Biochemical characterization, regulation, and inhibition of human transcription kinases CDK12 and CDK13 and human cell cycle-related kinase CDK14

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Sofia Dust
aus Peking, China

Bonn, 2019

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. rer. nat. Matthias Geyer
2. Gutachterin: Prof. Dr. rer. nat. Diana Imhof

Tag der Promotionsprüfung: 13.11.2019
Erscheinungsjahr: 2019

Parts of this thesis are published in:

Johnson, S.F., Cruz, C., Greifenberg, A.K., Dust, S., Stover, D.G., Chi, D., Primack, B., Cao, S., Bernhardy, A.J., Coulson, R., Lazaro, J.-B., Kochupurakkal, B., Sun, H., Unitt, C., Moreau, L.A., Sarosiek, K.A., Scaltriti, M., Juric, D., Baselga, J., Richardson, A., Rodig, S., D'Andrea, A.D., Balman J., Johnson, N., Geyer, M., Serra, V., Lim, E. and Shapiro, G.I. 2016. CDK12 Inhibition Reverses De Novo and Acquired PARP Inhibitor Resistance in BRCA Wild-Type and Mutated Models of Triple-Negative Breast Cancer. *Cell Reports*, **17**: 2367–2381.

Krajewska, M., Dries, R., Grassetti, A.V., **Dust, S.**, Gao, Y., Huang, H., Sharma, B., Day, D.S., Kwiatkowski, N., Pomaville, M., Dodd, O., Chipumuro, E., Zhang, T., Greenleaf, A.L., Yuan, G.-C., Gray, N.S., Young, R.A., Geyer, M., Gerber, S.A. and George, R.E. 2019. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nature Communications*, **10**(1):1757.

Insco, M.L., Abraham, B.J., Dubbury, S.J., **Dust, S.**, Chen, K.Y., Liu, D., Wu, C., Ludwig, C.G., Fabo, T., Henriques, T., Adelman, K., Geyer, M., Sharp, P.A., Young, R.A., Boutz, P.L. and Zon, L.I. submitted. CDK13 Mutations Drive Melanoma via Accumulation of Prematurely Terminated Transcripts. *Manuscript submitted for publication*.

Table of contents

List of Abbreviations I
List of Figures V
List of Tables/Equations IX

1 Abstract 1

2 Introduction 3
2.1 DNA Organization 3
2.2 Regulation of gene expression in eukaryotes 4
2.3 Transcription in eukaryotes 5
2.3.1 Transcription cycle 6
2.3.2 C-terminal domain of RNA polymerase II 9
2.3.3 CTD kinases 11
2.3.3.1 Cyclin dependent kinases 12
2.3.3.2 Other CTD Kinases 13
2.4 Structural features of Cyclin dependent kinases 15
2.5 CDK12/Cyclin K and CDK13/Cyclin K 17
2.5.1 Domain architecture of CDK12, CDK13 and Cyclin K 18
2.5.2 Structural features of CDK12 and CDK13 kinase complex 19
2.5.3 CDK12 and CDK13 in disease 21
2.6 Understudied CDKs 23
2.6.1 CDK14/Cyclin Y 24

3 Materials and Methods 27
3.1 Materials 27
3.1.1 Chemicals 27
3.1.2 Nucleotides 27
3.1.3 Kits 27
3.1.4 Marker/size standards 28
3.1.5 Consumables 28
3.1.6 Vectors 28
3.1.7 Oligonucleotides 29
3.1.8 Enzymes 29
3.1.9 Cell line and bacterial strains 29
3.1.10 Cell culture media and reagents 30
3.1.11 Chromatography columns/column materials 31
3.1.12 Devices 31

3.1.13 Commonly used buffers. 32

3.2 Molecular biological methods. 34

3.2.1 Electro competent cells. 34
3.2.2 Chemical competent cells. 34
3.2.3 Polymerase chain reaction (PCR). 34
3.2.4 Agarose gel electrophoresis. 37
3.2.5 Restriction digest. 37
3.2.6 Ligation. 38
3.2.7 Transformation of *E. coli* cells. 38
3.2.8 DNA-sequencing. 39
3.2.9 Generation of multi-gene expression cassettes via Cre-Lox recombination. 39
3.2.10 Transposition of transfer vectors into baculoviral genome. 40
3.2.11 Isolation of recombinant bacmid DNA. 40
3.2.12 Cultivation of Sf9 insect cells. 41
3.2.13 Transfection of Sf9 insect cells. 41
3.2.14 Virus amplification. 41

3.3 Biochemical methods. 42

3.3.1 Protein expression in *E. coli* cells. 42
3.3.2 Protein expression in *Sf9* insect cells. 42
3.3.3 Cell lysis. 43
3.3.4 Protein purification from inclusion bodies. 43
3.3.5 Affinity chromatography. 44
3.3.6 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). 46
3.3.7 Concentration of protein samples. 47
3.3.8 Determination of protein concentration. 48
3.3.9 Size-exclusion chromatography (SEC). 49
3.3.10 Mass spectrometry. 50
3.3.11 Kinase assays. 52
3.3.12 Pull-down assays. 53
3.3.13 Crystallization. 54

4 Results. 57

Transcriptional kinases – CDK12 and CDK13

4.1 Identification and characterization of potential CDK12 and CDK13 small molecule inhibitors 57

4.1.1 Impact assessment of dinaciclib. 57
4.1.1.1 Identification of potential CDK12 and CDK13 inhibitor by *in silico* modeling. 57
4.1.1.2 Analysis of the inhibitory activity of dinaciclib in comparison with flavopiridol 59

4.1.1.3 Study the effect of C-terminal extension helix mutations on the inhibitory potency of dinaciclib and flavopiridol. 63
4.1.1.4 Crystallographic analysis of co-crystallization of dinaciclib with CDK12. 67
4.1.2 Characterization of non-covalent and covalent CDK12 inhibitors under different conditions. 68
4.1.3 Structural elucidation of CDK12/Cyclin K co-crystallized with inhibitor. 70

4.2 Validation of pre-mRNA processing factors as new CDK12 and CDK13 substrates. 76

4.3 Characterization of CDK13 in human melanoma. 83

4.3.1 Structural and biochemical analysis of melanoma-associated CDK13 mutations. 83
4.3.2 Validation of CDK13 binding protein identified in human melanoma. 87

Cell cycle-related kinase – CDK14

4.4 Structural and biochemical analysis of CDK14/Cyclin Y. 90

4.4.1 Crystallization of the CDK14/Cyclin Y complex 90
4.4.2 Identification of CDK14/Cyclin Y binding proteins required for kinase activity 97
4.4.3 Study of CDK14–Cyclin Y–14-3-3 interactions in kinase complex formation 102

5 Discussion. 107

Transcriptional kinases – CDK12 and CDK13

5.1 Protein structure-based approach assisting in understanding of protein-ligand interaction. 107

5.2 Interpretation of the potency of non-covalent and covalent kinase inhibitors. 109

5.3 Phosphorylation activities of CDK12 and CDK13 contribute to co-transcriptional RNA processing. 111

5.4 Implication and substrate specificity of CDK13 in melanoma. 113

Cell cycle-related kinase – CDK14

5.5 Regulatory interplay between CDK14, Cyclin Y, and 14-3-3 in mediating CDK14/CycY association. 115

6 References. 121

7 Appendix. 133

8 Danksagung. 135

List of Abbreviations

Å	Ångström (10^{-10} m)
aa	amino acids
ATP	adenosine triphosphate
APS	ammonium persulfate
bp	base pair
BRCA1	breast cancer 1
Brd4	bromodomain-containing protein 4
β-ME	β-mercaptoethanol
CAK	Cyclin-dependent kinase activating kinase
CDC2	cell division cycle 2
CDK	Cyclin dependent kinase
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage stimulatory factor
CTD	C-terminal domain
CV	column volumes
Da	Dalton
DDR	DNA damage repair
DDR	DNA damage response
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DSIF	DRB sensitivity-inducing factor
DTE	Dithioerythrit
DTT	Dithiothreitol
DYRK	dual-specificity tyrosine-regulated kinase
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FGCZ	Functional Genomics Center Zurich
GSH	Glutathione
Gsk3ß	glycogen synthase kinase-3 beta
GST	Glutathione-S-Transferase
GTF	general transcription factor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HR	homologous recombination
IC_{50}	half-maximal inhibitory concentration
IP	Immunoprecipitation
IPTG	isopropyl β-D-1-thiogalactopyranoside
K_i	non-covalent binding constant
K_{inact}	inactivation constant
kb	kilo base
KD	kinase domain

kv	kilo volt
LB	lysogeny broth
LMW	low molecular weight
LRP6	low-density lipoprotein receptor-related protein 6
MALDI	matrix-assisted laser desorption/ionization
MAPK	mitogen-activated protein kinase
MBP	maltose-binding-protein
MES	2-(N-morpholino)ethanesulfonic acid
mRNA	messenger RNA
MS	mass spectrometry
m/z	mass-to-charge
MWCO	molecular weight cut off
NB	neuroblastoma
NBD	nucleotide-binding domain
NDSB	1-(3-Sulfopropyl)-Pyridinium Hydroxide
ncRNA	noncoding RNA
NELF	negative elongation factor
NTA	nitrilotriacetic acid
OD	optical density
pA	polyadenylation
PARP	poly (ADP-ribose) polymerase
PBS	phosphate buffered saline
PCPA	premature cleavage and polyadenylation
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PIC	pre-initiation complex
PMF	peptide mass fingerprint
PMSF	Phenylmethylsulforylfluorid
Plk3	Polo-like kinase 3
PRM	proline-rich motif
P-TEFb	positive transcription elongation factor b
PTM	posttranslational modification
RB	retinoblastoma protein
RMSD	root-mean-square deviation
rpm	revolutions per minute
RNA	ribonucleic acid
rRNA	ribosomal RNA
snRNA	small nuclear RNA
tRNA	transfer RNA
RNAP	RNA polymerase
RNAPI	RNA polymerase I
RNAPII	RNA polymerase II
RNAPIII	RNA polymerase III
RT	room temperature
SD	standard deviation

SDS–PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
Sf	*Spodoptera frugiperda*
SH3	Src-homology 3
SILAC	stable isotope labeling with amino acids in cell culture
SMART	simple modular architecture research tool
TCEP	Tris(2-carboxyethyl)phosphin
TCGA	The Cancer Genome Atlas
TEMED	tetramethylethylenediamine
TEV	Tobacco Etch Virus
Tris	tris(hydroxymethyl)aminomethane
TSS	transcription start site
TTS	transcription termination site
WT	wild type

Regular units of measurement as well as amino acids are abbreviated according to the IUPAC nomenclature.

List of Figures

Figure 1: The organization of DNA within the chromatin structure. 4
Figure 2: Overview of eukaryotic gene expression. 5
Figure 3: Overview of the eukaryotic transcription cycle. 8
Figure 4: Schematic illustration of the RNAPII CTD. 9
Figure 5: Schematic overview of the functional unit of the CTD. 10
Figure 6: Changes of CTD modifications at different sites of the heptad repeat along the transcription cycle. 11
Figure 7: Overview of the CDK family. 12
Figure 8: Schematic overview of the domain structure of individual mammalian CDK and their evolutionary relationships among each other based on comparison of the human kinase domains. 15
Figure 9: Characteristic structural features of cyclin dependent kinases. 16
Figure 10: Model of CDK activation exemplified by monomeric CDK2 and CDK2/CycA. . 16
Figure 11: Domain architecture of CDK12, CDK13 and Cyclin K. 19
Figure 12: Protein structure of human CDK12/CycK and CDK13/CycK complex. 20
Figure 13: Association of the C-terminal extension helix with the bound nucleotide and the kinase core domain of CDK13 and CDK12. 21
Figure 14: Genomic alterations of the CDK12 gene across TCGA. 22
Figure 15: Schematic illustration of the reported CDK13 pathogenic mutations on CDK13 protein sequence. 23
Figure 16: Schematic overview of the cell cycle-dependent Wnt signaling pathway induced by CDK14/CycY mediated activation of the LRP6 receptor during the G_2/M transition. 24
Figure 17: Domain architecture of CDK14 and Cyclin Y. 25
Figure 18: Sequence alignment of the human transcription regulating kinases in composed of CDK9, CDK12 and CDK13 and superimposition of the CDK9 crystal structure with the crystal structure of CDK12. 58
Figure 19: Modeling of flavopiridol and dinaciclib into the ATP-binding site of CDK12. . .59
Figure 20: *In vitro* kinase assays without and with addition of 10-fold and 1000-fold dinaciclib using purified recombinant kinase complexes of CDK9/CycT1, CDK12/CycK, and CDK13/CycK. 60
Figure 21: Dose-dependent inhibition measurements of dinaciclib with different transcription regulating kinases. 60
Figure 22: Dose-dependent inhibition measurements of dinaciclib with different cell cycle-related kinases. 61
Figure 23: Dose-dependent inhibition measurements of dinaciclib and flavopiridol with CDK9 and CDK12. 62
Figure 24: Site directed mutagenesis of C-terminal extension helix of CDK12 and CDK13. 63
Figure 25: *In vitro* kinase assay for kinase activity validation of CDK12/13 mutants. . . . 64
Figure 26: Analysis of the effect of specific CDK12 C-terminal extension helix mutations on the inhibitory potency of dinaciclib and flavopiridol. 66
Figure 27: Crystallographic approach of co-crystallization of dinaciclib with CDK12 H1040G mutant. 67

Figure 28: Dose-dependent inhibition measurements of non-covalent reversible and covalent irreversible CDK12 inhibitors under different conditions. 69
Figure 29: Dose-dependent inhibition measurements of CMP-175 with CDK12/13. 71
Figure 30: Purification of recombinant CDK12/CycK complex for co-crystallization of CDK12/CycK with CMP-175. 72
Figure 31: Examples of obtained protein crystals of CDK12/CycK co-crystallized with CMP-175. 72
Figure 32: Electron density maps showing incorporation of CMP-175 into CDK12 via covalent targeting. 73
Figure 33: CMP-175 co-crystal structure with CDK12/CycK. 75
Figure 34: Overview of generated constructs of CDK12/13 top candidate substrates provided by phosphoproteomic analysis. 77
Figure 35: *In vitro* kinase time course measurements of CDK12/CycK- and CDK13/CycK-mediated phosphorylation of the indicated candidate substrates. 78
Figure 36: *In vitro* kinase time course measurements of transcription elongation kinases-mediated phosphorylation of GST-SPT6H (1323-1534) 79
Figure 37: Dose-dependent inhibition measurements of CDK12/CycK and CDK13/CycK by THZ531 with different potential CDK12/13 substrates. 80
Figure 38: Mass spectrometry spectra of CDC5L (370-505), SF3B1 (113-462) and SPT6H (1323-1534) after incubation with CDK12/CycK. 81
Figure 39: Representation of melanoma-associated CDK13 kinase domain mutations predicted to affect kinase activity. 84
Figure 40: Overview of the generated CDK13 constructs containing melanoma-associated single mutation. 86
Figure 41: *In vitro* kinase assay of wild type and melanoma patient-mutated CDK13. 86
Figure 42: Overview of the generated constructs of ZC3H14 protein. 87
Figure 43: *In vitro* kinase time course measurement of CDK13/CycK-mediated phosphorylation of ZC3H14 (312-736) and MBP-ZC3H14 (591-736) 88
Figure 44: *In vitro* kinase time course measurement of CDK9/CycT1-mediated phosphorylation of ZC3H14 (312-736) and MBP-ZC3H14 (591-736) 89
Figure 45: Homology modeling of CDK14 using SWISS-MODEL. 91
Figure 46: Overview of CDK14 protein expression constructs intended for crystallization trials. 92
Figure 47: Homology modeling of Cyclin Y using SWISS-MODEL. 93
Figure 48: Overview of Cyclin Y protein expression constructs intended for crystallization trials. 93
Figure 49: Overview of selected combinations of truncated CDK14 and Cyclin Y constructs intended for crystallization trials. 94
Figure 50: ESI-Mass spectrometry analysis of unlabeled CDK14 and CDK14 labeled with compound. 95
Figure 51: Examples of generated protein crystals of the two CDK14/CycY complexes. . . .96
Figure 52: Overview of generated fusion complex of full length CDK14/Cyclin Y. 97
Figure 53: *In vitro* kinase activity assay to test the activity of differently formed CDK14/CycY complexes. 98
Figure 54: ESI-Mass spectrometry analysis of CDK14 (125-469) 99
Figure 55: *In vitro* kinase activity assay of CDK14/Cyclin Y with various inhibitor concentrations in association with increasing pre-incubation times. 100

Figure 56: Illustration of the identified proteins by peptide mass finger print analysis. . . .101
Figure 57: Overview of full length CDK14 and Cyclin Y separately expressed in *Sf9* and *E. coli* cells. 102
Figure 58: GST-pull-down assay of GST-Cyclin Y (*E. coli*) with different 14-3-3 isoforms. 103
Figure 59: GST-pull-down assay of GST-CDK14 with Cyclin Y without and with binding of 14-3-3 proteins. 104
Figure 60: Different full length Cyclin Y proteins were sent to peptide mass finger print analysis to identify specific phosphorylation sites. 104
Figure 61: Overview of identified phosphorylation sites in different Cyclin Y constructs by peptide mass fingerprint analyses. 105
Figure 62: Co-expression of full length Cyclin Y(S100A) mutant with full length CDK14 as fusion construct in *Sf9* cells. 106
Figure 63: Comparison of the action of a target-specific inhibitor under non-covalent reversible and covalent irreversible binding condition. 110
Figure 64: Structure of a 14-3-3 dimer in complex with phosphorylated 14-3-3 binding motif. 115
Figure 65: Possible regulatory role of 14-3-3 dimer contributing to CDK14/CycY association. 117
Figure 66: Illustration of the potential location of S100 and S326 on the homology model of Cyclin Y. 118
Figure 67: Proposed model of the regulatory interplay between CDK14, Cyclin Y and 14-3-3 in mediating CDK14/CycY association. 120

List of Tables

Table 1: Overview of the CTD kinases and their function in modifying RNAPII CTD. . . .14
Table 2: Nucleotides and manufacturers. 27
Table 3: Kits and manufacturers. 27
Table 4: Marker/size standards and manufacturers. 28
Table 5: Consumables and manufacturers. 28
Table 6: Vectors and manufacturers. 28
Table 7: Enzymes and manufacturers. 29
Table 8: Bacterial strains and cell line. 29
Table 9: Cell culture media and reagents. 30
Table 10: Chromatography columns/column materials and manufacturers. 31
Table 11: Devices and manufacturers. 31
Table 12: Commonly used buffers and their composition. 32
Table 13: Composition of PCR reaction mixture. 35
Table 14: Standard PCR running conditions. 35
Table 15: Composition of colony PCR reaction mixture. 36
Table 16: SDS-PAGE composition. 47
Table 17: Proteins applied to crystallization trials, including applied protein concentration and the used precipitant agent. 55
Table 18: Overview of the determined IC_{50} values of dinaciclib for various CDK family members. 61
Table 19: Overview of determined IC_{50} values of dinaciclib and flavopiridol for CDK12 mutants and wild type protein. 66
Table 20: Overview of determined IC_{50} values for a non-covalent reversible and for a covalent irreversible CDK12 inhibitor under different conditions. 69
Table 21: Comparative overview of identified phosphorylation sites of CDK12/CDK13 candidate substrates by performing phosphoproteomic analyses of neuroblastoma cells treated with THZ531 and peptide mass finger print analyses of recombinant proteins *in vitro*. 82
Table 22: Overview of the predicted implications of six melanoma-associated CDK13 kinase domain mutations. 85

List of Equations

Equation 1: Calculation of ligation mixture. 38
Equation 2: Lambert-Beer's law. 48
Equation 3: Calculation of molar absorption coefficient. 48

1 Abstract

The transcription process is the first step in the molecular information flow from genome to proteome. In eukaryotes, the RNA polymerase II (RNAPII) transcribes genes that code for proteins by synthesizing pre-mature messenger RNAs (mRNAs). The Cyclin-dependent kinases CDK12 and CDK13 in association with Cyclin K are transcription regulating kinases by modulating the function of RNAPII. Both CDK/Cyclin pairs regulate transcriptional elongation as well as processes occurring co-transcriptionally through phosphorylation of the C-terminal domain (CTD) of RNAPII. They selectively affect the expression of genes involved in DNA damage response (DDR) and mRNA processing, respectively. CDK12 and CDK13 are involved in numerous types of cancer, representing potential targets for novel cancer treatments. Small molecule inhibitors that selectively target CDK12 and CDK13 in an ATP-competitive manner are used to investigate the consequences of their inhibition in healthy cells and cancer cells.

In this thesis, using a protein structure-based approach, dinaciclib, a small molecule inhibitor of CDKs 1, 2 and 5, was identified as a potent inhibitor of CDK12 and CDK13. In comparison to flavopiridol, a known transcription-regulating CDK inhibitor, inhibition of CDK12 by dinaciclib was significantly more potent than by flavopiridol. By performing mutagenesis experiments of residues in the C-terminal extension helix of CDK12, we show that the histidine residue of the DCHEL motif causes a sterical hindrance that may reduce the inhibitory activity of flavopiridol. We also conclude that dinaciclib may readily accommodate its pyridine ring at the histidine binding site of the kinase leading to highly potent inhibition of CDK12 and CDK13. This might be achieved by a displacement mode of binding rather than by performing a base stacking interaction with the histidine imidazole ring. In a further approach, non-covalent reversible and covalent irreversible ATP-competitive small molecule inhibitors were characterized at different ATP concentrations and pre-incubation times. The result indicates that the potency of irreversible inhibitors cannot be easily characterized by conventional IC_{50}-values as the kinetics of the covalent interaction involves a time dependency. In order to understand how small molecule compounds selectively target the kinase hinge region in CDK12, we determined the three-dimensional complex structure of CDK12/CycK co-crystallized with an irreversible ATP-competitive small molecule inhibitor at 2.6 Ångstrom resolution.

In a validation study, our findings derived from *in vitro* kinase activity assays and peptide mass finger print analyses confirmed that CDK12 and CDK13 in complex with Cyclin K directly phosphorylate the pre-mRNA processing proteins CDC5L, SF3B1, CSTF64 and SPT6H. The validation of pre-mRNA processing factors as new CDK12 and CDK13 substrates sheds light on the diverse regulatory function of the kinases in the process of co-transcriptional RNA processing. Structural and biochemical analysis of melanoma-associated CDK13 mutations supported the

assumption that mutations near the kinase ATP-binding site may represent the leading cause for the development of cutaneous melanoma by inactivating the catalytic center of the kinase. The zinc-finger containing protein ZC3H14, which was supposed as a CDK13 interacting protein in human melanoma cells, could not be confirmed as a CDK13/CycK phosphorylation substrate in our kinase assay. However, using CDK9 in complex with Cyclin T1, phosphorylation activity was present against ZC3H14. This suggests that in melanoma cells, Cyclin T1 may be required to generate an active CDK13 kinase complex. This assumption is further supported by the fact that Cyclin T1 was identified as the cyclin binding partner for CDK13 in human melanoma and was demonstrated to be required for the CDK13-mediated melanomagenesis.

Besides the CDKs that regulate progression through the cell cycle and different stages of the transcription process, there is a third CDK subfamily consisting of the understudied kinases CDK14-CDK18. Human CDK14 belongs to the family of cell division cycle 2 (CDC2)-related kinases and is associated with Cyclin Y, which contains an N-terminal myristoylation signal for plasma membrane targeting. The kinase complex of CDK14/CycY is implicated in the Wnt/ß-catenin signaling pathway by phosphorylation of the transmembrane receptor protein involved in Wnt signaling. The present structural and biochemical analysis of the human CDK14/CycY kinase complex identifies 14-3-3 proteins as CDK14/CycY binding proteins which are essential for the kinase activity. Our data provides a possible new mechanistic insight into the activation of CDK14 by proposing a phosphorylation-dependent regulatory interplay between CDK14, Cyclin Y, and 14-3-3 proteins that enhances the kinase-cyclin association and thus facilitates the kinase activity.

2 Introduction

2.1 DNA Organization

In most organisms, the genetic information is stored in the form of DNA. 'Higher' organisms (eukaryotes), such as yeast and humans, package their DNA inside the nucleus, in a separate compartment within the cell. In humans, the nucleus of each cell contains 6 billion base pairs of DNA distributed over 23 pairs of chromosomes (Tjio and Levan, 2010), which is known collectively as the human genome (International Human Genome Sequencing Consortium, 2004). 22 of these pairs are autosomes, which are located in somatic cells and look the same in both males and females. The 23rd pair, which are found in germ cells, consists of the two sex chromosomes, X and Y. Females have two copies of the X chromosome, while males have one X and one Y chromosome.

In eukaryotic cells the nuclear DNA is packed within the nucleus by association with proteins forming a 'beads-on-a-string' structure called chromatin (Fussner *et al.*, 2011; Woodcock *et al.*, 1976). The primary subunit of chromatin, also the lowest level of the DNA organization, is the nucleosome, which contains the histone proteins responsible for the vast degree of DNA packaging. The nucleosome architecture consists of approximately 146 bp of DNA wrapped in two superhelical turns around an octamer of core histones containing two copies of each of the four core histones (H2A, H2B, H3 and H4). This arrangement results in a five- to ten-fold compaction of the DNA (Kornberg, 1974). The nucleosomes are positioned every ~200 bp on the DNA strand and connected to its neighbors by a short segment of linker DNA (~10–80 bp in length). This polynucleosome string is folded into a compact fibre with approximately 30 nm in diameter, which proposed to be the early stage in compaction (Finch and Klug, 1976). Even though several models have been proposed to describe the general structure of the 30 nm chromatin fibre, the arrangement of nucleosomes within the fibre is unknown and even less is known about the structure levels beyond the fibre in which they are further packed to form highest-order structures (van Holde and Zlatanova, 1995). The organization of the DNA within the chromatin structure is shown in figure 1.

As the compact form of chromatin is inaccessible to many proteins involved in the transcription machinery, the chromatin structure can be altered by three general ways to regulate the accessibility of genes to transcription. These functions involve post-translational modifications of histone tails (Zhang and Reinberg, 2001), incorporation of histone variants by replacing one or more of the core histones (Ahmad and Henikoff, 2002; Redon *et al.*, 2002; Smith, 2002), and nucleosome remodeling, which typically requires energy derived from ATP hydrolysis (Becker and Hörz, 2002).

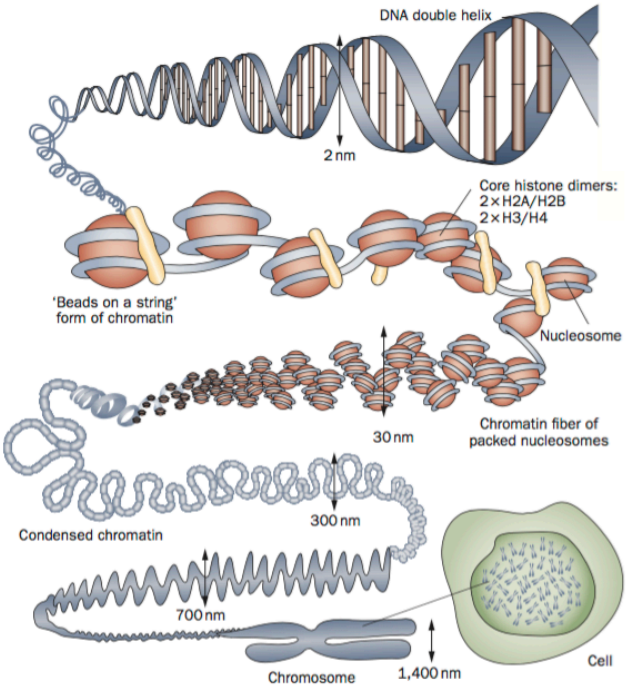


Figure 1: The organization of DNA within the chromatin structure. The lowest level of organization is the nucleosome, composed of 146 bp of genomic DNA (2 nm in diameter) wrapped around a histone octamer that consists of two copies of the core histones (H2A, H2B, H3 and H4). Nucleosomes are connected to one another by short linker DNA. At the next level of organization, the string of nucleosomes is folded into a chromatin fibre about 30 nm in diameter and these fibres are further folded into higher-order structures (Tonna *et al.*, 2010).

2.2 Regulation of gene expression in eukaryotes

The human genome is thought to contain ~20.000 protein-coding genes and at least as many genes of noncoding RNA (ncRNA) (Djebali *et al.*, 2012). Some of the ncRNAs contribute to control of gene expression through modulation of transcriptional or post-transcriptional processes (Bartel, 2009; Lee, 2012; Orom and Shiekhattar, 2011; Rinn and Chang, 2012), but not much is understood yet. The process of gen expression, in which the genetic information flowing from a DNA sequence to messenger RNA (mRNA) then to a protein inside cells, is a fundamental principle of the molecular biology that is often referred to as the central dogma of molecular biology, first stated by Francis Crick and Sydney Brenner in 1957 (Brenner and Crick, 1957; Cobb, 2017) (Figure 2). Gene expression involves two main stages: transcription and translation. Both processes are highly regulated in humans and take place in different cellular compartments. Regulation occurs at many points during the transcription and translation processes and involves proteins that can attach to DNA and influence the gene expression program. During the process of transcription, both coding and

noncoding regions of DNA are transcribed into mRNA. During the following mRNA processing step, genes that do not code for amino acids (introns) are removed. The spliced mRNA molecule is then prepared for export out of the nucleus to the cytoplasm through 5´-endcapping and polyadenylation. Once in the cytoplasm, the mRNA is used to direct protein synthesis, initiating the process of translation. Here, the mRNA binds to ribosomes, which synthesize a protein based on the mRNA sequence. In this study, the focus is on the regulation of the transcription process.

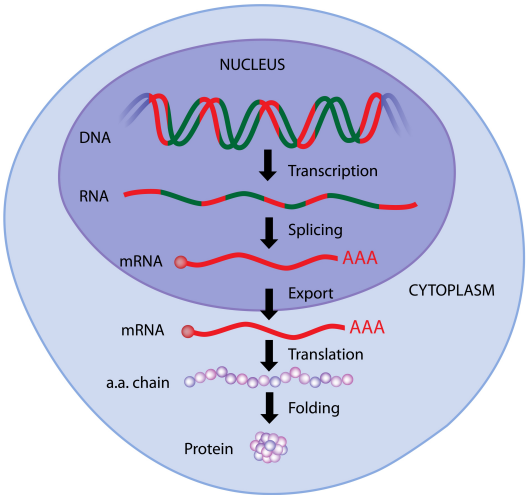


Figure 2: Overview of eukaryotic gene expression. Eukaryotic gene expression involves the successive processes of transcription and translation. Following transcription and RNA processing in the nucleus, protein translation takes place in the cytoplasm (Healio Learn Genomics, 2019).

2.3 Transcription in eukaryotes

Transcription is the first step in the universal pipeline of genetic information flow from genome to proteome. In the process of transcription, a particular portion of a single stranded DNA nucleotide sequence is copied into a single stranded RNA nucleotide sequence complementary to the DNA template. Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds. The main chemical differences between both polymers include the nucleotide backbone and one base exchange. Instead deoxyribose, RNA contains the sugar ribose. The base thymine (T) is replaced by the cognate base uracil (U), still allowing complementary base-pairing with adenine (A).

The transcription process is carried out by RNA polymerases (RNAP) along with several accessory proteins factors. When active on DNA, the RNAP is usually complexed with these protein factors,

which recruit cofactors and RNAP to the appropriate transcription site of the target genes by binding to enhancer elements and promoters (Lelli *et al.*, 2012; Ong and Corces, 2011; Spitz and Furlong, 2012). After opening and unwinding of a small portion of the DNA double helix to expose the bases for the transcription process, RNAP attaches to the template DNA strand for the synthesis of the complementary RNA strand. Here, the RNAP moves stepwise along the DNA and catalyzes the formation of the phosphodiester bonds in an enzymatically catalyzed reaction. In this way, the RNA strand is elongated one nucleotide at a time in the 5′-to-3′ direction. Just behind the region where the ribonucleotides are being added, the RNA chain is displaced from the hydrogen bond to the DNA and released as a single strand. The DNA returns to the double helix state hereafter.
In contrast to bacteria, which has only one RNAP, three different types of RNAP exist in eukaryotic cells: RNA polymerase I (RNAPI), RNA polymerase II (RNAPII), and RNA polymerase III (RNAPIII). While RNAPI and III transcribe genes that encode most of the ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) as well as various small regulatory RNAs, respectively, RNAPII transcribes genes that encode proteins by synthesizing mRNAs (Cramer *et al.*, 2008). Therefore, this thesis is focused on the RNAPII-dependent transcription process required for the synthesis of proteins.

