amplification (red) and homozygous deletion (dark blue) were determined using genetic data from >2,000 cancer cases obtained through cBioPortal for Cancer Genomics. Detailed information about each case and disease is accessible through the cBioPortal for Cancer Genomics. ACC, adrenocortical carcinoma;; AML, acute myeloid leukaemia; CCLE, Cancer Cell Line Encyclopaedia; CS, carcinosarcoma; GBM *Glioblastoma* multiforme; MBL, medulloblastoma; MM, multiple myeloma; NCI60, US National Cancer Institute (NCI) 60 human tumour cell line anticancer drug screen; SC, serous cystadenocarcinoma; squ, squamous [39, 40].

5.4 References

1. Malumbres, M., et al., *Cyclin-dependent kinases: a family portrait.* Nat Cell Biol, 2009. 11(11): p. 1275-6.
2. Ropers, H.H., *Genetics of early onset cognitive impairment.* Annu Rev Genomics Hum Genet, 2010. 11: p. 161-87.
3. Hu, H., et al., *X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes.* Mol Psychiatry, 2016. 21(1): p. 133-48.

4. Mikolcevic, P., et al., *Cyclin-dependent kinase 16/PCTAIRE kinase 1 is activated by cyclin Y and is essential for spermatogenesis.* Mol Cell Biol, 2012. 32(4): p. 868-79.

5. Cwiek, P., et al., *RNA interference screening identifies a novel role for PCTK1/CDK16 in medulloblastoma with c-Myc amplification.* Oncotarget, 2015. 6(1): p. 116-29.

6. Yanagi, T., J.C. Reed, and S. Matsuzawa, *PCTAIRE1 regulates p27 stability, apoptosis and tumor growth in malignant melanoma.* Oncoscience, 2014. 1(10): p. 624-33.

7. Yanagi, T. and S. Matsuzawa, *PCTAIRE1/PCTK1/CDK16: a new oncotarget?* Cell Cycle, 2015. 14(4): p. 463-4.

8. Yanagi, T., et al., *PCTAIRE1 phosphorylates p27 and regulates mitosis in cancer cells.* Cancer Res, 2014. 74(20): p. 5795-807.

9. Nam, S., et al., *Indirubin derivatives inhibit Stat3 signaling and induce apoptosis in human cancer cells.* Proc Natl Acad Sci U S A, 2005. 102(17): p. 5998-6003.

10. Anastassiadis, T., et al., *Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity.* Nat Biotechnol, 2011. 29(11): p. 1039-45.

11. Chan, W.W., et al., *Conformational control inhibition of the BCR-ABL1 tyrosine kinase, including the gatekeeper T315I mutant, by the switch-control inhibitor DCC-2036.* Cancer Cell, 2011. 19(4): p. 556-68.

12. Laquerre, S., et al., *Abstract B88: A selective Raf kinase inhibitor induces cell death and tumor regression of human cancer cell lines encoding B-RafV600E mutation.* Molecular Cancer Therapeutics, 2009. 8(12 Supplement): p. B88.

13. Conroy, A., et al., *SNS-032 is a potent and selective CDK 2, 7 and 9 inhibitor that drives target modulation in patient samples.* Cancer Chemother Pharmacol, 2009. 64(4): p. 723-32.

14. Ali, M.A., et al., *SNS-032 prevents tumor cell-induced angiogenesis by inhibiting vascular endothelial growth factor.* Neoplasia, 2007. 9(5): p. 370-81.

15. Brasca, M.G., et al., *Identification of N,1,4,4-tetramethyl-8-{[4-(4-methylpiperazin-1-yl)phenyl]amino}-4,5-dihydro-1H-py razolo[4,3-h]quinazoline-3-carboxamide (PHA-848125), a potent, orally available cyclin dependent kinase inhibitor.* J Med Chem, 2009. 52(16): p. 5152-63.

16. Chaikuad, A., S. Knapp, and F. von Delft, *Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens.* Acta Crystallogr D Biol Crystallogr, 2015. 71(Pt 8): p. 1627-39.

17. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. 66(Pt 2): p. 125-32.

18. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. 69(Pt 7): p. 1204-14.

19. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallogr D Biol Crystallogr, 2011. 67(Pt 4): p. 235-42.

20. Leslie, A.G.W. and H.R. Powell, *Processing diffraction data with mosflm*

Evolving Methods for Macromolecular Crystallography, R.J. Read and J.L. Sussman, Editors. 2007, Springer Netherlands. p. 41-51.

21. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. 40(Pt 4): p. 658-674.

22. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures.* Acta Crystallogr D Biol Crystallogr, 2011. 67(Pt 4): p. 355-67.

23. Cowtan, K., *Fitting molecular fragments into electron density.* Acta Crystallogr D Biol Crystallogr, 2008. 64(Pt 1): p. 83-9.

24. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. 66(Pt 4): p. 486-501.

25. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2010. 66(Pt 1): p. 12-21.

26. Taylor, S.S. and A.P. Kornev, *Protein kinases: evolution of dynamic regulatory proteins.* Trends Biochem Sci, 2011. 36(2): p. 65-77.

27. Shehata, S.N., et al., *Cyclin Y phosphorylation- and 14-3-3-binding-dependent activation of PCTAIRE-1/CDK16.* Biochem J, 2015. 469(3): p. 409-20.

28. Shehata, S.N., et al., *Analysis of substrate specificity and cyclin Y binding of PCTAIRE-1 kinase.* Cell Signal, 2012. 24(11): p. 2085-94.

29. Halling-Brown, M.D., et al., *canSAR: an integrated cancer public translational research and drug discovery resource.* Nucleic Acids Research, 2012. 40(Database issue): p. D947-D956.

30. van Rikxoort, M., et al., *Oncolytic Effects of a Novel Influenza A Virus Expressing Interleukin-15 from the NS Reading Frame.* PLoS ONE, 2012. 7(5): p. e36506.

31. Hao, H., et al., *Context-dependent roles of mutant B-Raf signaling in melanoma and colorectal carcinoma cell growth.* Mol Cancer Ther, 2007. 6(8): p. 2220-9.

32. Kido, K., et al., *Simultaneous suppression of MITF and BRAF V600E enhanced inhibition of melanoma cell proliferation.* Cancer Sci, 2009. 100(10): p. 1863-9.

33. McDermott, U., et al., *Identification of genotype-correlated sensitivity to selective kinase inhibitors by using high-throughput tumor cell line profiling.* Proc Natl Acad Sci U S A, 2007. 104(50): p. 19936-41.

34. Davis, M.I., et al., *Comprehensive analysis of kinase inhibitor selectivity.* Nat Biotechnol, 2011. 29(11): p. 1046-51.

35. Cichowski, K. and P.A. Janne, *Drug discovery: inhibitors that activate.* Nature, 2010. 464(7287): p. 358-9.

36. Yanagi, T., et al., *PCTAIRE1-knockdown sensitizes cancer cells to TNF family cytokines.* PLoS One, 2015. 10(3): p. e0119404.

37. Echalier, A., et al., *An integrated chemical biology approach provides insight into Cdk2 functional redundancy and inhibitor sensitivity.* Chem Biol, 2012. 19(8): p. 1028-40.

38. Patricelli, M.P., et al., *Functional interrogation of the kinome using nucleotide acyl phosphates.* Biochemistry, 2007. 46(2): p. 350-8.

39. Cerami, E., et al., *The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.* Cancer Discov, 2012. 2(5): p. 401-4.

40. Gao, J., et al., *Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.* Sci Signal, 2013. 6(269): p. pl1.

Chapter 6 Characterisation of CDK16 interactions with cyclins

6.1 Introduction

CDK14-18 display high sequence conservation across their kinase domains, but have large N- and C-terminal extensions (Fig. 6.1). Little has been described about their regulation, but they do contain a diverged cyclin binding motif in their αC helix comprising 'PFTAIRE' for CDK14-15 and 'PCTAIRE' for CDK16-18. Monomeric CDK16 appears inactive, as evidenced by its crystal structure (PDB ID: 3MTL), and is suggested to require a cyclin similarly to other classical CDKs for activity. The precise cyclin complement for CDK14-18 is unknown, although a number of potential cyclin binding partners have been suggested; for CDK14 CycY and CycD3, for CDK16 CycY, CycI. CycK, CycY-L1 and p35 and for CDK18 CycK [1-7].The novel cyclin, CycY, has emerged as the lead candidate cyclin partner in recent work, although it appears to be rather promiscuous with binding to both CDK14 and CDK16 [2-5]. CycY has only poor sequence conservation with classical cyclins and contains only a single cyclin box fold, whereas most cyclins contain two [8]. However, CycY also has a large N-terminal region of unknown function. CDK14 has also been found to bind to CycD3 in a yeast-two hybrid assay, while CDK16 failed to bind CycD1 in a pull-down assay [9].

In this chapter I tried to establish the cyclin specificity of CDK16, examining CycY and an oncogenic D-type cyclin, Kaposi's sarcoma-associated herpes virus (KSHV) viral K-cyclin (K-cyclin). Viral K-cyclin's properties overlap and extend beyond that of the D-type cyclins. K-cyclin was shown to form an active complex with CDK6 and showed enhanced kinase activity, compared to CDK6/CycD complexes, as well as broader substrate specificity [10].

I also tried to characterize the minimal domain requirements for binding between CDK16 and these cyclins with the view to determining the structures of these novel complexes as well as their mechanisms of kinase activation. For example, once established, purified CDK-cyclin complexes can enable a range of activity assays and inhibitor screens that can lead to a better understanding of CDK16 functions in signal transduction and disease.

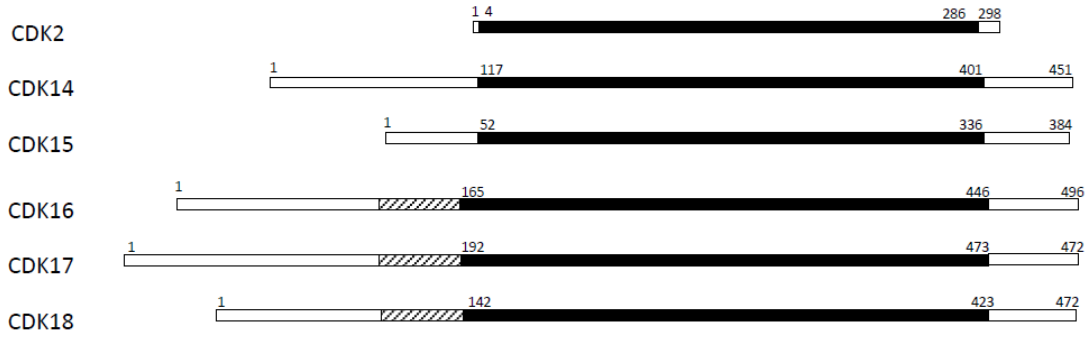


Figure 6.1. Schematic diagram showing an alignment of CDK14-18 in comparison to CDK2. Domain boundaries are marked as described by PFAM. Black bars represent the conserved kinase domain; patterned bars in CDK16-18 represents a region conserved amongst the 'PCTAIRE' kinases and white bars represent the more diverse N- and C-terminal extensions.

6.2 Results

6.2.1 Establishment of protocols for protein production

To investigate CDK/cyclin regulation, I sought to establish expression and purification protocols for a range of CDK and cyclin truncation constructs, and then to investigate their interaction as a precursor to structural studies and activity assays. In light of the CDK16 structure with indirubin bound solved in an inactive state, my initial focus was on examining CDK16 and its binding partners (PDB ID: 3MTL). This included a novel cyclin, CycY, and as well as the oncogenic D-type viral cyclin, K-cyclin, for which a

soluble construct was already available in the SGC (K-cyclin$^{1-257}$, a kind gift from Professor Ernest Laue, University of Cambridge).

Having verified existing protocols for bacterial expression of CDK16$^{163-478}$ (PDB ID: 3MTL) (Chapter 5) and K-cyclin (PDB ID: 1G3N), I investigated larger CDK16 fragments and paralogues. These constructs had been shown previously to be insolubly expressed in *E. coli*, so attempts were made to express multi-domain constructs of CDK16 and CycY in a baculoviral expression system using Sf9 insect cells.

6.2.1.1 Purification of structurally determined targets: CDK16 kinase domain and K-cyclin

Purification protocols for the minimal kinase domain of CDK16$^{163-478}$ with a phosphomimetic S319D mutation were described in chapter 5. This construct consistently gave a high yield of soluble protein, which was stable throughout purification (up to 46 mg/L) (appendix Table A2). Additional constructs that had been prepared at the SGC prior to my arrival are also shown in appendix Table A2.

K-cyclin$^{1-257}$ has an N-terminal hexahistidine tag with a cleavage site for enterokinase. The construct gave variable expression levels (0.9-2.7mg/L) and purification was problematic, often suffering losses due to precipitation at each stage (standard expression and purification protocols are provided in chapter 2).

6.2.1.2 Small scale expression and purification of CDK16 and CycY

Initial 3 mL test expressions of various CDK16 and CycY constructs were performed in insect Sf9 cells. Nickel-affinity purification showed relatively successful recovery of several multi-domain constructs of CDK16-18 and CycY that contained C-terminal hexahistidine and Flag tags, which were cloned and expressed in the SGC prior to my arrival (see appendix A9). Full-length CDK16 (CDK16$^{1-496}$) also showed good soluble expression, whereas full-length CycY failed to yield soluble protein.

Scale up of the expression and purification of the solubly expressed CycY constructs proved challenging, as detailed below. Therefore, additional CycY constructs were designed with N-terminal hexahistidine tags and different domain boundaries. These constructs were then cloned by staff at the SGC and provided to me. As detailed in appendix A9, at the 3 mL scale, similar numbers of constructs for each target were expressed and purified successfully. CDK14-18 and various other cyclins were included amongst this set, but here I focus on my work with CDK16.

6.2.1.3 Longer CDK16 and CycY constructs suffer from degradation

Upon subsequent large scale expression in *Sf9* cells, it became apparent that some longer CDK16 and CycY constructs suffered from degradation to varying degrees, as shown in Fig. 6.2 and summarised in appendix A10. For example, whereas CDK16$^{163\text{-}478}$ was stably expressed in *E. coli* to high yields, CDK16$^{1\text{-}496}$ from *Sf9* cells was recovered as multiple degraded products (Fig. 6.2). A co-expression trial of CDK16$^{1\text{-}496}$/CycY$^{1\text{-}331}$ in insect cells showed that both partners still suffered from similar degradation (Fig. 6.2). Degradation was slightly higher upon baculoviral expression for 72 hours compared to 48 hours, and could be minimised further by a rapid and immediate purification after expression (data not shown). All protein samples were supplemented with protease inhibitor cocktail set III (Calbiochem). Unfortunately, no soluble protein could be recovered for full length CycY.

N-terminal hexahistidine-tagged CDK16$^{109-496}$ showed good soluble expression and did not degrade to the same degree as the full-length protein. However, CycY$^{24-331}$ suffered from degradation similarly to the C-terminally tagged CycY$^{1-331}$ (Figure 6.2).