2.3.1 Transcription cycle

The cycle of eukaryotic transcription, from initiation to elongation and termination is regulated at multiple steps (Figure 3). Coordinated action of various regulatory factors regulates the transcriptional ability of RNAPII at different stages.

Transcription initiation

Unlike bacterial RNA polymerase, eukaryotic RNA polymerase is not able to initiate transcription on a DNA template without the help of additional proteins called general transcription factors (GTFs) (Samuels *et al.*, 1982). Together they form the transcription pre-initiation complex (PIC) including the GTFs TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH, ensuring the accurate initiation of transcription (Flores *et al.*, 1989, 1992; Reinberg and Roeder, 1987; Sawadogo and Roeder, 1985). PIC formation usually begins with TFIID binding to the TATA box, a specific core promoter sequence primarily composed of T and A nucleotides typically located 25 nucleotides upstream from the transcription start site (TSS). The TATA box is the most important DNA sequence that signals the start of transcription for most of the RNAPII promoters (Baumann *et al.*, 2010; Mathis and

Chambon, 1981). Binding of TFIID to the TATA box is followed by the assembly of other GTFs and RNAPII at the promoter through either a sequential assembly pathway (Fire *et al.*, 1984; Hawley and Roeder, 1985; Samuels and Sharp, 1986) or a preassembled RNAPII holoenzyme pathway (Chao *et al.*, 1996; Kim *et al.*, 1994; Ossipow *et al.*, 1995). After recruitment to the promoter, TFIIH, which contains a DNA helicase, breaks down the DNA double helix allowing access to the template strand at the TSS. In addition, TFIIH also catalyzes the promoter escape of RNAPII by phosphorylation of its C-terminal domain (CTD), as TFIIH contains the Cyclin dependent kinase 7 (CDK7) as one of its subunits (Cho *et al.*, 1997; Rodriguez *et al.*, 2000; Schroeder *et al.*, 2000). Besides the GTFs, transcription initiation further requires activator proteins. These proteins bind to the enhancer DNA elements and promote a sequential recruitment of the GTFs, RNAPII and the Mediator complex to the target gene promoter. The Mediator complex is essential for the formation of the transcription PIC by bridging the GTFs with RNAPII (Flanagan *et al.*, 1991). In addition, activator proteins also attract chromatin-remodeling complexes and histone acetylases in order to overcome the difficulty of binding to DNA packaged in high-order chromatin structures (Kuras and Struhl, 1999).

Transcription elongation

Once the recruited RNAPII begins to elongate the RNA transcript, it usually transcribes a short distance, typically 20–50 bp, until the promoter-proximal pausing of RNAPII takes place (Adelman and Lis, 2012). The paused RNAPII is controlled by the protein complexes DRB sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF), which bind to and inhibit the function of RNAPII (Vos *et al.*, 08 2018a). To facilitate the entry of RNAPII into the productive mode of elongation, pause release and subsequent formation of an activated RNAPII elongation complex is modulated through the kinase activity of the positive transcription elongation factor b (P-TEFb) and the elongation factors PAF1 complex (PAF) and SPT6. PAF is required to displace NELF from the RNAPII funnel for pause release. After phosphorylation of the RNAPII linker to the CTD by CDK9, which is the associated kinase of the P-TEFb complex, SPT6 binds to the phosphorylated sites and opens the RNA clamp formed by DSIF (Vos *et al.*, 8/2018b). The P-TEFb-mediated release of RNAPII from the pause sites allows the transcription elongation to proceed. At this point, most of the GTFs are released from the DNA. RNAPII traverses through the coding region of the gene and generates a RNA copy by adding nucleotides to the 3´ end of the growing RNA chain with a pace of approximately 22-25 nucleotides per second in eukaryotes (Izban and Luse, 1992).

Transcription termination

After completion of the pre-mRNA synthesis, the end of a transcription cycle is indicated by dissociation of the RNAPII from the DNA template. For the most RNAPII transcribed protein-coding

genes, transcription termination is functionally connected with cleavage and polyadenylation of the nascent transcript at the 3´ end (Zaret and Sherman, 1982). Most of the protein coding pre-mRNAs consist a highly conserved poly(A) signal, 5′-AAUAAA-3′, which is followed by a G/U-rich sequence towards the 3′ end. Transcription of the poly(A) signal triggers the endonucleolytic cleavage of the nascent transcript, result in an upstream cleavage product that is immediately polyadenylated (Colgan and Manley, 1997; Zhao *et al.*, 1999) and a downstream cleavage product that is rapidly degraded (Manley *et al.*, 1982). In eukaryotes, several protein complexes are responsible for recognizing the poly(A) signal and catalyzing the reactions of cleavage and polyadenylation. These include the cleavage and polyadenylation specificity factor (CPSF), the cleavage stimulatory factor (CstF) and poly(A) polymerase (Kuehner *et al.*, 2011). Here, the phosphorylated CTD of RNAPII acts as a scaffold for the recruitment of these termination complexes to the poly(A) signal (Ahn *et al.*, 2004).

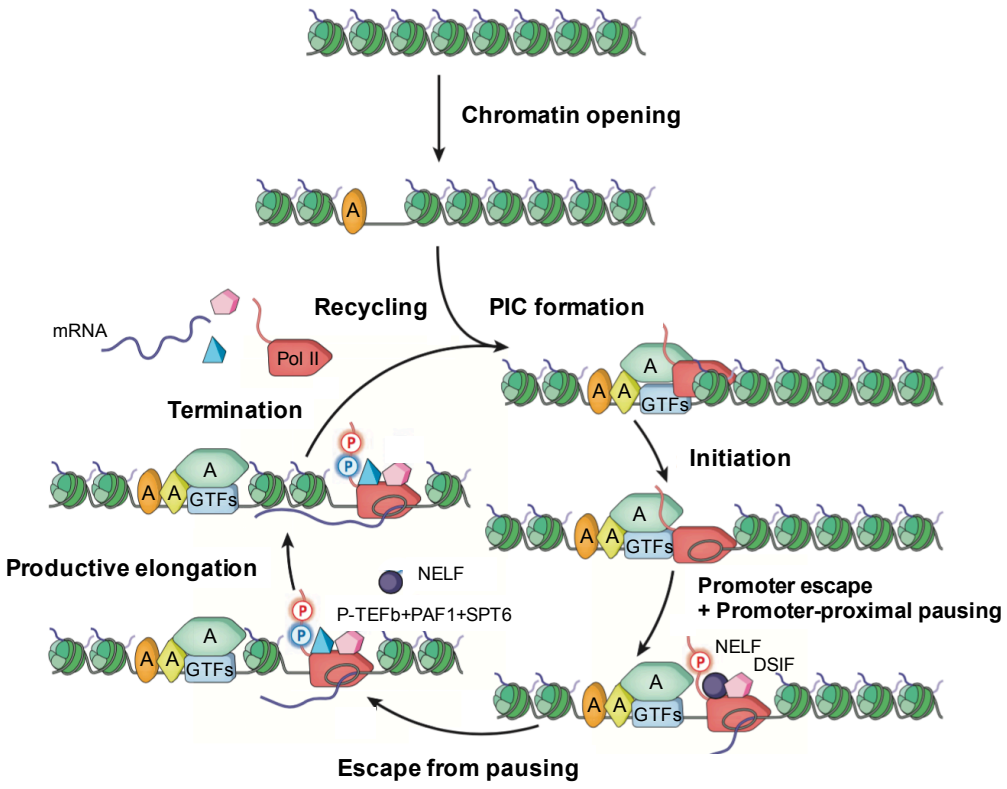


Figure 3: Overview of the eukaryotic transcription cycle. The transcription cycle is a multistep process including chromatin opening by binding of an activator protein; PIC formation by binding of additional activator proteins, facilitating GTFs and RNAPII to form the PIC; Transcription initiation by unwinding the DNA at the TSS; Promoter escape followed by promoter-proximal pausing after transcription of 20–50 bases downstream of the TSS; Escape from pausing by recruitment of the protein complex consisting of P-TEFb, PAF1 and SPT6; Productive elongation by the active RNAPII elongation complex transcribing through the gene; Transcription termination by cleavage and polyadenylation of the RNA transcript and removing of RNAPII from the DNA; Recycling of the freed RNAPII. Modified figure based on (Fuda *et al.*, 2009).

2.3.2 C-terminal domain of RNA polymerase II

As described above, the process of eukaryotic transcription and the concomitant pre-mRNA processing require a precise coordination and recruitment of specific sets of transcription-associated factors at specific stages of the transcription cycle. Research over the last decades has confirmed the C-terminal domain (CTD) of RNAPII as a selective and flexible platform for the recruitment of transcription factors, RNA processing factors, and chromatin modifiers, which affect the synthesis of the mRNA (Egloff and Murphy, 2008; Meinhart *et al.*, 2005; Phatnani and Greenleaf, 2006). In contrast to the other two eukaryotic RNA polymerases, Rpb, the largest subunit of RNAPII, carries a heptapeptide repeat structure of the consensus sequence Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 at its CTD (Corden *et al.*, 1985). The number of these heptad repeats varies from organism to organism and appears to correlate with genomic complexity. The length of the CTD differs markedly between mammals and *S. cerevisiae*, with 26 and 52 repeats, respectively (Figure 4) (Allison *et al.*, 1988; Corden, 1990).

Figure 4: Schematic illustration of the RNAPII CTD. Comparison of the CTDs between mammals (*Homo sapiens*) and *S. cerevisiae*. Divergences from the consensus heptad repeat sequence YSPTSPS are highlighted in blue (Eick and Geyer, 2013)**.**

The five amino acids Tyr1, Ser2, Thr4, Ser5, and Ser7 can all be reversibly modified by phosphorylation of the side chain hydroxyl group (Chapman *et al.*, 2007; Heidemann and Eick, 2012; Hintermair *et al.*, 2012; Mayer *et al.*, 2012). In addition, Thr4 (Kelly *et al.*, 1993), Ser5 and Ser7 (Ranuncolo *et al.*, 2012) also have been described to be O-glycosylated. Together with the two proline configurations, which can exist in either *cis*- or *trans*-configuration, a single heptad repeat contains 432 possible combination of post-translational modifications and proline configurations. In the functional unit of the CTD, which comprises one full repeat including the next four residues of the following repeat (Figure 5), the possible number of combinations is 10368.

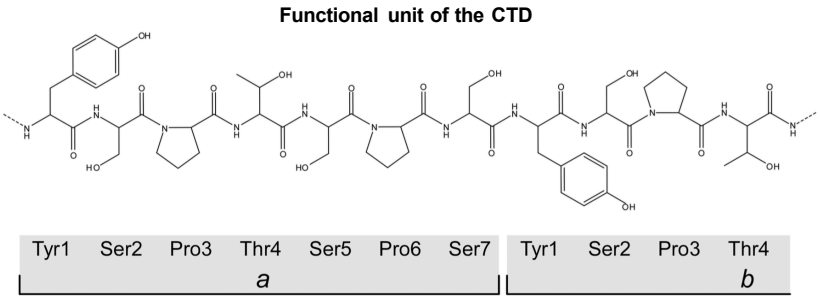


Figure 5: Schematic overview of the functional unit of the CTD. It comprises one full repeat including the next four residues of the following repeat (Eick and Geyer, 2013).

The dynamic change of the CTD phosphorylation patterns during the transcription cycle is regulated by a series of site-specific phosphorylation and dephosphorylation events, which occur primarily at CTD Ser2, Ser5, and Ser7 residues (Bataille *et al.*, 2012; Brookes and Pombo, 2009; Kim *et al.*, 2010; Mayer *et al.*, 2010; Odawara *et al.*, 2011; Tietjen *et al.*, 2010). In general, Ser5 and Ser7 phosphorylation (Ser5-P and Ser7-P) correlates with transcription initiation. Events that are functionally related to Ser5-P are for example the specific binding of enzymes operating on the 5'-capping process. Other protein factors of the initiation process such as histone modifying enzymes, also make use of the Ser5-P to determine its association with the CTD (Buratowski, 2009). While Ser5-P steadily decreases toward the termination, Ser2 phosphorylation (Ser2-P) increases during the elongation process and reaches the peak at the poly(A) site. It is known that Ser2-P correlates with the process of polyadenylation (Ahn *et al.*, 2004) and splicing (Yoh *et al.*, 2007). Ser7 phosphorylation (Ser7-P) is present at promoter and coding regions of active genes in mammalian cells (Chapman *et al.*, 2007) and is thought to occur together with Ser5-P and Ser2-P (Akhtar *et al.*, 2009; Tietjen *et al.*, 2010). Until now, the only event associated with Ser7-P is the recruitment of the integrator complex for the snRNAs 3'-processing in mammalian cells (Buratowski, 2009). Tyr1 phosphorylation (Tyr1-P) has been shown to favor the binding of elongation factors and prevent

binding of termination factors to the CTD (Mayer *et al.*, 2012). As RNAPII with a hypo-phosphorylated CTD, RNAPIIA, preferentially associates with the transcription PIC (Lu et al., 1991), all CTD phospho-marks are removed by phosphatases during the process of termination before starting a new transcription cycle. The changes of the modifications at the different sites of the CTD heptad repeat along the transcription cycle is visualized in figure 6 and is often referred to as the 'CTD code' (Buratowski, 2003; Eick and Geyer, 2013).

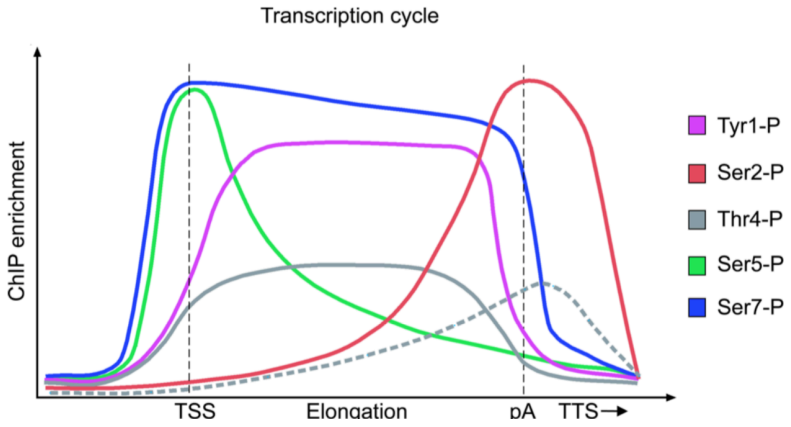


Figure 6: Changes of CTD modifications at different sites of the heptad repeat along the transcription cycle. The phosphorylation signals are visualized by ChIP experiments in *S. cerevisiae* (Bataille *et al.*, 2012; Kim *et al.*, 2010; Mayer *et al.*, 2010; Tietjen *et al.*, 2010). Dashed line shows Thr4-P marks in mammalian cells (Hintermair *et al.*, 2012). The transcription cycle is defined as the distance from the transcriptions start site (TSS) to the polyadenylation (pA) and transcription termination site (TTS) (Eick and Geyer, 2013).

2.3.3 CTD kinases

The phosphorylation states of the CTD and therefore its binding properties are modulated by the activity of various CTD kinases and phosphatases, allowing the CTD to undergo a dynamic cycle of post-translational modifications. The activity of a protein kinase involves transferring the gamma phosphate group of an adenosine triphosphate (ATP) molecule to specific amino acid residues trough covalent attachment. The free hydroxyl group of a serine, threonine or tyrosine residue provides the major phosphorylation sites for many of these kinases. Depending on their specificity towards the substrate amino acids, these protein kinases are known as: Serine/threonine-specific protein kinases, acting on both serine and threonine. Tyrosine-specific protein kinases, acting only on tyrosine, and the dual specificity kinases, acting on all three residues (Dhanasekaran and Reddy, 1998).

2.3.3.1 Cyclin dependent kinases

Cyclin dependent kinases (CDKs) belong to the serine/threonine kinases and partner with regulatory subunits known as cyclins, which control the kinase activity and substrate specificity (Peeper *et al.*, 1993; J Pines, 1995; Schulman *et al.*, 1998). The serine/threonine phosphorylation activity with preference for the (S/T)-P-X-(K/R) sequence (Errico *et al.*, 2010) is mostly proline-directed due to the presence of a hydrophobic pocket near the catalytic site (Echalier *et al.*, 2010). By comparing the sequence of the kinase domains, CDKs are determined to be members of the CMGC group of kinases (named after the initials of some members), along with mitogen-activated protein kinases (MAPKs), glycogen synthase kinase-3 beta (Gsk3β), members of the dual-specificity tyrosine-regulated kinase (DYRK) family and CDK-like kinases (Manning, 2002). Protein kinases belonging to the CDK family have been named as CDK1 through to CDK20 (Malumbres *et al.*, 2009).

The significant role of CDKs acting as a major eukaryotic protein kinase family involved in the modulation of the processes of transcription and cell cycle has been established many years before (Lim and Kaldis, 2013; Malumbres and Barbacid, 2009, 2005; Morgan, 1997). The phosphorylation states of the CTD is modulated by the kinase activity of the transcription-associated subset of the Cyclin-dependent kinases, known as transcriptional CDKs. In contrast to the cell cycle-related CDKs that bind multiple cyclins and regulate the cell cycle by displaying a periodic variation of activity in synchrony with the cell cycle, the transcriptional CDKs typically form complexes with a single specific cyclin, which are usually not regulated in a cell-cycle-dependent manner (Cao *et al.*, 2014; Liu and Kipreos, 2000) (Figure 7).

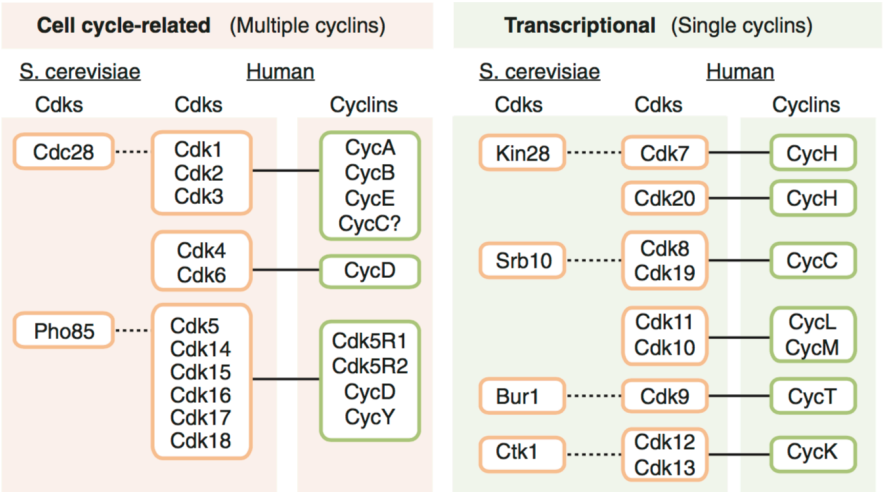


Figure 7: Overview of the CDK family. Comparison of yeast and mammalian CDKs within the group of cell cycle-related and transcriptional CDKs (Malumbres, 2014).

CDK7/Cyc H and its yeast ortholog Kin28/Ccl1 is the CTD kinase subunit of the general transcription factor TFIIH. CDK7 regulates the phosphorylation of CTD Ser5 and Ser7 residues during the processes of transcription initiation and promoter escape (Akhtar *et al.*, 2009; Glover-Cutter *et al.*, 2009; Hong *et al.*, 2009; Kanin *et al.*, 2007; Kim *et al.*, 2009; Viladevall *et al.*, 2009). The kinase/cyclin pair of mammalian CDK8 in association with Cyclin C (Srb10/Srb11 in yeast) is a subunit of the Mediator complex and has been described to have Ser2 and Ser5 phosphorylation activity *in vitro* (Liao *et al.*, 1995; Sun *et al.*, 1998). Although CDK8 has been considered as the only CTD kinase to be implicated in the repression of transcription and counteracts the function of CDK7 (Akoulitchev *et al.*, 2000; Hengartner *et al.*, 1998), findings also indicate that CDK8 plays a positive role in transcriptional activation (Galbraith *et al.*, 2010; Meinhart *et al.*, 2005). Unlike the other two kinases, CDK9 (Bur1 in yeast), the kinase subunit of P-TEFb in association with Cyclin T1, is not part of the transcription PIC but is recruited to the CTD after initiation. As mentioned above, CDK9 plays an essential role in the transition from transcription initiation to productive elongation by overcoming promoter-proximal pausing of RNAPII (Marshall *et al.*, 1996; Marshall and Price, 1995; Vos *et al.*, 8/2018b). For a long time, CDK9 was considered as the specific kinase for Ser2-P of the CTD. However, the characterization of CDK9 as a Ser2 CTD kinase was recently contradicted by several studies. Nowadays, CDK9 is supposed to preferentially phosphorylate Ser5 of the CTD rather than primarily conducts phosphorylation of Ser2 (Czudnochowski *et al.*, 2012; Ghamari *et al.*, 2013; Jones *et al.*, 2004; Ramanathan *et al.*, 2001, 1999). CDK9 also indicates a preference for substrates pre-phosphorylated at Ser7 (Czudnochowski *et al.*, 2012). Therefore, it is possible that the kinase activity of CDK9 mainly results in Ser5 phosphorylation at repeats that have already been phosphorylated at Ser7, producing doubly phosphorylated heptamers. Two CDKs, CDK12 and CDK13 in association with Cyclin K, the metazoan orthologs of yeast Ctk1, are additionally identified as CTD kinases. They have been found to display Ser2-P activity and are acknowledged as the primary kinases responsible for CTD Ser2-P in metazoan cells, modulating the process of transcriptional elongation. Unlike CDK7, 8 and 9, which are concentrated closer to the promoter region, CDK12 is distributed throughout the gene, supposed to maintain the Ser2-P process during the elongation phase (Bartkowiak *et al.*, 2010). Recently, CDK11 has also been supposed as Ser2-CTD kinase (Pak *et al.*, 2015).

2.3.3.2 Other CTD Kinases

Besides the transcriptional CDKs, the extracellular signal-regulated kinases, Erk1 and Erk2, from the MAP/Erk protein kinase family also shown phosphorylation activity towards Ser5 of the CTD *in*

vitro (Bellier *et al.*, 1997; Bonnet *et al.*, 1999). The bromodomain-containing protein 4 (Brd4), which is required to form the transcriptionally active P-TEFb complex by displacing negative regulators such as HEXIM1 and 7SK small nuclear RNA (snRNA) complex from P-TEFb (Jang *et al.*, 2005; Liu *et al.*, 2013, p. 4; Yang *et al.*, 2005), facilitates the phosphorylation of CTD Ser2, indicating a direct and active role for Brd4 in regulating transcription (Devaiah *et al.*, 2012; Zhang *et al.*, 2012). Polo-like kinase 3 (Plk3) has been described as a Thr4-specific kinase in mammalian cells (Hintermair *et al.*, 2012). Phosphorylation of Tyr1 is mediated by the proto-oncogene tyrosine kinases, c-Abl1 and c-Abl2, which phosphorylate the CTD to high levels *in vivo* and *in vitro* (Baskaran *et al.*, 1993, 1996) and inhibit the binding of the termination factor to RNAPII (Mayer *et al.*, 2012). An overview of the CTD kinases and their function in modifying the RNAPII CTD is shown in the table below.

Table 1: Overview of the CTD kinases and their function in modifying RNAPII CTD (Eick and Geyer, 2013).

Organism	**CTD Kinase**	**CTD modification**
Mammals	CDK7 (TFIIH)	Ser5-P, Ser7-P
	CDK8 (Mediator)	Ser2-P, Ser5-P
	CDK9 (P-TEFb)	Ser2-P, Ser5-P, Ser7-P
	CDK11	Ser2-P
	CDK12	Ser2-P
	CDK13	Ser2-P
	Plk3	Thr4-P
	cAbl1, cAbl2	Tyr1-P
	Erk1, Erk2	Ser5-P
	Brd4	Ser2-P
Saccharomyces cerevisiae	Kin28	Ser5-P, Ser7-P
	CDK8 (Srb10)	Ser2-P, Ser5-P
	Bur1	Ser2-P, Ser5-P, Ser7-P
	Ctk1	Ser2-P

2.4 Structural features of Cyclin dependent kinases

Cyclin dependent kinases comprise a conserved, central protein kinase domain, which is enclosed by N- and C-terminal sequences of variable length (Lim and Kaldis, 2013) (Figure 8). The CDK kinase fold, as first exemplified by the monomeric protein structure of CDK2 (De Bondt *et al.*, 1993), is composed of a smaller N-terminal lobe of twisted anti-parallel β-strands which is linked via a flexible hinge sequence to a larger C-terminal lobe, rich in α-helices. The kinase active site is located in-between the N- and C-terminal lobe. The N-terminal lobe contains a glycine-rich loop (sequence GXGXXG) and a unique major helix, the αC-helix. The C-terminal lobe comprises an activation domain, which spans from the Asp-Phe-Gly (DFG) motif to the Ala-Pro-Glu (APE) motif and includes the activation loop (also refereed to as T-loop) with the conserved phosphorylation-sensitive threonine residue (Figure 9).

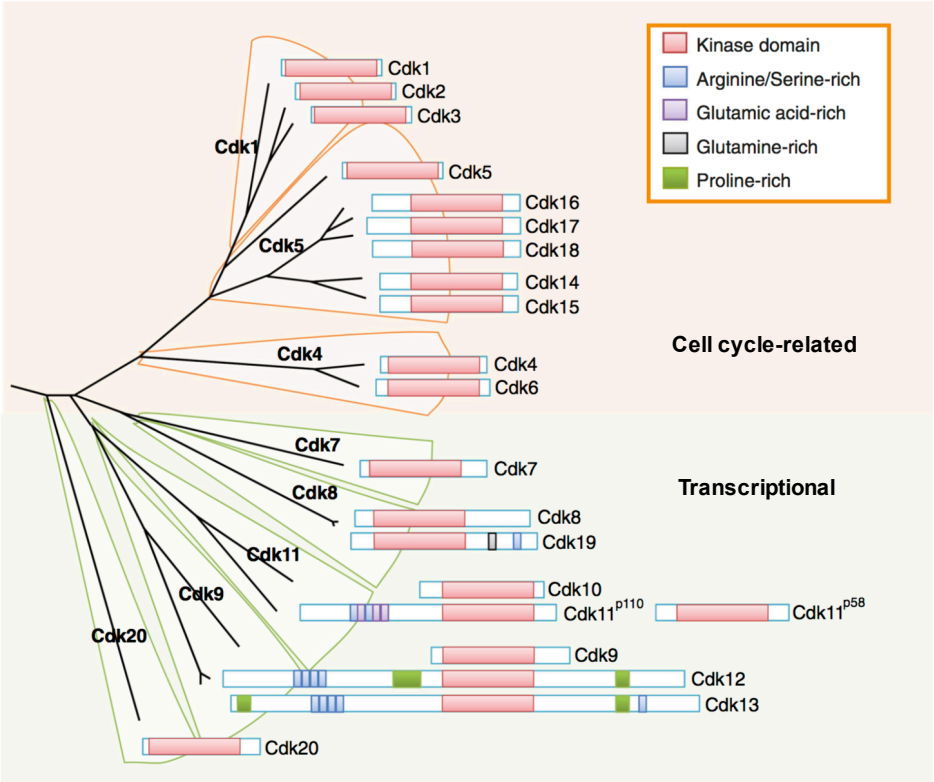


Figure 8: Schematic overview of the domain structure of individual mammalian CDK and their evolutionary relationships among each other based on comparison of the human kinase domains (Manning, 2002)**.** All member of the mammalian CDK family comprise a conserved, central protein kinase domain (red), which is enclosed by N- and C-terminal sequences of variable lengths. Besides the kinase domain, additional domains and motifs are depicted for some CDKs (Malumbres, 2014).

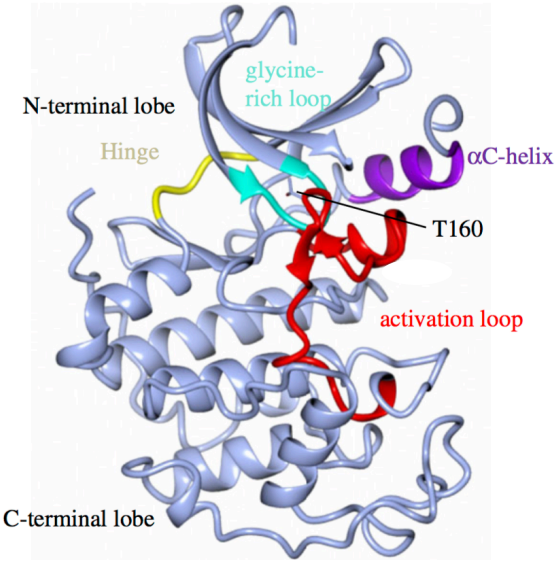


Figure 9: Characteristic structural features of cyclin dependent kinases. Structural features are exemplified by the monomeric structure of CDK2 (PDB: 1HCK; (De Bondt *et al.*, 1993). N-and C-terminal lobe are shown in light blue ribbon. Following structural features are highlighted in the CDK2 fold: glycine-rich loop (cyan), αC-helix (purple), hinge linker (yellow), activation loop with conserved threonine residue (red) (Wood and Endicott, 2018).

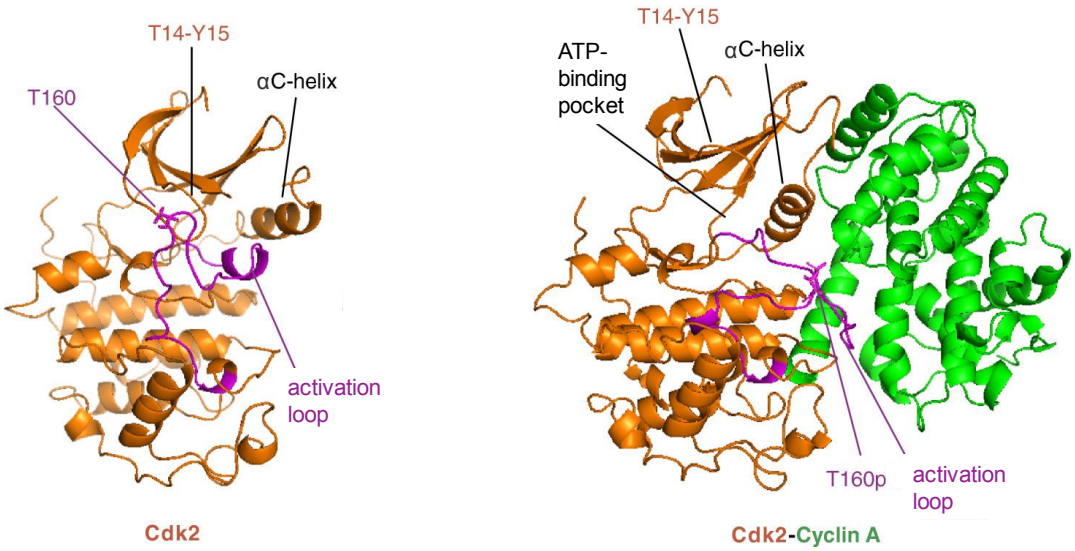


Figure 10: Model of CDK activation exemplified by monomeric CDK2 (PDB: 1HCL; (Schulze-Gahmen *et al.*, 1996) **and CDK2/CycA (PDB: 1JST;** (Russo *et al.*, 1996)**.** In monomeric CDK2 (left), the major αC-helix of the N-terminal lob and the activation loop (purple) are close, preventing access of ATP to the nucleotide binding pocket. Upon binding of Cyclin A (green), the αC-helix and the activation loop are pulled apart, allowing access to the ATP-binding pocket for enzymatic activity. The activated form of the kinase complex is further stabilized by phosphorylation of the conserved threonine residue (T160) in the activation loop of CDK2 (Malumbres, 2014).