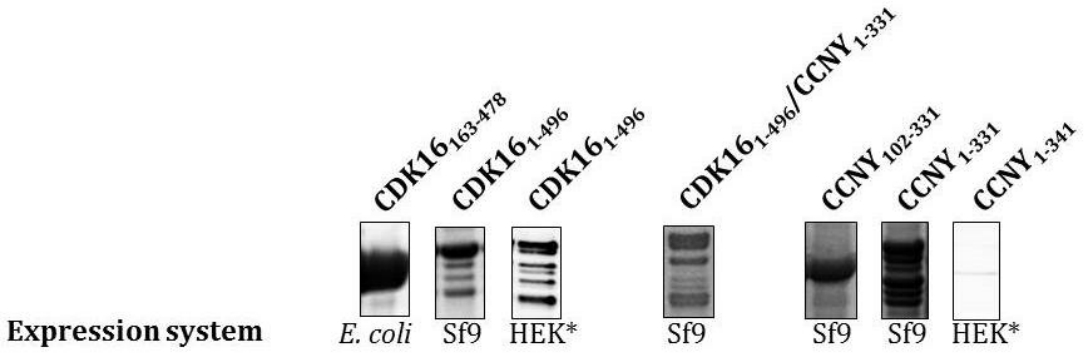


Figure 6.2. Upon scale up longer constructs of CDK16 and CycY suffered from degradation. Full-length CDK16 constructs suffered from degradation in insect (Sf9 cells) and mammalian cells (HEK293 cells), in contrast to CDK16$^{163-478}$ (kinase domain) construct expressed in *E. coli* cells. CycY$^{1-331}$ constructs suffered badly from degradation in insect cell expression. Co-expression failed to prevent CDK16 and CycY degradation in insect cells. Full-length CycY expressed poorly in mammalian cells (HEK293 cells) in preliminary investigations and had previously failed to express in insect cells. Figure shows Coomassie stained SDS PAGE gels of the eluted proteins from Nickel-affinity purification except those marked with an asterisk (*), which are from a Western blot using an antibody to the hexahistidine tag.

6.2.1.4 Expression trials in the HEK293 expression system

It has been shown that CDK16 and CycY protein when co-expressed in a mammalian expression system, such as HEK293, can bind and have activity against a suitable substrate [5, 11]. An initial test expression using the Bacmam viral expression system gave very low CycY$^{1-341}$ expression, whether alone or after co-expression with CDK16. In contrast, CDK16$^{1-496}$ expressed at a reasonable level in this system, but again showed a high degree of degradation (Figure 6). Co-expression of two viruses (which were highly concentrated to obtain the current yields) gave undesirable toxicity (data not shown).

6.2.1.5 Cloning and expression of viral K-cyclin with a Z-basic tag

Expression of viral K-cyclin$^{1-257}$ with an N-terminal hexahistidine tag and enterokinase cleavage site gave modest expression levels (0.9-2.7 mg/L) and the purified protein was prone to heavy precipitation. I therefore used PCR to amplify the viral K-cyclin DNA sequence and inserted this into the pNIC-Zb vector using ligation-independent cloning (LIC cloning). The pNIC-Zb vector provides an N-terminal hexahistidine tag followed by an additional Z-basic domain tag. The Z-basic domain is 54-amino acid sequence derived from protein A, which has been modified to have a high positive surface charge to achieve efficient recovery in anion exchange chromatography [12, 13]. A TEV cleavage site is included after the tags. This provided a more robust viral K-cyclin construct for purification to achieve better yields [12, 13]. This protein was used in subsequent crystallisation trials (Table 6.1) and in radiometric kinase activity assays (section 6.2.4.3).

6.2.2 CDK14, CDK16 and CycY are phosphorylated after baculoviral expression

Reversible phosphorylation of CDKs is an important mechanism of regulation, which may lead to kinase activation or inhibition. The interaction between CDK16 and CycY has been shown to be phosphorylation dependent [5]. To explore this post-translational modification, CDK16$^{1-496}$, CycY$^{102-331}$ and CycY$^{9-331}$ constructs were expressed in *Sf9* cells and purified by Ni-affinity purification. Evidence for phosphorylation was identified from intact mass spectrum analysis (Fig. 6.3). Phosphorylation mapping was undertaken by Rod Chalk at the SGC. Proteins were cleaved with trypsin and potential phosphopeptides were analysed by LC-MSMS. Three phosphorylation sites were identified in CDK16$^{1-496}$ and mapped to Ser78, Ser82 and Ser110, all outside the kinase domain. Ser78 and Ser82 had been identified in large-

proteomic analyses, but their influence upon kinase activity has not been examined [14, 15]. Phosphorylation of Ser110 has also been reported previously and the S110A mutant has been shown to have reduced kinase activity toward the generic substrate myelin basic protein [9, 16]. CycY$^{102-331}$ phosphorylation sites were mapped to Ser324 and Ser326. CycY$^{24-331}$ was also found to be phosphorylated at Ser324 and Ser74. All of these sites had been identified previously in large-scale proteomic studies, but not investigated for their functional effect [17, 18].

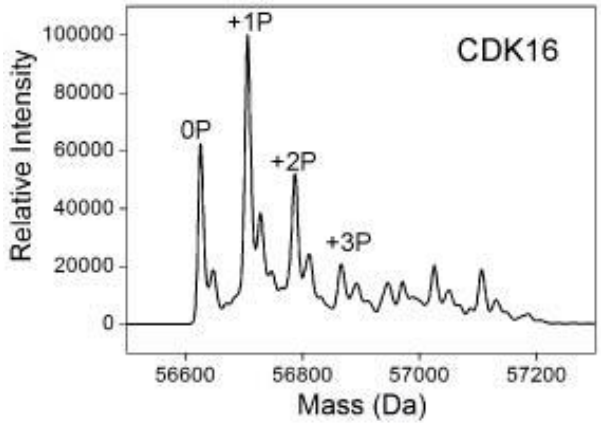


Figure 6.3. Representative intact mass spectrum of CDK16 purified from insect cells. At least three sites of phosphorylation were identified in full-length CDK16 by the intact mass spectrum. These sites were mapped by Rod Chalk (SGC, University of Oxford) to the N-terminal region at residues Ser78, Ser82 and Ser110.

6.2.3 Characterization of CDK/Cyclin binding interactions

The initial aim of this chapter was to characterise a potential new family of CDK/cyclin complexes. Therefore, different binding assays were explored to demonstrate theses interactions. For these studies my efforts initially focused on the bacterial expression constructs of CDK16$^{163-478}$ and K-cyclin$^{1-257}$ and then expanded to include full-length CDK16$^{1-496}$ as well as several CycY constructs.

For each complex, the CDK and cyclin partner were mixed and explored for their co-purification upon size-exclusion chromatography and ion-exchange chromatography. Binding was further probed using pull-down assays and isothermal titration calorimetry (ITC).

6.2.3.1 Viral K-cyclin$^{1-257}$ binds to CDK16

A nickel-affinity pull-down was performed using the N-terminal hexahistidine tagged K-cyclin as the 'bait' protein and a TEV-cleaved construct of wild-type CDK16$^{163-478}$ as the prey (cloned, expressed and purified similarly to the S319D mutant). Proteins were buffered in binding buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM imidazole)(Fig. 6.4a). Proteins at 1.1 mg/mL concentration were loaded onto a pre-equilibrated nickel-column either alone, as representative of input, or sequentially to test for a pull-down interaction. In the latter case, the cyclin bait was pre-loaded onto the nickel-sepharose beads and washed extensively before addition of the CDK16. Washes contained binding buffer supplemented with 30 mM imidazole. The remaining bound proteins were then eluted with a buffer supplemented with 250 mM imidazole. Analysis of the elutions on a SDS-PAGE gel showed the co-purification of K-cyclin$^{1-257}$ and CDK16$^{163-478}$, which represent the minimal kinase and cyclin domain folds (Fig.6.4a).

Co-purification of CDK16$^{163-478}$ and K-cyclin$^{1-257}$ was also demonstrated by size-exclusion chromatography on a HiLoad S200 Superdex 16/60 column (GE Healthcare) buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 0.5 mM TCEP (data not shown). The eluted protein was then passed through an anion exchange step using a 5 mL HiTrap Q HP column (GE Healthcare) buffered in 50 mM HEPES pH 7.5, 50 mM NaCl, 0.5 mM TCEP (Fig. 6.4b). A salt gradient elution was tested from 50 mM NaCl to 500 mM NaCl. CDK16$^{163-478}$ has a calculated pI value of 6.5, whereas K-cyclin$^{1-257}$ has a calculated pI

value of 5.5; the CDK16$^{163-478}$/K-cyclin$^{1-257}$ complex has a calculated pI value of 6.0. Again, the two proteins co-eluted further supporting their likely complex.

Finally, the CDK16$^{163-478}$/K-cyclin$^{1-257}$ complex interaction was analysed by ITC (Fig. 6.4c). The CDK16$^{163-478}$ and K-cyclin$^{1-257}$ proteins were prepared separately using nickel-affinity and size-exclusion chromatography. The two proteins were concentrated in a 10 kDa MWCO Amicon Ultra concentrator to 75 mg/mL and 5 mg/mL, respectively, and dialysed overnight at 4°C in D-tube dialyzer midi tubes (3.5 kDa molecular weight cut-off filter, Millipore) into a buffer containing 50 mM HEPES pH7.5, 150 mM NaCl, 1 mM DTT, 10 mM L-arginine and 10 mM L-glutamine. Samples were de-gassed before use and the concentration measured spectrometrically in triplicate (K-cyclin has a mass of 32,216 Da and an extinction co-efficient of 29,450 M^{-1} cm^{-1}. CDK16 has a mass of 37,252 Da and an extinction coefficient of 35,870 M^{-1} cm^{-1}). The CDK16$^{163-478}$ sample was titrated into K-cyclin$^{1-257}$ or dialysis buffer alone (blank run) at 10°C in a VP-ITC calorimeter (MicroCal). Data were fit to a 1:1 binding model using the Origin software package provided with the instrument (MicroCal). The data showed that K-cyclin$^{1-257}$ bound to CDK16$^{163-478}$ with a sub-micromolar dissociation constant (K_D = 625 nM, Fig. 6.4c). Collectively, these results indicate that the CDK16 kinase domain interacts with K-cyclin.

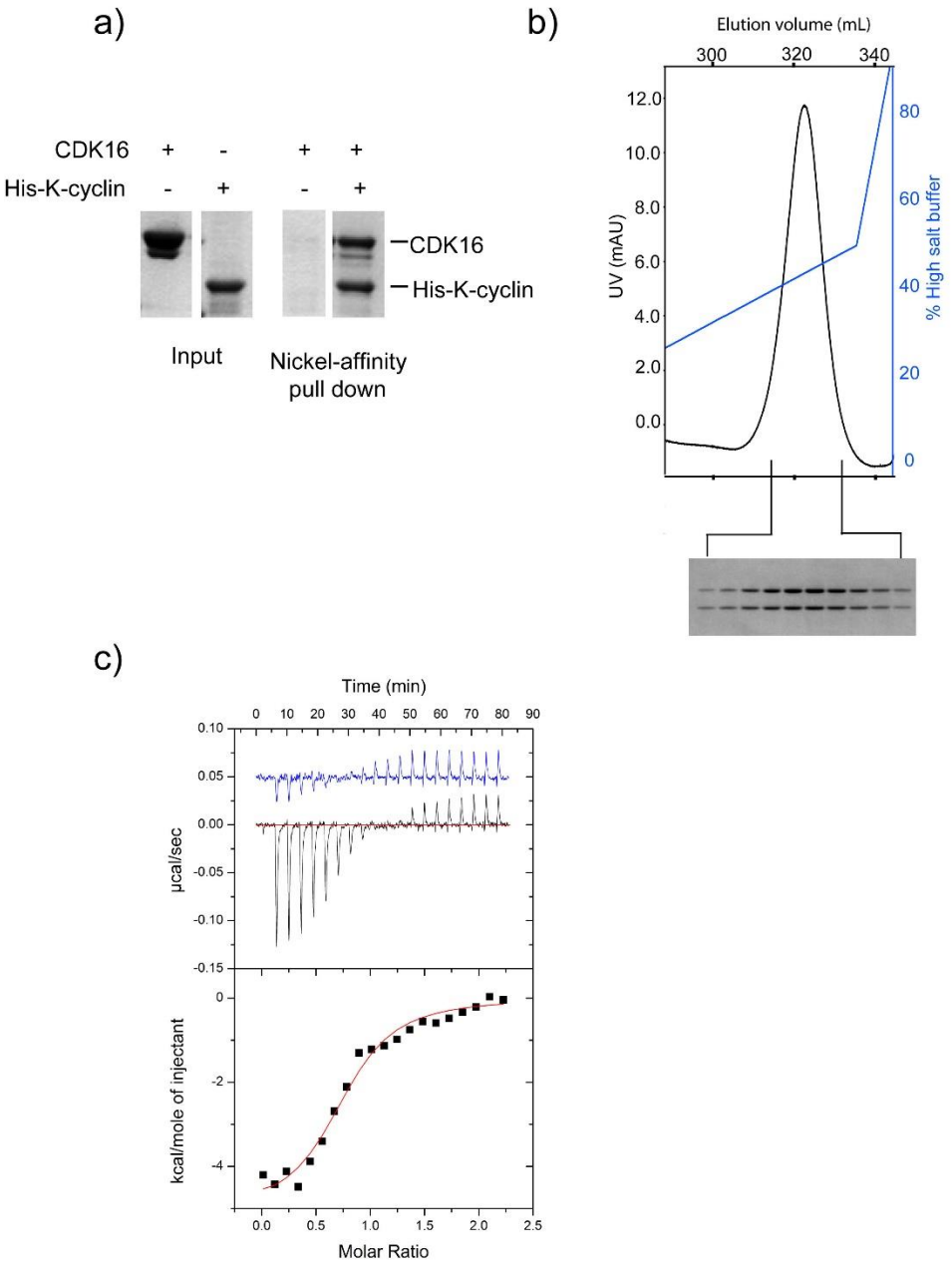


Figure 6.4. K-cyclin binds to the CDK16 kinase domain. a) SDS PAGE gel slices showing representative input and pull down elution fractions with indicated protein samples. Hexahistidine tagged K-cyclin$^{1\text{-}257}$ was the 'bait' and CDK16$^{163\text{-}478}$ was the 'prey'. **b)** Co-purification of CDK16$^{163\text{-}478}$ and K-cyclin$^{1\text{-}257}$ from anion exchange chromatography. Elution is shown monitored by UV signal (top) and by SDS PAGE (lower). **c)** ITC experiment showing binding of K-cyclin$^{1\text{-}257}$ to CDK16$^{163\text{-}478}$. CDK16 (70 µM) was titrated into buffer alone (blank titration, blue) or into K-cyclin$^{1\text{-}257}$ (7 µM). Data from the blank titration were subtracted from the experimental data. Data were fit to a single site 1:1 binding model and gave a K_D value of 625nM. The top panel shows the raw heats recorded during injections, whereas the lower panel shows the corresponding integrated heat signals.

To test if the full-length CDK16 protein had equivalent binding, CDK16$^{1-496}$ was later expressed in *Sf9* insect cells and purified similarly by nickel-affinity and size-exclusion chromatography. An ITC experiment was performed at 10°C titrating CDK16$^{1-496}$ (10 µM) into K-cyclin$^{1-257}$ (100 µM). Unfortunately, insufficient material and time were available for a blank control, but the last injections of the titration appeared equivalent to a baseline (saturated binding) and were used to estimate the heats of dilution for subtraction. The ITC data indicated that the longer CDK16$^{1-496}$ had a higher binding affinity for K-cyclin, with a K_D value of 57 nM (Fig. 6.5).

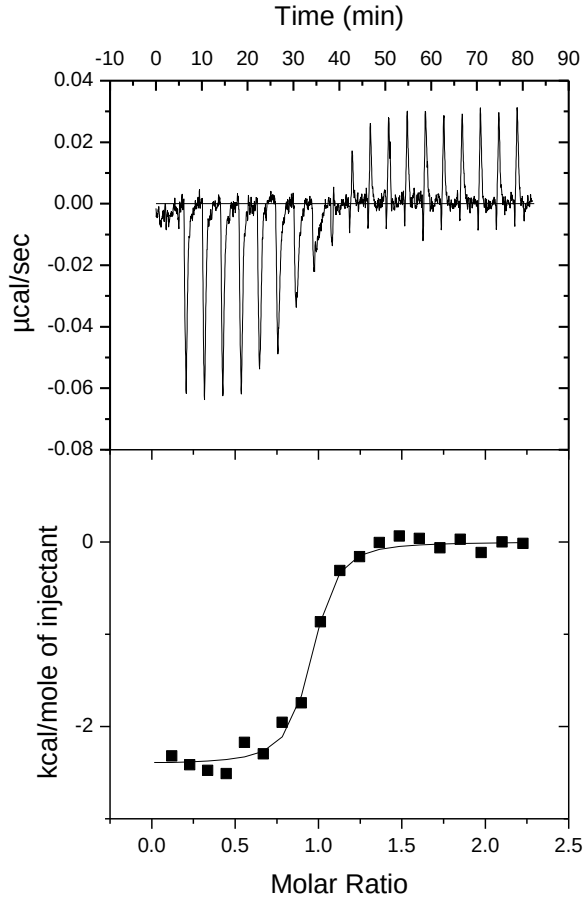


Figure 6.5. Full-length CDK16 displays high affinity for viral K-cyclin. ITC measurements following the titration of CDK16$^{1-496}$ (100µM) into K-cyclin$^{1-257}$ (10µM). Proteins were buffered at 10°C in 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT. A curve fit to a 1:1 binding model indicates a K_D value of 57 nM.

6.2.3.2 CDK16 kinase domain alone does not bind CycY

Initial investigations aimed to determine if CDK16 and CycY could assemble through the minimal kinase and cyclin box domains as observed for other CDK family complexes. An immunoprecipitation was performed using Flag-tagged $CycY^{102-331}$ from baculoviral expression as 'bait' and bacterially-expressed $CDK16^{163-478}$ as 'prey'. SDS PAGE analysis of the input and immunoprecipitated fractions indicated that there was no binding (Fig. 6.6a).

The interaction of these two proteins was then also tested by size-exclusion and anion exchange chromatography as performed previously with the viral K-cyclin (Fig. 6.6b). $CDK16^{163-478}$ has a calculated pI value of 6.5, whereas $CycY^{102-331}$ has a calculated pI value of 5.5; their complex is calculated pI value of 6.2. Again, these truncated constructs failed to co-purify, as clearly demonstrated by anion exchange with the appearance of $CDK16^{163-478}$ in the loading flow-through, whilst $CycY^{102-331}$ was collected in elution fractions (Fig. 6.6b).

An ITC experiment was also performed in which $CDK16^{163-478}$ (55 µM) was titrated into $CycY^{102-331}$ (5.5 µM) or dialysis buffer alone (blank run) (Fig. 6.6c). These binding measurements further indicated that $CycY^{102-331}$ did not bind to $CDK16^{163-478}$ with any detectable affinity, although analysis was complicated by undesirable heats of dilution, potentially indicative of aggregation effects (Fig. 6.6c). Overall, the ITC data were consistent with the lack of co-purification or co-precipitation of $CycY^{102-331}$ and $CDK16^{163-478}$.

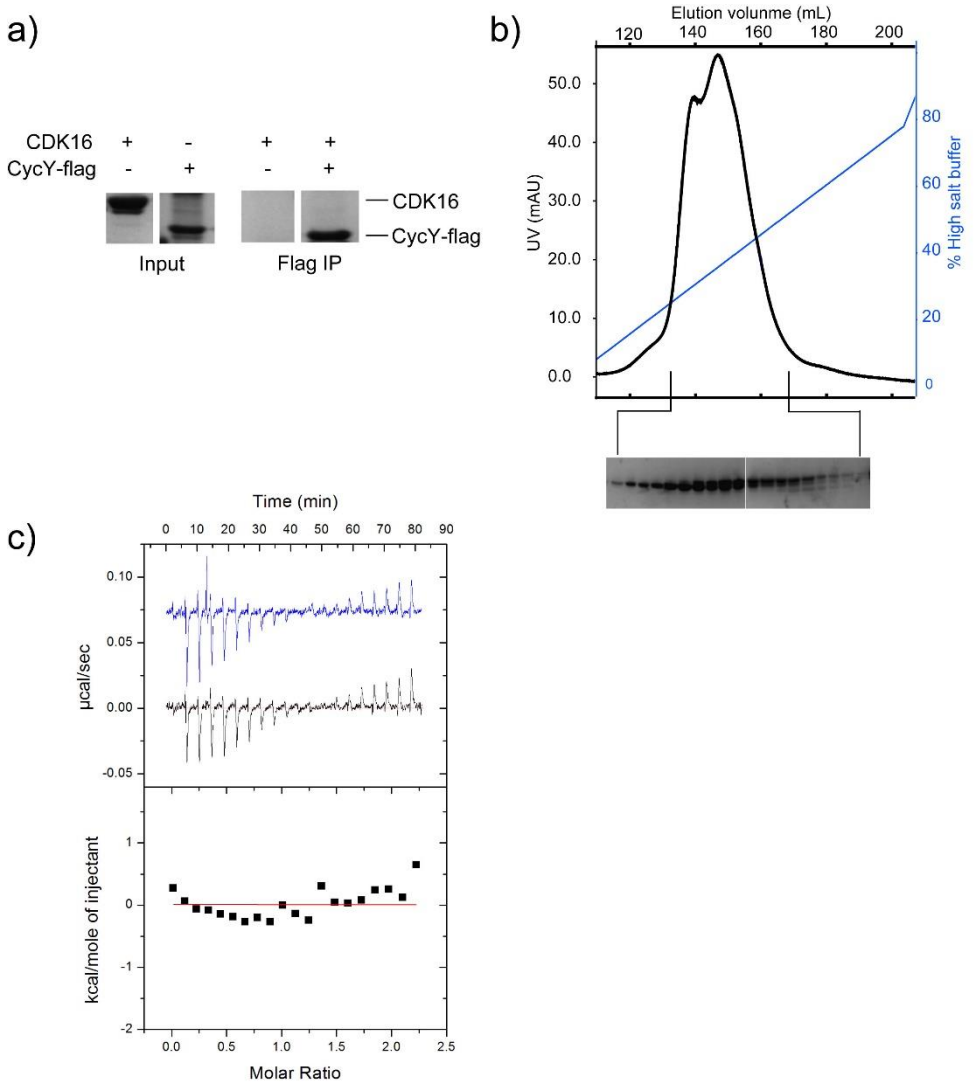


Figure 6.6. CycY$^{102-331}$ does not bind the CDK16 kinase domain. a) Flag-immunoprecipitation did not show interaction between Flag-tagged CycY$^{102-331}$ (the 'bait') and CDK16$^{163-478}$ (the 'prey'). **b)** Anion-exchange chromatography on a HiTrap Q HP column (5 mL) showed separate elution of CDK16$^{163-478}$ and CycY$^{102-331}$. **c)** ITC measurements also failed to detect interaction between CDK16$^{163-478}$ and CycY$^{102-331}$. CDK16$^{163-478}$ (55 µM) was titrated into dialysis buffer (blank titration, blue) or into CycY$^{102-331}$ (5 µM).

6.2.3.3 Multi-domain CDK16 and CycY proteins from baculoviral expression show no interaction

The accumulation of the data above for the truncated CDK16 and CycY constructs, coincided with a publication by Mikocelvic *et al*. indicating that the interaction between

CDK16 and CycY is via a novel mechanism requiring the CDK16 N-terminal extension [5]. To explore this, full-length CDK16 and CycY$^{1-331}$ were separately purified by nickel-affinity chromatography then tested for co-purification by size-exclusion chromatography on a HiLoad S200 Superdex 16/60 column (GE Healthcare) (Fig. 6.7a). The complex eluted at the same elution volume as monomeric CDK16 showing no evidence of complex formation. SDS PAGE showed that CycY$^{1-331}$ also suffered from degradation. Therefore, alternative slightly truncated constructs were also trialled in subsequent assays.

A nickel-affinity pull-down experiment was performed as previously described using N-terminally hexahistidine-tagged CycY$^{24-331}$ as the 'bait' protein and CDK16$^{109-496}$ as the 'prey'. A weak band for CDK16$^{109-496}$ was detected in the pull down using the bait protein suggestive of an apparent association (Fig. 6.7b). However, the control experiment indicated that the prey protein also bound weakly to the nickel-sepharose beads alone, showing that the interaction was non-specific.

An ITC experiment titrating CDK16$^{109-496}$ into CycY$^{24-331}$ also failed to show detectable binding (Fig. 6.7c). Similarly, no binding was observed by ITC using full length CDK16 and CycY$^{1-331}$ (Data not shown).

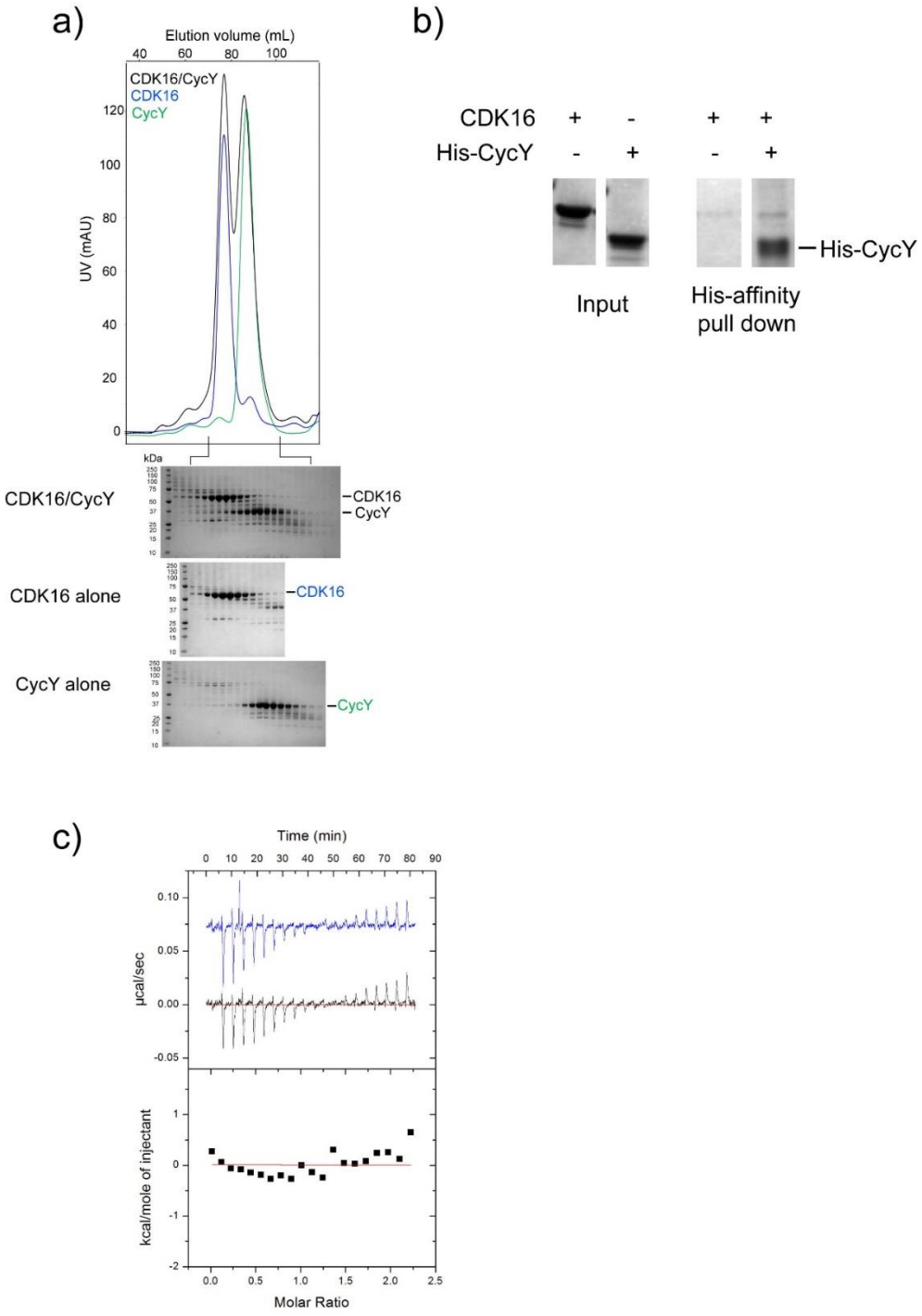


Figure 6.7. CDK16$^{109-496}$ fails to bind CycY $^{9-331/24-331}$. **a)** Size-exclusion chromatography experiments using 50 µM CycY$^{1-331}$ or 55 µM CDK16$^{1-496}$ alone or a mixture of the two proteins. The UV trace and SDS-PAGE gel show indicated eluted fractions. There was no evidence of binding. **b)** Nickel-affinity pull-down using hexahistidine-tagged CycY$^{24-331}$ ^{as} the 'bait' and CDK16$^{109-496}$ as the 'prey'. Weak pull down of CDK16 was observed in both the control and experimental samples suggestive of non-specific binding to the beads. c) ITC measurements showing titration of CDK16$^{109-496}$ (100µM) into dialysis buffer (blank titration, blue) or into CycY$^{9-331}$ (10µM). No binding was detected, but strong heats of dilution were observed, potentially resulting from sample aggregation or non-specific interactions with the instrument.

6.2.4 Kinase activity assays

To test for any kinase activity from the putative CDK/cyclin complexes, an ADP-Glo kinase assay (Promega) was employed, utilizing a range of substrates. This luminescent ADP detection assay measures kinase activity by quantifying the amount of ADP produced during a kinase reaction. Total signal may reflect autophosphorylation as well as substrate phosphorylation. Glycerol stock for a CDK2/CycA complex was provided by Dr Leila Alexander at the SGC as a positive control and tested against a known CDK2 substrate peptide (HHASPRK).

Until recently, the phosphorylation consensus sequence for CDK16 substrates was unknown. Previous studies have therefore used generic substrates, such as histone H1 and myelin basic protein (MBP), which can give inconclusive results due to potential contaminating kinases with low activity [6, 9, 16]. A recent investigation used a positional scanning peptide library to identify a preferred CDK16 substrate peptide consensus sequence SP(K/R)x(K/R/H) (where x is a small aliphatic amino acid) [11]. Following this study, the investigators generated an optimised 'PCTAIREtide' peptide (PKSPKARKKL) and showed that CDK16/CycY activity was 10-fold higher using this substrate compared to a CDK2 substrate peptide [11]. This result also provided good evidence of proper formation of a CDK16/CycY complex using full-length proteins purified from HEK293 cells.

6.2.4.1 CDK16 kinase domain shows no activity with K-cyclin and a range of substrates

The CDK16/K-cyclin complex identified previously offered potential for an alternative active kinase complex that could be expressed with high yield. The ADG-Glo

kinase assay was therefore performed using CDK16$^{163-478}$ and K-cyclin$^{1-257}$ proteins that were expressed separately in *E. coli* and then purified by nickel-affinity size -exclusion chromatography. CDK16$^{163-478}$ with, or without K-cyclin, showed no activity above baseline values against a range of substrates (casein, CDK2 peptide (HHASPRK purchased from GenScript) or PCTAIREtide (purchased from Pepceuticals Limited, UK) and luminescence results were similar in the presence or absence of a CDK16 inhibitor (Indirubin E804) (Fig. 6.8). The apparent interaction of CDK16 and K-cyclin, but lack of activity raised the possibility that this unusual cyclin might impart a distinct substrate preference. However, protein complexes submitted to Dr Scott Gerber (Dartmouth) also revealed no activity against a random peptide library (unpublished results) using a highly sensitive mass spectrometric method [19].