Activation of CDK requires binding of CDKs with their respective cyclins to generate a binary complex (Jeffrey *et al.*, 1995). Cyclins interact with the corresponding CDK partner via their highly conserved cyclin box domain (Brown *et al.*, 1995). In the absence of cyclin, the ATP-binding pocket is blocked by the flexible activation loop, preventing the access of the catalytic pocket for ATP. Upon binding of the cyclin, the CDK αC-helix associates with one specific helix in the cyclin through a surface characterized by extensive hydrophobic interactions. This interaction promotes a rotation in the axis of this segment, allowing the αC-helix and the activation loop to be pulled apart, thereby relieving the blocking of the ATP-binding pocket (Malumbres, 2014) (Figure 10). In addition, cyclin-binding also allows a conserved threonine residue on the activation loop to become accessible for activating phosphorylation by the Cyclin-dependent kinase activating kinase (CAK) (Krek and Nigg, 1992; Lorca *et al.*, 1992; Solomon *et al.*, 1992). This phosphorylated threonine stabilizes the activated form of the kinase heterodimer (Echalier *et al.*, 2010; Pavletich, 1999). Together, these activated processes result in conformational changes in and around the kinase active site of the CDKs, thus increasing the kinase complex activity and sets the stage for the full CDK activation.
It has to be noted though that the classical model of CDK activation exemplified by CDK2/Cyc A is not applicable to all CDKs. The extent of the CDK-cyclin interface varies in the structure of different CDKs. Contacts between CDK and cyclin can be located at both the N- and C-terminal lobes or only limited to the N-terminal lobe of the kinase (Day *et al.*, 2009; Echalier *et al.*, 2010; Takaki *et al.*, 2009). It is also the case that the cyclin not necessarily initiate an active conformation on the kinase (Day *et al.*, 2009; Takaki *et al.*, 2009). Some CDKs, such as CDK8 and CDK5, appear to lack a requirement for the activating phosphorylation by CAK, as they can obtain the needed conformation through other interactions (Echalier *et al.*, 2010). It is supposed in case of CDK8 that the negatively charged Asp residue, roughly at the position normally occupied by the Thr, operates as the phosphorylation substitute (Leclerc *et al.*, 1996). Moreover, a few CDKs, as CDK16 and CDK12/13, contain additional domains and motifs with functional relevance supplementary to the consensus kinase domain (Kohoutek and Blazek, 2012; Petra Mikolcevic *et al.*, 2012; Zi *et al.*, 2015).

2.5 CDK12/Cyclin K and CDK13/Cyclin K

Human CDK12 and CDK13 are large proteins with molecular weights of 164 and 165 kDa, respectively. They were first identified in cDNA screens for cell cycle regulators (Ko *et al.*, 2001; Marqués *et al.*, 2000). Human Cyclin K, the cyclin subunit of CDK12 and CDK13, is a 70 kDa protein which associates with CDK12 and CDK13 in two separate complexes. It was found to be

associated with RNAPII and possessing both CTD kinase activity and CAK activity (Edwards *et al.*, 1998). In functional studies, it has also been shown that Cyclin K activates transcriptional elongation (Lin *et al.*, 2002). The functional link between CDK12 and Cyclin K has been indicated by the overlapping set of genes affected by the absence of either CDK12 or Cyclin K (Blazek, 2012; Blazek *et al.*, 2011). Both CDK12/13 kinases and Cyclin K are localized in nuclear speckles, which are subnuclear structures enriched for splicing factors (Blazek *et al.*, 2011; Chen *et al.*, 2006; Ko *et al.*, 2001). In mammalian cells, CDK12 and CDK13 associate with a large number of splicing factors and positively regulate the expression of a large number of genes (Berro *et al.*, 2008; Chen *et al.*, 2006, 2007; Even *et al.*, 2006). It has been supposed that in addition to CTD phosphorylation, CDK12 and CDK13 may also affect RNA processing through interacting directly with RNA processing factors and by regulating their expression (Liang *et al.*, 2015).

2.5.1 Domain architecture of CDK12, CDK13 and Cyclin K

Human CDK12 and CDK13 shares similar protein domain architecture. The kinase domain of both proteins consists of about 300 amino acids with high sequence similarity (> 93%). They are localized in the center of the kinase sequence and characterized by a PITAIRE motif (Marqués *et al.*, 2000). In contrast to other transcriptional CDKs 7, 8, and 9, the expanded N-terminal regions of CDK12 and CDK13 contain additional arginine-serine (RS)-rich motifs, which spanning residues 130-380 in CDK12 and 200-435 in CDK13, respectively. These motifs are supposed to serve as docking sites for the assembly of splicing factors and regulators of splicing, thus linking the kinases to the RS protein family involved in the procedures of RNA processing and pre-mRNA splicing (Hertel and Graveley, 2005; Long and Caceres, 2009; Zhou and Fu, 2013). In addition, both kinases have proline-rich motifs (PRM) in their C-terminal region, whereas CDK12 carries one more in its central region and CDK13 has one additional PRM in its N-terminal region. Proteins with PRMs are recognized for their function in transcriptional regulation, RNA processing, and alternative splicing as they serve as binding sites for Src-homology 3 (SH3), WW (named for a conserved Trp-Trp motif) or profilin domain containing proteins (Ball *et al.*, 2005; Sudol *et al.*, 2001).

The domain architecture of Cyclin K is composed of a N-terminal cyclin-box domain involving two cyclin boxes followed by a wide-stretched proline-rich region consists of several PRMs with 42% proline content over 220 amino acids. Together, both cyclin-boxes encompass fifteen helices that mediate binding of the Cyclin to its CDK partner (Baek *et al.*, 2007). A schematic representation of the domain architecture of CDK12, CDK13 and Cyclin K is shown in figure 11.

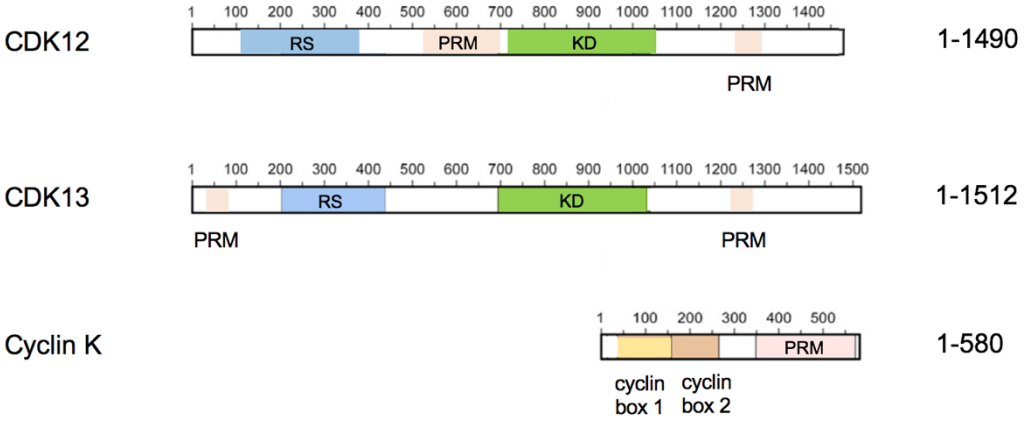


Figure 11: Domain architecture of CDK12, CDK13 and Cyclin K. For CDK12 and CDK13, following domains and motifs are depicted: kinase domain (KD), arginine-serine rich motifs (RS) and proline-rich motifs (PRM). For Cyclin K, two cyclin boxes and proline-rich motifs (PRM) are depicted (Greifenberg *et al.*, 2016).

2.5.2 Structural features of CDK12 and CDK13 kinase complex

Human CDK12 and CDK13 share a sequence identity of 43%. Superimposition of the three-dimensional protein structure of CDK12 (PDB: 4NST; (Bösken *et al.*, 2014) and CDK13 (PDB: 5EFQ; (Greifenberg *et al.*, 2016) clearly reflects their structural homology (Figure 12). Both CDK12 and CDK13 structures exhibit a typical kinase fold consisting of the N-terminal lobe (CDK12: aa 715–816; CDK13: aa 695-794) and the C-terminal lobe (CDK12: aa 817–1020; CDK13: aa 795–998). As described in chapter 2.4, binding of the cyclin subunit directs CDK12 and CDK13 in the active kinase conformation as characterized by the orientation of the αC-helix mediating contacts to Cyclin K. The DFG motif at the start of the activation segment and the activation segment itself adopt a conformation that allows access to the ATP-binding site.

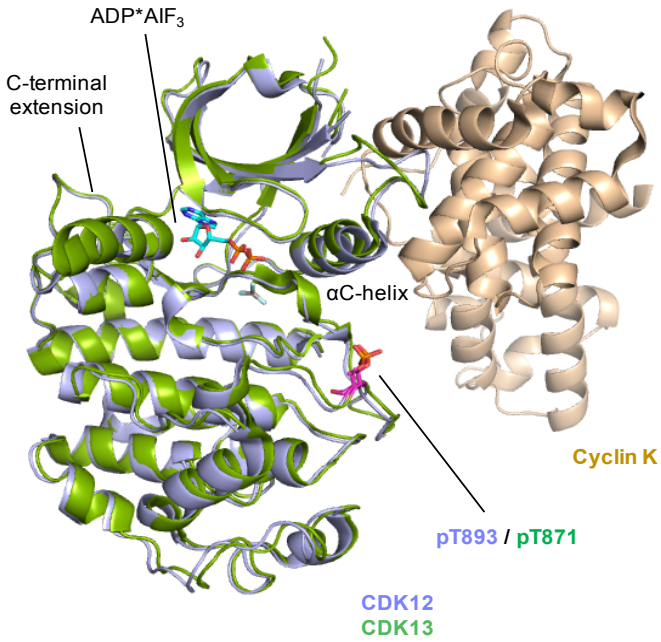


Figure 12: Protein structure of human CDK12/CycK (PDB: 4NST; (Bösken *et al.*, 2014) and CDK13/CycK complex (PDB: 5EFQ; (Greifenberg *et al.*, 2016). Structure superimposition of CDK12 and CDK13 reflects their high degree of homology.

Following the kinase domain, both kinases contain a C-terminal extension helix (CDK12: aa 1017-1039; CDK13: aa 999-1032) that associates in the ATP cleft with the ribose of the bound ATP as well as forms intra-molecular contacts with the kinase core domain through multiple water-mediated interactions. Here, the DCHEL motif within the C-terminal extension helix is the primary interaction site towards the bound nucleotide and the kinase domain. H1018 and E1019 of the DCHEL motif in CDK13 form water-mediated contacts to the ribose of the bound nucleotide, which is complemented by interactions of D1016 with the residues of the N- and C-terminal kinase lobes (Figure 13A). In contrast, H1040 of the DCHEL motif in CDK12 interacts directly with the adenine base of the bound nucleotide, which is also complemented by interactions of D1038 with residues of the N- and C-terminal kinase lobe (Figure 13B). The DCHEL motif is directly followed by a polybasic region consisting of six positively charged residues rich in lysine and arginine: ^{1045}KKRRRQR for CDK12 and ^{1023}KKRRRQK for CDK13. Association of this polybasic region in close proximity to the ATP binding site could facilitate an electrostatic interplay to the highly negatively charged CTD of RNAPII. Together, the DCHEL motif and the polybasic region constitute the conserved motifs within the C-terminal extension helix of CDK12 and CDK13.

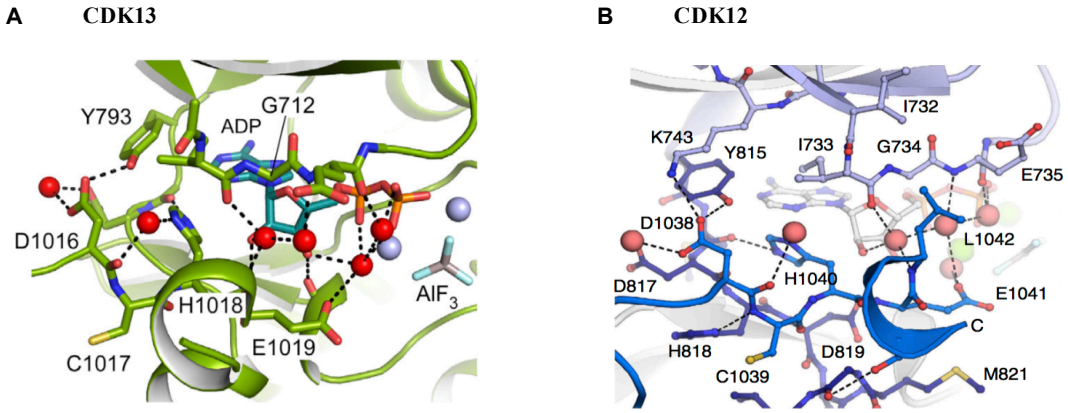


Figure 13: Association of the C-terminal extension helix with the bound nucleotide and the kinase core domain of **(A)** CDK13 (Greifenberg *et al.*, 2016) and **(B)** CDK12 (Bösken *et al.*, 2014).

2.5.3 CDK12 and CDK13 in disease

As CDK12 regulates the expression of several cancer-related genes, such as breast cancer 1 (BRCA1) gene (Blazek, 2012; Blazek *et al.*, 2011), dysregulation of CDK12 has been considered to be implicated in several types of cancer. Analysis of the CDK12 gene across The Cancer Genome Atlas (TCGA) revealed mutations, amplifications or deep deletions in 30/32 tumor types (Lui *et al.*, 2018) (Figure 14). Several evidences have pointed to an important role of CDK12 in the development of breast cancer (Iorns *et al.*, 2009; Kauraniemi *et al.*, 2001, 2003). Furthermore, in high-grade serous ovarian cancer, *CDK12* was defined to be one of the most frequently somatically mutated genes (The Cancer Genome Atlas Research Network, 2011). The majority of the CDK12 mutations identified in this type of ovarian carcinoma are point mutations located in the kinase domain that abrogate the kinase activity of CDK12 (Ekumi *et al.*, 2015; Joshi *et al.*, 2014). As CDK12/CycK is a master regulator of proteins specifically involved in DNA damage repair (DDR) by homologous recombination (HR) as well as represents an important player in cellular response to DNA damage, it has often been suggested that loss-of-function of CDK12 promote carcinogenesis by impairing HR. Accumulation of spontaneous and induced DNA alterations can foster an unstable genome, characteristic of all cancers (Blazek *et al.*, 2011; Chilà *et al.*, 2016). On the other hand, HR deficiency caused by CDK12 loss-of-function can enhance cell sensitivity to different anticancer agents, such as poly (ADP-ribose) polymerase (PARP) inhibitors. It has been shown that the broad utility of these inhibitors has been limited by their lack of activity in HR-proficient cancers, as well as acquired resistance of initially responding tumors, often mediated by restoration of HR (Bouwman and Jonkers, 2014). By this way, both screening human tumors for CDK12 loss-of-function and the

inhibition of CDK12 kinase activity can be used as a possible tool to increase the efficacy of cancer treatment.

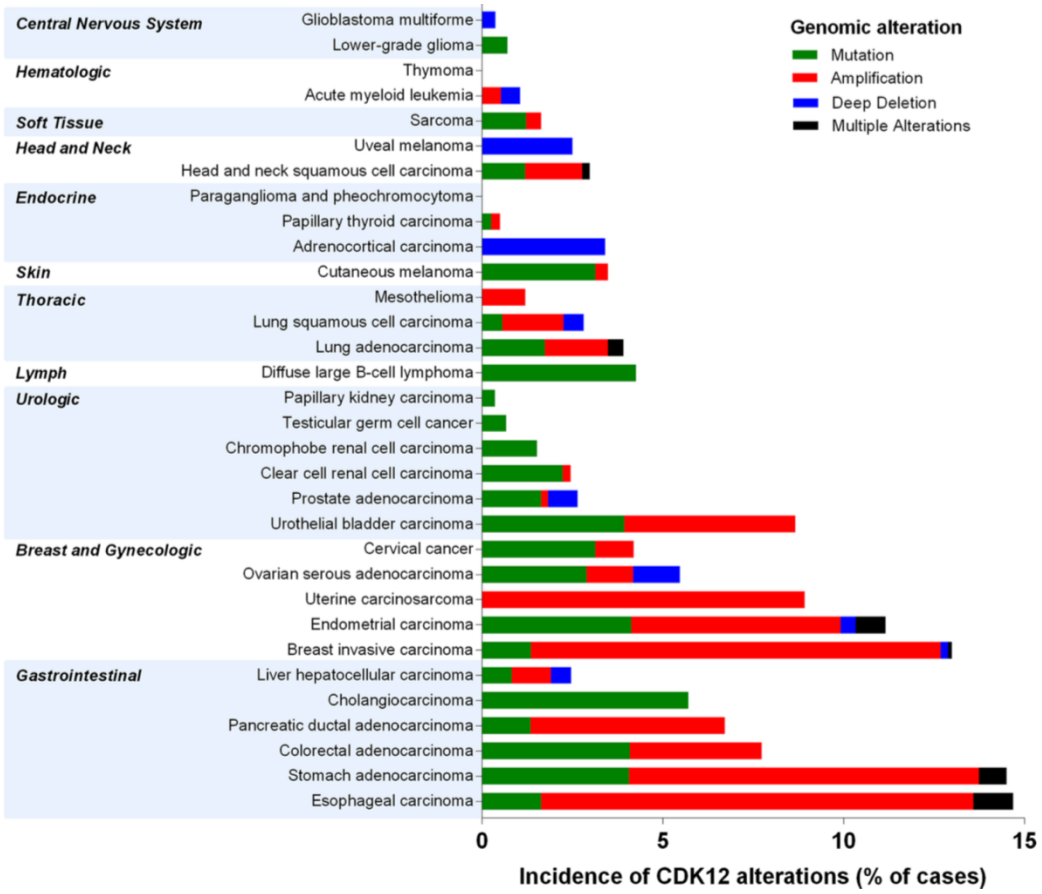


Figure 14: Genomic alterations of the CDK12 gene across TCGA (Lui *et al.*, 2018).

In comparison with CDK12, much less evidence exists for the clinical significance of CDK13. Nevertheless, considerable progress in understanding the clinical and genotypic spectrum of CDK13-related disorder has been become more apparent with numerous reported cases of CDK13-related disorder. Since the first evidence linking CDK13 to human disease reported in 2016, detailed clinical data have been reported on 44 individuals with likely pathogenic mutations in CDK13 until now (Bostwick *et al*., 2017; Carneiro *et al*., 2018; Hamilton *et al*., 2018; Sifrim A. *et al*., 2016; Uehara *et al*., 2018; van den Akker *et al*., 2018). 36 of the 44 clinically characterized cases were identified to have missense mutations affecting the protein kinase domain of CDK13 (Figure 15). It has been recently shown that missense mutations affecting the protein kinase domain of CDK13 cause a recognizable syndrome of intellectual disability (Hamilton *et al*., 2018; van den Akker *et al*., 2018).

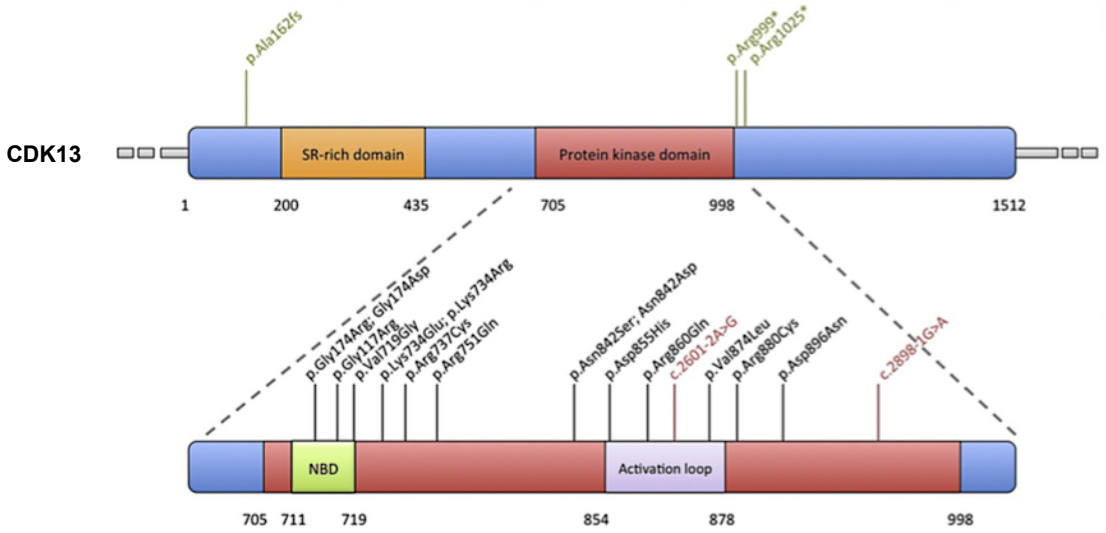


Figure 15: Schematic illustration of the reported CDK13 pathogenic mutations on CDK13 protein sequence. The pathogenic mutations are divided into three subgroups: missense variants (black), splice site variants (red) and frame shift or nonsense variants (green). NBD: nucleotide-binding domain; SR-rich domain: serine-arginine-rich domain (Hamilton and Suri, 2019).

2.6 Understudied CDKs

Besides the CDKs that regulate progression through the cell cycle and different stages of the transcription process, there is a third CDK subfamily consisting of the highly conserved CDK14-CDK18. According to their common 'TAIRE' sequence motif, they are better known as PFTAIRE (PFTK) kinases referring to CDK14 and 15 and PCTAIRE (PCTK) kinases referring to CDK16–18 (Malumbres et al., 2009). In contrast to cell cycle-related and transcriptional CDKs, the function and regulation of this CDK subfamily are poorly characterized (Malumbres, 2014). In the recent years, substantial progresses have been made in understanding their nature. The ''TAIRE'' sequence motif has been identified to be essential for binding to the G_2/M cyclins, Cyclin Y and Cyclin Y-like 1 proteins. Kinase activation takes place by binding to Cyclin Y, a highly conserved cyclin protein consists of one cyclin box and a N-myristoylation signal allowing targeting to the plasma membrane. Thus, upon binding to Cyclin Y, the kinase will be targeted to the plasma membrane (Jiang *et al.*, 2009; Petra Mikolcevic *et al.*, 2012). In this study, the focus is on the characterization of the CDK14 kinase in association with its cyclin partner, Cyclin Y.

2.6.1 CDK14/Cyclin Y

Human CDK14 belongs to the family of cell division cycle 2 (CDC2)-related kinases and was identified as a CDK involved in cell cycle regulation (Shu *et al.*, 2007; Yang and Chen, 2001). Findings have shown that the kinase complex of CDK14/CycY participates in Wnt signaling through activation of the low-density lipoprotein receptor-related protein 6 (LRP6) transmembrane receptor protein, a key regulatory protein involved in the Wnt/ß-catenin pathway. CDK14/CycY activates the LRP6 receptor by phosphorylation in a cell cycle dependent manner. When dividing cells enter the G_2/M phase, CDK14/CycY accumulates at the plasma membrane and phosphorylates LRP6 at its PPPSP sites, resulting in receptor priming. The primed LRP6 is hypersensitive to incoming Wnt signal-induced phosphorylation by CK1 at sites adjacent to the phosphorylated PPPSP motifs. This activation procedure finally results in stabilization of cytoplasmic ß-catenin and transcriptional activation of the target genes (Davidson *et al.*, 2009; Davidson and Niehrs, 2010; Wang *et al.*, 2016) (Figure 16).

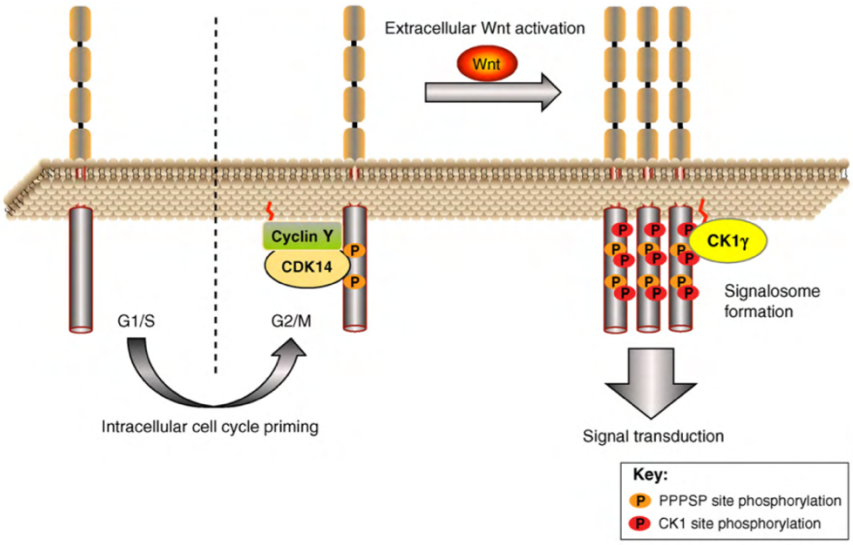


Figure 16: Schematic overview of the cell cycle-dependent Wnt signaling pathway induced by CDK14/CycY mediated activation of the LRP6 receptor during the G_2/M transition (Davidson and Niehrs, 2010).

Considering the domain architecture of CDK14 and Cyclin Y (Figure 17), CDK14 contains many predicted domains and motifs typically conserved in protein kinases. Besides the ATP-binding region and several other binding/active-sites, the kinase domain of CDK14 contains the specific PFTAIRE sequence motif. In contrast, the domain composition of Cyclin Y differs from that of the conventional cyclins. Unlike the conventional cyclins, Cyclin Y only contains a single cyclin box, which appears to be sufficient for its activity (Liu *et al.*, 2010). It is unknown how Cyclin Y binds to and activates

its kinase partner with a single cyclin box as usually two cyclin boxes are required for proper cyclin activities (Noble *et al.*, 1997). It has been shown that the interaction between CDK14 and Cyclin Y requires the cyclin box in Cyclin Y and the PFTAIRE motif in CDK14. But it was also noted that the kinase domain of CDK14 is not sufficient for Cyclin Y binding as additional regions outside the CDK domain, in the range of the N-terminal extensions, are also required (Jiang *et al.*, 2009). In addition, recent findings indicate that binding of 14-3-3 proteins enhances the association of the CDK14/CycY kinase complex and that Ser100 and Ser326 residues in Cyclin Y are crucial for 14-3-3 binding *in vitro* and *in vivo* (Li *et al.*, 2014). 14-3-3 proteins are a family of homologous proteins that consist of seven isoforms (β, γ, ε, η, ζ, σ, and τ) in mammals and exist as homo- and heterodimers (Aitken, 1995; Rosenquist *et al.*, 2000). They are known to interact with proteins containing specific pSer/pThr motifs (Furukawa *et al.*, 1993; Muslin *et al.*, 1996) and regulate a wide range of biological processes (Datta *et al.*, 2000; Fantl *et al.*, 1994; Fu *et al.*, 2000; Skoulakis and Davis, 1998; Tzivion and Avruch, 2002; Zha *et al.*, 1996). Although protein structures are available for many CDK monomers as well as CDK/cyclin dimers, the protein structure of neither CDK14 nor Cyclin Y has been elucidated yet.

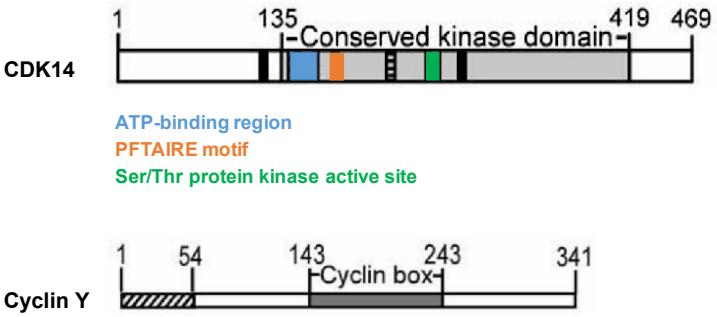


Figure 17: Domain architecture of CDK14 and Cyclin Y. For CDK14, following domains and motifs are highlighted within the conserved kinase domain: ATP-binding region (blue), PFTAIRE motif (orange) and Ser/Thr protein kinase active site (green). For Cyclin Y, the single cyclin box is depicted (Jiang *et al.*, 2009).

The role of Wnt signaling in carcinogenesis is well established. Dysregulation of Wnt signaling has most prominently been described for colorectal cancer, but aberrant Wnt signaling is also observed in many more cancer entities (Novellasdemunt *et al.*, 2015; Zhan *et al.*, 2017). As the kinase complex of CDK14/CycY is implicated in the regulation of the WNT signaling pathway and overexpression of CDK14 was identified in human colorectal cancer patients which was associated with poor prognosis (W. Zhang *et al.*, 2016), CDK14 has been considered as a potential therapeutic target for different types of cancer, especially for colorectal cancer. By this way, CDK14 came recently into the focus of many studies intending to develop small molecule inhibitors specifically targeting this kinase (Ferguson *et al.*, 2019).

3 Materials and Methods

3.1 Materials

3.1.1 Chemicals

The chemicals used in this study were purchased from the following manufacturers:
Abcam (Cambridge, UK), Agilent Technologies, Inc. (Santa Clara, USA), AppliChem (Darmstadt, Germany), BioRad (Munich, Germany), Boehringer (Mannheim, Germany), Carl Roth GmbH (Karlsruhe, Germany), Cayman Chemicals (Ann Arbor, USA), Eppendorf (Hamburg, Germany), Fermentas (St. Leon- Rot, Germany), FMC Bioproducts (Vallensbaek Strand, Denmark), GE-Healthcare (Munich, Germany), Gerbu (Gaiberg, Germany), Invitrogen (Karlsruhe, Germany), J. T. Baker / Mallinckrodt Baker B.V. (Deventer, Netherlands), Merck (Darmstadt, Germany), Roche (Basel, Switzerland), Millipore (Schwalbach, Germany), MWG Biotech (Ebersberg, Germany), New England Biolabs (Frankfurt, Germany), Novagen, Merck (Darmstadt, Germany), Qiagen (Hilden, Germany), Sarstedt (Nümbrecht, Germany), Serva (Heidelberg, Germany), Sigma-Aldrich (Munich, Germany), Thermo Fisher Scientific Inc. (Waltham, USA), Waters Corporation (Milford, USA).

3.1.2 Nucleotides

Table 2: Nucleotides and manufacturers.

Nucleotides	Manufacturers
ATP, 100 mM stock solution	Jena Bioscience, Jena, Germany
dNTPs	Roche, Mannheim, Germany
[γ-^{32}P]-ATP, 5 µCi/µL, 3000 mCi/mmol	PerkinElmer, Boston, USA

3.1.3 Kits

Table 3: Kits and manufacturers.

Kits	Manufacturers
QIAprep Spin Miniprep Kit	Qiagen, Hilden, Germany
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
QIAquick PCR Purification Kit	Qiagen, Hilden, Germany
Plasmid Midi Kit	Qiagen, Hilden, Germany

3.1.4 Marker/size standards

Table 4: Marker/size standards and manufacturers.

Marker/size standards	Manufacturers
LMW marker	GE Healthcare, Munich, Germany
Page Ruler Plus Prestained protein ladder	Thermo Fisher Scientific, Waltham, USA
DNA ladder 100bp / 1kb	CarlRoth, Karlsruhe, Germany

3.1.5 Consumables

Table 5: Consumables and manufacturers.

Consumables	Manufacturers
Cellulose Spectra/Por dialysis membrane	Carl Roth GmbH, Karlsruhe, Germany
Micro tubes (0.5 mL, 1 mL and 2 mL), Pipette tips	Eppendorf, Munich, Germany
Serological pipettes (sterile, 5 mL, 10 mL, 25 mL and 50 mL), Screw cap tubes (15 mL and 50 mL), plastic-cuvettes, Filtropur S (0.2/0.45 µM)	Sarstedt, Nümbrecht, Germany
Electroporation cuvettes	PeqLab Biotechnology GmbH, Erlangen
Bottle Top Filters	Nalgene Labware, Roskilde, Dänemark
Amicon Ultra Centrifugal filters (MWCO 3kDa, 10kDa, 30kDa)	Sigma-Aldrich, Munich, Germany
Whatman P81 Zellulose Optitran BA-S85 reinforced membrane	Omnilab, Bremen Whatman, Maidstone, UK
Optitran BA-S85 reinforced membrane	Whatman, Maidstone, UK
PD-10 desalting columns	GE Healthcare, Munich, Germany

3.1.6 Vectors

Table 6: Vectors and manufacturers.

Vectors	Expression tag	Expression species	Manufacturers
pGEX-4T1-tev (mod.),	GST	*E. coli*	GE Healthcare, Munich, Germany
pGEX6PN-tev (mod.)	GST	*E. coli*	GE Healthcare, Munich, Germany
pET28a-GST-tev (mod.),	GST	*E. coli*	Novagen, Darmstadt, Germany
pET28a-MBP-tev (mod.),	MBP	*E. coli*	Novagen, Darmstadt, Germany
pET23d	His	*E. coli*	Novagen, Darmstadt, Germany
pPROEX HTb	His	*E. coli*	GE Healthcare, Munich, Germany
pACEBac1-GST-tev (mod.),	GST	*Sf9*	ATG Biosynthetics, Merzhausen,
pIDK-GST-tev (mod.),	GST	*Sf9*	Germany
pIDK, pIDC	-	*Sf9*	

3.1.7 Oligonucleotides

All primers used in this study were synthesized by Metabion International AG (Planegg, Steinkirchen, Germany). The primers were dissolved in Millipore water to a concentration of 10 pmol/µL and stored at -20°C.