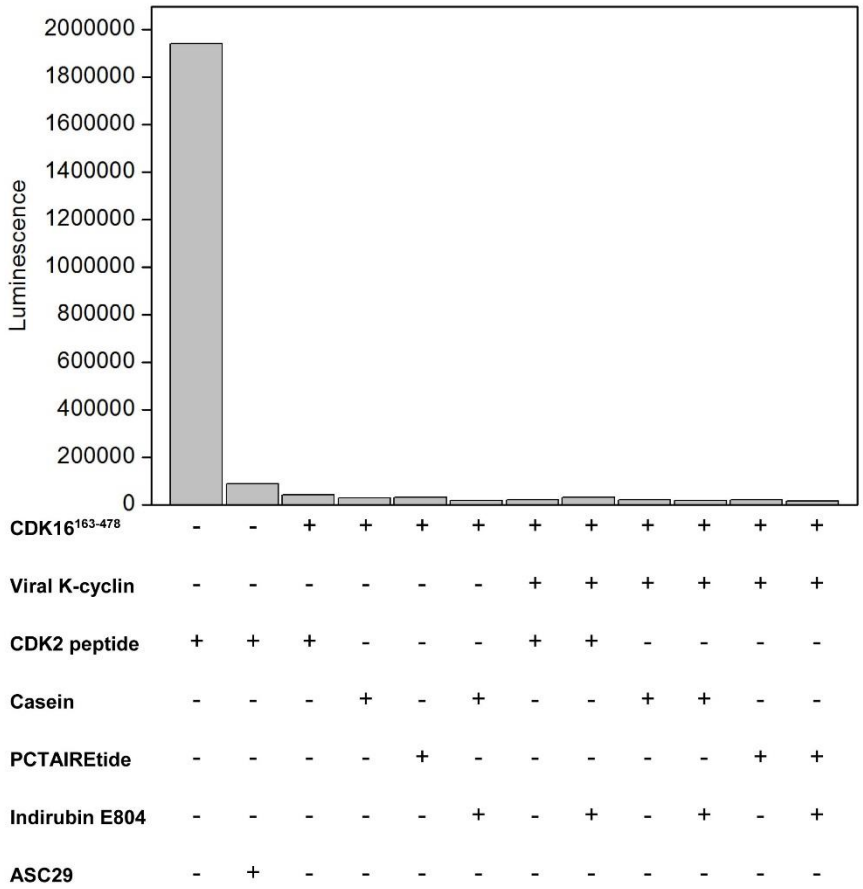


Figure 6.8. CDK16 kinase domain and K-cyclin show no kinase activity. Bar chart showing results from an ADP-Glo kinases assay. All Proteins used at a final concentration of 0.01 mg/mL. PCTAIREtide and casein were used at a final concentration of 10 μM, whilst CDK2 peptide was used at a final concentration of 200 μM. The kinase activity assay was run at room temperature for 30 minutes in a buffer containing 50 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mg/mL BSA and a final volume of 100 μL. Each condition was run in triplicate and the error bars represent ±S.D. ADP-detection was used as a measurement of activity and was detected by the ADP-Glo kinase activity assay (Promega). No activity above background was observed for CDK16 samples. CDK2/CycA with or without a pan-kinase inhibitor ("ASC29") acted as positive and negative controls, respectively. Indirubin E804 was used as a CDK16 inhibitor.

6.2.4.2 CDK16 shows no activity with CycY from insect expression

The full length CDK16 and isolated kinase domain both failed to interact with recombinant CycY I had purified from *Sf9* cells. Nonetheless, the available proteins were trialled in an ADP-Glo kinase activity assay to check for any weaker or transient

interactions that might allow activity. Full-length CDK16 and CDK16$^{163-478}$ were expressed in *Sf9* cells and *E. coli*, respectively, as described previously. CycY$^{1-331}$ and CycY$^{102-331}$ were also expressed separately in *Sf9* insect cells. Each protein was purified by nickel-affinity chromatography prior to use in the activity assay. None of the potential CDK16/CycY complexes tested appeared to show any significant activity towards PCTAIREtide relative to the background luminescent signal, whereas the CDK2/CycA positive control showed high activity (Fig. 6.8).

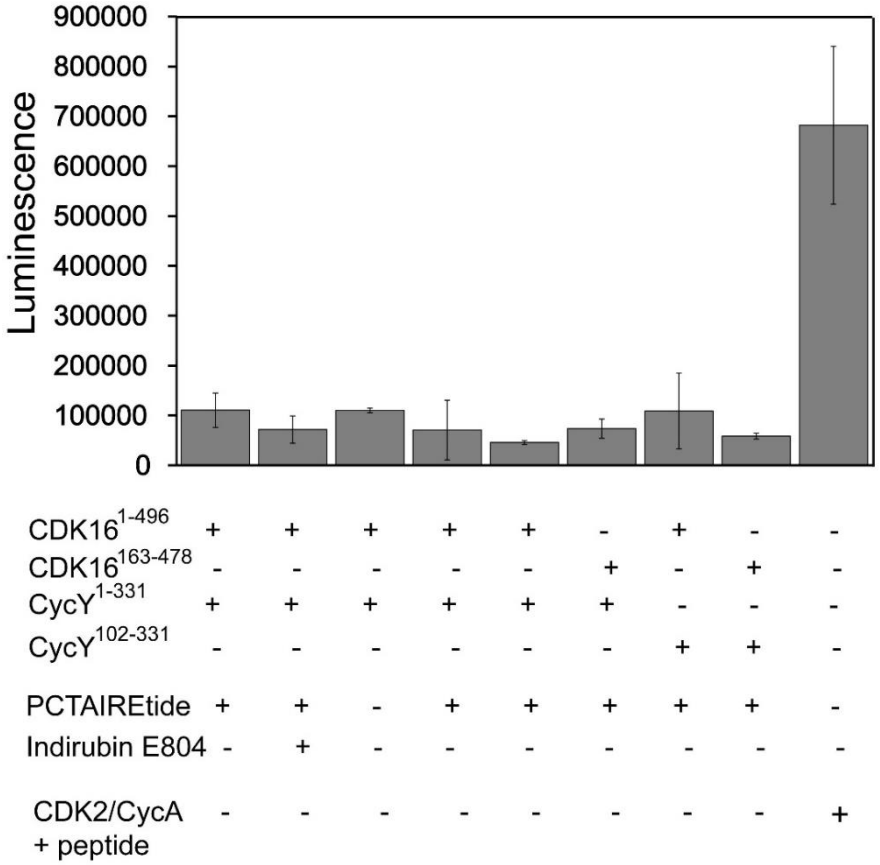


Figure 6.9. CDK16 and baculoviral expression constructs of CycY show no apparent activity. The ADP-Glo kinase activity assay (Promega) was used to monitor kinase activity from different CDK16/CycY combinations. In this assay, ADP formation is detected via luminescence. Indicated proteins were used at a final concentration of 2.5 µM. PCTAIREtide was used at a final concentration of 10 µM. The inhibitor Indirubin E804 was used at a concentration of 6 µM. The kinase activity assay was run at room temperature for 30 minutes in a buffer containing 50 mM HEPES, pH 7.5, 100 mM NaCl, 20 mM $MgCl_2$, 0.1 mg/mL BSA and a final volume of 5 µL. Each condition was run in triplicate and the error bars represent ±S.D. No activity above background was seen. CDK2/CycA acted as a positive control with the substrate peptide HHASPRK ("peptide"). The same results were obtained in a duplicate experiment.

6.2.4.3 Radiometric *in vitro* kinase assays using CycY from mammalian expression

Latterly during this work, a number of investigations confirmed CDK16/CycY as a viable active complex. Shehata *et al.* demonstrated that CDK16 binds to cyclin Y and could increase CDK16 activity towards PCTAIREtide over 100-fold [11]. Mikolcevic *et al.* also confirmed that the minimum cyclin binding region of CDK16 extended beyond the kinase domain using CDK16 immunoprecipitated from HEK293A cells or testis lysate [5].

The family of 14-3-3 proteins has been amongst those partners isolated as CDK16 interacting proteins in yeast-two hybrid screens, but no functional significance could be ascribed [9]. These proteins bind to phosphorylated serine or threonine residues and can serve to regulate the function of the target protein [20].

Most recently, Shehata *et al.* demonstrated that CycY phosphorylation at Ser100 and Ser326 could induce 14-3-3 binding that was critical for the interaction and activation of CDK16 [21]. Further, they established protocols to reconstitute an active CDK16 complex for use in an *in vitro* radiometric kinase assay (Chapter 5) [21]. I undertook a visit to this laboratory to test my samples in this assay following a kind invitation from Professor Kei Sakamoto (Lausanne, Switzerland). As well as the IC_{50} assays discussed in chapter 5, some preliminary studies were also undertaken using my recombinant $CDK16^{163\text{-}478}$ and viral K-cyclin proteins in their sensitive radiometric *in vitro* kinase assay. Due to visitor health and safety concerns the experimental procedures in Lausanne were performed by Saif Shehata.

As reported, their full length CDK16 was only active when mixed with CycY protein that was purified from COS-1 cells in complex with 14-3-3 (Fig. 6.10). However, no kinase activity was observed when the same CycY/14-3-3 complex was mixed with my

bacterially expressed CDK16$^{163-478}$ protein. Similar results were obtained with CDK16$^{165-446}$ protein that was prepared by Saif Shehata (Fig. 6.10). These results appear to confirm reports that the minimal kinase domain is insufficient for CycY binding.

Additional assays were performed in which CycY was replaced with my viral K-cyclin. Again, no activity was seen with the full-length or truncated CDK16 even in the presence of added 14-3-3 protein (Fig. 6.10). Interestingly, the high activity of the full length CDK16/CycY/14-3-3 complex was reduced significantly if the CDK16 monomer was pre-incubated with my K-cyclin. This result suggested that the K-cyclin was able to bind to CDK16 and inhibit its subsequent assembly with CycY. Figure 6.11 provides a schematic overview of the proposed hypothesis of a competition seen between CycY and viral K-cyclin for the binding of CDK16, from activity data collected. Notably, this dominant negative effect was diminished if the K-cyclin was added subsequently to the CycY sample (Fig. 6.10).

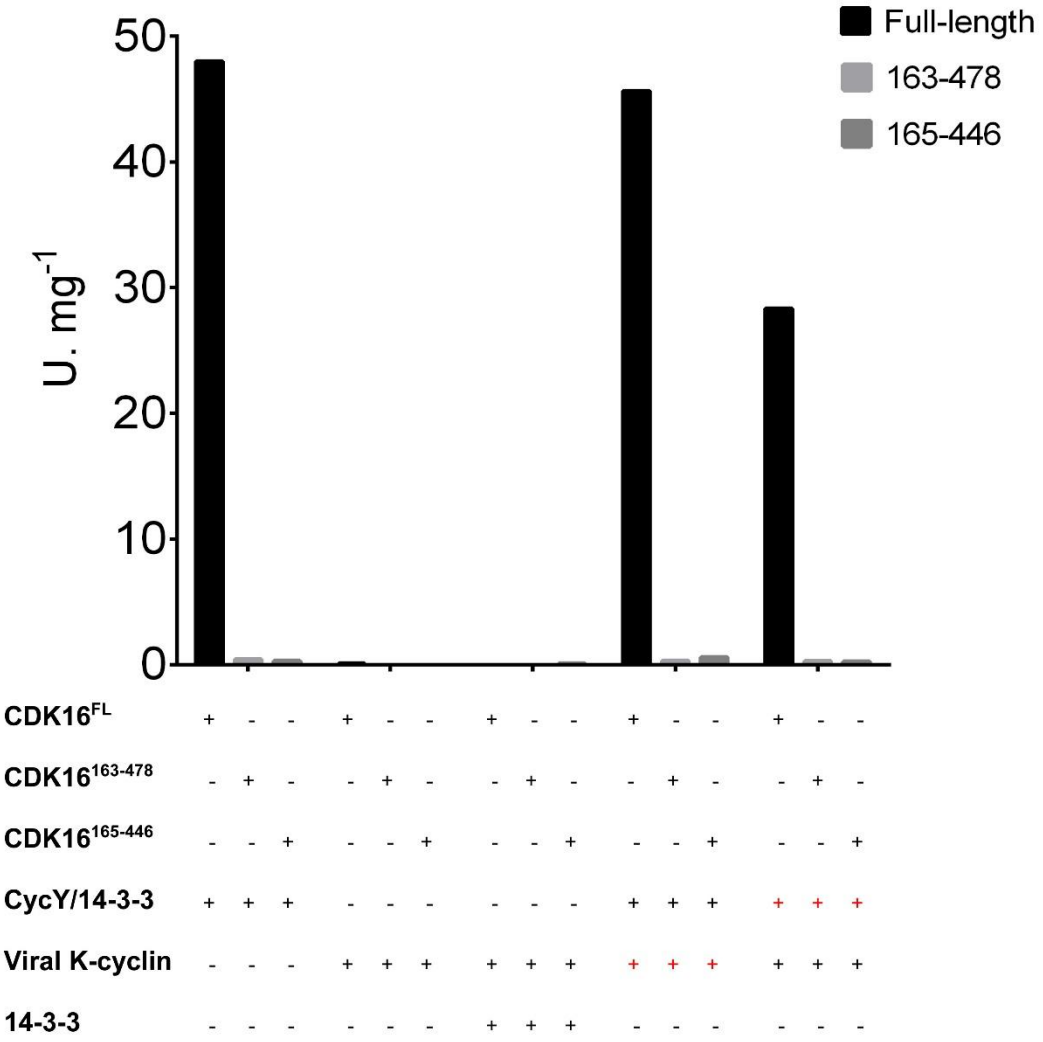


Figure 6.10. CDK16 kinase activity requires CycY and 14-3-3. A radiometric kinase assay was performed by Saif Shehata (Nestle Institute, Lausanne, Switzerland). Activity was only observed for complexes containing full length CDK16/CycY/14-3-3. The indicated purified proteins were each assayed for phosphotransferase activity in a final assay volume of 50 µL containing 50 mM HEPES, pH 7.5, 0.1 mM EGTA, 10 mM magnesium acetate, 0.1 mM [γ-^{32}P]ATP (300 c.p.m./pmol), 0.03% Brij 35, 1 mM DTT and 50 µM PCTAIREtide (PKSPKARKKL). Reactions were incubated at 30°C and terminated by spotting onto P81 paper and immersion in 75 mM H_3PO_4. P81 filters were washed three times for 10 min with H_3PO_4, rinsed with acetone, air-dried and incorporation of γ-^{32}P determined by Cherenkov counting. CycY was expressed in COS-1 cells and thus contains the necessary phosphorylations to bind 14-3-3, which was also present. The 14-3-3 sample was a purified, recombinant 14-3-3 expressed in *Sf21* insect cells. The '+' indicates that this cyclin was added after a 5 minute pre-incubation of the other proteins.