3.1.8 Enzymes

Table 7: Enzymes and manufacturers.

Enzymes	**Manufacturers**
Restriction endonucleases	New England Biolabs, Ipswich, USA
T4 DNA polymerase	New England Biolabs, Ipswich, USA
Taq Expand High Fidelity	Roche, Mannheim
Q5 High Fidelity DNA polymerase	New England Biolabs, Ipswich, USA
T4 DNA Ligase	New England Biolabs, Ipswich, USA
Cre Rekombinase	New England Biolabs, Ipswich, USA
TEV Protease	UKB, Bonn (in-house production)
PreScission	UKB, Bonn (in-house production)

3.1.9 Cell line and bacterial strains

Table 8: Bacterial strains and cell line.

***E. coli* strains**	**Genotype**
Top10	F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 nupG λ-
NEB 10-beta	Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)
XL1 Gold	TetrD(mcrA)183 D(mcrCBhsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac The [F'proAC lacIqZDM15 Tn10 (Tetr) Amy Camr
BL21 (DE3)	F–, ompT, hsdSB(rB-, mB-), dcm, gal, λ(DE3)
BL21 (DE3) R3-Rosetta	F–, ompT, hsdSB(rB-, mB-), dcm, gal, λ(DE3), pRARE
DH10 MultiBac^{Turbo}	F– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK λ– rpsL nupG /pMON14272 v-cath::Ampr chiA::LoxP */ pMON7124

Cell line	Manufacturers
Sf9 insect cells (clonal isolates derived from *Spodoptera frugiperda* cell line IPLB-*Sf21*-AE)	Thermo Fisher Scientific, Waltham, USA

3.1.10 Cell culture media and reagents

Table 9: Cell culture media and reagents.

Cell culture medium and reagents	Compositions / Manufacturers
Sf-900^{TM} III SFM 1x	Invitrogen, Karlsruhe, Germany
Cellfectin II	Thermo Fisher Scientific Inc., Waltham, USA
	Roth, Karlsruhe, Germany
LB Medium (Lennox)	UKB, Bonn (in-house production):
LB Medium	1% (w/v) bactotrypton,
	0.5% (w/v) yeast extract
	1% (w/v) NaCl, pH 7.4
	UKB, Bonn (in-house production):
TB medium	900 mL
	1.2% (w/v) bactotrypton
	2.4% (w/v) yeast extract
	0.04% (v/v) glycerol
	100 mL
	0.17 M KH_2PO_4; 0.72 M K_2HPO_4
	UKB, Bonn (in-house production):
Differential LB-agar plates for	LB agar with
E. coli DH10	100 µg/mL Ampicillin
	50 µg/mL Kanamycin
	7 µg/mL Gentamycin
	10 µg/mL Tetracyclin
	100 µg/mL X-Gal
	40 µg/mL IPTG
	UKB, Bonn (in-house production):
Antibiotics	Ampicillin (100 mg/L medium)
	Chloramphenicol (34 mg/L medium)
	Kanamycin (50 mg/L medium)
	Gentamycin (7 mg/L medium)
	Tetracyclin (10 mg/L medium)

3.1.11 Chromatography columns/column materials

Table 10: Chromatography columns/column materials and manufacturers.

Chromatography columns/ column materials	Manufacturers
HisTrap FF	GE Healthcare, Munich, Germany
Ni-NTA Superflow	Qiagen, Hilden, Germany
GSTrap FF	GE Healthcare, Munich, Germany
GSH 4 Fast Flow	GE Healthcare, Munich, Germany
MBPTrap HP	GE Healthcare, Munich, Germany
Amylose resin	New England Biolabs, Frankfurt, Germany
HiLoad 16/600 Superdex 75 prep grade	GE Healthcare, Munich, Germany
HiLoad 16/600 Superdex 200 prep grade	GE Healthcare, Munich, Germany

3.1.12 Devices

Table 11: Devices and manufacturers.

Devices	Manufacturers
Agarose gel chamber, DNA-SUB-Cell	Bio-Rad, Munich, Germany
Autoclave 5075 EL	Systec, Linden, Germany
Centrifuge, Avanti J265 XP	Beckman Coulter, Brea, California, USA
Centrifuge, Eppendorf 5810	Eppendorf, Munich, Germany
Centrifuge, Eppendorf 5810R	Eppendorf, Munich, Germany
Centrifuge Rotor, JA-25.50	Beckman Coulter, Brea, California, USA
Centrifuge Rotor, JLA-8.1000	Beckman Coulter, Brea, California, USA
Electroporator	Eppendorf, Munich, Germany
FPLC-System Äkta prime plus	GE Healthcare, Munich, Germany
FPLC-System Äkta Start	GE Healthcare, Munich, Germany
Freezer (-80°C)	Thermo Scientific, Corston, UK
Incubator Heratherm	Thermo Scientific, Waltham, USA
Incubator Innova 40	New Brunswick Scientific, Jersey, USA
Incubator Multitron pro	Infors HT, Bottmingen, Switzerland
Incubator, Heraeus CO2 Auto zero	Heraeus Instruments, Hanau, Germany
Magnetic stirrer MR 2000	Heidolph, Schwabach, Germany
Magnetic stirrer MR 3002	Heidolph Instruments, Schwabach, Germany
Mastercycler nexus SX1	Eppendorf, Munich, Germany
Micro balance CPA324S	Sartorius, Göttingen, Germany
Microscope Axiolab	Carl Zeiss, Jena, Germany
Millipore	Thermo Scientific, Corston, UK
Mini protean Tetra Cell	Bio-Rad, Munich, Germany
Nanodrop 2000/c UV-Spectrometer	Thermo Scientific, Corston, UK
Odyssey, gel documentation system	Li-Cor Biosciences, Bad Homburg, Germany
pH meter lab 850	Schott Instruments, Mainz, Germany
Photometer Eppendorf Biophotometer	Eppendorf, Munich, Germany
Scintillation counter	Beckman Coulter, Brea, California, USA
Sonifier Vibra cell	Sonics, Newton, USA

Sonifier W-250	Branson, Danbury, USA
Tabletop centrifuge Eppendorf 3424	Eppendorf, Munich, Germany
Tabletop centrifuge Eppendorf 3424 R	Eppendorf, Munich, Germany
Tabletop centrifuge Eppendorf 5804	Eppendorf, Munich, Germany
Thermomixer comfort	Eppendorf, Munich, Germany
Ultracentrifuge Optima Max-TL	Beckman and Coulter, Brea, California, USA
Ultracentrifuge Optima L-80 XP	Beckman, Fullerton, USA
Vortexer Vortex Genie 2	Bender & Hobein, Bruchsal, Germany
Waage Sartorius BP110S	Sartorius, Göttingen, Germany
Water bath Julabo 5	Julabo, Seelbach, Germany

3.1.13 Commonly used buffers

Table 12: Commonly used buffers and their composition.

Application	Buffer	Composition
Protein purification GSTrap FF	Lysis buffer	50 mM Hepes pH 7.8; 300 mM NaCl; 10% Glycerol; 5 mM β-ME
	Wash buffer	50 mM Hepes pH 7.8; 1 M NaCl; 10% Glycerol; 5 mM β-ME
	Elution buffer	50 mM Hepes pH 7.8; 300 mM NaCl; 10% Glycerol; 5 mM β-ME; 10 mM GSH
	Column regeneration	6 M Gdn-HCl
Protein purification HisTrap FF	Lysis buffer	50 mM Hepes pH 7.8; 100 mM NaCl; 5 mM β-ME; 5 mM Imidazole
	Wash buffer	50 mM Hepes pH 7.8; 300 mM NaCl; 5 mM β-ME; 10 mM Imidazole
	Elution buffer	50 mM Hepes pH 7.8; 100 mM NaCl; 5 mM β-ME; 300 mM Imidazole
	Column regeneration	6 M Gdn-HCl; 200 mM acetic acid
Protein purification Inclusion bodies HisTrap FF	Lysis buffer	50 mM Hepes pH 7.6; 100 mM NaCl; 5 mM β-ME;
	Wash buffer	50 mM Hepes pH 7.6; 100 mM NaCl; 5 mM β-ME; 1 M Urea; 0.5% Triton-X
	Extraction buffer	50 mM Hepes pH 7.6; 100 mM NaCl; 5 mM β-ME; 8 M Urea
	Refolding buffer I	50 mM Hepes pH 7.6; 100 mM NaCl; 5 mM β-ME; 4 M Urea
	Refolding buffer II	50 mM Hepes pH 7.6; 100 mM NaCl; 5 mM β-ME; 10 mM Imidazole
	Elution buffer	50 mM Hepes pH 7.6; 100 mM NaCl; 5 mM β-ME; 300 mM Imidazole
Protein purification MBPTrap HP	Lysis buffer	50 mM Tris pH 8.0; 300 mM NaCl; 10% Glycerol; 1 mM DTT (or DTE)
	Wash buffer	50 mM Tris pH 8.0; 300 mM NaCl; 1 mM DTT (or DTE)
	Elution buffer	20 mM Tris pH 8.0; 300 mM NaCl;

		1 mM DTT (or DTE); 15 mM Maltose
	Column regeneration	0.5 M NaOH
Size exclusion chromatography	Standard buffer	50 mM Hepes pH 7.5; 150 mM NaCl; 10% glycerol; 2 mM TCEP
	Crystallization buffer I	20 mM Hepes pH 7.8; 400 mM NaCl; 5% glycerol; 2 mM TCEP
	Crystallization buffer II	20 mM Hepes pH 7.5; 300 mM NaCl; 1 mM TCEP
	Crystallization buffer III	20 mM Hepes pH 7.8; 150 mM NaCl; 2 mM TCEP
SDS-PAGE	Separation gel buffer	1.5 M Tris pH 8.8; 0.4% (w/v) SDS
	Stacking gel buffer	0.5 M Tris pH 6.8; 0.4% (w/v) SDS
	Running buffer (10x)	250 mM Tris; 1.94 M glycine; 1% (w/v) SDS
	SDS sample buffer (4x)	240 mM Tris pH 6.8; 8% (w/v) SDS; 40% (v/v) glycerol; 0.04% bromphenol blue; 5% (v/v) β-ME
	Coomassie staining solution	40% (v/v) ethanol; 10 % (v/v) acetic acid; 0.1% (w/v) Coomassie Brilliant Blue R 250
	Coomassie destaining solution	10% (v/v) ethanol; 5% (v/v) acetic acid
Agarose gel electrophoresis	DNA loading buffer	10% (w/v) Ficoll; 0.025% (w/v) Bromphenol blue or 0.025 % (w/v) Xylenxyanol in 1x TBE buffer
	TBE buffer (10x)	89 mM Tris; 89 mM Boric acid; 0.9 mM EDTA
	TAE buffer (50x)	2 M Tris; 1 M acetic acid; 50 mM EDTA pH 8
TEV digestion	TEV cleavage buffer (10x)	500 mM Tris pH 8.0; 1.5 M NaCl; 5 mM EDTA
Cell harvesting	PBS	150 mM NaCl; 25 mM sodium phosphate; 2 mM EDTA pH 8.0

All buffers were prepared using MilliPore filtered H_2O.

3.2 Molecular biological methods

3.2.1 Electro competent cells

Fresh starter culture of desired strain including appropriate antibiotics was grown overnight at 37°C and 130 rpm. 100 mL of LB media including antibiotics was inoculated with 10 mL starter culture and incubated in 37°C shaker until OD_{600} reached 0.5-0.6. The cells were harvested by centrifugation at 4000 rpm for 10 min at 4°C. The supernatant was discarded and the pellet was washed for three times with 5% ice cold glycerol by centrifugation. Finally, the pellet was resuspended in 10% ice cold glycerol. 50 µL aliquot of the cell suspension was pipetted into cold sterile 1.5 mL microfuge tubes and snap frozen with liquid nitrogen. The frozen cells were stored at -80°C.

3.2.2 Chemical competent cells

Fresh starter culture of desired strain including appropriate antibiotics was grown overnight at 37°C and 130 rpm. 100 mL of LB media including antibiotics was inoculated with 10 mL starter culture and incubated in 37°C shaker. When the OD_{600} reached 0.5-0.6, the cell culture was put on ice for 20 min. The cells were harvested by centrifugation at 4000 rpm for 6 min at 4°C. The supernatant was discarded and the pellet gently resuspended in 10 mL ice cold 0.1 M $CaCl_2$. The cell suspension was kept on ice for 20 min then harvested by centrifugation at 4000 rpm for 6 min at 4°C. After discarding of the supernatant, the pellet was gently resuspended in 2.8 mL ice cold 0.1 M $CaCl_2$, and 1.2 mL ice cold 80% glycerol. 100 µL aliquot of the cell suspension was pipetted into cold sterile 1.5 mL microfuge tubes and snap frozen with liquid nitrogen. The frozen cells were stored at -80°C.

3.2.3 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) was widely applied for the amplification of specific double stranded DNA fragments from a DNA template. Specific forward and reverse primers of complementary oligonucleotides (20-30 bp) are required to flank the DNA region of interest. In a 3-step process, the double stranded DNA was separated at 98°C first (denaturation). In the second step (annealing), the temperature was reduced to 50-65°C. The annealing temperature was typically about 3-5°C below the melting temperature (T_m) of the primers used, allowing proper annealing of the primers to each of the single stranded DNA templates. In the final step (elongation), the heat

stable DNA polymerase synthesizes a new DNA strand complementary to the template strand by extending the primer in 5' to 3' direction using free dNTPs. Both the temperature and time required for the elongation is depending on the DNA polymerase used. For Q5 High-Fidelity DNA polymerase (New England Biolabs, Frankfurt, Germany), the elongation step was carried out at 72°C for 30 sec/ 1000 bp. Typically, 35 cycles of denaturation, annealing and elongation were carried out. The components using in each PCR are the following:

Table 13: Composition of PCR reaction mixture.

Component	**Final concentration**
dsDNA template	100-500 ng
primer (forward)	1 µM
primer (reverse)	1 µM
dNTPs	200 µM
Q5 reaction buffer (5x)	1x
Q5 DNA polymerase	0.02 U/µL
ddH_2O	ad 50 µL

The DNA fragments of interest were amplified using a thermocycler (Eppendorf, Hamburg, Germany), following the standard PCR program shown in table 14.

Table 14: Standard PCR running conditions.

Step	**Temperature**	**Time**	
initial denaturation	98°C	30 sec	
denaturation	98°C	10 sec	35 cycles
annealing	50-65°C	30 sec	
elongation	72°C	30 sec/1000 bp	
final elongation	72°C	2 min	

The amplified DNA fragments were purified using a PCR purification kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Besides the standard PCR procedures, following PCR applications were also carried out in this study using the standard PCR protocol.

Quick change site-directed mutagenesis PCR

The method of quick change site-directed mutagenesis PCR enabled generation of DNA point mutations using only one primer containing the desired mutation(s). Following the annealing step, the primer is extended by the DNA polymerase resulting in a mutated plasmid DNA with nicked circular strand. To select the mutation-containing synthesized DNA, the methylated parental DNA template was digested using the restriction endonuclease Dpn I. The nicked circular DNA was repaired after transformation into competent *E. coli* cells.

Colony PCR

Colony PCR of transgene bacterial colonies is a common method used to confirm the success of cloning. A single *E. coli* colony was transferred into 4 ml LB medium with appropriate antibiotics and incubated in a 37°C shaker over night. The bacterial culture was put directly into the PCR reaction mixture. Performing colony PCR, color Taq DNA polymerase (Roboklon, Berlin, Germany) was used, which contains purple colored dye mix allowing direct gel loading. Positive clones were subsequently detected by agarose gel electrophoresis. The components using in each colony PCR are the following:

Table 15: Composition of colony PCR reaction mixture.

Component	Final concentration
bacterial culture	1 µL
primer (forward)	1 µM
primer (reverse)	1 µM
dNTPs	200 µM
buffer Pol B (10x)	1x
color Taq DNA polymerase	0.01 U/µL
ddH_2O	ad 20 µL

3.2.4 Agarose gel electrophoresis

The agarose gel electrophoresis is a standard method used to separating DNA fragments of varying sizes. Based on the negatively charged polyphosphate backbone, DNA molecules loaded onto an agarose gel migrate towards the positively charged anode of the gel chamber according their size and conformation. While small DNA fragments migrate fast, migration of larger fragments through the gel is slow.

Depending on the size of the DNA fragments, 1-2% (w/v) Tris-Borate-EDTA (TBE) agarose gels were produced which is supplemented with 1:20.000 PeqGreen (VWR International, Vienna, Austria), a fluorescence dye to visualize DNA using UV light. Before loading onto an agarose gel in 1x TBE buffer, the DNA samples were mixed with 6x loading dye. A standard DNA ladder (100 bp or 1kb) (CarlRoth, Karlsruhe, Germany) was used to determine the relative DNA fragment size. Electrophoresis was performed at 115 V for 30-40 min depending on the size of expected DNA fragments. Visualization of the DNA fragments was carried out using UV light. In case of preparative agarose gel electrophoresis, the separated DNA band of interest was excised from the gel and further purified using a gel extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

3.2.5 Restriction digest

Restriction endonucleases are enzymes that recognize a specific DNA sequence and are responsible to cleave DNA molecules within or adjacent to that restriction site. A specific methylation pattern is applied by prokaryotes to distinguish foreign from self DNA.

Restriction enzymes mainly differ regarding their cleavage position and sequence specificity.

In this study, DNA was cleaved by type II restriction endonucleases. These enzymes recognize DNA based on a palindromic sequence and catalyze the hydrolytic cleavage of the phosphodiester bond. The resulting sticky or blunt ends can be further used to ligate the cleaved DNA fragment into the desired vector in a directive manner ensuring the proper orientation of the DNA insert.

Restriction digests were carried out at 37°C for 1h using high fidelity restriction enzymes (New England Biolabs, Frankfurt, Germany) with 10x fast digestion buffer. The cleaved DNA fragments were purified using a PCR purification kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

3.2.6 Ligation

DNA ligation is the procedure of enzymatic joining of DNA fragments by T4 DNA ligase from T4 bacteriophage origin. It is used to ligate the insert DNA fragment into the target vector following digestion by restriction enzymes. The molar ratio of insert DNA and the target vector used for the ligation mixture was calculated according to the following equation, assuming a 5:1 ratio of insert to vector.

Equation 1: Calculation of ligation mixture

$$\text{amount of insert [ng]} = \frac{5}{1} \times \frac{\text{amount of vector [ng]} \times \text{insert length [bp]}}{\text{vector length [bp]}}$$

The ligation was carried out using T4 DNA ligase (New England Biolabs, Frankfurt, Germany) with 10x T4 DNA ligase buffer over night at 16°C.

3.2.7 Transformation of *E. coli* cells

The transformation of plasmid DNA into competent *E. coli* cells was performed using either heat-shock or electroporation. For DNA amplification, NEBß10, Top10, XL1gold, DH10 cells were transformed. BL21 (DE3) Rosetta cells were used for *E. coli* protein expression.
Prior to transformation, the competent cells were thawed on ice and then incubated with 50-100 ng DNA for 15 min. In case of heat-shock transformation, the reaction mixture was heat shocked at 42°C for 1 min before incubated on ice for 2 min. Transformation via electroporation was carried out using an electroporation cuvette in an Eporator (Eppendorf, Hamburg, Germany), applying a pulsation of 1.5 kV. The transformed cells were admitted into 1 mL LB medium and incubated for 60 min at 37°C, 1000 rpm. After incubation, a cell suspension of ca. 100 µL was streaked out on a LB agar plate supplemented with the appropriate antibiotics. The plate was incubated at 37°C over night. To isolate plasmid DNA, single bacterial colonies were transferred in 4 mL medium supplemented with appropriate antibiotics and incubated overnight at 37°C and 130 rpm. Extraction of the plasmid DNA was achieved by using a plasmid Miniprep kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Isolated plasmid DNA was stored at -20°C.

3.2.8 DNA-sequencing

After isolation of the plasmid DNA, the ligation efficiency was analyzed by restriction digest using appropriate restriction endonucleases. Samples showing expected restriction bands were sent for Sanger sequencing to GATC Biotech AG (Konstanz, Germany).

3.2.9 Generation of multi-gene expression cassettes via Cre-Lox recombination

In contrast to prokaryotic based expression system, the baculovirus / insect cell expression system is well-suited for the production of eukaryotic protein complexes containing many subunits. The baculoviral multi-gene expression system MultiBac^{Turbo} (ATG:Biosynthetics) consists of a single engineered baculovirus that has been optimized for multi protein complex expression into which several genes of interest can be inserted. This tailor-made baculovirus is then used to infect insect cells (Bieniossek *et al.*, 2012).

At the core of this expression system are specifically designed transfer vectors, named as acceptor (pACEBac1, pACEBac2) and donor vectors (pIDC, pIDK, pIDS). Genes of interest which form the subunits of a protein complex are assembled into the transfer vectors via the standard restriction-ligation cloning strategy. Before incorporation into the baculoviral DNA, a fusion of acceptor and donor vectors was created. Since both acceptor and donor vectors contain single loxP sites, acceptor-donor fusions can be generated by *in vitro* Cre-LoxP recombination using Cre recombinase. These fusions always contain one acceptor vector while one to several donor vectors. During the Cre-loxP recombination, the fusion activity of the Cre recombinase will result in an equilibrium state where coexistence of single vectors and all possible fusions is possible. To select desired acceptor-donor fusions, a conditional origin of replication that depends on a pir positive background is present in donor vectors. Transformation of the Cre-reaction in pir negative cloning strains (e.g. TOP10) including using agar with appropriate antibiotics ensure selection of the desired acceptor-donor fusions. For a total volume of 20 µL Cre reaction mixture, 800 ng of each educt was mixed with 2 µl 10x Cre buffer and 1 µL Cre recombinase (1U) (New England Biolabs), incubated at 37°C for 60 min. 10-15 µL of the Cre reaction mixture was transformed in 50 µl chemically competent *E. coli* cells as described in chapter 3.2.7. The correct composition of the acceptor-donor fusion vector of each single clones from colonies present after overnight incubation at 37°C were confirmed by restriction digest referred to chapter 3.2.5.

3.2.10 Transposition of transfer vectors into baculoviral genome

The baculovirus exists as a bacterial artificial chromosome in *E. coli* DH10 MultiBac^{Turbo} cells. The baculoviral genome contains a Tn7 transposition acceptor site (mini-attTn7) embedded in a lacZα gene accessible for bacterial Tn7 transposase. The Tn7 transposase is expressed in the *E. coli* DH10 MultiBac^{Turbo} cells from a helper plasmid (pMON7124). After transformation into *E. coli* DH10 MultiBac^{Turbo} cells, acceptor vector containing Tn7 transposition sequences were integrated into the baculovirus via Tn7 transposition. Insertion via Tn7 transposition disrupts the expression of the lacZ peptide. Successful transposition was indicated by blue/white screening in presence of X-gal and IPTG plus selective antibiotic resistance markers. White colonies indicate bacmid DNA carrying acceptor vector with integrated multi-gene cassettes.
After generation of the recombinant plasmid DNA, 0.5 µg DNA was transformed in *E. coli* DH10 MultiBac^{Turbo} cells and incubated for 4 h at 37°C and 1000 rpm. Cell suspension was then plated on agar plates supplemented with 100 µg/mL Ampicillin, 50 µg/mL Kanamycin, 7 µg/mL Gentamycin, 10 µg/mL Tetracycline, 100 µg/mL X-Gal and 40 µg/mL IPTG and incubated for 24-48 h at 37°C.

3.2.11 Isolation of recombinant bacmid DNA

The recombinant bacmid DNA was used for transfection of *Sf9* insect cells to generate the initial virus stock V0. Single white colonies, selected to isolate the recombinant bacmid DNA, were used to inoculate 4 mL LB medium supplemented with Ampicillin (100 µg/mL), Kanamycin (50 µg/mL), Gentamycin (7 µg/mL) und Tetracyclin (10 µg/mL) and incubated for 24 h at 37°C and 130 rpm. Cell suspensions of transformed *E. coli* DH10 MultiBac^{Turbo} cells were pelleted by centrifugation at 5000 rpm for 3 min and then subjected to alkaline lysis followed by precipitation of cell components using a standard plasmid Miniprep kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. For DNA precipitation, the supernatant was mixed with 800 µL ice cold isopropanol and centrifuged for 1 h at 4°C and 14000 rpm after keeping on ice for 15 min. The resulting pellet was washed twice with 500 µL ice cold 70% ethanol for 15 min at 4°C and 14000 rpm. The bacmid DNA pellet was then air dried and resuspended in 10 µL of sterile H_2O.

3.2.12 Cultivation of *Sf9* insect cells

The insect cell lines *Sf9* and *Sf21,* originate from the cell line PLBSF-21 from the pupal ovarian tissue of the fall army worm *Spodoptera frugiperda,* are used as standard for transfection and amplification of recombinant baculovirus.
Cultivation of *Sf9* insect cells was carried out as suspension culture in *Sf*-900^{TM} III SFM medium (Thermo Fisher Scientific Inc., Waltham, USA) without antibiotics in Erlenmeyer or Fernbach flasks at 27°C and 80 rpm. The cell density and viability were consistently monitored using an automated cell counter EVE (VWR International, Vienna, Austria) and continuously maintained at 0.3-2.0x10^6 cells/mL.

3.2.13 Transfection of *Sf9* insect cells

For transfection of the isolated recombinant bacmid DNA, 2 mL of *Sf9* cells with a cell density of 0.3x10^6 cells/mL was plated onto a 6-well plate and incubated for 30 min at 27°C. The transfection reagent Cellfectin (Thermo Fisher Scientific Inc., Waltham, USA) was used to transfect the bacmid DNA into *Sf9* insect cells. Both 8 µL of the Cellfectin reagent and 10 µL of the resuspended bacmid DNA were mixed with 100 µL of *Sf*-900^{TM} III SFM medium each. Diluted Cellfectin reagent and bacmid DNA were combined and incubated for 30 min at RT. The mixture was gently added to the plated *Sf9* cells and incubated for 3-5 h at 27°C. After exchange of the medium, the cells were incubated for 72 h at 27°C. The supernatant, which contained the initial virus stock V_0, was filtered using a 0.22 µm pore filter. The virus stock was stored at 4°C.

3.2.14 Virus amplification

To amplify the baculovirus, 2 mL of the initial baculovirus stock V_0 was used to infect 50 mL of *Sf9* cells with a cell density of 0.6x10^6 cells/mL. The cell suspension was incubated at 27°C and 80 rpm for 96 h. The supernatant containing the amplified baculovirus stock V_1 was harvested by centrifugation at 500 rpm for 20 min and filtered using a 0.22 µm filter. The virus stock was stored at 4°C. To further amplify the baculovirus, 100 mL of *Sf9* cells with a cell density of 1.0x10^6 cells/mL was infected with 1 mL of the V_1 baculovirus stock and incubated at 27°C and 80 rpm for 72 h. Again, the cell suspension was harvested by centrifugation at 500 rpm for 20 min. The supernatant containing the amplified baculovirus stock V_2 was filtered using a 0.22 µm pore filter and stored at 4°C.

3.3 Biochemical methods

3.3.1 Protein expression in *E. coli* cells

The bacterium *E. coli* is often the first choice when expressing recombinant proteins using bacterial expression systems. As a well-established host, *E. coli* offers significant benefits over other expression systems, including ease-of-us, scale, short culturing time, and low cost media.
E. coli BL21 (DE3) Rosetta cells are generally used for expression of recombinant proteins. A preculture, necessary for the following large-scale expression, was generated by inoculating 100-150 mL LB medium supplemented with appropriate antibiotics with a single colony and incubated over night at 37°C and 150 rpm. For the large-scale expression, 3-5 L LB medium with appropriate antibiotics was inoculated with the preculture to an $OD_{600\ nm}$ of 0.1 and incubated at 37°C and 110 rpm until an $OD_{600\ nm}$ of 0.6-0.8 was reached. Protein expression was induced by addition of 0.3-0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Typically, the cell cultures were then incubated either at 37°C for 3 h or at 16°C over night. The cells were harvested by centrifugation at 4000 rpm for 10 min. The cell pellets were subsequently washed with chilled PBS and centrifuged again at 4000 rpm for 30 min. The cell pellets were frozen in liquid nitrogen and stored at -20°C.

3.3.2 Protein expression in *Sf9* insect cells

In case the *E. coli* expression system has failed to yield a protein of desired quality or post-translational modifications (PTMs) of proteins are required, more complex protein expression systems like insect cells were used.
Sf9 insect cells with a cell density of $1.5x10^6$ cells/mL were typically used for recombinant protein expressions. The cell suspension cultures were infected with 2-3% of the baculovirus V_2 and incubated for 72-96 h at 27°C and 80 rpm at a rotational orbit of 25 mm. The cells were harvested at 2000 rpm for 10 min, washed with chilled PBS, centrifuged again at 2000 rpm for 30 min and stored at -20°C after frozen in liquid nitrogen.

3.3.3 Cell lysis

The cell pellets resulting from protein expressions in *E. coli* and *Sf9* cells were resuspended on ice in 30 mL lysis buffer supplemented with 1 mM phenylmethylsulforylfluorid (PMSF). The cells were lysed on ice by sonication using a Sonifier W-250 (Branson, Danbury, USA) at an amplitude of 40% for 3x2 min with a pulse of 5 sec on/off. To separate insoluble cell particles and cell membranes from the soluble fraction, the lysate was centrifuged at 25.000 rpm or 45.000 rpm and 4°C for 45-60 min. Prior to administration of the lysate to an affinity column, the supernatant was filtered through a 0.45 µm syringe filter to further clarify the lysate from cell membranes and other insoluble components.

3.3.4 Protein purification from inclusion bodies

Overexpression of recombinant proteins in *E. coli* often leads to the formation of insoluble protein aggregation commonly named as inclusion bodies. In order to obtain biologically active and soluble protein, inclusion bodies have to be isolated, denatured and refolded *in vitro*.
After cell lysis as described in chapter 3.3.3, the lysate was centrifuged at 15.000 rpm and 4°C for 30 min. The pellet was resuspended with 70 mL wash buffer and centrifuged at 15.000 rpm and 4°C for 30 min. This procedure was iterated three times. The pellet was then resuspended in lysis buffer and centrifuged at 15.000 rpm and 4°C for 30 min. Following this procedure, the pellet was resuspended in 80 mL extraction buffer and incubated at RT for 30-60 min. Both wash and extraction buffer contained urea, functioning as protein denaturant. After centrifugation at 15.000 rpm and RT for 60 min, protein dialyses were carried out to perform buffer exchanges. Protein dialysis was typically performed using cellulose Spectra/Por dialysis membranes with a molecular weight cut off (MWCO) of 3000 Da or 14000 Da (Carl Roth GmbH, Karlsruhe, Germany). Prior to filling the protein solution into a dialysis membrane, the membrane was boiled twice in H_2O. The first dialysis was carried out in refolding buffer I at RT, over night, which was followed by a second dialysis in refolding buffer II at 4°C for 6 h. The protein solution was then centrifuged at 15.000 rpm and 4°C for 30 min. After filtration using syringe filter (0.45 µM) (Sarstedt, Nümbrecht, Germany), the protein solution was loaded onto affinity columns for further purification steps as described in the following chapters.

3.3.5 Affinity chromatography

The affinity chromatography is a chromatographic separation method commonly used to separate the protein of interest from crude extracts. This method based on the specific reversible interaction between the specific tag fused to the target protein and the ligand coupled to the stationary phase of the chromatography column. Thus, components of the solution not or unspecific binding to the ligand were separated from the target protein by washing steps. The bound protein was stripped from the stationary phase in the elution step. Elution can be specific by using such as a competitive ligand (chemicals or proteins), or nonspecific, such as changing pH, ionic strength, or polarity. The elution procedure resulted in a purified and concentrated target protein.
In this thesis, polyhistidine- (His_6-), Glutathione-S-Transferase- (GST-), and maltose-binding-protein- (MBP-) tag were used for affinity purifications. Between the protein of interest and the tag, a TEV (*tobacco etch virus*) or a PreScission protease site was usually inserted in order to cleave off the tag after the affinity purification by proteolytic digest, if required.