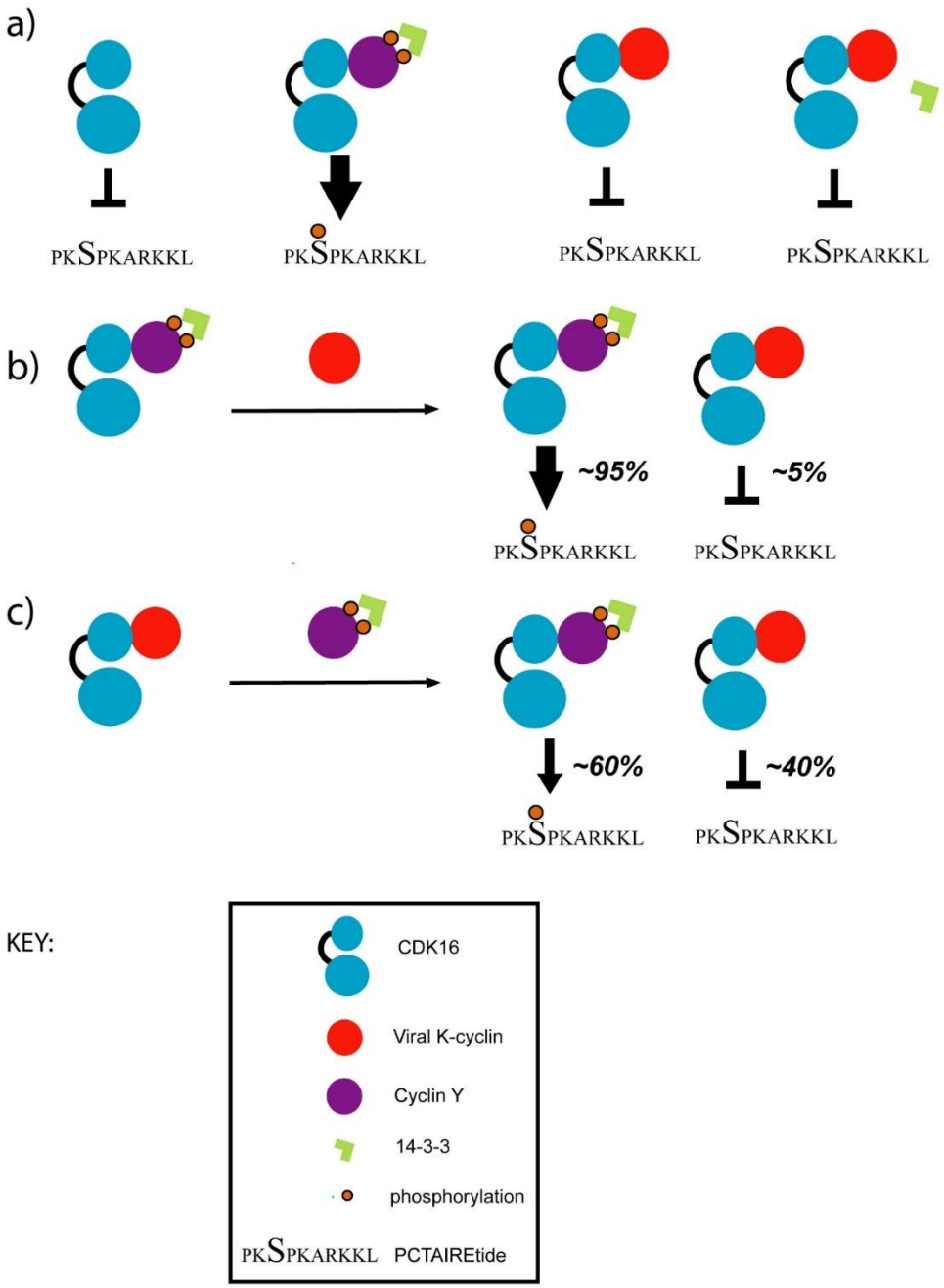


Figure 6.11. Schematic to show the interactions between CDK16, CycY, 14-3-3 and Viral K-cyclin and the activity against the PCTAIREtide consensus peptide. a) CDK16 alone is inactive. CDK16/ CycY/14-3-3 forms an active complex and is able to phosphorylate the PCTAIREtide peptide (PKSPKARKKL). Viral K-cyclin binds CDK16 with high affinity but the complex is inactive, even in the presence of 14-3-3. **b)** When viral K-cyclin was added to CDK16/CycY/14-3-3 after a 5 minute pre-incubation, there was a slight reduction in activity seen. **c)** When CycY and 14-3-3 are added to the CDK16/Viral K-cyclin complex after 5 minute pre-incubation, there was a reduction in overall activity output, presumably due to the inability for CycY/14-3-3 to outcompete the bound Viral K-cyclin. Percentages represent the percentage of the each form seen, derived from the activity data in Fig. 6.10 in comparison to the activity of the CDK16/CycY/14-3-3 complex alone.

6.2.5 Structural studies

Crystallisation trials were undertaken for CDK16$^{163-478}$/CycK$^{1-257}$ complexes and CycY constructs alone. Table 6.1 shows those constructs and complexes entered into crystal trials and the corresponding crystallisation screens tested. Granular precipitation of a CDK16$^{163-478}$/CycK$^{1-257}$ complex with Indirubin E804 was seen using a reservoir condition containing 0.1 M HEPES, pH 7.5, 10% polyethylene glycol 8000, 8% ethylene glycol (Fig. 6.12), but no mountable crystals were identified. The S319D mutant CDK16$^{163-478}$ construct that previously crystallised was also put into crystal trials with the PCTAIREtide substrate peptide, but this complex also failed to give any mountable crystals. Some additional crystal screens were performed with other members of the extended CDK14-18 family where suitable yields of soluble protein could be obtained, but similarly failed to give any crystals (see Appendix A11). Degradation of some CycY constructs rendered them unsuitable for crystallisation trials.

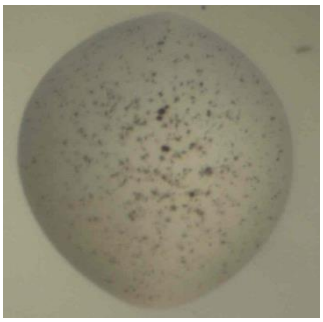


Figure 6.12. Example granular precipitation of a CDK16$^{163-478}$/CycK$^{1-257}$ complex with Indirubin E804 (1 mM). Image shows one example sitting drop from a Hampton Crystal Screen (HCS) condition containing 0.1 M HEPES, pH 7.5, 10% polyethylene glycol 8000, 8% ethylene glycol.

Table 6.1. Summary of crystallization trials for CDK16 and cyclins.

Protein(s)	Protein Concentration (mg/mL)	Coarse screens	Fine screen	Additions	Temperature (°C)
$CycY^{102-331}$	8.4	LFS, JCSG		10 µg/mL chymotrypsin	20
	8.9	All 4		-	4 or 20
$CycY^{9-331}$	10.5	LFS, JCSG		-	4
	10.5	HCS		1 mM FMP-31762;10 µg/mL chymotrypsin	20
$CDK16^{163-478}$	26.3	BCS	Sodium-formate fine screen	1 mM AMP-PNP;2 mM PCTAIREtide;0.1 mM $MgCl_2$	20
	26.3	BCS	Sodium-formate fine screen	1 mM Indirubin E804; 2 mM PCTAIREtide	20
	5-7.8	HIN1, JCSG, LFS		1 mM Indirubin E804; 2 mM PCTAIREtide	20
$CDK16^{163-478}$/ $K\text{-}cyclin^{1-257}$	7.6	HIN, JCSG, LFS		1 mM ADP + 50 mM $MgCl_2$	4 or 20
	5.8	All 4		1 mM Indirubin E804	4 or 20
	7.6	HIN, JCSG, LFS		1 mM ADP + 50 mM $MgCl_2$	4 or 20
	5.8	All 4		1 mM Indirubin E804	4 or 20
	40	All 4, MIDAS		1 mM ADP + 50 mM $MgCl_2$	4 or 20

("All 4 refers to HCS, HIN, LFS and JCSG screens")

6.3 Discussion

CDK14-18 form an extended subfamily of cyclin-dependent kinases. These kinases are structurally diverse from classical CDKs having large N and C-terminal extensions and putative binding to an atypical membrane-associated CycY. CDK16 has been ascribed diverse roles in the brain and testis where it is in the highest abundance. It is implicated in regulation of neurite outgrowth, spermatogenesis, muscle differentiation and glucose homeostasis. However, data are lacking on the structure and regulation of CDK16/CycY.

This chapter aimed to identify the novel structural mechanisms that determine activation of putative CDK16/cyclin complexes. Initially, a series of deletion mutants from recombinant baculoviral expression were characterized to determine the minimum domain requirements for an active CDK/cyclin complex. However, all of the CDK16/CycY combinations tested showed no interaction. This was subsequently explained by the work of Mikolcevic *et al.* and Shehata *et al*. which identified a requirement for a phosphorylated CycY subunit and a 14-3-3 partner [5, 11, 21]. Interestingly, I found that CDK16 could alternatively bind with high affinity to a smaller D-type cyclin, a KSHV viral K-cyclin. While this complex was inactive, the viral K-cyclin appeared capable of blocking the interaction of CycY and therefore exhibited dominant negative behaviour. Further, a D-type cyclin, CycD3, also has reported binding to CDK14 [1] and it would be interest to explore the binding and activity of this complex.

Both the viral K-cyclin and the CycY appear to have only one cyclin box domain. However, the different requirements for association of K-cyclin and CycY are indicative of a different mode of binding. For example, the binding of the K-cyclin required only the kinase domain of CDK16, whereas the active CycY complex appeared additionally dependent on a large N-terminal extension in CDK16. Notably, ITC measurements showed

that the affinity of the K-cyclin interaction was increased using full length CDK16. Therefore, it remains of significant interest to determine whether the N-terminal kinase extension carries additional structural elements.

CDK family members have been shown to have interchangeable binding surfaces that allow many different cyclin–CDK combinations to form, such as CDK2/CycA, CDK2/CycE and CDK1/CycA [22]. The structures of CDK/cyclin complexes reveal significant variability in their interface size and character, as well as the cyclin position. For example, CycA binding spans both the N and C-lobes of CDK2, whereas the CycT1 interface with CDK9 is reduced to the kinase N-lobe [23]. The crystal structure of CDK6 in complex with K-cyclin showed that this cyclin bound predominantly to the 'PSTAIRE' motif of CDK6 through the first cyclin repeat and N-terminal helix [24].

Overall, the requirements of CDK16 for an N-terminal kinase extension and the atypical CycY are highly indicative of a novel regulatory mechanism. The requirement of domains beyond the conserved central kinase fold for CDK regulation is not unique to CDK16. Both CDK9 and CDK12 have been shown to require a C-terminal extension, although this region promotes ATP binding and catalysis and is dispensable for cyclin interaction [25, 26]. Future work should further define the minimal domains required for CDK16/CycY complex formation and facilitate much needed structural determination of a CDK16/CycY complex, potentially including bound 14-3-3 protein.

6.4 References

1. Shu, F., et al., *Functional characterization of human PFTK1 as a cyclin-dependent kinase.* Proc Natl Acad Sci U S A, 2007. 104(22): p. 9248-53.

2. Davidson, G., et al., *Cell cycle control of wnt receptor activation.* Dev Cell, 2009. 17(6): p. 788-99.

3. Jiang, M., et al., *Cyclin Y, a novel membrane-associated cyclin, interacts with PFTK1*. FEBS Lett, 2009. 583(13): p. 2171-8.

4. Ou, C.Y., et al., *Two cyclin-dependent kinase pathways are essential for polarized trafficking of presynaptic components.* Cell, 2010. 141(5): p. 846-58.

5. Mikolcevic, P., et al., *Cyclin-dependent kinase 16/PCTAIRE kinase 1 is activated by cyclin Y and is essential for spermatogenesis.* Mol Cell Biol, 2012. 32(4): p. 868-79.

6. Cheng, K., et al., *Pctaire1 interacts with p35 and is a novel substrate for Cdk5/p35.* J Biol Chem, 2002. 277(35): p. 31988-93.

7. Rual, J.F., et al., *Towards a proteome-scale map of the human protein-protein interaction network.* Nature, 2005. 437(7062): p. 1173-8.

8. Mikolcevic, P., J. Rainer, and S. Geley, *Orphan kinases turn eccentric: a new class of cyclin Y-activated, membrane-targeted CDKs.* Cell Cycle, 2012. 11(20): p. 3758-68.

9. Charrasse, S., et al., *PCTAIRE-1: characterization, subcellular distribution, and cell cycle-dependent kinase activity.* Cell Growth Differ, 1999. 10(9): p. 611-20.

10. Verschuren, E.W., N. Jones, and G.I. Evan, *The cell cycle and how it is steered by Kaposi's sarcoma-associated herpesvirus cyclin.* J Gen Virol, 2004. 85(Pt 6): p. 1347-61.

11. Shehata, S.N., et al., *Analysis of substrate specificity and cyclin Y binding of PCTAIRE-1 kinase.* Cell Signal, 2012. 24(11): p. 2085-94.

12. Hedhammar, M. and S. Hober, *Z(basic)--a novel purification tag for efficient protein recovery.* J Chromatogr A, 2007. 1161(1-2): p. 22-8.

13. Savitsky, P., et al., *High-throughput production of human proteins for crystallization: the SGC experience.* J Struct Biol, 2010. 172(1): p. 3-13.

14. Daub, H., et al., *Kinase-selective enrichment enables quantitative phosphoproteomics of the kinome across the cell cycle.* Mol Cell, 2008. 31(3): p. 438-48.

15. Oppermann, F.S., et al., *Large-scale proteomics analysis of the human kinome.* Mol Cell Proteomics, 2009. 8(7): p. 1751-64.

16. Graeser, R., et al., *Regulation of the CDK-related protein kinase PCTAIRE-1 and its possible role in neurite outgrowth in Neuro-2A cells.* J Cell Sci, 2002. 115(Pt 17): p. 3479-90.

17. Beausoleil, S.A., et al., *Large-scale characterization of HeLa cell nuclear phosphoproteins.* Proc Natl Acad Sci U S A, 2004. 101(33): p. 12130-5.

18. Cantin, G.T., et al., *Combining protein-based IMAC, peptide-based IMAC, and MudPIT for efficient phosphoproteomic analysis.* J Proteome Res, 2008. 7(3): p. 1346-51.

19. Kettenbach, A.N., et al., *Rapid determination of multiple linear kinase substrate motifs by mass spectrometry.* Chem Biol, 2012. 19(5): p. 608-18.

20. Tzivion, G., et al., *14-3-3 proteins as potential oncogenes.* Semin Cancer Biol, 2006. 16(3): p. 203-13.

21. Shehata, S.N., et al., *Cyclin Y phosphorylation- and 14-3-3-binding-dependent activation of PCTAIRE-1/CDK16.* Biochem J, 2015. 469(3): p. 409-20.

22. Hochegger, H., S. Takeda, and T. Hunt, *Cyclin-dependent kinases and cell-cycle transitions: does one fit all?* Nat Rev Mol Cell Biol, 2008. 9(11): p. 910-6.

23. Echalier, A., J.A. Endicott, and M.E. Noble, *Recent developments in cyclin-dependent kinase biochemical and structural studies.* Biochim Biophys Acta, 2010. 1804(3): p. 511-9.

24. Jeffrey, P.D., L. Tong, and N.P. Pavletich, *Structural basis of inhibition of CDK-cyclin complexes by INK4 inhibitors.* Genes Dev, 2000. 14(24): p. 3115-25.

25. Baumli, S., et al., *The CDK9 Tail Determines the Reaction Pathway of Positive Transcription Elongation Factor b.* Structure, 2012. 20(10): p. 1788-95.

26. Dixon-Clarke, S.E., et al., *Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding.* Sci Rep, 2015. 5.

Chapter 7 Conclusions and Future Perspectives

The aims of this thesis were to characterise novel members of the extended CDK family. I was successful in developing methods to express and purify the human CDK12/CycK complex as well as a putative complex between CDK16 and a viral D-type cyclin. Further, using X-ray crystallography I was able to elucidate various complex structures for both CDK12 and CDK16, including complexes with small molecule inhibitors that are of particular interest for cancer studies.

CDK12 in complex with CycK is thought to play an important role in gene transcription, particularly elongation, by phosphorylating Ser2 within the heptad repeats of the C-terminal domain in RNA Polymerase II. It was therefore of specific interest determine the crystal structure of the CDK12/CycK complex and to compare this to that of the CDK9/CycT1 complex, which displays overlapping, but non redundant functions in mRNA elongation.

Initial investigations showed that CDK12 could only be solubly expressed in the presence of CycK, which has also previously been suggested as a cyclin partner for CDK9. Unfortunately, the requirement for co-expression prevented titration experiments to compare the relative binding affinities of CDK12 and CDK9 for CycK. Instead, it may be of interest to try baculoviral co-expression of a putative CDK9/CycK complex and to explore the relative stability of this complex during purification and perhaps crystallisation. My crystal structure of the phosphorylated CDK12/CycK complex revealed that the overall assembly was closely conserved with that of the CDK9/CycT1 complex. However, several sites in the protein interface were identified that were diverged in CDK9, but highly conserved in the CDK13, including a large β4-β5 loop insertion that drapes across the top of the cyclin. These data support suggestions that CycK may be a specific cyclin partner for CDK12 and CDK13, whereas CDK9 may be specific for CycT1 and CycT2. As

observed for CDK9, the activation loop in CDK12 could be phosphorylated *in vitro*, or in cells, using a yeast CAK enzyme. In contrast, I could not detect autophosphorylation of CDK12 using an *in vitro* kinase assay.