Immobilized metal chelates affinity chromatography

Immobilized metal chelate affinity chromatography (IMAC) is a standard method performed to purify polyhistidine-tagged proteins. The underlying principle of IMAC based on the characteristic of the histidine to act as an electron donor by reversibly binding to divalent transition metal ions. An appropriate stationary phase was prepared by coupling of nitrilotriacetic acid (NTA), acting as a chelator, to agarose-resin which was then loaded with Ni^{2+}-ions. Ni^{2+}-NTA matrix coupled to a solid support resin is commercially available. This matrix coordinates the Ni^{2+}-ions through four coordination sites while leaving two coordination sites of the Ni^{2+}-ions exposed to interact with histidine residues in the affinity tag. Coordination bonds are formed between the electron donor groups on the histidine imidazole ring and the Ni^{2+}-ion. For protein purification, the cell lysate of the N-terminally His_6-tagged protein was loaded onto a HisTrap FF crude column (GE Healthcare, München, Germany). The column was pre-equilibrated with the corresponding lysis buffer containing 5 mM imidazole in order to reduce unspecific interactions with the Ni^{2+}-NTA-resin. The reagent imidazole acts a competitive ligand of the histidine residues. To further remove unspecific bindings, the column was washed with 10-15 column volumes (CV) of wash buffer containing 10 mM of imidazole. The elution step was carried out with a 20 CV gradient of 20 to 300 mM imidazole in elution buffer in order to elute the protein from the column slowly. Fractions showing absorbance at 280 nm in the chromatogram were further analyzed using SDS-PAGE (referred to 3.3.6). Fractions containing the desired protein were pooled and subsequently concentrated (referred to 3.3.7).

GST-affinity chromatography

Inclusion of the Glutathione S-transferase (GST) protein in the vector of recombinant proteins often promotes expression and solubility of the proteins since GST folds rapidly into a stable and highly soluble protein upon translation. The purification of a N-terminally GST-tagged, recombinant protein was performed using a column consisting of sepharose resins coupled to Glutathione (GSH). Immobilized GSH can be used to capture GST-tagged proteins via the enzyme-substrate binding reaction based on the high binding affinity of GST to its substrate GSH.
The cell lysate containing the GST-tagged protein was loaded onto a GSTrap FF column (GE Healthcare, Munich, Germany), which was pre-equilibrated with the corresponding lysis buffer. The washing procedure was carried out with the wash buffer, which was followed by an additional washing step with the lysis buffer. The wash buffer contained a high salt content of 1 M NaCl which has the function to reduce unspecific interactions of proteins with the stationary phase. The column was washed with at least 10 CV of the wash and lysis buffer, respectively. The protein of interest was eluted from the GSH resin by competitive displacement with a buffer containing 10 mM GSH. Fractions showing absorbance at 280 nm in the chromatogram were further analyzed using SDS-PAGE (referred to 3.3.6). Fractions containing the protein of interest were pooled and subsequently concentrated (referred to 3.3.7).

Maltose binding protein (MBP)-affinity chromatography

Maltose binding protein (MBP) is a commonly used tag for purification of recombinant proteins. First, MBP facilitates the proper folding and solubility of the target protein. Second, due to its specific interaction with maltodextrins, MBP allows the use of a simple capture affinity step on sepharose resin coupled to dextrin or amylose. Since MBP belongs to the maltose system, it is responsible for the uptake and metabolism of maltodextrins. Purification of MBP-tagged recombinant proteins result often in a protein with higher purity than GST-tagged proteins. The cell lysate containing the MBP-tagged protein was loaded onto a MBP-trap column (GE Healthcare, Munich, Germany), which was pre-equilibrated with the corresponding lysis buffer. The wash step was carried out by a gradient wash over 20 CV with lysis buffer and wash buffer to slowly remove glycerol. The protein was subsequently eluted with a buffer containing 15 mM maltose. Maltose interacted directly with the MBP-tag by competing for the interaction sites MBP used for binding to the dextrin-sepharose matrix. Fractions showing absorbance at 280 nm in the chromatogram were further analyzed using SDS-PAGE (referred to 3.3.6). Fractions containing the protein of interest were pooled and subsequently concentrated (referred to 3.3.7).

3.3.6 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Polyacrylamide gel electrophoresis is a standard biochemical technique used to separate proteins in a polyacrylamide gel in an electric field depending on their molecular weight. Smaller proteins migrate faster through the gel than bigger proteins. The separation efficiency depends on the percentage of applied polyacrylamide, which determines the pore sizes. The higher the percentage, the smaller the pore size. The choice is depending on the size of the protein of interest. Polymerization of the polyacrylamide is induced and catalyzed by ammonium persulfate (APS) and tetramethylethylenediamine (TEMED). In this study, a discontinuous electrophoresis system was used. A lower percentage stacking gel is on top of the higher percentage separating gel. Thus, samples were concentrated first before separation. Electrophoretic separation of proteins was carried out under denaturing conditions by using of the anionic detergent sodium dodecyl sulfate (SDS). SDS disrupts the tertiary structure of proteins and upon binding to the denatured proteins the proteins are uniformly negatively charged. Since the SDS-coated proteins have the same charge to mass ratio, migration through the gel only based on the molecular weight of the proteins.

Electrophoresis was performed using the Mini Protean Tetra Cell system (Bio-Rad, München, Germany) at 160 V for 60-90 min at RT. After the run, the gels were stained using Coomassie Brilliant Blue staining solution and destained using Coomassie destain solution. The protein markers LMW marker (GE Healthcare, München, Germany) and Page Ruler Plus Prestained protein ladder (Thermo Fisher Scientific, Waltham, USA) were used to calibrate the migration behavior of the protein and estimate its molecular weight. Acrylamide concentration of the separating gel varies between 12-18%, depending on the size of the protein of interest. The stacking gel consists of 5% acrylamide. The recipes for commonly used 12%, 15% and 18% separating gels and 5% stacking gel are shown in the table below.

Table 16: SDS-PAGE composition.

	12% running gel	**15% running gel**	**18% running gel**
	4 gels	**4 gels**	**4 gels**
H_2O	6.75 mL	4.65 mL	1.65 mL
30% acrylamide	8.40 mL	10.50 mL	13.50 mL
running gel buffer	5.85 mL	5.85 mL	5.85 mL
10% APS	235.50 µL	235.50 µL	235.50 µL
TEMED	7.10 µL	7.10 µL	7.10 µL

	5% stacking gel
	4 gels
H_2O	3.60 mL
30% acrylamide	0.90 mL
running gel buffer	0.70 mL
10% APS	53.0 µL
TEMED	5.30 µL

3.3.7 Concentration of protein samples

After the affinity and size exclusion chromatography, collected fractions containing the purified protein of interest were pooled and concentrated via ultrafiltration using Amicon Ultra Centrifugal filters (Sigma-Aldrich, Munich, Germany) available in different molecular weight cut-offs (3, 10, 30, 100 kDa). Appropriate molecular weight cut-off keeps the target protein that exceeded the cut-off inside the filter unit by allowing remaining solution and small molecules pass. Thus, the protein solution can be concentrated up to a volume in the microliter range until the desired protein concentration or volume was reached. Prior to use, the filters were washed with H_2O and equilibrated with the corresponding buffer by centrifugation. The protein samples were concentrated by centrifugation at 4°C according to the manufacturer's instructions.

3.3.8 Determination of protein concentration

The protein concentration was measured using a NanoDrop 2000/c Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, USA). The aromatic residues of the amino acids tyrosine, tryptophan, phenylalanine and also the disulfide bonds formed between cysteine residues absorb UV light at a wavelength of 280 nm. This characteristic is common used to determine protein concentrations by utilizing the Lambert-Beer´s law:

Equation 2: Lambert-Beer's law

$$A_{280} = \varepsilon \times c \times d \quad \Leftrightarrow \quad c = \frac{A_{280}}{\varepsilon \times d} \left[\frac{\text{mol}}{\text{L}}\right]$$

A: absorption at 280 nm
ε: molar extinction coefficient at 280 nm [$M^{-1}cm^{-1}$]
c: protein concentration [mol L^{-1}]
d: cuvette thickness [cm]

The specific extinction coefficient ε of a protein, depending on the amino acid composition of each protein, describes the weighed sum of the molar absorptivities of the amino acids tryptophan, tyrosine, cysteine and phenylalanine at the wavelength of 280 nm. At an absorption of 280 nm, the molar extinction coefficient is calculated as following:

Equation 3: Calculation of molar absorption coefficient

$$\varepsilon_{280} = n_{Trp} \times 5500\ [M^{-1}cm^{-1}] + n_{Tyr} \times 1490\ [M^{-1}cm^{-1}] + n_{SS} \times 125\ [M^{-1}cm^{-1}]$$

The specific extinction coefficient for each protein was calculated using an online tool of protein extinction coefficient calculator (BioMol.net, 2019).

3.3.9 Size-exclusion chromatography (SEC)

After affinity chromatography, the proteins were subjected to further purification using size-exclusion chromatography (SEC). The SEC, also known as gel filtration, is a hydrodynamic technique that separates macromolecules in solution according to their size and shape. This separation method is based on the relative abilities of macromolecules to penetrate a bed of porous beads with well-defined pore sizes, which determines the range of fractionation. The SEC column is composed of a stationary phase made of a porous matrix and a mobile phase of aqueous buffer system. Molecules larger than the fractionation range are completely excluded from entering the pores. They elute first at the void volume (V_0), which is the volume of the mobile phase in a column. Within the fractionation range, small molecules are able to diffuse into the pores of the beads, thus they move through the bed more slowly than bigger molecules and elute last. In general, the smaller the molecule, the greater is its chance to enter the pores of the matrix and thus the greater is its retention time. The separation resolution can be influenced by diverse factors such as the bed material, the column size, amount of the protein as well as the flow rate.

Preparative size-exclusion chromatography

Preparative SEC was performed as the second purification step of recombinant proteins after the affinity chromatography procedure. Typically, a HiLoad 16/600 Superdex 75 prep grade column (GE Healthcare, Munich, Germany) was used to separate proteins of a size ranging from 3 kDa to 70 kDa. Proteins in a range from 10 kDa to 600 kDa were separated using a HiLoad 16/600 Superdex 200 prep grade (GE Healthcare, Munich, Germany). Depending on their volume, the protein samples were injected via a 1-5 mL loop onto the column, which was pre-equilibrated with the gelfiltration buffer. According to the manufacturer´s instructions, the flow rate was adjusted between 0.2-1 mL/min. Fractions showing absorbance at 280 nm in the chromatogram were further analyzed using SDS-PAGE (referred to 3.3.6). Fractions containing the protein of interest were pooled and subsequently concentrated (referred to 3.3.7). The protein sample was frozen immediately in aliquots using liquid nitrogen and stored at -80°C.

For applications such as buffer exchange and removal of low molecular weight compounds, pre-packed PD-10 desalting gel filtration columns (GE Healthcare, Munich, Germany) were used. The columns contain Sephadex G-25 medium, which allows rapid separation of high molecular weight substances from substances with low molecular weight. The separation was performed using the gravity protocol according the manufacturer's instructions.

3.3.10 Mass spectrometry

Mass spectrometry (MS) is a sensitive analytical technique widely used in biological laboratories that addresses different fields of applications by providing both qualitative and quantitative information on the molecules of interest after their conversion to ions. It can be used to determine the molecular mass of the analyte molecules, especially to confirm PTMs. It can also be used to identify and characterize proteins/peptides by performing peptide mass fingerprint (PMF) analysis.
Prior accessing the mass analyzer, the molecules are first ionized to acquire positive or negative charges by introducing the sample into the ionization source of the mass spectrometer. The ions are separated according to their mass-to-charge (m/z) ratio. A mass spectrum displays the signals by showing the relative abundance of the signals versus their m/z ratio.

Determination of the molecular mass of proteins

In this study, electrospray ionization mass spectrometry (ESI-MS) was performed to determine the molecular mass of the protein of interest. The procedure of ESI induces soft ionization, resulting in low fragmentation of the sample. Thus, it is particularly suitable for thermally labile biomolecules.
ESI uses electrical energy to assist the transfer of ions from solution into the gas phase. The diluted sample solution is injected through a capillary at low flow rate. A high voltage of 2–6 kV is applied to the tip of the metal capillary. The strong electric field causes a dispersion of the sample solution into an aerosol of highly charged droplets. On the way towards the mass analyzer, the droplets are continuously reduced in size by solvent evaporation, leading to an increase of surface charge density. If the electric field strength within the charged droplets reaches a critical point, ions are ejected from the droplets into the gas phase. They were further accelerated into the mass analyzer for subsequent analysis of molecular mass and measurement of ion intensity according to their m/z ratio.
The ESI-MS analysis to determine the protein´s molecular mass was performed by the Protein Analysis Group of the Functional Genomics Center Zurich (FGCZ) (Zurich, Switzerland). Prior to the ESI-MS analysis, the samples were desalted using a C4 ZipTip. The sampling cone energy was set at 40 V. The m/z data were then deconvoluted into MS-data using the MaxEnt1 Software.

Peptide mass fingerprinting (PMF) analysis

Peptide mass fingerprinting (PMF) analysis is an analytical technique used in the field of mass spectrometry which not only allows a precise determination of the molecular mass of peptides but also the determination of their sequences. Together, this information can be used for protein identification, de novo sequencing and identification of specific posttranslational modification sites. Proteins to be characterized were first digested into peptides by trypsin protease. The absolute mass of the peptides was analyzed typically by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). Like ESI, it also induces soft ionization by using a laser energy absorbing matrix to create ions with minimal fragmentation. In PMF analysis, MALDI is more preferred, as it has a wide sequence coverage. The peptide mass was then compared to databases containing the theoretical peptide masses of all known protein sequences. The identity of the protein was finally determined by the match with the highest score.

PMF analysis to identify (unknown) proteins and their specific phosphorylation sites were performed both by the Protein Analysis Group of the FGCZ (Zurich, Switzerland) and the Bioanalytical Mass Spectrometry Group of the Max Planck Institute for Biophysical Chemistry, Göttingen, Germany (Prof. Dr. H. Urlaub). SDS-PAGE separated and Coomassie stained protein gel bands were stored in sterile Eppendorf tubes and sent to the respective laboratories.

The procedure carried out in the FGCZ was the following: gel bands were cut into small pieces and were washed two times with 100 µL 100 mM NH_4HCO_3/50% acetonitrile and one time with 50 µL acetonitrile. Subsequently, the samples were digested with10 µL trypsin (5 ng/µL in 10 mM Tris/2 mM $CaCl_2$, pH 8.2) in 30 µL buffer (10 mM Tris/2 mM $CaCl_2$, pH 8.2) at 60°C for 34 min in microwave. The supernatant was removed and the gel pieces were extracted with 150 µL 0.1% TFA/50% acetonitrile for 15 min by ultrasonic. The samples were then dried and dissolved in 20 µL 0.1% TFA. 4 µL of the solution were desalted using a C18 ZipTip and eluted in 4 µL 50% ACN including 0.1% TFA. Of this 1 µL was mixed in 1:1 ratio with the matrix solution (1.4 mg/ml HCCA in 85% ACN, 0.1% TFA, 1mM $NH_4H_2PO_4$) and spotted onto the target. The samples were analyzed by MALD-MS. Database searches were performed by using the database search programs Mascot (Matrix Science Inc., Boston, USA) and Peaks Studio (Bioinformatics Solutions Inc., Waterloo, Canada).

3.3.11 Kinase assays

Determination of the phosphorylation activity of different kinases was performed in *in vitro* assays using recombinantly expressed and purified kinases, different kinase substrate peptides/proteins and adenosine triphosphate (ATP) in presence or absence of low molecular weight compounds as potential kinase inhibitors. The kinase reaction, characterized by the quantity of transferred phosphate groups of ATP to substrate, was examined via different methods such as filter binding assays using radioactive-labeled [γ-^{32}P]-ATP, ESI-MS, as well as PMF analysis.

Radioactive [γ-^{32}P]-ATP kinase assay

Radioactive *in vitro* kinase assays utilizes ATP, which is labeled with the beta-radiator ^{32}P on the gamma phosphate group. In this assay, the kinase transfers a radioactive phosphate group from ATP to the target substrate. Filter binding assays with radioactive ATP were performed using nitrocellulose paper as filter matrix, because of its negatively charged nature. It is well suited for immobilizing proteins, which have usually a net positive charge. The negatively charged ATP, however, can not stick to the filter paper by itself. Only the phosphorylated substrate will stick. In this way, free radiolabeled ATP is separated from radiolabeled substrate. The amount of ^{32}P-phosphorylated substrate is then quantified by measuring the amount of radioactivity on the filter paper using a scintillation counter.

Radioactive kinase reactions were typically carried out in a total volume of 30 µL containing kinase buffer (40 mM Hepes, pH 7.5, 34 mM NaCl, 34 mM KCl, 10 mM $MgCl_2$, 5% Glycerol), 0.2 µM kinase, 10-100 µM substrate, cold ATP (to a final concentration of 0.2 or 1 mM) and 0.6 or 3 mCi [γ-^{32}P]-ATP. Various small molecular weight compounds, acting as potential ATP-competitive kinase inhibitors, were used as described in the respective sections. Except in the case of time-dependent measurements, the reaction mixture was incubated at 30°C and 350 rpm for 30 min by default and terminated by adding EDTA to a final concentration of 50 mM. Aliquots of 15 µL each were spotted onto P81 Whatman paper squares. Paper squares were washed three times for 5 min with 0.75% (v/v) phosphoric acid, with at least 5 ml washing solution per paper square. Radioactivity was quantitated in a Beckman Scintillation Counter (Beckman-Coulter) for max. 5 min. Measurements were performed in duplicate and are represented as mean with standard deviation.

Detection of phosphorylation via mass spectrometry

The purpose of the radioactive kinase assay is to provide a qualitative indication of the kinase phosphorylation activity as well as of the inhibitory activity of potential kinase inhibitors by defining IC_{50}-valuses. By performing PMF analysis, as described in chapter 3.3.10, the absolute number of phosphorylations as well as specific phosphorylation sites on substrates can be determined.
The standard reaction mixture of the kinase assay for PMF analysis contained 0.2 µM kinase, 10-100 µM substrate and 1 mM cold ATP. The reaction was incubated at 30°C and 350 rpm. After 4 h, the reaction was stopped by addition of 5 µL SDS sample buffer, allowing a direct separation of the protein of interest from other impurities using SDS-PAGE. Protein gel bands were further analyzed by mass-spectrometer as described.

3.3.12 Pull-down assays

The pull-down assay is a small-scale affinity purification method, used to determine physical interactions between two or more proteins. It is a very valuable tool for both confirming the existence of a predicted protein-protein interaction or identifying previously unknown protein–protein interactions as an initial screening assay. GST- and His_6-tagged fusion proteins are commonly used to perform pull-down experiments. Typically, the affinity system consists of a GST- or His_6-tagged protein (bait protein) that is captured by agarose beads coated a GSH- or Ni^{2+}-chelate complex. Acting as bait protein, the immobilized fusion-tagged protein is able to capture potential interacting partner (prey protein). The immobilized bait protein is then incubated with the recombinant highly purified prey protein, intending to confirm previously suspected protein–protein interaction. After the prescribed washing steps, selective protein elution was performed using competitive analytes depending on the affinity ligand. The elution sample as well as the supernatant received after protein incubation were further analyzed by SDS-PAGE as described in chapter 3.3.6. In order to demonstrate that the final result truly based on the interaction of bait and prey protein, each pull down assay has to be properly controlled. Thus, bait- and prey-free controls have to be performed to both confirm the absence of bait–prey interactions in the bait lysate and that the prey protein does not bind to the affinity ligand coated agarose beads.
Prior to load the bait protein, 100 µL of the desired agarose beads were first sedimented and then equilibrated by washing with the gelfiltration buffer five times by centrifugation at 2000 rpm and 4°C for 3-5 min. 250 µg of the bait protein was added to the beads and the reaction mixture was filled up to 500 µL. After incubation at 4°C for 2-3h, the beads were washed with the gelfiltration buffer three times in order to remove unbound proteins. 100 µL gelfiltration buffer was then added to the

washed beads and divided into 50 µL aliquots. The protein coated beads were stored at -80°C after freeze by liquid nitrogen. 300 µg of the prey protein was added to the prepared aliquot of beads and the reaction mixture was filled up to 500 µL. After incubation at 4°C for 1h, the beads were sedimented by centrifugation at 2000 rpm and 4°C for 5 min. The obtained supernatant was kept for further SDS-PAGE analysis (20 µL supernatant + 5 µL 5x SDS sample buffer). After washing the beads with the gel filtration buffer three times in order to remove unbound proteins, the potential protein-protein interaction complex was eluted by adding of 30 µL appropriate elution buffer. The reaction mixture was incubated on ice for 15 min and then centrifuged at 2000 rpm and 4°C for 5 min. The supernatant was used for SDS-PAGE analysis.

3.3.13 Crystallization

As protein crystallization requires the protein to archive the state of supersaturation, the purified protein samples were concentrated to the highest possible degree without any sign of aggregation or precipitation (usually 8-20 mg/mL for CDK kinases). Crystallization was conducted using the hanging drop vapor diffusion method. A drop containing a mixture of precipitant agent and protein solution in a one to one ratio is sealed upside down in a well, filled with a reservoir of the same precipitant agent of higher concentration. Usually, the precipitating agent consists of salt of some kind and a buffer to set the pH. In some cases, additives may be added specifically for the protein of interest. The incubation temperature was generally between 4°C and room temperature. Over time, water diffuses out of the hanging drop until the osmolarity of the drop and the precipitant are equal. By this way, the protein concentration will increase and cause precipitation and potentially leading to crystallization of the protein. Different compositions of the precipitant agent and incubation temperatures were usually tested to identify the optimal crystallization condition for the protein of interest (Dessau and Modis, 2011).

Crystallographic trials were performed by Dr. Kanchan Anand at the Institute of Structural Biology, Bonn, Germany.

In the following, the concentration of the proteins applied to crystallization trials including the respectively used precipitant agent leading to protein crystallization are listed:

Table 17: Proteins applied to crystallization trials, including applied protein concentration and the used precipitant agent.

Protein	**Precipitant agent**
CDK12H1040G/CycK + Dinaciclib (8 mg/mL)	0.1 M HEPES pH 7.0 + 10 % PEG8000, 0.2 M $MgCl_2$
CDK12/CycK + CMP-175 (18 mg/mL)	0.1 M MES pH 6.0 + 0.2 M $MgCl_2$ + 30 % PEG mix 4 (mix of PEG 3350 + PEG 4k + PEG 6K) + 0.3 M NDSB
CDK14 (125-469)/CycY (73-341) (9 mg/mL)	0.1 M sodium acetate pH 4.6 + 0.1-0.2 M $LiSO_4$ + 0.2 M sodium acetate pH 8.4 + 60 % Peg 400
CDK14 (133-447)/CycY (92-341) (8 mg/mL)	same as above
CDK14 (125-469)/CycY (73-341) + compound (9 mg/mL)	same as above
CDK14 (133-447)/CycY (92-341) + compound (8 mg/mL)	same as above

4 Results

Transcriptional kinases – CDK12 and CDK13

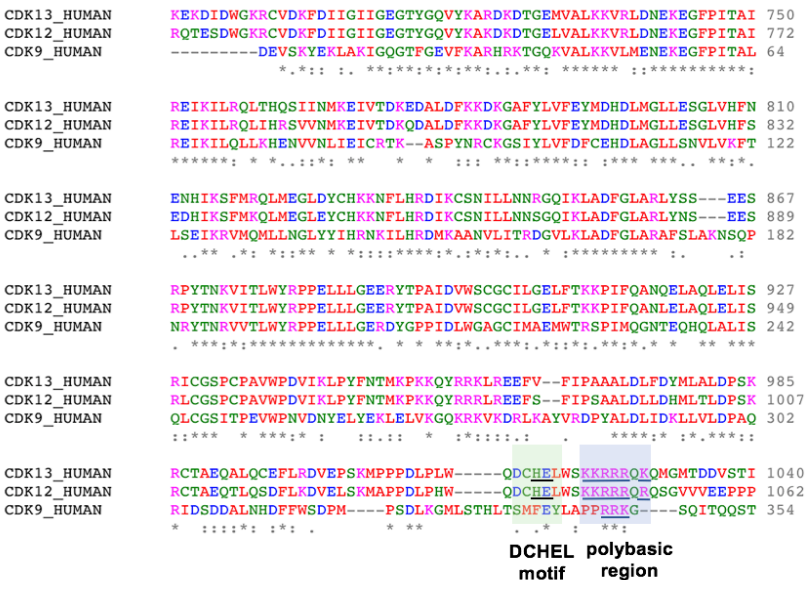
4.1 Identification and characterization of potential CDK12 and CDK13 small molecule inhibitors

4.1.1 Impact assessment of dinaciclib

4.1.1.1 Identification of potential CDK12 and CDK13 inhibitor by *in silico* modeling

Although CDK12 and CDK13 share extensive primary sequence homology (Figure 18A) as well as tertiary structure homology with CDK9 within their kinase domains (Figure 18B), a panel of well known small molecule CDK9 inhibitors was previously shown to have substantially reduced potency against CDK12 in *in vitro* biochemical assays (Bösken *et al.*, 2014). Analysis of the C-terminal extension helix of each kinase domain reveals significant variance between CDK12/13 and CDK9. While CDK12/13 harbour an initiating $^{1038/1016}$ DCHEL motif, CDK9 does not share this motif in its C-terminal extension helix. Since this structural variation occurs in close proximity to the site of binding for small molecule inhibitors of CDK9, we hypothesized that the presence of the DCHEL motif in CDK12/13 may be responsible for the modest inhibitory activity of the previously tested CDK9 inhibitors against CDK12.

A



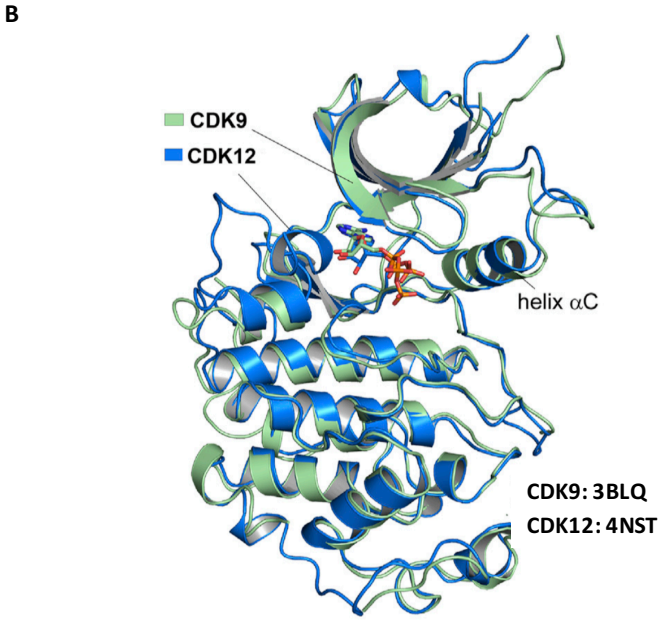


Figure 18: (A) Sequence alignment of the human transcription regulating kinases in composed of CDK9, CDK12 and CDK13. In contrast to CDK12/13, the C-terminal extension helix of CDK9 does not contain a DCHEL motif and only a shortened basic cluster. **(B)** Superimposition of the CDK9 crystal structure (PDB: 3BLQ; (Baumli *et al.*, 2008) with the crystal structure of CDK12 (PDB: 4NST; (Bösken *et al.*, 2014). The two kinases share extensive tertiary structural homology in the kinase domain with a root-mean-square deviation (RMSD) of 0.83 Å.

To address this hypothesis, flavopiridol, a well studied potent small molecule inhibitor of CDK9, was modeled into the ATP-binding site of CDK12 by superimposition of the CDK9-flavopiridol crystal structure (PDB: 3BLR; (Baumli *et al.*, 2008) with the crystal structure of CDK12-ADP (PDB: 4NST; (Bösken *et al.*, 2014) (Figure 19A). Structural modeling of flavopiridol into the ATP-binding site of CDK12 revealed a significant steric clash between the benzene ring of bound flavopiridol with the imidazole ring of H1040.

To confirm whether this occlusion was a shared feature of other known inhibitors of CDK9, we modeled dinaciclib, a known small molecule inhibitor of both CDK9 and cell cycle-related CDKs 1, 2, 5 (Parry *et al.*, 2010) that had not been tested against CDK12 *in vitro*, into the ATP-binding site of CDK12 (Figure 19B). The model was generated by superimposition of the CDK2-dinaciclib crystal structure (PDB: 4KD1; (Martin *et al.*, 2013) with the crystal structure of CDK12-ADP (PDB: 4NST; (Bösken *et al.*, 2014). In contrast to flavopiridol, modeling of dinaciclib did not show potential steric hindrance between the aromatic ring of H1040 with dinaciclib. Beside the missing sterical clash, this model additionally revealed possible overlapping of the pyridine-N-oxide ring of dinaciclib with the aromatic ring of H1040. This event may result in a favorable π-π stacking interaction of the aromatic ring systems that may stabilizes the interaction as well as contribute to the binding specificity between dinaciclib and CDK12.

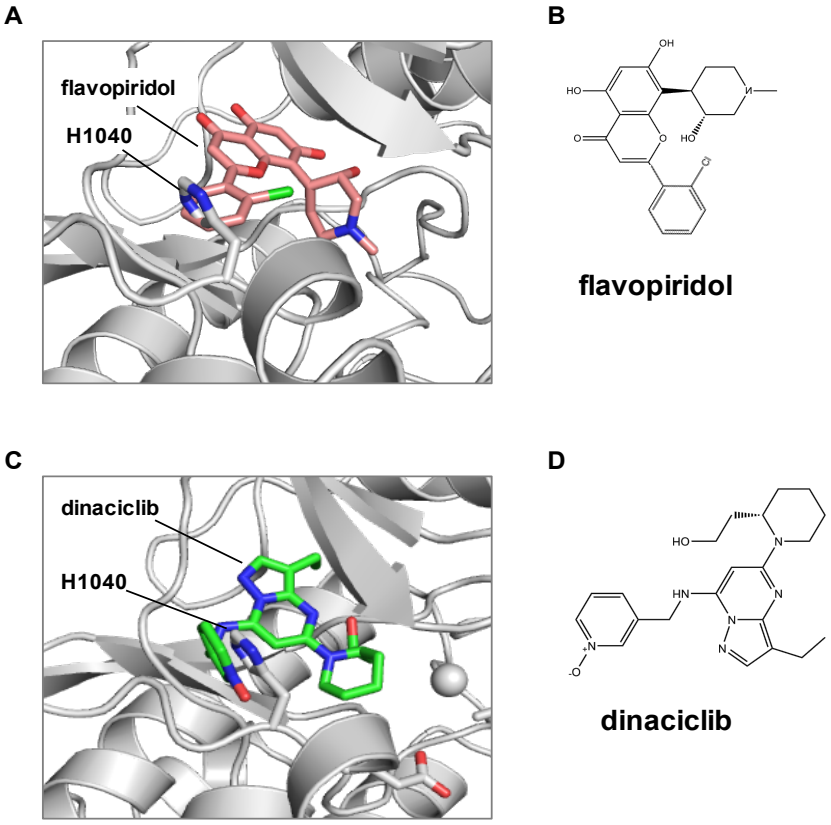


Figure 19: Modeling of flavopiridol and dinaciclib into the ATP-binding site of CDK12. (A) Modeling of flavopiridol into the CDK12 ATP-binding site by superimposition of CDK9-flavopiridol crystal structure (PDB: 3BLR; (Baumli *et al.*, 2008) with the crystal structure of CDK12-ADP (PDB: 4NST; (Bösken *et al.*, 2014). **(B)** Chemical structure of flavopiridol. **(C)** Modeling of dinaciclib into the CDK12 ATP-binding site by superimposition of CDK2-dinaciclib crystal structure (PDB: 4KD1; (Martin *et al.*, 2013) with the crystal structure of CDK12-ADP (PDB: 4NST; (Bösken *et al.*, 2014). **(D)** Chemical structure of dinaciclib.

4.1.1.2 Analysis of the inhibitory activity of dinaciclib in comparison with flavopiridol

Based on the result from the computational modeling, we assumed that the absence of steric clash as well as the presence of possible favorable base-stacking interaction could afford dinaciclib to be a potent inhibitor of CDK12 and CDK13. For the *in vitro* kinase assay using radioactive labelled [^{32}P]-γ-ATP, purified recombinant protein complex of CDK9, CDK12, and CDK13 containing the respective kinase domains (CDK9 (1-372), CDK12 (714-1063), CDK13 (694-1039) and their corresponding cyclin subunits were tested for inhibition with dinaciclib. In accordance with the outcome of the *in vitro* kinase assay of CDK9/Cyclin T1, addition of 10-fold and 1000-fold concentration of dinaciclib to 0.2 µM purified recombinant CDK12/CycK, CDK13/CycK

holoenzyme complexes at 1 mM ATP concentration significantly reduced the kinase activity by approximately 10- to 20-fold (Figure 20).