The most interesting structural feature lay in the CDK12 C-terminal extension that was identified in structures of the longer CDK12$^{715\text{-}1052}$ construct. This region was shown to be flexible adopting distinct conformations in the two CDK12/CycK complexes that were present in the crystallographic asymmetric unit. In an AMP-PNP bound CDK12 chain the C-terminal extension was folded across the front of the kinase where it formed an additional αK–helix. Interestingly, this position was distinct from the αK site in CDK9. The C-terminal extension in CDK12 formed multiple contacts with both the kinase N- and C-lobes as well as the bound nucleotide. Strikingly, the second CDK12 chain displayed a different path for the C-terminal extension, which was looped instead around the back of the kinase, whilst retaining the αK-helix. No evidence for the binding of AMP-PNP was observed in this structure suggesting that the C-terminal extension may act as a gating mechanism for the binding of the ATP.

CDK12 also has been implicated in regulation of mRNA splicing in which the proposed cyclin binding partner is CycL1 or CycL2. Unfortunately, it was not possible to establish an expression system for the soluble production of CycL1 and a cDNA was not available for CycL2. However, it remains of significant interest to pursue these complexes further in future work. Overall, the co-expression strategy developed for CDK12/CycK appears to be a favourable method to rescue the soluble expression of otherwise insoluble constructs.

The precise structural and functional roles of the large N and C-terminal domains found in full length CDK12 also remain to be determined, as well as a large C-terminal extension in full length CycK. The full length CDK12/CycK complex was 10-fold more

active than my crystallised constructs, suggesting that these regions fulfil a critical interaction. Several protein-protein interaction regions are predicted including proline-rich repeats in both proteins and a RS-rich domain in CDK12.

Small differences in the transcriptional roles of CDK12/CycK and CDK9/CycT1 remain to be elucidated. Moreover, there is an urgent need to understand and address the dysregulation of CDK12 in several cancers, including ovarian, breast and gastric cancer. Whilst loss of CDK12 activity is apparent in some of these cancers, there is also apparent synthetic lethality when CDK12 loss is combined with other inhibitors of DNA repair, including PARP inhibitors. Lack of a chemical probe for CDK12 has hindered our ability to investigate its precise role in transcription as well as its potential as a therapeutic target.

Following the elucidation of the CDK12/CycK structures, I was able to characterise a number of covalent CDK12 inhibitors under development in the group of Professor Nathanael Gray (Dana Farber Cancer Institute, Boston). Moreover, I was able to solve the crystal structure of CDK12/CycK in complex with the highly selective inhibitor THZ531. While most covalent kinase inhibitors target reactive cysteine residues found within the ATP pocket, the THZ531 inhibitor targets a distal cysteine present in the C-terminal kinase extension, Cys1039. This represents a novel strategy for selective kinase inhibitor design following the recent success of the CDK7 covalent inhibitor THZ1.

In the co-crystal structure of CDK12 with THZ531, the flexibility inherent in both the C-terminal extension and the THZ531 linker resulted in two distinct inhibitor binding conformations. The αK helix was disrupted in both binding conformers, which appears a necessary rearrangement for correct positioning of the covalently bound Cys1039 side chain. My structural models suggest that the C-terminal cysteines in CDK7 and CDK12 are folded at slightly different positions on the kinase C-lobe allowing the length and geometry of the covalent linker to influence inhibitor selectivity. Further modifications to

the linker in THZ531 may be considered in future work. For example, the current linker allows multiple inhibitor binding conformations and appears to push the C-terminal kinase extension away from the C-lobe surface. Potentially, shorter linkers may reduce unwanted reactions with more distant cysteines in other systems.

In my second project, I investigated the structure and inhibition of the PCTAIRE family kinase CDK16. This work was aided by the available structure of the kinase domain of CDK16, which was solved in complex with the type I inhibitor, indirubin E804 (PDB ID: 3MTL). CDK16 was recently shown to be upregulated in various forms of cancer and suggested as a novel therapeutic target based on gene knockdown or use of indirubin E804 [1]. Interestingly, knockdown of CycY did not suppress cancer cell proliferation in the same way as CDK16 knockdown [2]. However, CycY-like1, encoded by the CCNYL1 gene, is suggested as alternative cyclin binding partner for CDK16 function in spermatogenesis [2].

My screening of a collection of clinical kinase inhibitors led to the identification of rebastinib and dabrafenib as more potent, if not selective, CDK16 inhibitors. The co-crystal structure of CDK16 in complex with rebastinib revealed a similar binding mode to ABL, the intended target of rebastinib. The structure also identified several sequence changes that could be exploited in future for the development of more selective CDK16 inhibitors. Previous work had suggested CDK16 as a potentially actionable target in two difficult to treat cancer models, including medulloblastoma and melanoma [1, 3]. My preliminary studies supported this conclusion with the finding that rebastinib could inhibit the proliferation of cell lines from both cancers in the low nanomolar range. Ideally, with time permitting, further siRNA experiments should be performed in parallel to chemical inhibitors to determine if these effects are due to the targeting of CDK16, or potentially due to a combination of other targets.

The first CDK16 structure with indirubin E804 had revealed considerable disorder in the kinase N-lobe, consistent with the requirement for a cyclin partner. By contrast, the type II inhibitor binding mode of rebastinib yielded a highly ordered, but inactive structure revealing the proper α-helical folding of the PCTAIRE motif for the first time. Parallel efforts to solve the structure of a CDK16/cyclin complex were unsuccessful. More recently, collaborators have demonstrated that the CDK16/CycY complex requires specific phosphorylation of CycY in addition to the presence of an adaptor 14-3-3 protein [4]. This presents a considerable challenge for structural studies as the activating kinase remains unknown. Moreover, the precise minimal domains required for complex formation are still to be determined. By contrast, I found that the minimal kinase domain of CDK16 was sufficient to bind with high affinity to the bacterially expressed KSHV viral K-cyclin. It would be interesting in future to determine if the overexpression of this viral cyclin induces CDK16 interaction in cells and whether this disrupts CDK16 function. It is particularly curious that CDK16 may bind KSHV viral K-cyclin, but form an inactive complex. The arsenal of viruses' strategies to circumvent the host's immune system and to manipulate molecular machinery to their own means is boundless, for example the HIV accessory protein Tat specifically binds to CycT1 to modulate CDK9 activity [5].

In addition to these projects, I have identified several starting points for expression and structural studies of CDK14 and CDK17 (see appendix). Methods applied to CDK16 may be of value for work on these kinases too. For example, rebastinib has been shown to inhibit CDK14 in addition to CDK16. I also explored other covalent inhibitor strategies that may be helpful to co-crystallise CDK12 or CDK16 with a substrate peptides. These included bisubstrate-analogue inhibitors, which were available for the PIM1 kinase [6, 7]. These compounds combine an ATP mimetic inhibitor with a covalently linked fragment of substrate peptide. Using PIM1 and CDK2 as a models, I also tested a bivalent inhibitor

developed by Professor Kevan Shokat (University of California) that covalently tethers the inhibitor to both the kinase domain catalytic lysine and a Cys-substituted substrate peptide [7]. Both of these techniques require further optimisation of the protocols and may require new reagents for CDK12/16.

Together, my studies reveal some of the novel structural mechanisms that control kinase activity in two groups of atypical CDKs. It will be interested to see the conservation of these features in future work and to further understand the level of redundancy between CDK12-13 as well as between CDK16-18. The identified kinase inhibitors should provide valuable reagents to investigate these matters.

7.1 References

1. Cwiek, P., et al., *RNA interference screening identifies a novel role for PCTK1/CDK16 in medulloblastoma with c-Myc amplification.* Oncotarget, 2015. 6(1): p. 116-29.
2. Zi, Z., et al., *CCNYL1, but Not CCNY, Cooperates with CDK16 to Regulate Spermatogenesis in Mouse.* PLoS Genet, 2015. 11(8): p. e1005485.
3. Yanagi, T., J.C. Reed, and S. Matsuzawa, *PCTAIRE1 regulates p27 stability, apoptosis and tumor growth in malignant melanoma.* Oncoscience, 2014. 1(10): p. 624-33.
4. Shehata, S.N., et al., *Cyclin Y phosphorylation- and 14-3-3-binding-dependent activation of PCTAIRE-1/CDK16.* Biochem J, 2015. 469(3): p. 409-20.
5. Tahirov, T.H., et al., *Crystal structure of HIV-1 Tat complexed with human P-TEFb.* Nature, 2010. 465(7299): p. 747-51.

6. Kestav, K., et al., *Bisubstrate inhibitor approach for targeting mitotic kinase Haspin.* Bioconjug Chem, 2015. 26(2): p. 225-34.
7. Statsuk, A.V., et al., *Tuning a three-component reaction for trapping kinase substrate complexes.* J Am Chem Soc, 2008. 130(51): p. 17568-74.

Appendix

Appendix A1. The crystallization conditions plate contents of a fine screen based created, based around HIN coarse screen conditions in wells D6 and D7

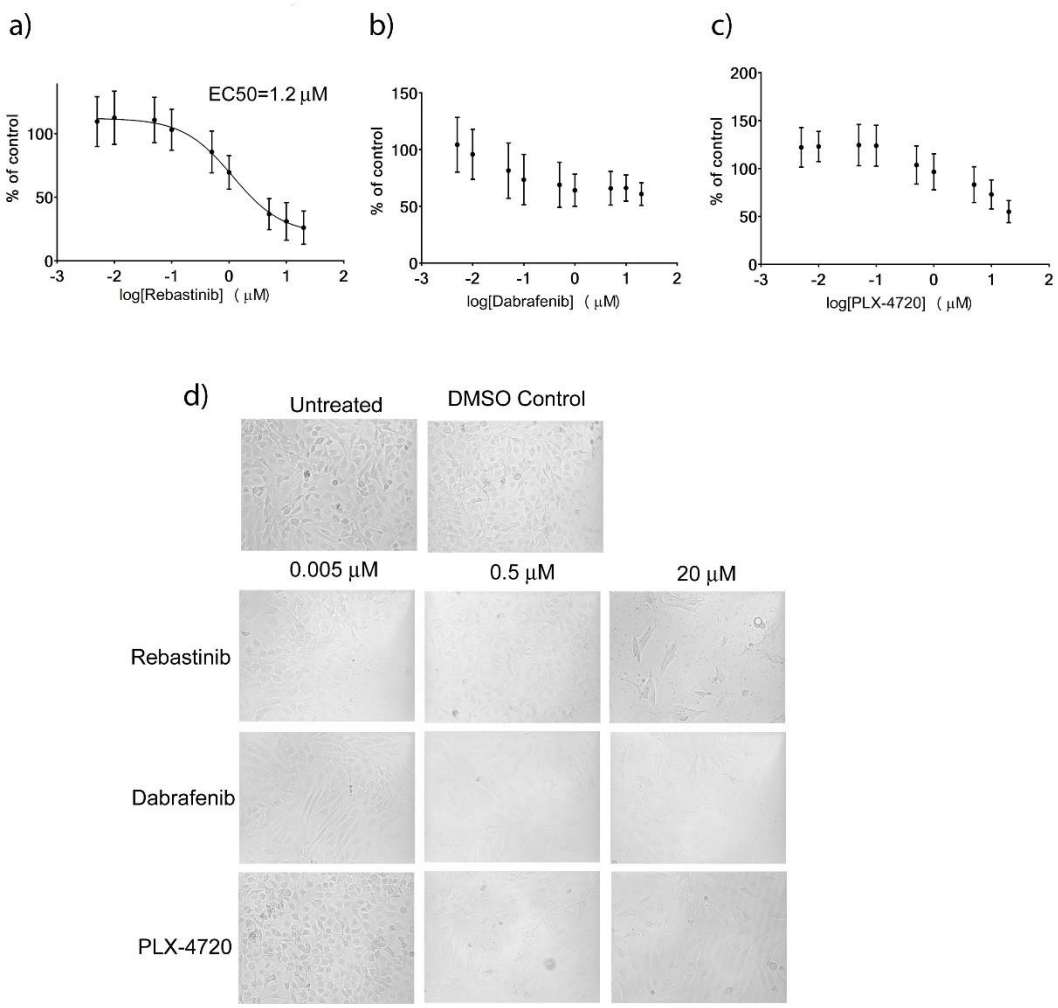
	1	2	3	4	5	6	7	8	9	10	11	12
A	15% PEG3.3K 0.1M bis-tris pH 5.5	17% PEG3.3K 0.1M bis-tris pH 5.5	19% PEG3.3K 0.1M bis-tris pH 5.5	21% PEG3.3K 0.1M bis-tris pH 5.5	23% PEG3.3K 0.1M bis-tris pH 5.5	25% PEG3.3K 0.1M bis-tris pH 5.5	25% PEG3.3K 0.1M bis-tris pH 5.5	27% PEG3.3K 0.1M bis-tris pH 5.5	29% PEG3.3K 0.1M bis-tris pH 5.5	31% PEG3.3K 0.1M bis-tris pH 5.5	33% PEG3.3K 0.1M bis-tris pH 5.5	35% PEG3.3K 0.1M bis-tris pH 5.5
B	15% PEG3.3K 0.1M bis-tris pH 5.75	17% PEG3.3K 0.1M bis-tris pH 5.75	19% PEG3.3K 0.1M bis-tris pH 5.75	21% PEG3.3K 0.1M bis-tris pH 5.75	23% PEG3.3K 0.1M bis-tris pH 5.75	25% PEG3.3K 0.1M bis-tris pH 5.75	25% PEG3.3K 0.1M bis-tris pH 5.75	27% PEG3.3K 0.1M bis-tris pH 5.75	29% PEG3.3K 0.1M bis-tris pH 5.75	31% PEG3.3K 0.1M bis-tris pH 5.75	33% PEG3.3K 0.1M bis-tris pH 5.75	35% PEG3.3K 0.1M bis-tris pH 5.75
C	15% PEG3.3K 0.1M bis-tris pH 6	17% PEG3.3K 0.1M bis-tris pH 6	19% PEG3.3K 0.1M bis-tris pH 6	21% PEG3.3K 0.1M bis-tris pH 6	23% PEG3.3K 0.1M bis-tris pH 6	25% PEG3.3K 0.1M bis-tris pH 6	25% PEG3.3K 0.1M bis-tris pH 6	27% PEG3.3K 0.1M bis-tris pH 6	29% PEG3.3K 0.1M bis-tris pH 6	31% PEG3.3K 0.1M bis-tris pH 6	33% PEG3.3K 0.1M bis-tris pH 6	35% PEG3.3K 0.1M bis-tris pH 6
D	15% PEG3.3K 0.1M bis-tris pH 6.25	17% PEG3.3K 0.1M bis-tris pH 6.25	19% PEG3.3K 0.1M bis-tris pH 6.25	21% PEG3.3K 0.1M bis-tris pH 6.25	23% PEG3.3K 0.1M bis-tris pH 6.25	25% PEG3.3K 0.1M bis-tris pH 6.25	25% PEG3.3K 0.1M bis-tris pH 6.25	27% PEG3.3K 0.1M bis-tris pH 6.25	29% PEG3.3K 0.1M bis-tris pH 6.25	31% PEG3.3K 0.1M bis-tris pH 6.25	33% PEG3.3K 0.1M bis-tris pH 6.25	35% PEG3.3K 0.1M bis-tris pH 6.25
E	15% PEG3.3K 0.1M bis-tris pH 6.5	17% PEG3.3K 0.1M bis-tris pH 6.5	19% PEG3.3K 0.1M bis-tris pH 6.5	21% PEG3.3K 0.1M bis-tris pH 6.5	23% PEG3.3K 0.1M bis-tris pH 6.5	25% PEG3.3K 0.1M bis-tris pH 6.5	25% PEG3.3K 0.1M bis-tris pH 6.5	27% PEG3.3K 0.1M bis-tris pH 6.5	29% PEG3.3K 0.1M bis-tris pH 6.5	31% PEG3.3K 0.1M bis-tris pH 6.5	33% PEG3.3K 0.1M bis-tris pH 6.5	35% PEG3.3K 0.1M bis-tris pH 6.5
F	15% PEG3.3K 0.1M bis-tris pH 6.75	17% PEG3.3K 0.1M bis-tris pH 6.75	19% PEG3.3K 0.1M bis-tris pH 6.75	21% PEG3.3K 0.1M bis-tris pH 6.75	23% PEG3.3K 0.1M bis-tris pH 6.75	25% PEG3.3K 0.1M bis-tris pH 6.75	25% PEG3.3K 0.1M bis-tris pH 6.75	27% PEG3.3K 0.1M bis-tris pH 6.75	29% PEG3.3K 0.1M bis-tris pH 6.75	31% PEG3.3K 0.1M bis-tris pH 6.75	33% PEG3.3K 0.1M bis-tris pH 6.75	35% PEG3.3K 0.1M bis-tris pH 6.75
G	15% PEG3.3K 0.1M bis-tris pH 7	17% PEG3.3K 0.1M bis-tris pH 7	19% PEG3.3K 0.1M bis-tris pH 7	21% PEG3.3K 0.1M bis-tris pH 7	23% PEG3.3K 0.1M bis-tris pH 7	25% PEG3.3K 0.1M bis-tris pH 7	25% PEG3.3K 0.1M bis-tris pH 7	27% PEG3.3K 0.1M bis-tris pH 7	29% PEG3.3K 0.1M bis-tris pH 7	31% PEG3.3K 0.1M bis-tris pH 7	33% PEG3.3K 0.1M bis-tris pH 7	35% PEG3.3K 0.1M bis-tris pH 7
H	15% PEG3.3K 0.1M bis-tris pH 7.25	17% PEG3.3K 0.1M bis-tris pH 7.25	19% PEG3.3K 0.1M bis-tris pH 7.25	21% PEG3.3K 0.1M bis-tris pH 7.25	23% PEG3.3K 0.1M bis-tris pH 7.25	25% PEG3.3K 0.1M bis-tris pH 7.25	25% PEG3.3K 0.1M bis-tris pH 7.25	27% PEG3.3K 0.1M bis-tris pH 7.25	29% PEG3.3K 0.1M bis-tris pH 7.25	31% PEG3.3K 0.1M bis-tris pH 7.25	33% PEG3.3K 0.1M bis-tris pH 7.25	35% PEG3.3K 0.1M bis-tris pH 7.25