This result was further confirmed by dose-dependent inhibition measurements in order to determine the half-maximal inhibitory concentration (IC_{50}) values of dinaciclib against CDK12/CycK, CDK13/CycK, and CDK9/CycT1. As shown in figure 21, dinaciclib demonstrated robust inhibitory activity against each of the three kinases, with IC_{50} values in the 40-80 nM range.

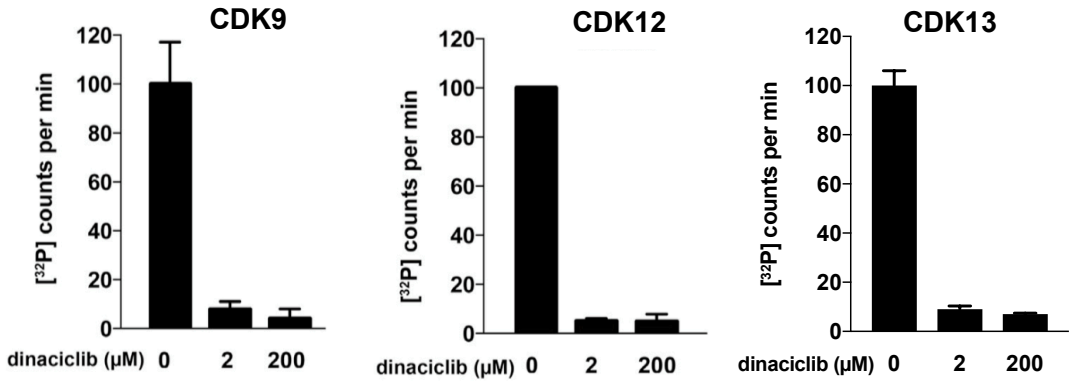


Figure 20: *In vitro* kinase assays without and with addition of 10-fold and 1000-fold dinaciclib using purified recombinant kinase complexes of CDK9/CycT1, CDK12/CycK, and CDK13/CycK at 0.2 µM kinase concentration with 1 mM ATP. Data are shown as mean ±SD; n=2.

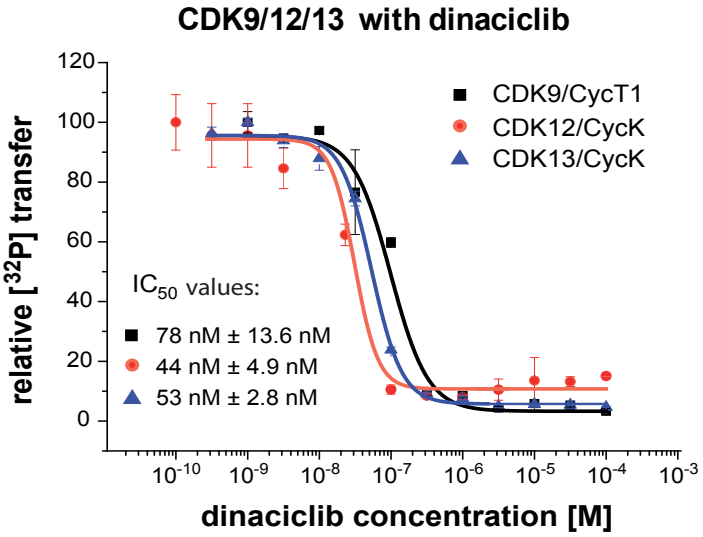


Figure 21: Dose-dependent inhibition measurements of dinaciclib with different transcription regulating kinases. 0.2 µM of purified recombinant kinase complexes of CDK9/CycT1, CDK12/CycK and CDK13/CycK were used. The counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are shown as mean ±SD; n=3.

In addition to CDK9 and CDK12/13 inhibition, dinaciclib is also known to potently inhibit CDK1/2/5 (Parry *et al.*, 2010). To confirm the inhibitory activity of dinaciclib against other members of the CDK family, further dose-dependent inhibition measurements were performed to determine IC_{50} values of dinaciclib for CDK1/2/5 (Figure 22). Using purified recombinant kinases purchased from ProQinase (Freiburg, Germany), we determined IC_{50} values of 632 nM ± 99 nM for CDK1/CycB1, 144 nM ± 26 nM for CDK2/CycA2 and 64 nM ± 25 nM for CDK5/p35NCK, respectively. In comparison with the IC_{50} values of dinaciclib obtained for CDK12 and CDK13 (44 nM and 53 nM, respectively), inhibition of CDK12 and CDK13 is even more potent than the inhibition of CDK1/2/5 by dinaciclib. In table 18, the determined IC_{50} values of dinaciclib for the evaluated CDK family members are summarized.

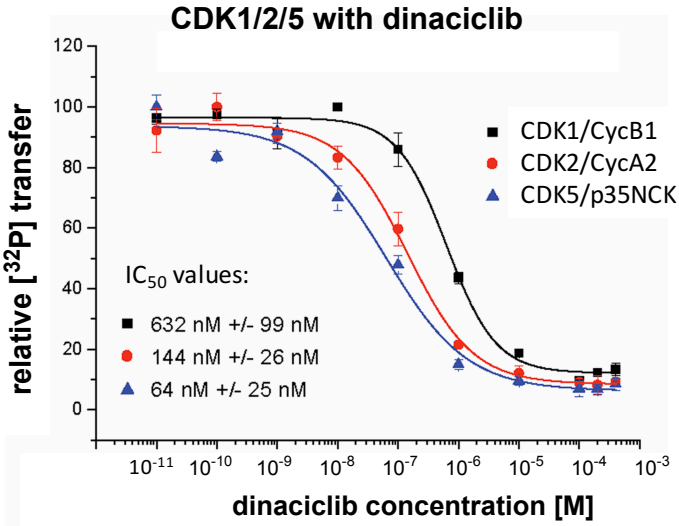


Figure 22: Dose-dependent inhibition measurements of dinaciclib with different cell cycle-related kinases. 0.2 µM of purified recombinant kinase complexes of CDK1/CycB1, CDK2/CycA2 and CDK5/p35NCK were used. Protein complexes were purchased from ProQinase (Freiburg, Germany). The counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are shown as mean ±SD; n=3.

Table 18: Overview of the determined IC_{50} values of dinaciclib for various CDK family members.

	CDKs	IC_{50} value
dinaciclib	CDK1/CycB1	632 ± 99 nM
	CDK2/CycA2	144 ± 26 nM
	CDK5/p35NCK	64 ± 25 nM
	CDK9/CycT1	78 ± 13.6 nM
	CDK12/CycK	44 ± 4.9 nM
	CDK13/CycK	53 ± 2.8 nM

In order to compare dinaciclib with flavopiridol regarding their inhibitory potency against CDK9 and CDK12, dose-dependent inhibition measurements of dinaciclib and flavopiridol were performed for both kinases. Whereas flavopiridol had only modest activity against CDK12, with potency compared to CDK9 reduced by more than 100-fold (Figure 23B and D), dinaciclib clearly demonstrated potent inhibitory activity against CDK12. Obtained IC_{50} values for dinaciclib against CDK9 and CDK12 were almost equivalent in 40-60 nM range (Figure 23A and C).

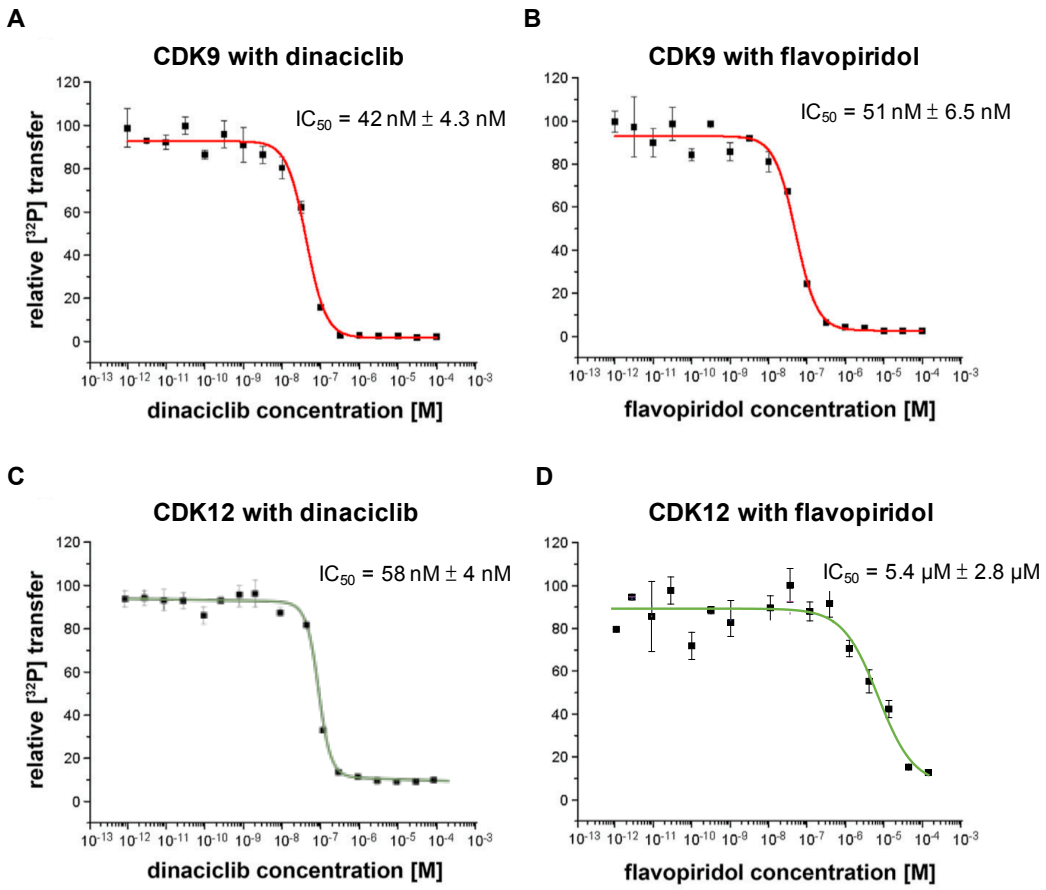


Figure 23: Dose-dependent inhibition measurements of dinaciclib and flavopiridol with CDK9 and CDK12. IC_{50} measurements of **(A)** dinaciclib and **(B)** flavopiridol with 0.2 µM CDK9/CycT1 kinase concentration. IC_{50} measurements of **(C)** dinaciclib and **(D)** flavopiridol with 0.2 µM CDK12/CycK kinase concentration. The counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are shown as mean ±SD; n=3.

4.1.1.3 Study the effect of C-terminal extension helix mutations on the inhibitory potency of dinaciclib and flavopiridol

Previous results from both *in silico* modeling and *in vitro* kinase assays indicated potential correlation between specific residues of the C-terminal extension helix characteristic for CDK12/13 and the inhibitory potency of dinaciclib and flavopiridol. To confirm the interaction of compound with the ATP-binding site of CDK12/13, site directed mutagenesis was performed for both kinases. Under the assumption that a π-π stacking interaction may favor the inhibitory potency of dinaciclib, we hypothesized, that mutation of H1040 or H1018 of CDK12/13 and/or mutation of the surrounding residues such as the entire DCHEL motif or the nearby polybasic region might result in a protein that is less amenable of binding to dinaciclib. In contrast, based on the previous modeling and kinase activity assay results, removal of these aforementioned residues may confer sensitivity of CDK12 and CDK13 to flavopiridol. Thus, we mutated CDK12 H1040 and CDK13 H1018 to glycine (H1040G/H1018G) as well as the entire $^{1038/1016}$DCHEL motif and the nearby polybasic region $^{1045/1023}$KKRRRQR/K to alanines (referred to as 'DCHEL to Ala' and 'polybasic to Ala') (Figure 24A). In *Sf9* cells expressed and purified recombinant kinase mutants of CDK12 and CDK13 in complex with Cyclin K are shown in figure 24B compared to analogue wild type kinase.

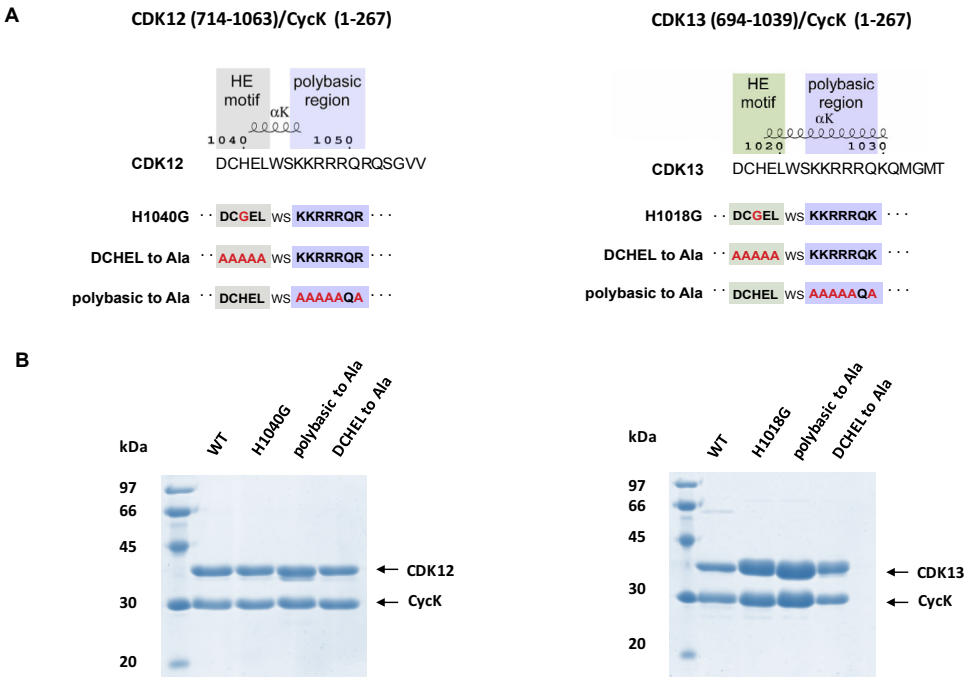


Figure 24: Site directed mutagenesis of C-terminal extension helix of CDK12 and CDK13. (A) Three mutation variants were generated to perform kinase activity assay with inhibitor. Mutation of CDK12 H1040/CDK13 H1018 to glycine, mutation of DCHEL motif and the adjacent polybasic region to alanine. **(B)** SDS-PAGE of purified recombinant kinase mutants of CDK12 and CDK13 in complex with Cyclin K compared to analogue wild type kinase.

To examine if introduced mutations may affect the natural kinase activity of CDK12 and CDK13, *in vitro* kinase assays using radioactive labelled [^{32}P]-γ-ATP were performed with established CDK12/13 substrate: cMyc protein and full length RNAPII CTD (CTD_{52}). Outcome received from kinase mutants was compared to the result received from wild type kinase (Figure 25). Negative control was performed with substrate only without addition of kinase. Typically, as shown in figure 25, cMyc protein is a much more potent target for both kinases than CTD_{52}. Nevertheless, kinase activity was detected for all generated CDK12/13 mutants comparable with the wild type kinase. Thus, this result indicates clearly, that the introduced mutations do not have considerable influence on the kinase activity by interrupting ATP-binding and can be further used for inhibitor testing.

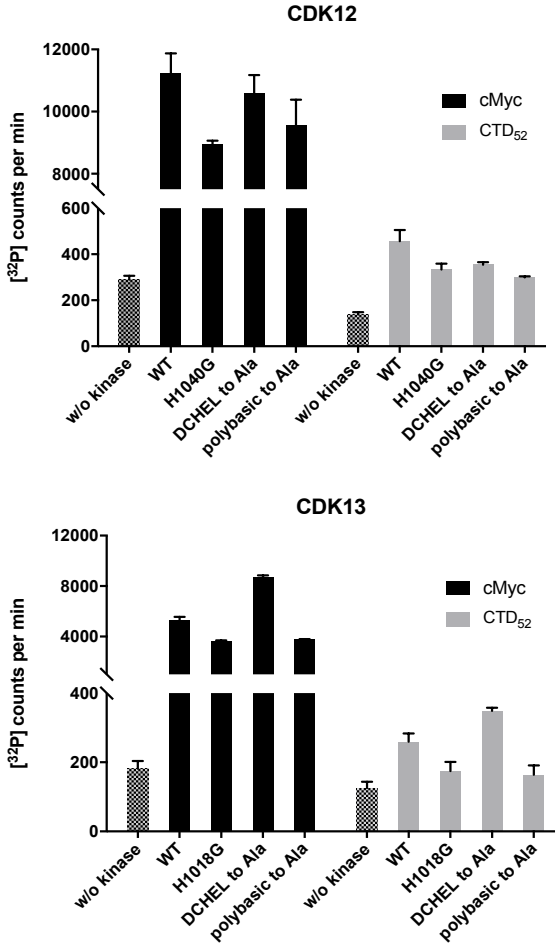


Figure 25: *In vitro* kinase assay for kinase activity validation of CDK12/13 mutants. 0.2 µM of purified recombinant kinase mutants were used with cMyc protein (50 µM) and CTD_{52} (10 µM) as substrate, and 1 mM ATP. Data are shown as mean ±SD; n=2.

The three mutants of CDK12 were chosen to study the effect of the mutations on the inhibitory potency of dinaciclib and flavopiridol. First, the inhibitory activity of the inhibitors against CDK12

mutants were analyzed by addition of 5-fold or 500-fold concentration of each of dinaciclib and flavopiridol to 0.2 µM purified recombinant kinase mutants and wild type kinase. As a next step, dose-dependent inhibition measurements of dinaciclib and flavopiridol were performed with the mutant proteins and wild type CDK12, respectively.

As shown in figure 26A and C, the inhibitory potency of dinaciclib against CDK12 mutants H1040G, DCHEL to Ala, and polybasic to Ala was comparable with the activity against the wild type protein. Addition of 5- or 500-fold concentration of dinaciclib over the kinase concentration reduced the kinase activity substantially by approximately 50- to 100-fold. The measured IC_{50} values of dinaciclib for the three kinase mutants did not show any significant discrepancy in comparison to the IC_{50} value measured for the wild type kinase. Inhibition of the two mutant variants DCHEL to Ala and polybasic to Ala by dinaciclib was even slightly more potent compared to the wild type protein, which might result from the better accessibility of the ATP-binding site upon loss of the association between the C-terminal extension helix and the kinase core domain. Based on the aforementioned results, it can be suggested that both the absence of the histidine residue at position 1040 and its surrounding residues in the C-terminal extension helix of CDK12 has no considerable effect on the inhibitory activity of dinaciclib.

Interestingly, while the inhibitory potency of dinaciclib was not impaired by the introduced mutations, the inhibitory potency of flavopiridol for the CDK12 mutants was significantly increased compared to the wild type kinase (figure 26B and D). As in the case of dinaciclib, addition of 500-fold concentration of flavopiridol almost completely reduced the kinase activity of CDK12 mutants as well as wild type protein. However, addition of 5-fold concentration of flavopiridol only reduced the kinase activity of the CDK12 mutants significantly, whereas the kinase activity of the wild type protein was just reduced by approximately 1.5-fold. Mutation of these residues also significantly decreased the IC_{50} values of flavopiridol up to 40-fold compared to the wild type protein. This result supports the assumption derived from *in silico* modeling that release of the steric clash between flavopiridol and H1040 by mutation to glycine increases the accessibility of the kinase active site for the inhibitor, thus increasing its inhibitory potency against CDK12. Moreover, the DCHEL to Ala mutant was more potently inhibited by flavopiridol than the polybasic to Ala mutant, which might loosen the association of the C-terminal extension helix to the kinase core domain.

Together, these results indicate that as opposed to the wild type protein, mutation of the H1040 site to glycine, or mutation of either the entire DCHEL motif or the adjacent polybasic region to alanine conferred sensitivity of CDK12 to flavopiridol. In contrast, theses introduced mutations of the C-terminal extension helix of CDK12 had no effect on the inhibitory potency of dinaciclib. The determined IC_{50} values of dinaciclib and flavopiridol for CDK12 mutant proteins and wild type protein are summarized in table 19.

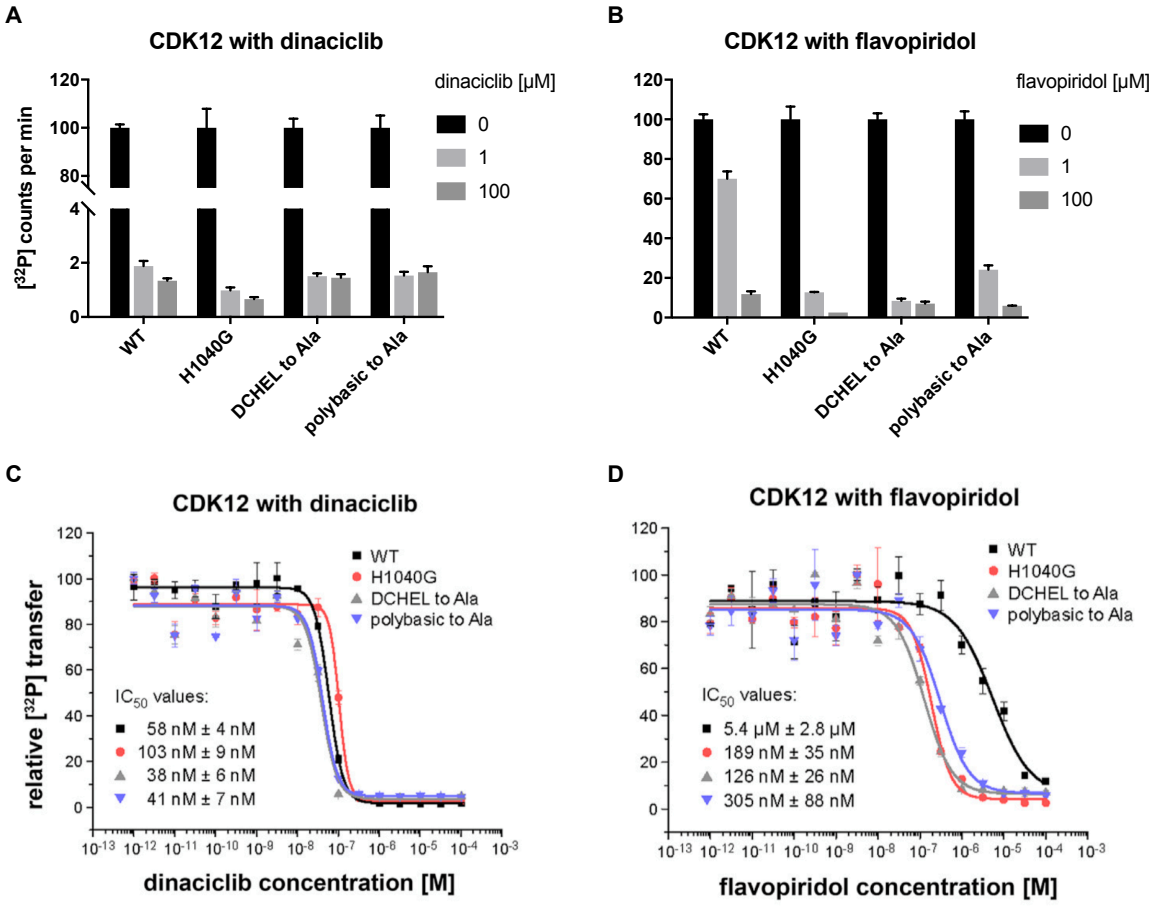


Figure 26: Analysis of the effect of specific CDK12 C-terminal extension helix mutations on the inhibitory potency of dinaciclib and flavopiridol. (A), (B) *In vitro* kinase activity assay of CDK12 mutant proteins and wild type protein alone or upon addition of 5- or 500-fold concentration of dinaciclib or flavopiridol, respectively. **(C), (D)** Dose-dependent inhibition measurement of dinaciclib or flavopiridol for CDK12 mutant proteins and wild type protein. Data are shown as mean ±SD; n=3.

Table 19: Overview of determined IC_{50} values of dinaciclib and flavopiridol for CDK12 mutants and wild type protein.

	CDK12/CycK	**IC_{50} value**
dinaciclib	WT	58 ± 4 nM
	H1040G	103 ± 9 nM
	DCHEL to Ala	38 ± 6 nM
	polybasic to Ala	41 ± 7 nM
flavopiridol	WT	5400 ± 2800 nM
	H1040G	189 ± 35 nM
	DCHEL to Ala	126 ± 26 nM
	polybasic to Ala	305 ± 88 nM

4.1.1.4 Crystallographic analysis of co-crystallization of dinaciclib with CDK12

As superimposition of resolved protein structures only provides a computational model of how interactions between dinaciclib and CDK12 could take place, we decided to perform crystallographic approaches to gain a better understanding on a structural level.

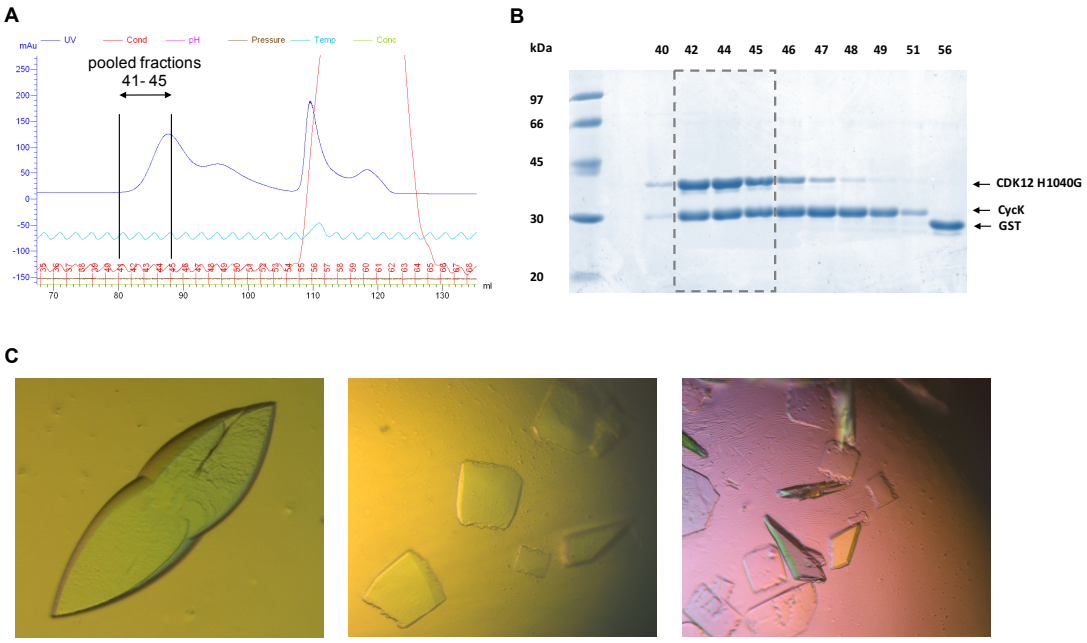


Figure 27: Crystallographic approach of co-crystallization of dinaciclib with CDK12 H1040G mutant. **(A)** Gel filtration chromatogram of purified CDK12 H1040G/CycK protein complex used in crystallization trials. **(B)** SDS-PAGE of CDK12 H1040G/CycK protein complex purified by gel filtration. Fraction samples showing an UV absorbance at 280 nm were loaded for visualization of the protein purity and ratio. **(C)** Examples of obtained protein crystals.

In the first instance, co-crystallization trials of dinaciclib with wild type CDK12/CycK protein complex have failed in initiation of protein crystallization. Based on the findings that CDK12 proteins with C-terminal extension helix mutation are irrelevant for the inhibitory potency of dinaciclib and therefore have no influence on its binding into the ATP-binding site, we started a new attempt using CDK12 (H1040G) mutant. The idea here is that mutation of histidine to glycine may increase the accessibility of the ATP-binding site for compound binding, which may reduce protein flexibility by stabilizing the ATP-binding pocket and thus generate a more stable protein essential to initiate protein crystallization. Besides the stability, the purity of the protein sample is a further essential factor for the ability of protein crystallization. The kinase complex of human CDK12 (H1040G) (714-1063) with Cyclin K (1-267) expressed in *Sf9* insect cells can be purified to

homogeneity without significant protein contamination or degradation (Figure 27A and B). Therefore, to improve protein crystallizability, CDK12 (H1040G) mutant was used for co-crystallization approach with dinaciclib. Fractions only containing kinase and Cyclin in 1:1 ratio were collected and subjected to crystallization trials after shock freezing. Initial hits were obtained after treatment with an excess of dinaciclib (Figure 27C). Crystals were optimized for a time of over 1.5 years. However, diffraction of the crystals measured at the SLS synchrotron could only be improved to 5 Å. Thus, an adequate structure determination of CDK12 in complex with dinaciclib was not achieved by the implemented measures.

4.1.2 Characterization of non-covalent and covalent CDK12 inhibitors under different conditions

CDK12 has been shown to play a manifold role in human health and diseases (Lui *et al.*, 2018). Thus, there is a continuous great interest in the design and development of synthetic small molecule inhibitors targeting this kinase. Small molecule inhibitors are typically divided into two classes: covalent and reversible binders. The binding of inhibitors that form a covalent attachment to the active site of their target protein is irreversible, e.g. by forming a covalent bound to a cysteine residue. In contrast, the binding of conceptually distinct non-covalent inhibitors is reversible, relying on non-covalent interactions such as hydrogen bonds, electrostatic interactions, hydrophobic interactions and/or Van der Waal interactions.

To characterize the mode of action of covalent and non-covalent CDK12 inhibitors, *in vitro* kinase activity assays of dose-dependent inhibition measurement were performed using various ATP concentrations as well as varying pre-incubation times (Figure 28). ATP concentration of 0.2 mM and 1 mM was used, while the latter mimics the cellular ATP concentration which is normally maintained in the range of 1 to 5 mM. Concurrently, the IC_{50} values were determined without and with 30 min pre-incubation time of inhibitor and kinase. The determined IC_{50} values are summarized in table 20. The data in this study are generated from unpublished CDK12 inhibitors, thus for illustration purposes only. In the following, the non-covalent reversible inhibitor is referred to as CMP-9001 and the covalent irreversible inhibitor is referred to as CMP-9088. Both inhibitors are sharing very similar basic chemical structural formula. However, in case of the non-covalent inhibitor CMP-9001, the electrophilic acrylamide present in the covalent inhibitor CMP-9088 is replaced with a propyl amide incapable of covalent reaction.

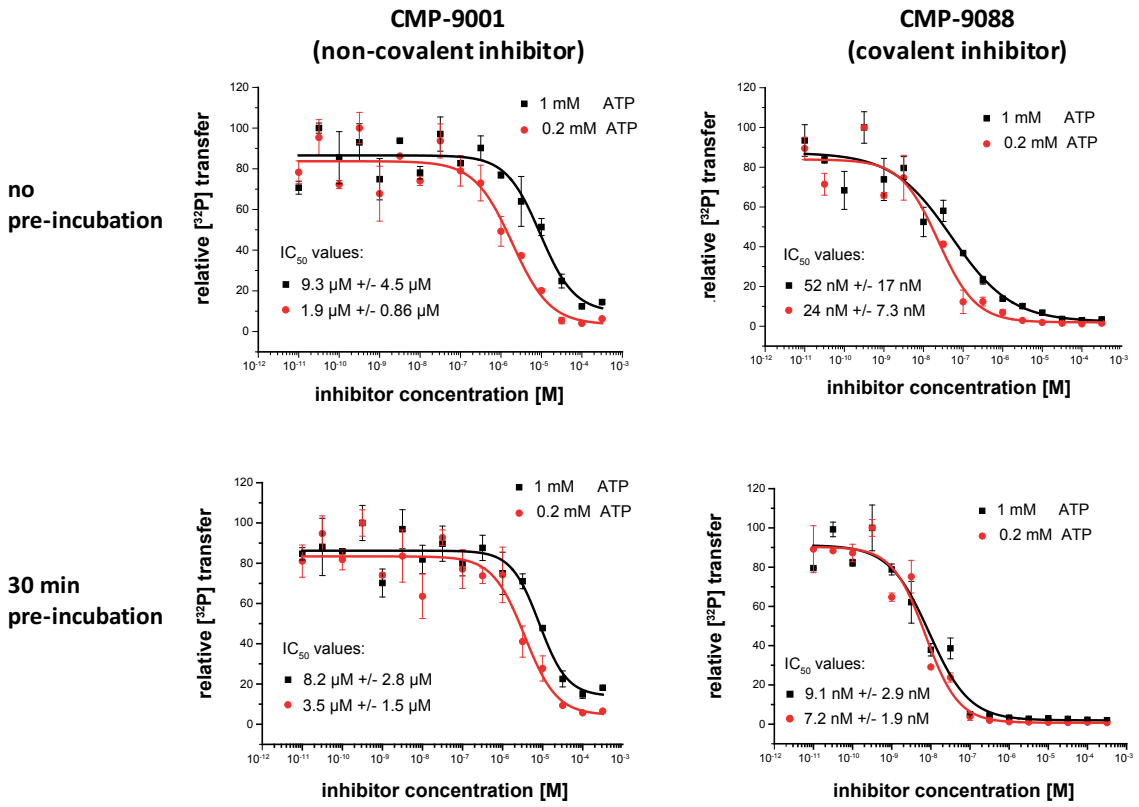


Figure 28: Dose-dependent inhibition measurements of non-covalent reversible and covalent irreversible CDK12 inhibitors under different conditions. Assays were performed using 0.2 mM and 1 mM ATP concentration as well as without and with 30 min pre-incubation of inhibitor and kinase. The counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are shown as mean ±SD; n=3.

Table 20: Overview of determined IC_{50} values for a non-covalent reversible and for a covalent irreversible CDK12 inhibitor under different conditions.

Non-covalent inhibitor		
pre-incubation	**0.2 mM ATP**	**1 mM ATP**
-	1.9 ± 0.86 µM	9.3 ± 4.5 µM
30 min	3.5 ± 1.5 µM	8.2 ± 2.8 µM
Covalent inhibitor		
pre-incubation	**0.2 mM ATP**	**1 mM ATP**
-	24 ± 7.3 nM	52 ± 17 nM
30 min	7.2 ± 1.9 nM	9.1 ± 2.9 nM

Overall, the covalent inhibitor has been shown to be much more potent against CDK12 than the non-covalent inhibitor, with IC_{50} values in nanomolar range towards IC_{50} in micromolar range. Without pre-incubation time, IC_{50} values of both non-covalent and covalent inhibitors were depending on the ATP concentration. The lower the ATP concentration, the higher the inhibitory potency of the inhibitor. Compared to the IC_{50} values measured at 1 mM ATP concentration, IC_{50} values of non-covalent and covalent inhibitor were reduced 5- and 2-fold at 0.2 mM ATP concentration, whereby ATP reliance appeared to be more pronounced in case of non-covalent inhibitor.
When considering the data obtained after 30 min pre-incubation time, IC_{50} values of non-covalent inhibitor hardly changed either at 0.2 mM nor at 1 mM ATP concentration, similar to the data received from measurement without pre-incubation. Thus, pre-incubation time had no significant impact on the activity of reversible binding inhibitors. In contrast, pre-incubation of covalent inhibitor with kinase caused reduction of the IC_{50} value at both ATP concentrations by approximately 3- to 6-fold. Interestingly, without pre-incubation, the IC_{50} value of covalent inhibitor at 1 mM ATP concentration was 3-fold higher than the IC_{50} value obtained at 0.2 mM ATP concentration. This gap, however, has almost disappeared after pre-incubation. These results indicated that irreversible inhibitors become more potent over time, depending on the length of the pre-incubation time. In addition, the significance of ATP concentration has been displaced by preceding pre-incubation procedure, making IC_{50} values independent of the ATP concentration.
Together, these results suggest that pre-incubation of inhibitor with kinase has an enormous impact on the the inhibitory potency of covalent irreversible inhibitor and thus plays a crucial role in the determination of IC_{50}. In contrast, inhibitory potency of a non-covalent reversible inhibitor will be only affected by the ATP concentration in this context.

4.1.3 Structural elucidation of CDK12/Cyclin K co-crystallized with inhibitor

In collaboration with the laboratory of Nathanael Gray from Boston, USA, who is mainly interested in the development of small molecule inhibitors against the transcriptional CDK kinases, we co-crystallized their current lead compound (CMP-175) in the field of covalent CDK12/13 targeting with our recombinant protein complex CDK12 (714-1063)/CycK (1-267). CMP-175 is supposed to be a selective ATP-competitive inhibitor of CDK12 and CDK13 by targeting the cysteine of the DCHEL motif proximal to the ATP-binding site of the kinase domain (C1039 for CDK12) and (C1017 for CDK13). As no other transcriptional CDK contains this unique cysteine at a similar position lying outside the conserved kinase domain (T. Zhang *et al.*, 2016), CMP-175 is presumed to target CDK12 and CDK13 selectively. In order to understand how this small molecule inhibitor

selectively targets CDK12 and CDK13, we determined the co-crystal structure of CDK12/CycK with CMP-175.

First, we performed an *in vitro* kinase activity assay using radioactive labelled [^{32}P]-γ-ATP in order to confirm the inhibitory activity of CMP-175 against both transcriptional kinases CDK12 and CDK13. Here, the ability of 0.2 µM purified recombinant CDK12 and CDK13 to phosphorylate 50 µM c-Myc substrate protein in the presence of 0.2 mM ATP was determined in an inhibitor dose-dependent manner. The inhibitor was pre-incubated with the respective kinase for 30 min. As demonstrated in figure 29, CMP-175 potently inhibited CDK12 and CDK13 after a 30 min measurement period with IC_{50}-values of 28 nM and 75 nM, respectively.

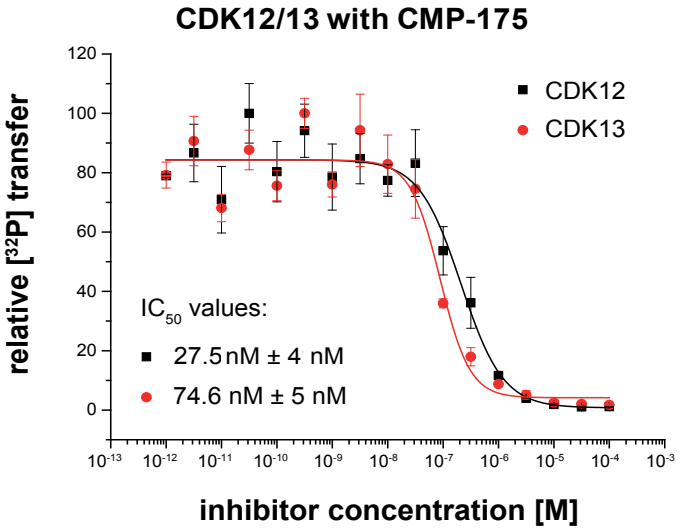


Figure 29: Dose-dependent inhibition measurements of CMP-175 with CDK12/13. 0.2 µM of purified recombinant kinase complexes of CDK12/CycK and CDK13/CycK were used in the presence of 0.2 mM ATP. The inhibitor was pre-incubated with the respective kinase for 30 min. The counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are shown as mean ±SD; n=3.

After confirming that CMP-175 potently inhibits the kinase activity of CDK12 and CDK13, the next step was to generate a recombinant protein complex of CDK12/CycK. The protein sample has to be purified to homogeneity, or as close as possible, as in biological samples the homogeneity of the protein preparation is a key factor in obtaining crystals that diffract to high resolution (Bergfors, 1999; McPherson, 1999). As shown in figure 30, the protein complex of CDK12/CycK was able to be purified to homogeneity. Only fractions containing the homogenous protein complex were pooled (fractions 43-46). Before applying to crystallization trials, the purified recombinant CDK12/CycK complex at 29 µM concentration was incubated overnight at 4°C with 43.5 µM CMP-175 in a ratio of 1:1.5. Excess free compound was removed by size-exclusion chromatography using a disposable

PD10 desalting column (Merck, Darmstadt, Germany). The complex of CDK12-CMP-175/CycK was finally concentrated up to 18 mg/mL, buffered in 50 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP and subsequently applied to crystallization trials. Crystals were typically grown in hanging drops. At the beginning, they were incubated at 10°C and then moved to 15°C after a week. Within a period of 10-12 days the crystals were of optimal size for flash freezing with 15% ethylene glycol in mother liquor as cryoprotectant. Examples of obtained protein crystals are shown in figure 31.

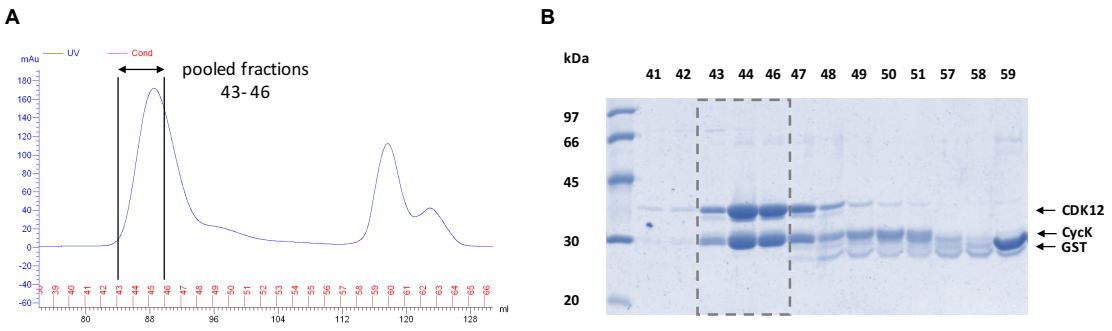


Figure 30: Purification of recombinant CDK12/CycK complex for co-crystallization of CDK12/CycK with CMP-175. (A) Gel filtration chromatogram of the recombinant CDK12/CycK complex. **(B)** SDS-PAGE of purified CDK12/CycK complex after gel filtration. Fraction samples showing an UV absorbance at 280 nm were loaded for visualization of the protein purity and ratio. Fractions 43-46 were pooled for crystallization trials.

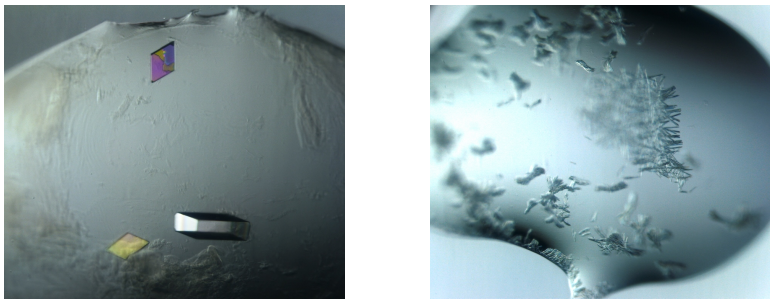


Figure 31: Examples of obtained protein crystals of CDK12/CycK co-crystallized with CMP-175.

Totally, 13 crystals were measured at the SLS synchrotron, Villigen, Switzerland, at the PX II beam line, which is not as strong or focused as the other beam line in order to get an overall good electron density map. The initial diffraction was up to 2.2 Å, however, the crystals melt away upon recording a full data set. Thus, we started with 2.2 Å and end after 5 min at 4.5 to 6 Å. In the end, six full data sets were recorded with the structure was varying between 2.8 and 3.5 Å. After processing the first

two data sets by molecular replacement followed by refinement procedures, the structure of the protein complex of CDK12-CMP-175/CycK was finally refined to a resolution of 2.6 Å.
As shown in figure 32, CMP-175 fits very well into the generated electron density map, indicating the linkage of the compound's flexible linker to C1039 in the C-terminal ending of CDK12.

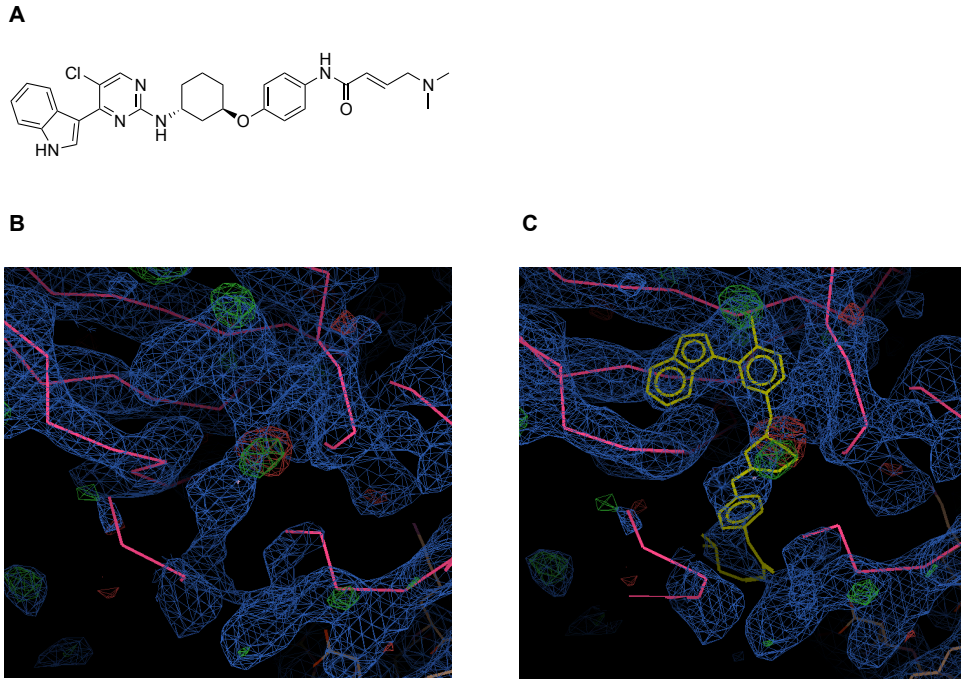
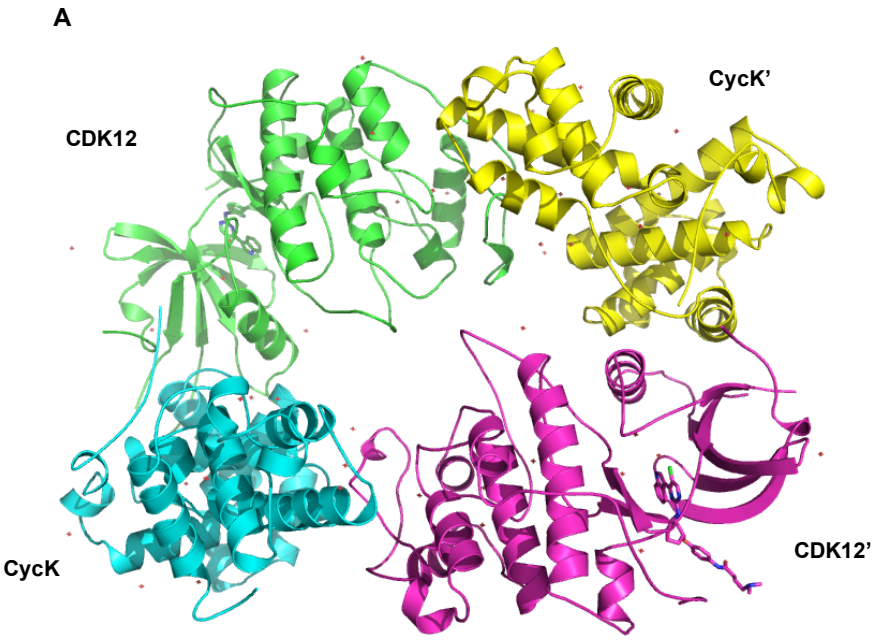
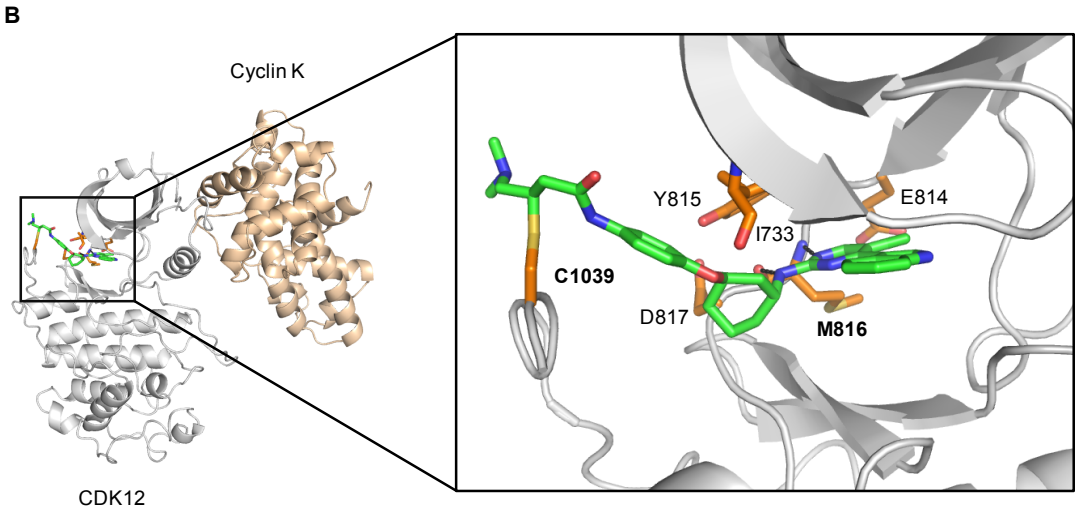


Figure 32: Electron density maps showing incorporation of CMP-175 into CDK12 via covalent targeting. **(A)** Structural formula of CMP-175. **(B), (C)** Magnified view of the electron density of CMP-175 without and with line presentation of the compound.

Overall, two CDK12/CycK complexes were found in the asymmetric unit, each covalently bind to CMP-175 in the same manner (Figure 33A). As illustrated in the magnified view in figure 33B, the C-terminal linker of the kinase is fully free and not attached to the kinase core domain. The flexible linker allows C1039 to orient towards the kinase active site for cross-linking to the compound via a sulfur-linkage. Besides the linkage to C1039 via its flexible linker, CMP-175 further interacts with the kinase hinge region by targeting M816 via formation of two hydrogen bonds between the aminopyrimidine of CMP-175 and the backbone of M816. Interestingly, in contrast to THZ531, CMP-175 only associates with the N-terminal lobe of the kinase. THZ531 instead, a well known covalent inhibitor of CDK12/13 with highly similar structural formula to CMP-175, targets the kinase active site in two conformations by associating both with the N- and C-terminal lobe of the kinase (T. Zhang *et al.*, 2016), Protein surface display of CDK12 with CMP-175 binding to the kinase active site is shown in figure 33C.


A
CDK12
CycK'
CycK
CDK12'


B
Cyclin K
CDK12
Y815
E814
I733
C1039
D817
M816

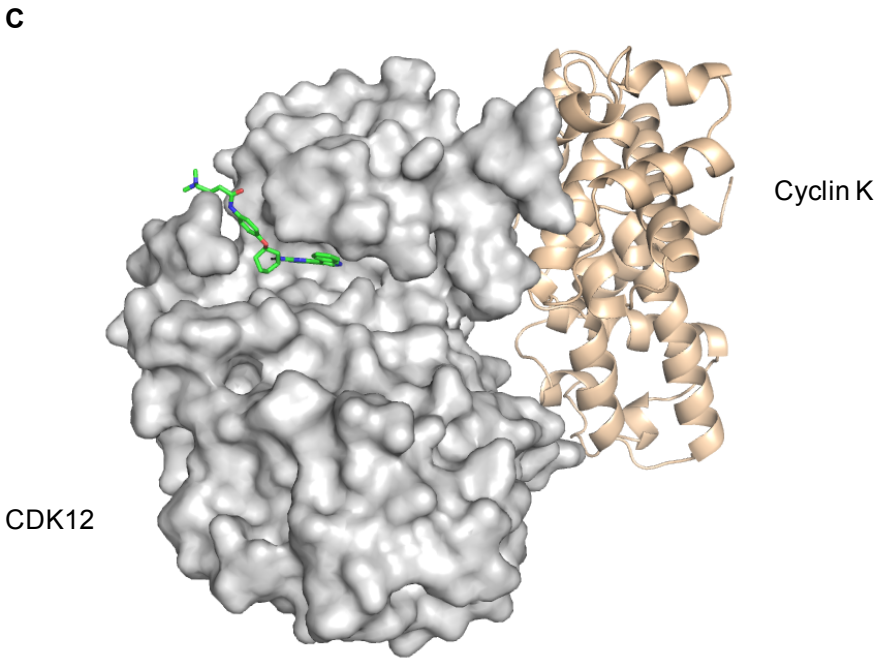


Figure 33: CMP-175 co-crystal structure with CDK12/CycK. (A) Asymmetric unit showing two CDK12/CycK complexes, each binding to CMP-175 in the same manner. Each kinase and cyclin is displayed in different color. **(B)** CMP-175 binds to M816 in the kinase hinge region through polar interactions and connects to C1039 via the compound's flexible linker. CMP-175 is displayed in green with surrounding amino acids illustrated in orange. **(C)** Protein surface display of CDK12 with CMP-175 binding to the kinase active site.

4.2 Validation of pre-mRNA processing factors as new CDK12 and CDK13 substrates

In close collaboration with the laboratory of Rani George from Boston, USA, we have recently published that inhibition of CDK12 and CDK13 by THZ531 leads to transcriptional elongation defect through premature cleavage and polyadenylation (PCPA) as a consequence of both decreased transcriptional elongation and lower activation levels of 3' end processing factors (Krajewska *et al.*, 2019). Given that the transcriptional activity of RNA polymerase II (RNAPII) and processing of pre-mRNA occur simultaneously (Bentley, 2014), we assumed that CDK12 and/or CDK13 kinase may regulate the phosphorylation of targets other than the RNAPII CTD, thereby contributing to co-transcriptional RNA processing additional to regulating transcription elongation. Using *in vivo* phosphoproteomic analysis of cells treated with and without THZ531, our collaboration partners indicated that CDK12/13 phosphorylates several pre-mRNA processing and splicing factors, including those regulating PCPA. Through stable isotope labeling with amino acids in cell culture (SILAC) analysis, they identified following top candidate targets that were affected by CDK12/13 inhibition: CDC5L, complex partner of the Prp19 protein, is implicated in RNA splicing and genomic stability (Grote *et al.*, 2010; Mu *et al.*, 2017); SF3B1, a component of the splicing machinery that is involved in pre-mRNA splicing (Wahl *et al.*, 2009); CSTF64, one of the multiple factors required for polyadenylation and 3'-end cleavage of mammalian pre-mRNAs (Takagaki and Manley, 1997); and SPT6H, which controls the transcription elongation rate and release of paused RNAPII into productive elongation (Vos *et al.*, 8/2018b). To validate the above mentioned phosphorylation target proteins of CDK12/13 *in vitro* as well as to distinguish between the role of CDK12 and CDK13 in this context, we performed *in vitro* kinase activity assays of time course measurement series with both kinases using ^{32}P-labeled ATP.

First, in order to archive stable and soluble substrate proteins *in vitro* with high-yield expression and low contamination, truncated protein domains were generated to meet these requirements. Based on the phosphorylation sites identified in the phosphoproteomic analysis as well as performing protein domain analysis using the ‘simple modular architecture research tool’ (SMART), constructs of truncated substrate proteins were designed, containing functional characteristic domains (Figure 34A). CDC5L (370-505), SF3B1 (113-462), CSTF64 (509-577), SPT6H (1323-1534) and SPT6H (1434-1544) were expressed as GST-fusion proteins in *E. coli* and purified to homogeneity (Figure 34B). In case of SPT6H, only the protein construct encompassing the full tandem SH2 domain with the C-terminally flanking residues (1323-1534) could be successfully expressed. As indicated by the SDS-PAGE gel, only the protein sample of SF3B1 contains significant impurities, even after purification by gel filtration. The GST-tag of CDC5L was cleaved off by TEV digestion, otherwise GST was needed to keep the proteins stable.

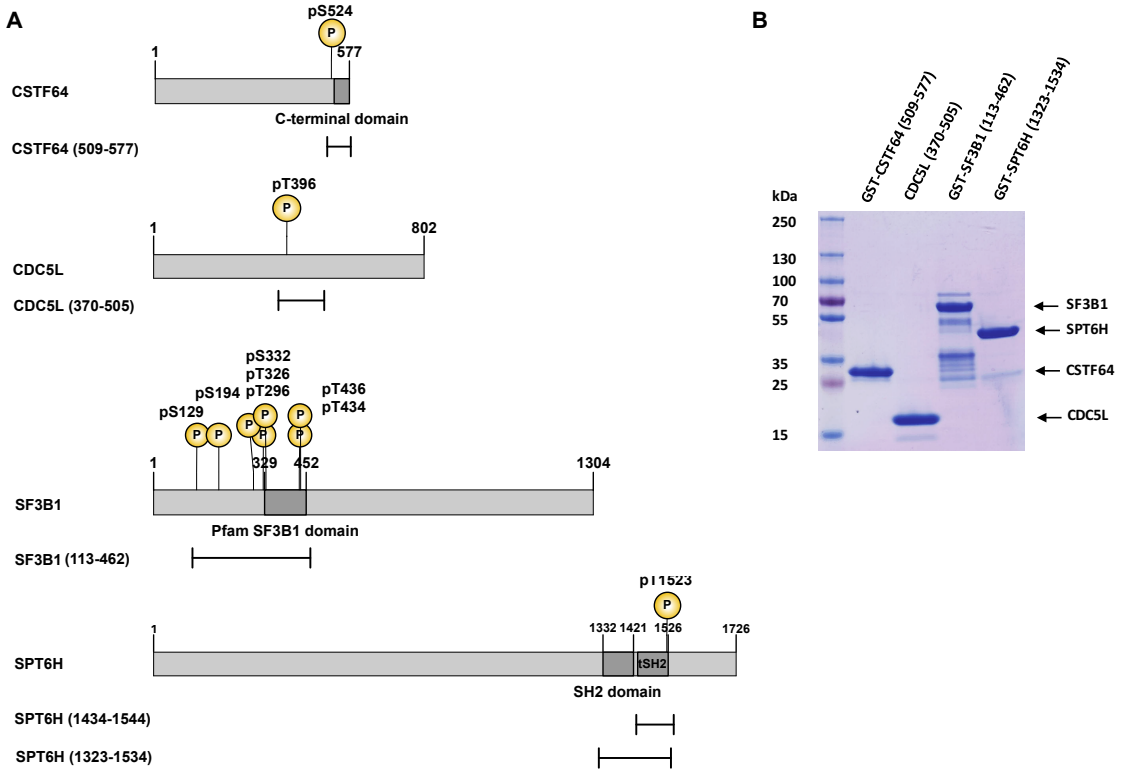


Figure 34: Overview of generated constructs of CDK12/13 top candidate substrates provided by phosphoproteomic analysis. (A) CSTF64 (509-577), CDC5L (370-505), SF3B1 (113-462), SPT6H (1434-1544) and SPT6H (1323-1534). Constructs of truncated CDK12/13 substrate proteins illustrating characteristic domains and phosphorylation sites identified in the SILAC analysis (Krajewska *et al.*, 2019). **(B)** SDS-PAGE of purified recombinant CDK12/13 substrate proteins.

In vitro kinase time course measurements were performed using 0.2 µM purified recombinant CDK12/CycK or CDK13/CycK with 50 µM each of purified recombinant candidate substrate and 1 mM ATP, over a time period of 6 h (Figure 35). The kinase activity was measured at seven time points: 0, 15, 30, 60, 120, 240 and 360 min. Additional, negative control measurements without the addition of either CDK12 or CDK13 kinase were performed, respectively. As shown in figure 35, both CDK12 and CDK13 phosphorylate the three substrates CDC5L, SF3B1 and CSTF64 in a time dependent manner. The measured radioactive counts in the control experiments of CDC5L, SF3B1 and CSTF64 were significantly below the values measured in the presence of active kinases. Of note, phosphorylation of the candidate substrates by CDK12 resulted in a higher rate of [^{32}P] incorporation for CDC5L, SF3B1, and CSTF64. The time course measurement with SPT6H as candidate substrate, however, showed no increase in radioactive counts neither by CDK12 nor CDK13 compared to the control measurement without addition of kinase.

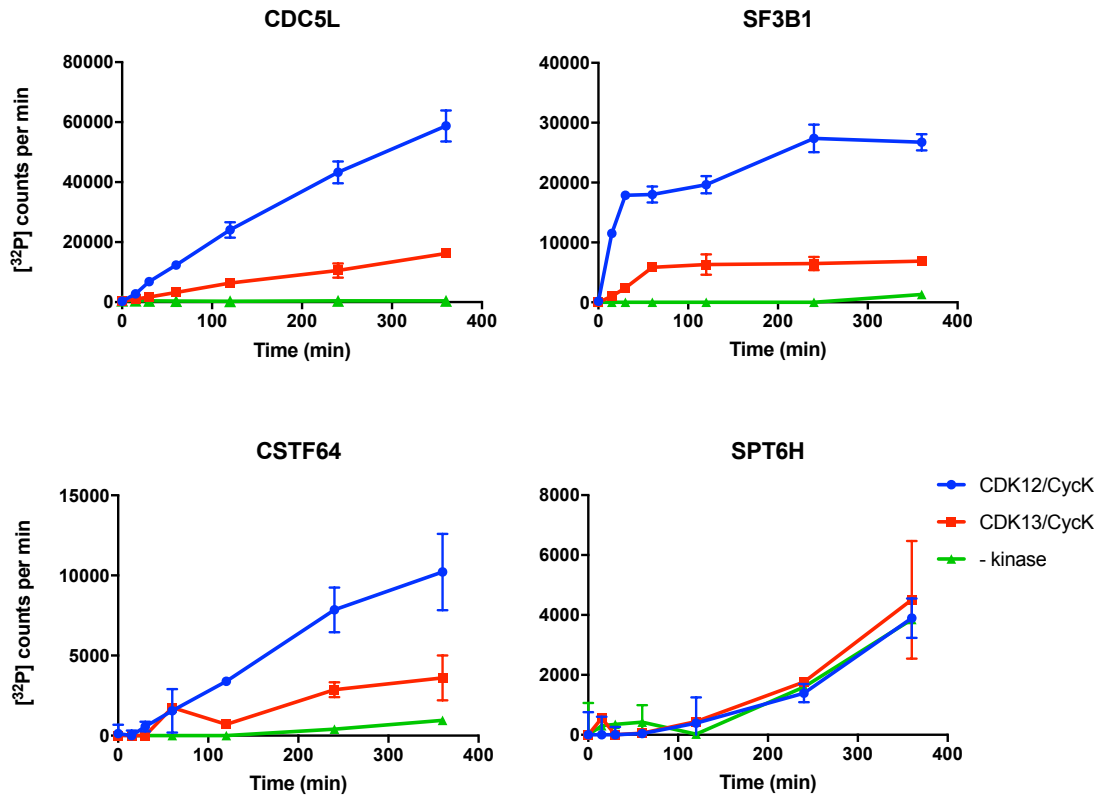


Figure 35: *In vitro* kinase time course measurements of CDK12/CycK- and CDK13/CycK-mediated phosphorylation of the indicated candidate substrates. Time course measurements over 6h were performed with CDC5L (370-505), GST-SF3B1 (113-462), GST-CSTF64 (509-577), and GST-SPT6H (1323-1534). Negative controls without kinase is shown in green. Data are shown as mean ±SD; n=3.

To further investigate the potential of SPT6H to be a target protein of transcription kinases, the time course measurement of SPT6H was repeated under the same condition with an other well-described RNAPII Ser2 elongation kinase, CDK9/CycT1. Similarly to CDK12 and CDK13, CDK9 is associated with transcription elongation and is able to phosphorylate the RNAPII CTD at Ser5 of the heptad repeat (Czudnochowski *et al.*, 2012; Marshall *et al.*, 1996). Interestingly, in contrast to CDK12/13, CDK9 significantly phosphorylates SPT6H in a time dependent manner (Figure 36), indicating that the recombinant truncated SPT6H protein is a suitable protein for this kinase.

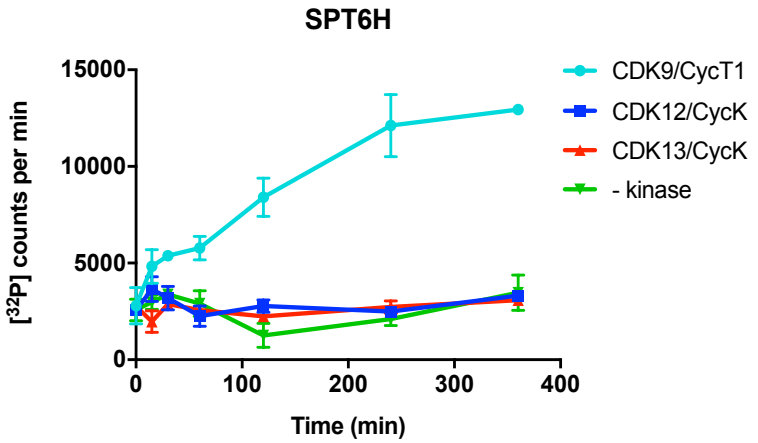


Figure 36: *In vitro* kinase time course measurements of transcription elongation kinases-mediated phosphorylation of GST-SPT6H (1323-1534). Compared to CDK12 and CDK13, CDK9 significantly phosphorylates SPT6H in a time dependent manner. Negative control without kinase is shown in blue. Data are shown as mean ±SD; n=3.