Appendix A2. The crystallization conditions plate contents of a fine screen based created, based around HIN coarse screen conditions in wells D10 (20% PEG5000MME, 0.1M bis-tris pH 6.5) and E10 (45% polypropylene glycol 400, 0.1M bis-tris pH 6.5)

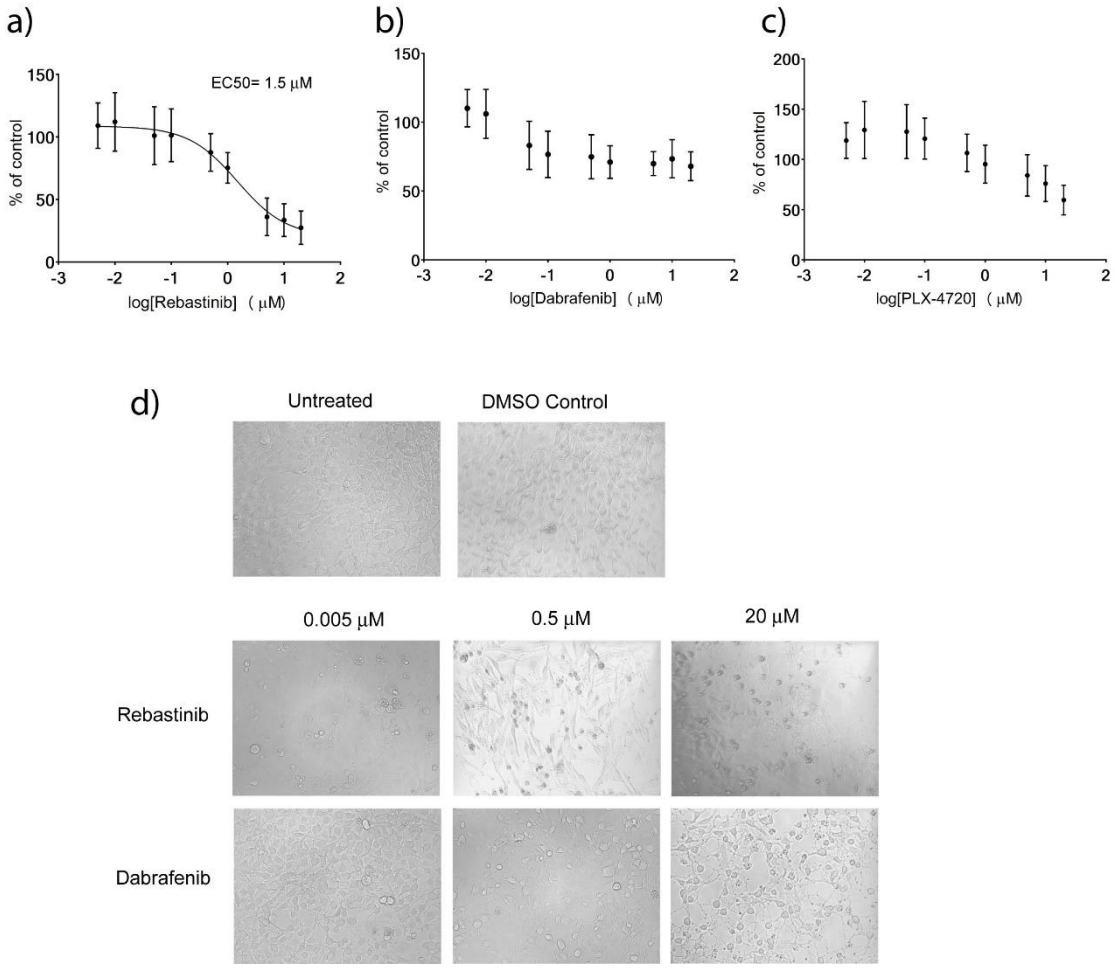
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.1M bis-tris pH 5.5 5% PEGm5K	0.1M bis-tris pH 5.5 10% PEGm5K	0.1M bis-tris pH 5.5 15% PEGm5K	0.1M bis-tris pH 5.5 20% PEGm5K	0.1M bis-tris pH 5.5 25% PEGm5K	0.1M bis-tris pH 5.5 30% PEGm5K	0.1M bis-tris pH 5.5 30% PPG400	0.1M bis-tris pH 5.5 35% PPG400	0.1M bis-tris pH 5.5 40% PPG400	0.1M bis-tris pH 5.5 45% PPG400	0.1M bis-tris pH 5.5 50% PPG400	0.1M bis-tris pH 5.5 55% PPG400
B	0.1M bis-tris pH 5.75 5% PEGm5K	0.1M bis-tris pH 5.75 10% PEGm5K	0.1M bis-tris pH 5.75 15% PEGm5K	0.1M bis-tris pH 5.75 20% PEGm5K	0.1M bis-tris pH 5.75 25% PEGm5K	0.1M bis-tris pH 5.75 30% PEGm5K	0.1M bis-tris pH 5.75 30% PPG400	0.1M bis-tris pH 5.75 35% PPG400	0.1M bis-tris pH 5.75 40% PPG400	0.1M bis-tris pH 5.75 45% PPG400	0.1M bis-tris pH 5.75 50% PPG400	0.1M bis-tris pH 5.75 55% PPG400
C	0.1M bis-tris pH 6 5% PEGm5K	0.1M bis-tris pH 6 10% PEGm5K	0.1M bis-tris pH 6 15% PEGm5K	0.1M bis-tris pH 6 20% PEGm5K	0.1M bis-tris pH 6 25% PEGm5K	0.1M bis-tris pH 6 30% PEGm5K	0.1M bis-tris pH 6 30% PPG400	0.1M bis-tris pH 6 35% PPG400	0.1M bis-tris pH 6 40% PPG400	0.1M bis-tris pH 6 45% PPG400	0.1M bis-tris pH 6 50% PPG400	0.1M bis-tris pH 6 55% PPG400
D	0.1M bis-tris pH 6.25 5% PEGm5K	0.1M bis-tris pH 6.25 10% PEGm5K	0.1M bis-tris pH 6.25 15% PEGm5K	0.1M bis-tris pH 6.25 20% PEGm5K	0.1M bis-tris pH 6.25 25% PEGm5K	0.1M bis-tris pH 6.25 30% PEGm5K	0.1M bis-tris pH 6.25 30% PPG400	0.1M bis-tris pH 6.25 35% PPG400	0.1M bis-tris pH 6.25 40% PPG400	0.1M bis-tris pH 6.25 45% PPG400	0.1M bis-tris pH 6.25 50% PPG400	0.1M bis-tris pH 6.25 55% PPG400
E	0.1M bis-tris pH 6.5 5% PEGm5K	0.1M bis-tris pH 6.5 10% PEGm5K	0.1M bis-tris pH 6.5 15% PEGm5K	0.1M bis-tris pH 6.5 20% PEGm5K	0.1M bis-tris pH 6.5 25% PEGm5K	0.1M bis-tris pH 6.5 30% PEGm5K	0.1M bis-tris pH 6.5 30% PPG400	0.1M bis-tris pH 6.5 35% PPG400	0.1M bis-tris pH 6.5 40% PPG400	0.1M bis-tris pH 6.5 45% PPG400	0.1M bis-tris pH 6.5 50% PPG400	0.1M bis-tris pH 6.5 55% PPG400
F	0.1M bis-tris pH 6.75 5% PEGm5K	0.1M bis-tris pH 6.75 10% PEGm5K	0.1M bis-tris pH 6.75 15% PEGm5K	0.1M bis-tris pH 6.75 20% PEGm5K	0.1M bis-tris pH 6.75 25% PEGm5K	0.1M bis-tris pH 6.75 30% PEGm5K	0.1M bis-tris pH 6.75 30% PPG400	0.1M bis-tris pH 6.75 35% PPG400	0.1M bis-tris pH 6.75 40% PPG400	0.1M bis-tris pH 6.75 45% PPG400	0.1M bis-tris pH 6.75 50% PPG400	0.1M bis-tris pH 6.75 55% PPG400
G	0.1M bis-tris pH 7 5% PEGm5K	0.1M bis-tris pH 7 10% PEGm5K	0.1M bis-tris pH 7 15% PEGm5K	0.1M bis-tris pH 7 20% PEGm5K	0.1M bis-tris pH 7 25% PEGm5K	0.1M bis-tris pH 7 30% PEGm5K	0.1M bis-tris pH 7 30% PPG400	0.1M bis-tris pH 7 35% PPG400	0.1M bis-tris pH 7 40% PPG400	0.1M bis-tris pH 7 45% PPG400	0.1M bis-tris pH 7 50% PPG400	0.1M bis-tris pH 7 55% PPG400
H	0.1M bis-tris pH 7.5 5% PEGm5K	0.1M bis-tris pH 7.5 10% PEGm5K	0.1M bis-tris pH 7.5 15% PEGm5K	0.1M bis-tris pH 7.5 20% PEGm5K	0.1M bis-tris pH 7.5 25% PEGm5K	0.1M bis-tris pH 7.5 30% PEGm5K	0.1M bis-tris pH 7.5 30% PPG400	0.1M bis-tris pH 7.5 35% PPG400	0.1M bis-tris pH 7.5 40% PPG400	0.1M bis-tris pH 7.5 45% PPG400	0.1M bis-tris pH 7.5 50% PPG400	0.1M bis-tris pH 7.5 55% PPG400

Appendix A3. The crystallization conditions plate contents of a fine screen based created, based around JCSG coarse screen conditions in well D1 (0.1 M HEPES pH 7, 25% PEG1000, 20% glycerol)

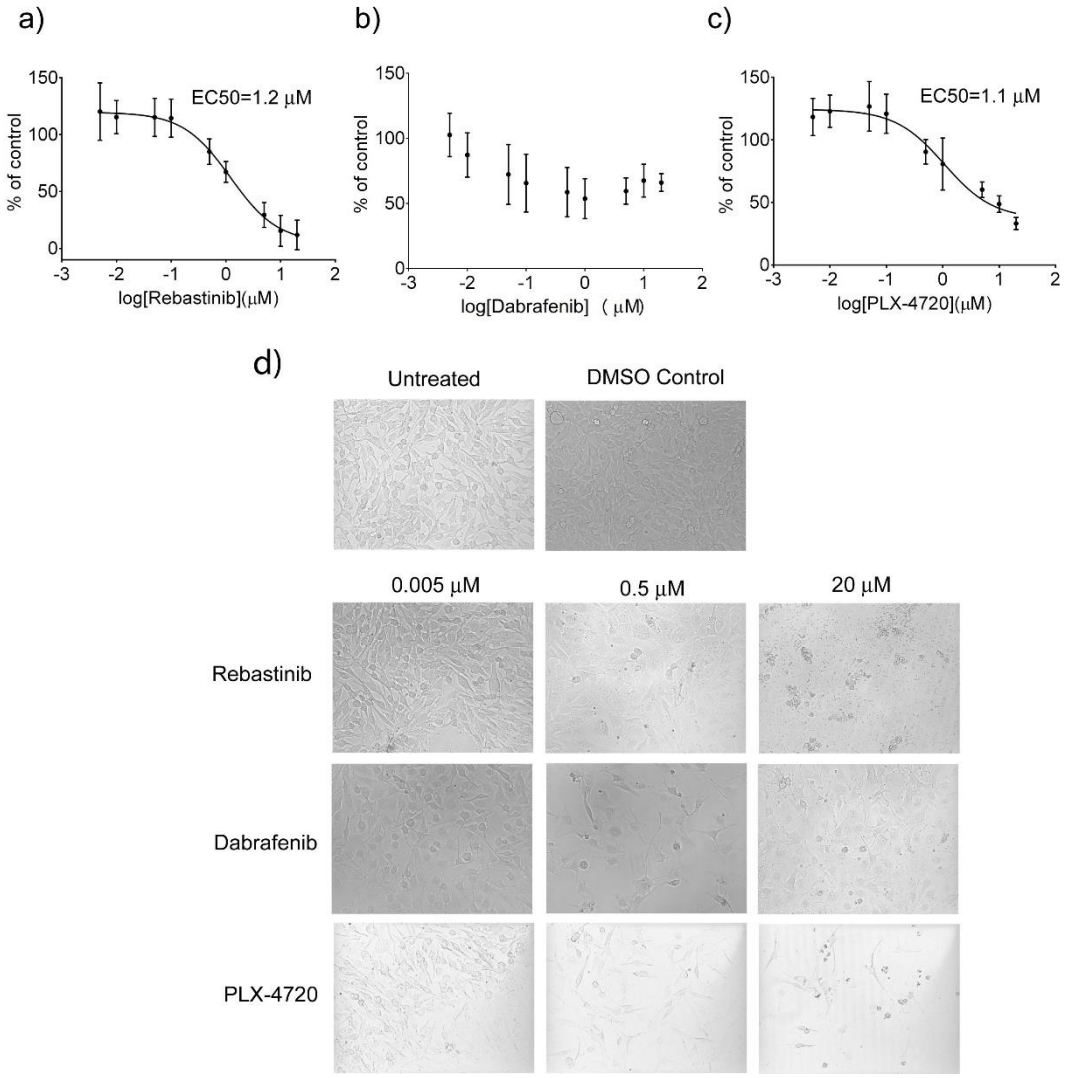
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.1M HEPES pH 7 20%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 14%(v/v) glycerol
B	0.1M HEPES pH 7 20%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 17%(v/v) glycerol
C	0.1M HEPES pH 7 20%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 20%(v/v) glycerol
D	0.1M HEPES pH 7 20%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 22%(v/v) glycerol
E	0.1M HEPES pH 7 20%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 14%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 14%(v/v) glycerol
F	0.1M HEPES pH 7 20%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 17%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 17%(v/v) glycerol
G	0.1M HEPES pH 7 20%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 20%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 20%(v/v) glycerol
H	0.1M HEPES pH 7 20%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 21%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 22%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 23%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 24%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 25%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 26%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 27%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 28%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 29%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 30%(w/v) PEG 1k 22%(v/v) glycerol	0.1M HEPES pH 7 31%(w/v) PEG 1k 22%(v/v) glycerol



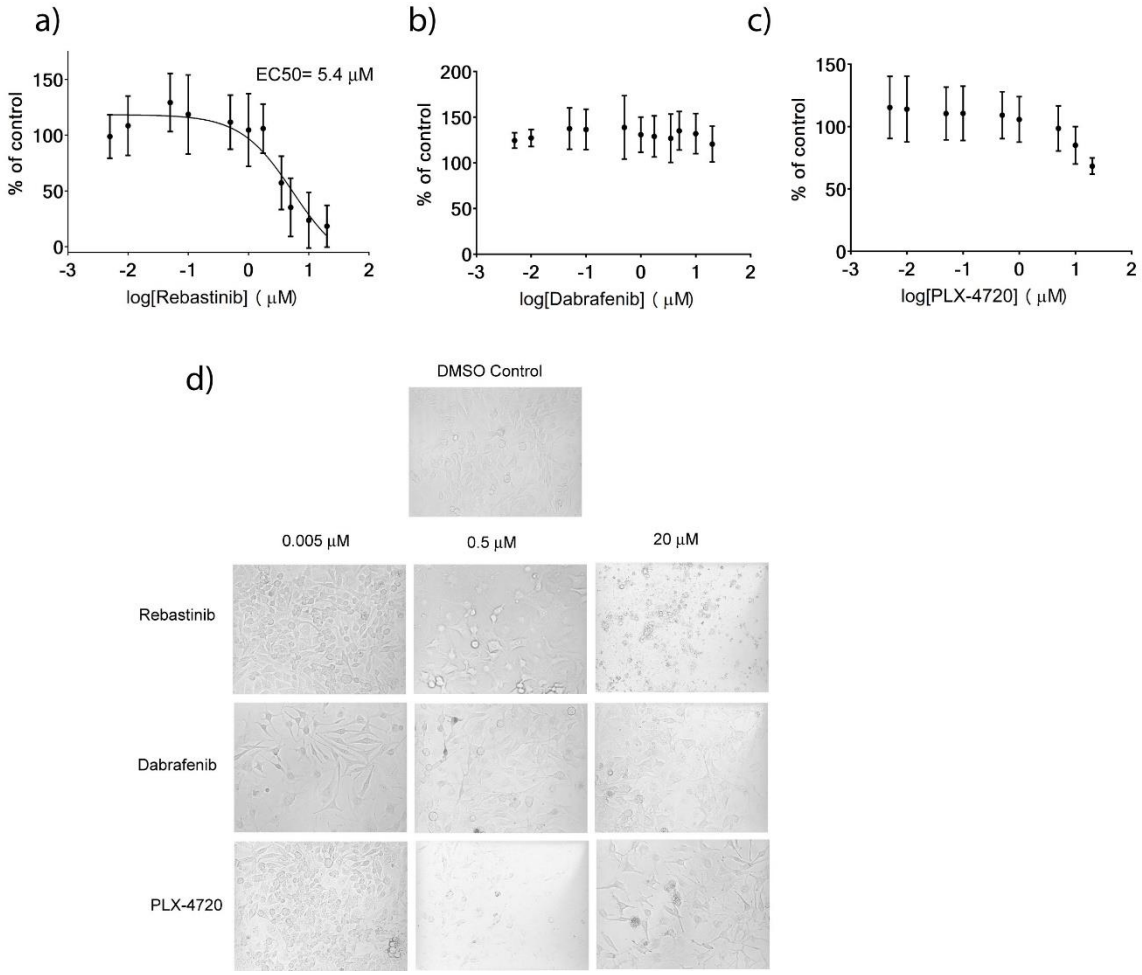
Appendix A4. IGR-37 cell viability was more sensitive to rebastinib whereas dabrafenib had no effect. a)b)c) for reference from section 5.8.2.8. Cell viability was determined for IGR-37 cells treated for 72 h with rebastinib, dabrafenib or PLX-4720, as indicated. Data represent mean value±S.D. of three independent experiments performed in triplicates and normalized to untreated control. Cell viability was assessed using a luminescence-based readout for ATP (CellTiter Glo) throughout. d) Images of treated and untreated cell lines were taken at 72 hours using the Floid Cell Imaging Station light microscopy (Thermofisher).



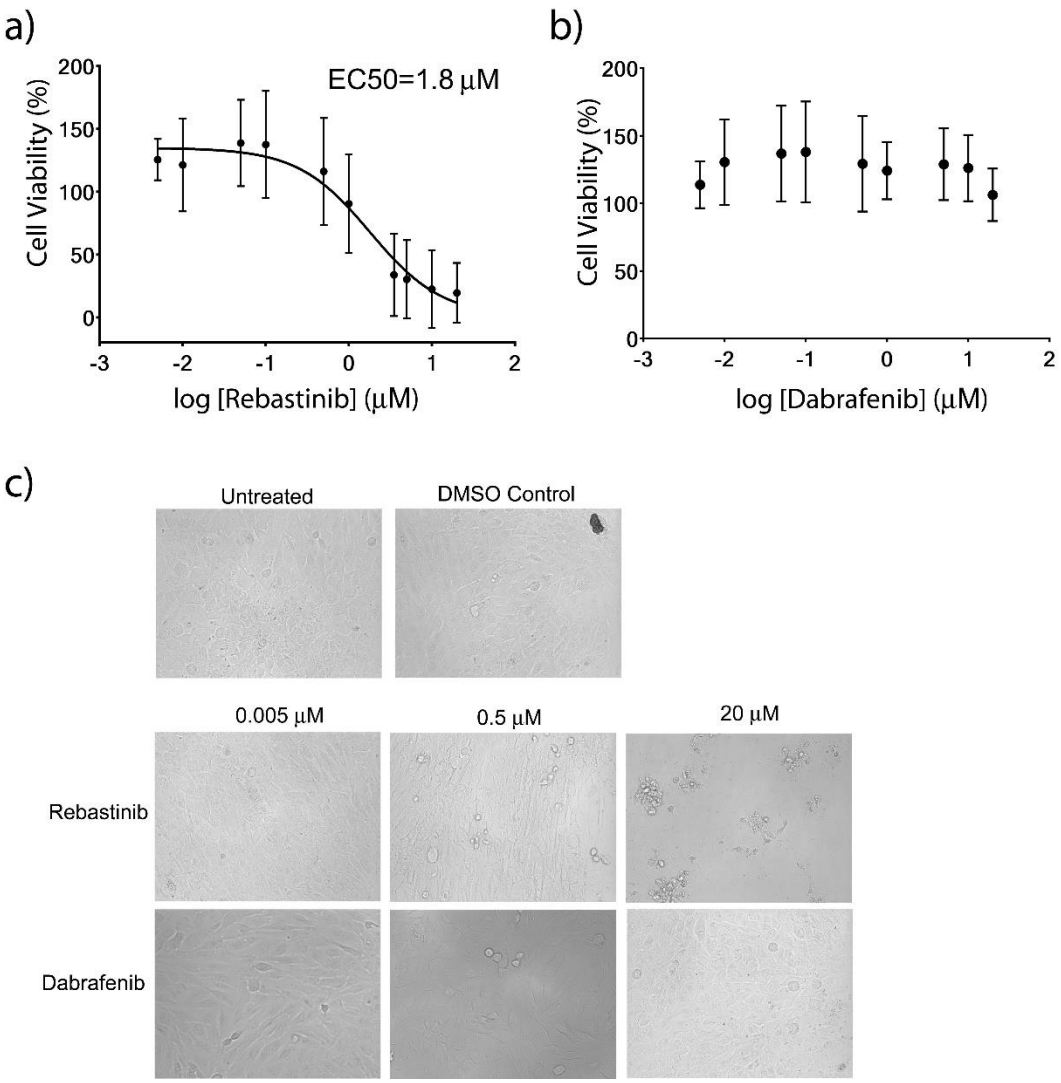
Appendix A5. IGR-39 cell viability was more sensitive to rebastinib whereas dabrafenib had no effect. a)b)c) for reference from section 5.8.2.8. Cell viability was determined for IGR-39 cells treated for 72 h with rebastinib, dabrafenib or PLX-4720, as indicated. Data represent mean value±S.D. of three independent experiments performed in triplicates and normalized to untreated control. Cell viability was assessed using a luminescence-based readout for ATP (CellTiter Glo) throughout. d) Images of treated and untreated cell lines were taken at 72 hours using the Floid Cell Imaging Station light microscopy (Thermofisher).



Appendix A6. SK-Mel-28 cell viability was more sensitive to rebastinib whereas dabrafenib had no effect. a)b)c) for reference from section 5.8.2.8. Cell viability was determined for SK-Mel-28 cells treated for 72 h with rebastinib, dabrafenib or PLX-4720, as indicated. Data represent mean value±S.D. of one experiment performed in triplicates and normalized to untreated control. Cell viability was assessed using a luminescence-based readout for ATP (CellTiter Glo) throughout. d) Images of treated and untreated cell lines were taken at 72 hours using the Floid Cell Imaging Station light microscopy (Thermofisher).



Appendix A7. 501-Mel cell viability was more sensitive to rebastinib whereas dabrafenib had no effect. a)b)c) for reference from section 5.8.2.8. Cell viability was determined for 501-Mel cells treated for 72 h with rebastinib, dabrafenib or PLX-4720, as indicated. Data represent mean value±S.D. of three independent experiments performed in triplicates and normalized to untreated control. Cell viability was assessed using a luminescence-based readout for ATP (CellTiter Glo) throughout. d) Images of treated and untreated cell lines were taken at 72 hours using the Floid Cell Imaging Station light microscopy (Thermofisher).



Appendix A8. DAOY cell viability was more sensitive to rebastinib whereas dabrafenib had no effect. a)b) for reference from section 5.8.2.8. Cell viability was determined for DAOY cells treated for 72 h with rebastinib or dabrafenib, as indicated. Data represent mean value±S.D. of three independent experiments performed in triplicates and normalized to untreated control. Cell viability was assessed using a luminescence-based readout for ATP (CellTiter Glo) throughout. d) Images of treated and untreated cell lines were taken at 72 hours using the Floid Cell Imaging Station light microscopy (Thermofisher).

Appendix A9. Insect cell test expression and purification (3ml scale) of various constructs of different targets, highlighting solubility. Constructs were expressed using baculovirus infected Sf9 cells at 3ml scale.* where N-T corresponds to N-terminal and C-T corresponds to C-terminal. Vector constructs provided in the appendix, Figure A1.

Target	N^{o} cloned constructs	N^{o} soluble constructs	Vector	Tag*
CDK14	6	0	pFB-CT10HF-LIC	C-T, His & FLAG
	9	9	pFB-LIC-Bse	N-T, His
CDK16	2	2	pFB-CT10HF-LIC	C-T, His & FLAG
	2	2	pFB-LIC-Bse	N-T, His
CDK17	4	2	pFB-CT10HF-LIC	C-T, His & FLAG
	4	3	pFB-LIC-Bse	N-T, His
CDK18	4	2	pFB-CT10HF-LIC	C-T, His & FLAG
	4	3	pFB-LIC-Bse	N-T, His
CycY	10	4	pFB-CT10HF-LIC	C-T, His & FLAG
	5	5	pFB-LIC-Bse	N-T, His
CycY-L1	9	0	pFB-CT10HF-LIC	C-T, His & FLAG
	2	0	pFB-LIC-Bse	N-T, His
CycY-L2	4	0	pFB-CT10HF-LIC	C-T, His & FLAG
	1	0	pFB-LIC-Bse	N-T, His
CycY-L3	9	0	pFB-CT10HF-LIC	C-T, His & FLAG
	1	0	pFB-LIC-Bse	N-T, His
CycD1	1	0	pFB-LIC-Bse	N-T, His
CycD3	3	0	pFB-LIC-Bse	N-T, His
CycI	3	0	pFB-LIC-Bse	N-T, His
CycL1	2	0	pFB-LIC-Bse	N-T, His
CycK	1	1	pFB-LIC-Bse	N-T, His

Appendix A10. Summary of proteins purified ≥ 1L scale. Where N-T corresponds to N-terminal and C-T corresponds to C-terminal. * Ni-affinity purification only, unless specified otherwise. ** Following gel filtration. Has a S319D mutation in the activation loop (PDB ID 3MTL). †† Very variable expression and final concentration following purification at later stages. BV corresponds to baculovirus.

Target	Construct sequence	Expression System	Expression Vector	Affinity Tag	Purified Yield (mg/L)*	Degradation
CDK14	1-451	BV	pFB-LIC-Bse	N-T, His (Tev)	13.0	+
	114-437	BV	pFB-LIC-Bse	N-T, His (Tev)	4.0	+
CDK16	163-478†	*E. coli*	pNIC-CH	C-T, His (Non-cleavable)	46.7	-
	163-478	*E. coli*	pNIC28-Bsa4	N-T, His (Tev)	9.6**	-
	1-496	BV	pFB-CT10HF-LIC	C-T, His & FLAG (Tev)	9.4	++
	1-477	BV	pFB-CT10HF-LIC	C-T, His & FLAG (Tev)	8.0	++
	109-496	BV	pFB-LIC-Bse	N-T, His (Tev)	10.9	-
CDK17	188-523	BV	pFB-CT10HF-LIC	C-T, His & FLAG (Tev)	2.3	-
	189 -505	BV	pFB-CT10HF-LIC	C-T, His & FLAG (Tev)	1.9**	-
CDK18	86-472	BV	pFB-LIC-Bse	N-T, His (Tev)	2.3	+
CycY	1-331	BV	pFB-CT10HF-LIC	C-T, His & FLAG (Tev)	4.4**	++
	102-331	BV	pFB-CT10HF-LIC	C-T, His & FLAG (Tev)	6.2**	-
	9-331	BV	pFB-LIC-Bse	N-T, His (Tev)	6.0	++
	24-331	BV	pFB-LIC-Bse	N-T, His (Tev)	4.1	++
CCNK	11-267	BV	pFB-LIC-Bse	N-T, His (Tev)	3.2	-
Kcyclin	1-257	*E. coli*	pRSETA A	N-T, His (Thrombin)	0.9-2.7††	-
	1-257	*E. coli*	pNIC-Zb	N-T, His (Tev)	5.9	-

Appendix A11. Summary of crystallization trials for CDK 14 with or without CycY and CDK17, with various inhibitors.

Protein(s)	Protein Concentration (mg/mL)	Coarse screens	Additions	Temperature (°C)
$CDK14^{1-451}$	8.7	JCSG, LFS	SU951	20
$CDK14^{114-437}$	12.2	HIN	1 mM Indirubin E804	4
$CDK14^{114-437}$/ $CycY^{9-331}$	6.1	All 4	1 mM Indirubin E804	4 or 20
$CDK14^{1-451}$/ $CycY^{9-331}$	9.9	All 4	10ug/ml chymotrypsin	4 or 20
	9.9	All 4	1 mM ASC29	4 or 20
	10.5	HIN, HCS, JCSG	1 mM SU9516	4 or 20
$CDK17^{188-523}$	6.3	LFS	1 mM Indirubin E804	20
$CDK17_{189-505}$	7	JCSG, LFS	1 mM Cdk/Crk Inhibitor	4
	6.8	All 4	0.4 mM ASC29; 0.9 mM CDK2 peptide substrate	4 or 20
$CDK17_{137-505}$	7.2	All 4	-	4 or 20
$CDK17_{137-505}$	7.2	All 4	0.8 mM ponatinib*	4 or 20

*Identified in a DSF assay. Data not shown.