Next, dose-dependent inhibition measurements of CDK12/CycK and CDK13/CycK were performed with CDC5L, SF3B1 and CSTF64 as substrate proteins. Here, THZ531 was used as kinase inhibitor, which is a well-described covalent kinase inhibitor that binds to an unique cysteine in the characteristic C-terminal extension helix outside the canonical kinase domains of CDK12 (Cys1039) and CDK13 (Cys1017), resulting in their prolonged and irreversible inactivation (T. Zhang *et al.*, 2016). The determined IC_{50} values, respectively, indicted clearly the strong inhibition of both CDK12 and CDK13 by THZ531. After the pre-incubation time of 30 min for both CDK/Cyclin complex with THZ531, phosphorylation of all three substrate proteins was reduced with increasing concentrations of the inhibitor in a dose-dependent manner (Figure 37). Notably, the IC_{50} values of THZ531 obtained from the *in vitro* assay with CDK12 and recombinant protein substrates [SF3B1 (85 nM ± 18 nM) and CSTF64 (91 nM ± 32 nM)] were comparable to those obtained from the assay with CDK13 [SF3B1 (62 nM ± 19 nM) and CSTF64 (106 nM ± 2 µM)]. However, it should be noted that the illustrated inhibition curves of CDK12/13 using CSTF64 as substrate protein is less convincing, particularly when using CDK13 as kinase. Nevertheless, the results received from the dose-dependent inhibition measurements of CDK12/CycK and CDK13/CycK by THZ531 clearly confirmed CDC5L, SF3B1, and to some extent also CSTF64, as phosphorylation targets of CDK12/13.

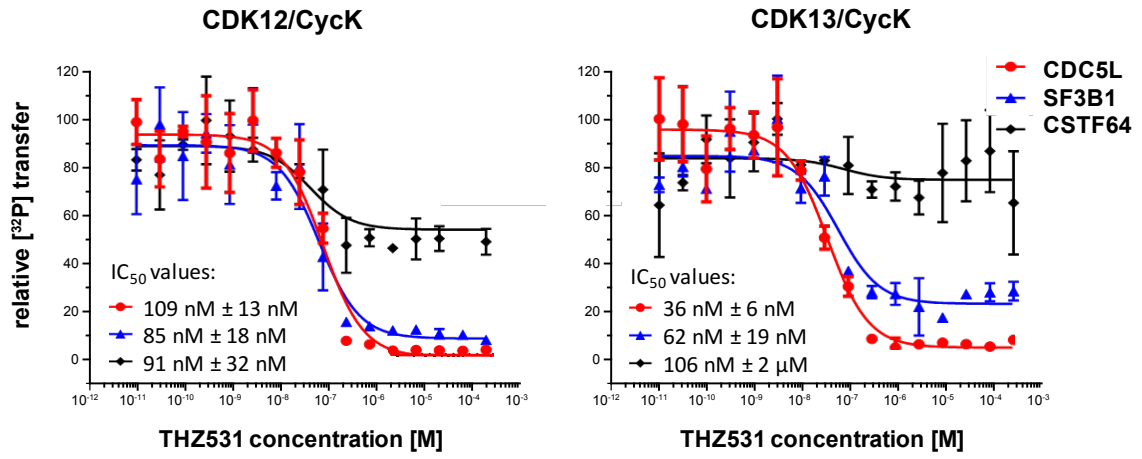


Figure 37: Dose-dependent inhibition measurements of CDK12/CycK and CDK13/CycK by THZ531 with different potential CDK12/13 substrates: CDC5L (370-505), GST-SF3B1 (113-462) and GST-CSTF64 (509-577). Radioactive kinase reactions were performed after 30 min pre-incubation time with increasing concentrations of THZ531. The counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are reported as mean ± SD; n = 3.

To identify the sites phosphorylated by CDK12/13 in the *in vitro* kinase assays as well as to validate the phosphorylation sites identified in the phosphoproteomic analysis *in vivo*, we performed peptide mass fingerprint analyses for the tested protein substrates after incubation with kinase. The *in vitro* phosphorylation reaction was conducted using 0.2 µM purified recombinant CDK12/CycK with 50 µM each of purified recombinant candidate substrate and 1 mM ATP over an incubation time of 4 h. Reaction mixture of each substrate was divided into two samples: A and B. Protein sample A was incubated without kinase, used as control, while protein sample B was incubated with CDK12. Peptide mass fingerprint analysis of CDC5L and SF3B1 indicated multiple serine and threonine phosphorylation sites. In case of CDC5L, nine phosphorylation sites were detected, including the site, pT396, which was also identified in the phosphoproteomic analysis in the cells (Figure 38A). For SF3B1, besides other detected phosphorylation sites, three of the seven sites identified in the phosphoproteomic analysis, S129, T326 and S332, could be confirmed (Figure 38B). Interestingly, by performing peptide mass fingerprint analysis of the SPT6H construct, which has not been shown to be phosphorylated by CDK12/13 in the previous kinase assay, a single phosphorylation site at T1523 has been confirmed (Figure 38C). T1523 was also the SPT6H phosphorylation site identified in the phosphoproteomic analysis. Unfortunately, phosphorylation sites of the CSTF64 construct could not be identified by performing peptide mass fingerprint analysis. A comparative overview of the phosphorylation sites identified in the phosphoproteomic and peptide mass finger print analysis is shown in table 21.

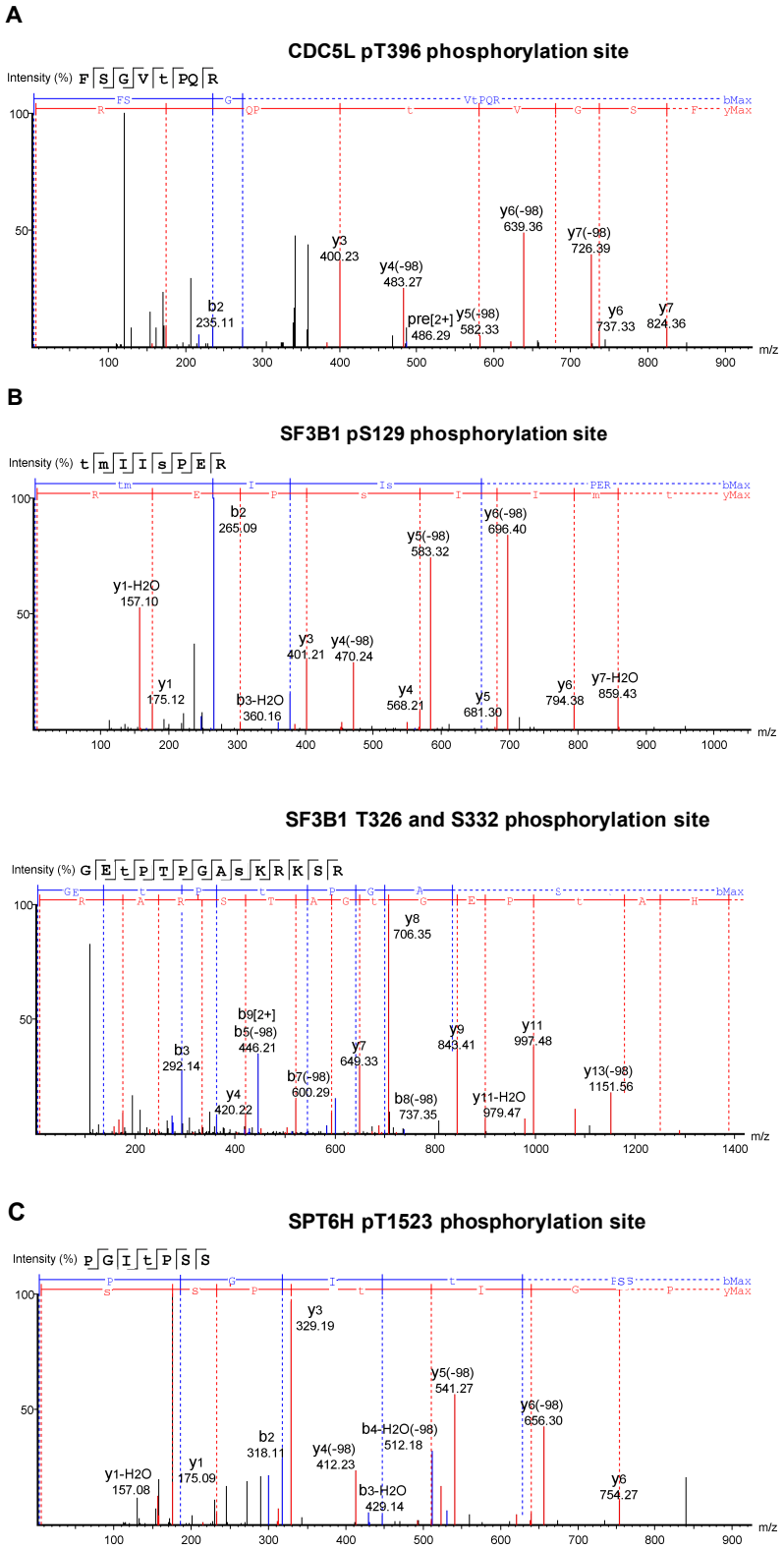


Figure 38: Mass spectrometry spectra of CDC5L (370-505), SF3B1 (113-462) and SPT6H (1323-1534) after incubation with CDK12/CycK. The spectra were generated using Peaks software (Bioinformatics Solutions Inc.) Illustration of the identified phosphorylation sites of **(A)** CDC5L (pT396), **(B)** SF3B1 (pS129, pT326 and pS332) and **(C)** SPT6H (pT1523).

Table 21: Comparative overview of identified phosphorylation sites of CDK12/CDK13 candidate substrates by performing phosphoproteomic analyses of neuroblastoma (NB) cells treated with THZ531 and peptide mass finger print analyses of recombinant proteins *in vitro*. Phosphorylation sites identified in both methods are highlighted in red. For CDC5L, SF3B1 and SPT6H phosphorylation sites could be detected by performing peptide mass finger print analysis, but this was not achieved for CSTF64.

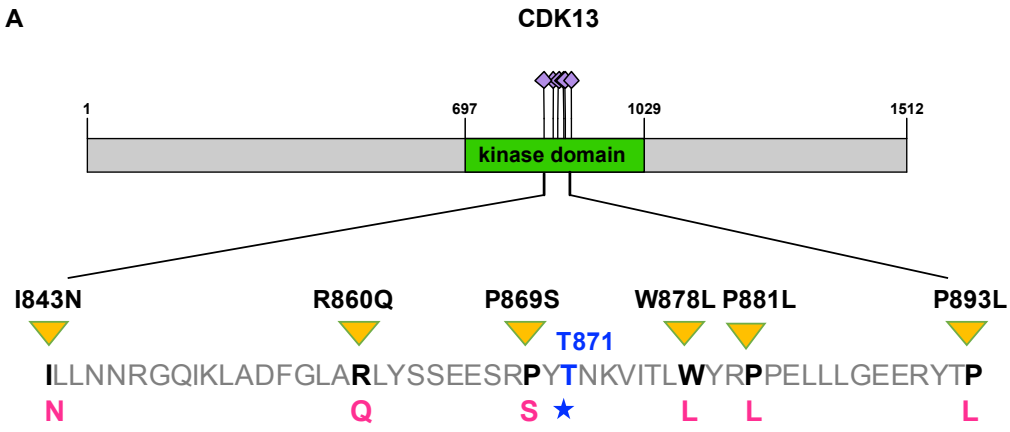
CDK12/13 candidate substrates	Phosphoproteomic analysis	Peptide mass fingerprint analysis
CDC5L	**pT396**	pT385, pS390, pS393, **pT396**, pT404, pS410, pT411, pT438, pT442
SF3B1	**pS129**, pS194, pT296, **pT326**, **pS332**, pT434, pT436	**pS129**, pT244, **pT326**, pT328, **pS332**, pT426
SPT6H	**pT1523**	**pT1523**
CSTF64	pS524	—

4.3 Characterization of CDK13 in human melanoma

4.3.1 Structural and biochemical analysis of melanoma-associated CDK13 mutations

In the current study by Megan Insco and co-workers, CDK13 has been shown to be mutated in 4% of human melanomas, which is associated with poor overall survival (Insco *et al.*, submitted). Patient melanoma mutations located in the CDK13 kinase domain near the activating phosphorylation site T871 in the T-loop of the protein were supposed to affect the kinase activity. Using a zebrafish model system, Megan Insco and co-workers first analyzed CDK13 mutations in cancer, occurring especially in human melanoma.

In close collaboration with Megan Insco, we investigated the potential impact of theses disease-related CDK13 mutations on the kinase activity. First, we plotted each of these single point mutations onto our CDK13 wild type crystal structure (PDB: 5EFQ; (Greifenberg *et al.*, 2016), displaying them as they might appear in the structure by replacing the natural AS residues. Overlay of the modelled structure with the wild type structure as well as the structure surface display indicated that six of these mutations (I843N, R860Q, P869S, W878L, P881L and P893L) may have an impact on the kinase activity (Figure 39).



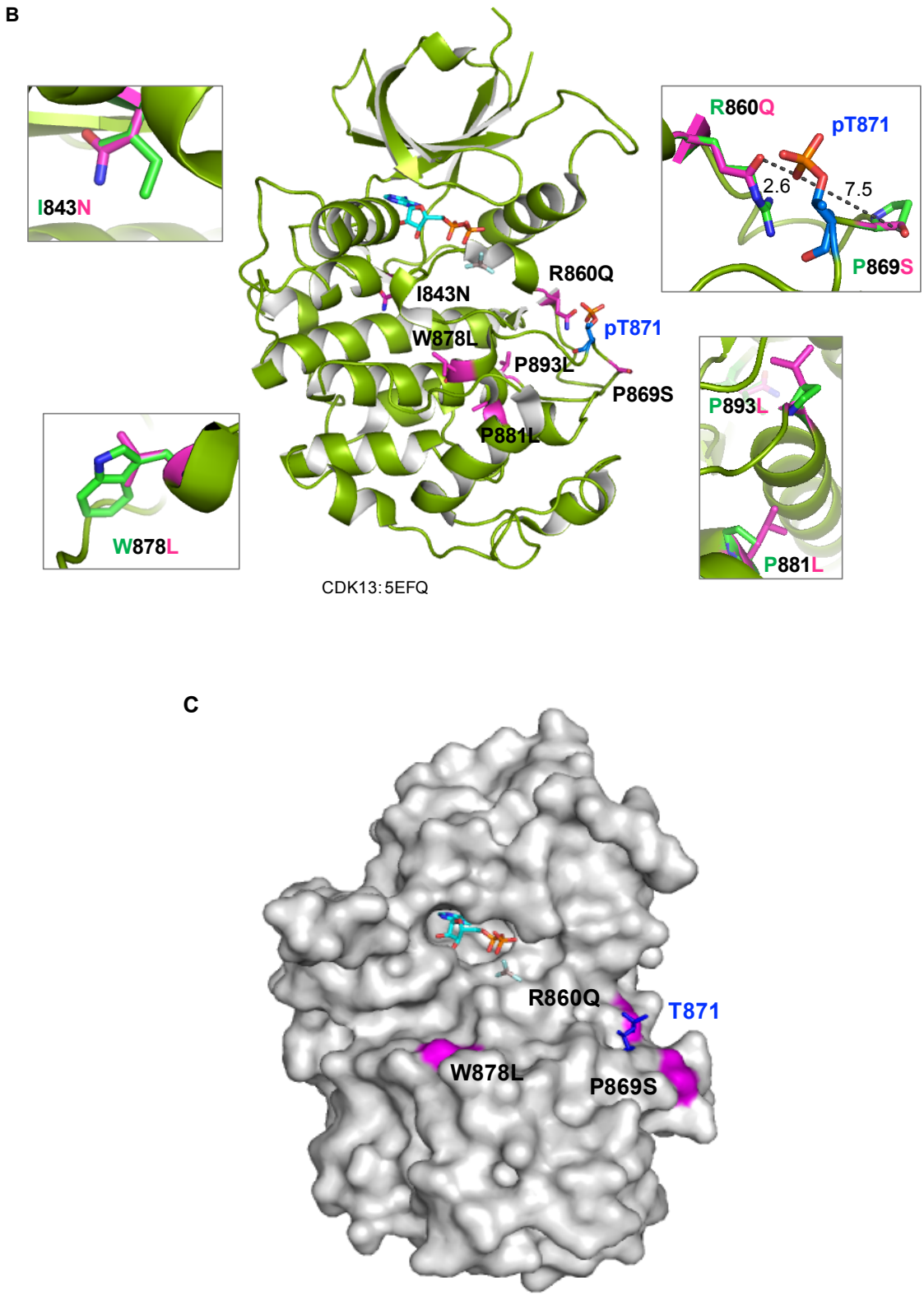


Figure 39: Representation of melanoma-associated CDK13 kinase domain mutations predicted to affect kinase activity. (A) Mapping of CDK13 melanoma mutations in the amino acid sequence of the CDK13 kinase domain. **(B)** Mapping of CDK13 melanoma mutations on CDK13 crystal structure (PDB: 5EFQ; (Greifenberg *et al.*, 2016). Magnified view of each mutation is illustrated as overlay of modelled structure (pink) with wild type structure (green). **(C)** Surface display of CDK13 melanoma mutations.

Both R860Q and P869S mutants are located in close proximity to the activating phosphorylation site of the kinase T-loop as well as fully on the protein surface. It is obviously that they may have an impact on the kinase activity as they may disrupt the T-loop activation by preventing phosphorylation of T871. The electro-negatively charged carbonyl group of the glutamine residue of the R860Q mutant, located in short distance to the phosphate group of pT871 (2.6 Å), may disrupt the T-loop phosphorylation by inducing electrostatic repulsion. Similarly, in case of the P869S mutant, the kinked, hydrophobic amino acid is replaced by linear, polar amino acid. The negatively charged hydroxyl group of the serine residue may affect the T871 phosphorylation by inducing electrostatic repulsion and may influence the conformation of the T-loop as well. As shown in the structure surface display, the W878L mutant is fully located on the surface of the substrate binding groove of the kinase. Replacement of the characteristic aromatic ring of tryptophan by the less representative leucine residue may induce significant conformational changes of the substrate binding pocket, thus affecting the recognition of substrates. The last three mutants I843N, P881L and P893L are buried inside the protein core. They seem to disrupt the kinase domain structure and may affect the overall protein stability through replacing planar aromatic rings by flexible residues as in case of P881L and P893L. Overview of the predicted implications of the described CDK13 kinase domain mutations R860Q, P869S, W878L, I843N, P881L and P893L is shown in table below.

Table 22: Overview of the predicted implications of six melanoma-associated CDK13 kinase domain mutations.

CDK13 kinase domain mutations	Predicted implication
R860Q, P869S	disrupting T-loop activation
W878L	changing substrate binding pocket
I843N, P881L, P893L	disrupting kinase domain structure

To analyze the implications of these mutations, *in vitro* kinase activity assays using radioactive labelled [^{32}P]-γ-ATP were performed. Here, kinase assays were conducted with the two CDK13 mutants R860Q and W878L, as representatives. As a control, we included the catalytically-dead CDK13 K734R mutant, which replaces the critical lysine residue required for the kinase catalytic activity by binding alpha and beta phosphates of ATP. The three recombinant kinase mutants of CDK13 in complex with Cyclin K were expressed in *Sf9* insect cells and purified to homogeneity (Figure 40).

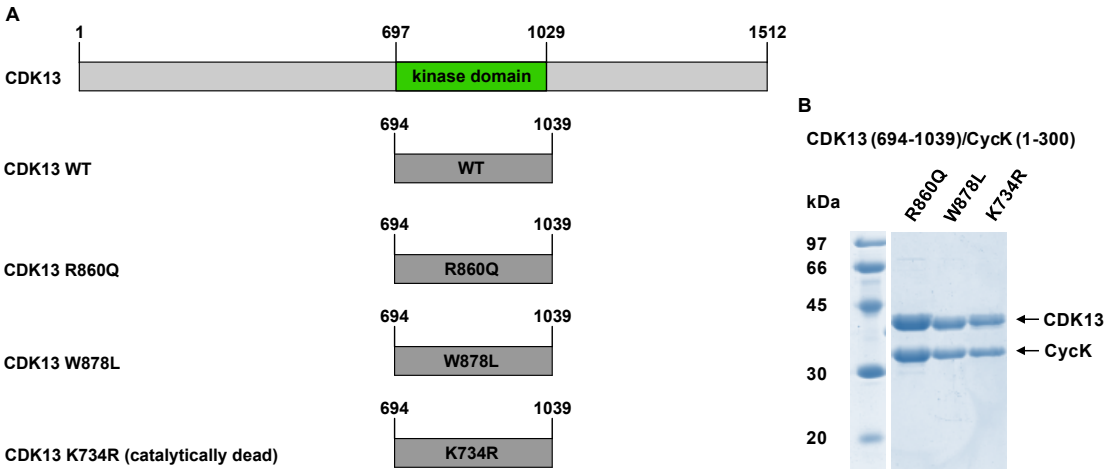


Figure 40: (A) Overview of the generated CDK13 constructs containing melanoma-associated single mutation R860Q and W878L. CDK13 wild type and K734R mutant were used as control. **(B)** SDS-PAGE of purified recombinant CDK13 mutant proteins in complex with Cyclin K.

The *in vitro* kinase assays were performed using 0.2 µM purified recombinant wild type CDK13 and CDK13 mutants with CTD_{52} (10 µM) and cMyc protein (50 µM) as substrates, and 1 mM ATP after 30 min incubation (Figure 41). While CDK13 wild type had robust kinase activity as displayed for both substrates, the R860Q, W878L, and K734R mutants failed to phosphorylate the full length RNAPII CTD as well as the cMyc protein. The obtained [^{32}P] counts per minute were in accordance with the outcome resulting from the control without kinase. Thus, this result confirmed the prediction derived from the structural analysis that CDK13 melanoma-associated mutations may cause inactivation of the kinase activity.

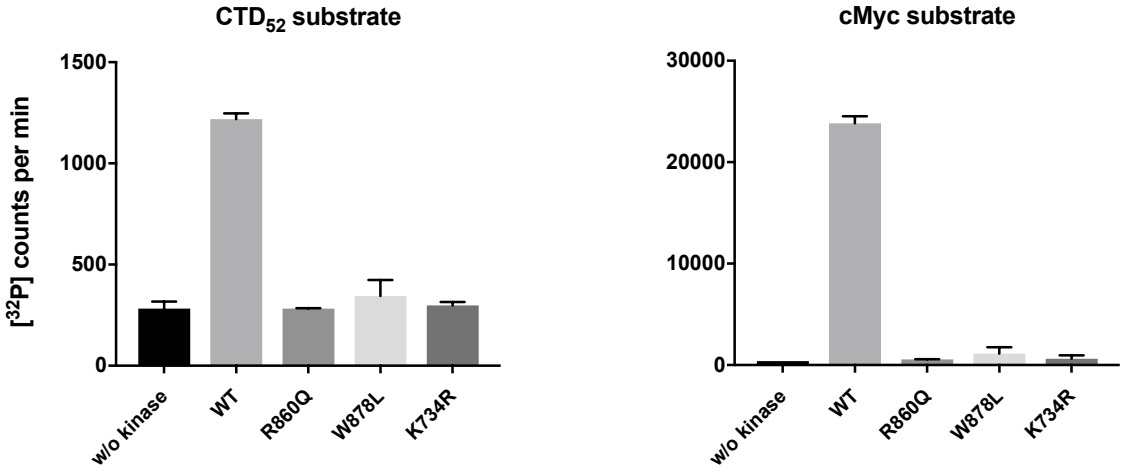


Figure 41: *In vitro* kinase assay of wild type and melanoma patient-mutated CDK13 (R860Q and W878L). Catalytically-dead CDK13 (K734R) mutant was used as negative control. CTD_{52} and cMyc protein were used as substrate. Data are shown as mean ±SD; n=2.

4.3.2 Validation of CDK13 binding protein identified in human melanoma

Based on the assumption that the CDK13 W878L mutant is capable of disrupting the kinase substrate binding, this melanoma patient mutation was chosen by Insco and co-workers to determine binding partners of CDK13 in human melanoma. Performing immunoprecipitation (IP) mass spectrometry experiments, they found that ZC3H14, a zinc finger protein implicated in negative regulation of polyA tail length (Pak *et al.*, 2011; Rha *et al.*, 2017), was the only identified interacting protein missing in the outcome of the CDK13 W878L mutant in comparison to the wild type kinase. This result suggested that ZC3H14 may be a substrate of CDK13 *in vivo*. To confirm this assumption, recombinant full length and truncated ZC3H14 proteins were generated in order to perform *in vitro* kinase activity assay to examine whether CDK13 wild type in complex with Cyclin K is able to phosphorylate ZC3H14 *in vitro.*

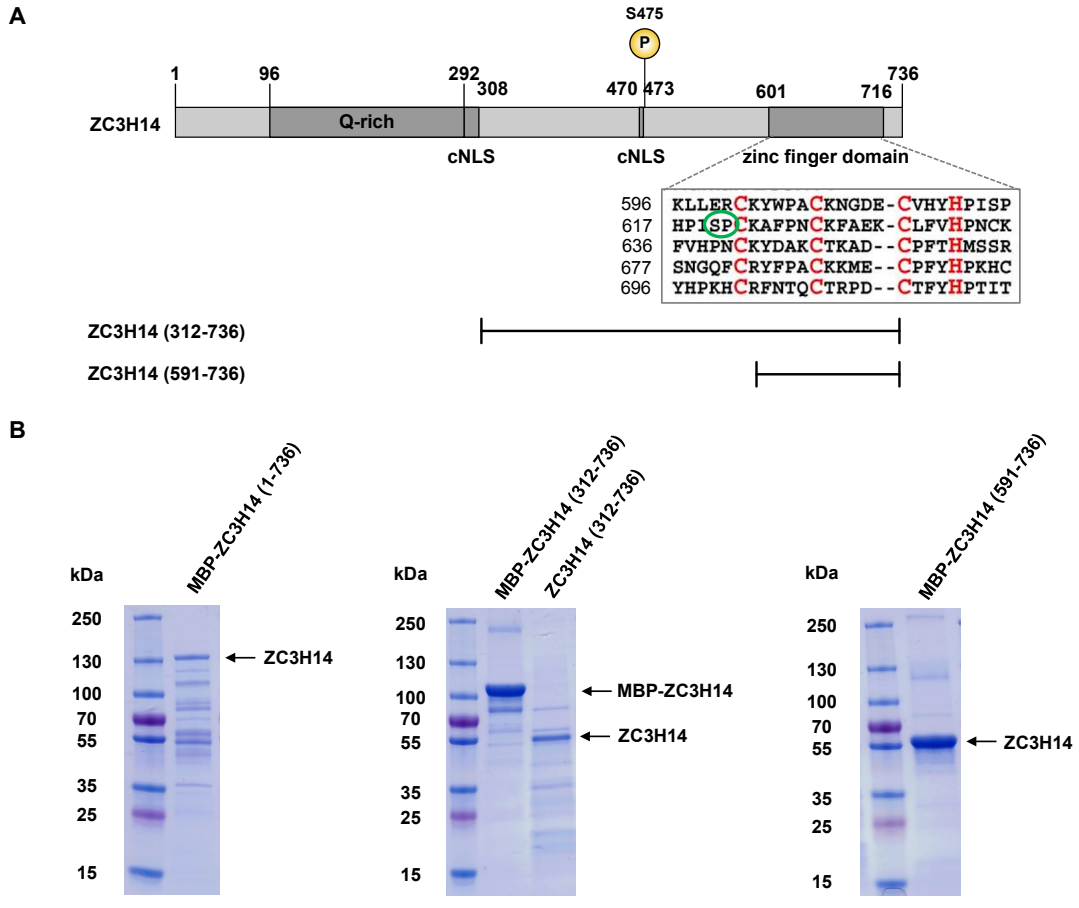


Figure 42: Overview of the generated constructs of ZC3H14 protein. (A) Constructs of full length ZC3H14 and truncated ZC3H14 protein (312-736 / 591-736) with characteristic domains. The zinc finger domain is displayed in detail. **(B)** SDS-PAGE of purified recombinant ZC3H14 proteins.

Three distinct constructs of ZC3H14 protein were generated: ZC3H14 full length and two truncated form of ZC3H14 by excluding the N-terminal glutamine-rich region, which may impair the protein expression (Figure 42A). Boundaries of the first construct were set from aa 312 to aa 736. This choice based on the IP mass spectrometry result that phosphorylation of ZC3H14 at S475, which was present in the control cell line, but was missing in the inactive CDK13 R860Q melanoma cell line. By this way, S475 of ZC3H14 was supposed to be the phosphorylation site targeted by CDK13. The second construct contained the zinc finger domain only, which provided a promising phosphorylation site due to the SP-motif at position 620. ZC3H14 proteins were expressed in *E. coli* cells and purified via the incorporated N-terminal MBP-tag. As shown in figure 42B, in contrast to the truncated ZC3H14 proteins, protein sample containing the full length ZC3H14 protein was highly instable. Hence, the full length ZC3H14 protein was neglected in the following kinase assays. The MBP-tag was only cleaved from ZC3H14 (312-736). In case of ZC3H14 (591-736), MBP was required to maintain the stability of the zinc finger domain.

The *in vitro* kinase assay was performed using ^{32}P-labeled ATP in a time course dependent manner. Each measurement was conducted with 0.2 µM purified recombinant CDK13/CycK, 50 µM of purified recombinant ZC3H14 protein and 1 mM ATP, over a time period of 4 h. Control measurements in absence of the kinase were performed as well. As shown in figure 43, both ZC3H14 constructs were not phosphorylated by CDK13/CycK, as no significant deviation between the values measured in the presence of the kinase and the values from control measurement could be observed at each indicated time point. Based on this result, the recombinant truncated ZC3H14 protein cannot be considered as a substrate of CDK13/CycK *in vitro*.

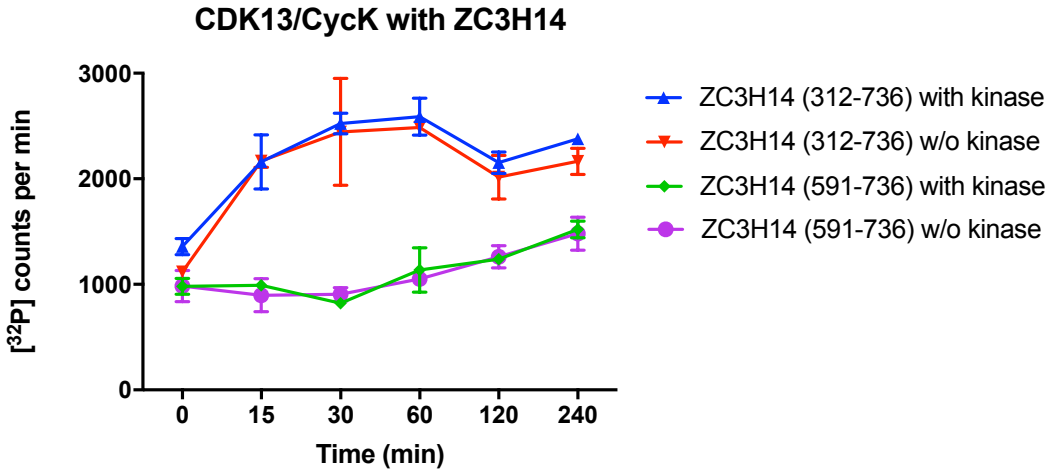


Figure 43: *In vitro* kinase time course measurement of CDK13/CycK-mediated phosphorylation of ZC3H14 (312-736) and MBP-ZC3H14 (591-736). Negative control without kinase is shown in red for ZC3H14 (312-736) and in purple for ZC3H14 (591-736). Data are shown as mean ±SD; n=3.

Surprisingly, using the same IP system, Insco and co-workers cannot detect Cyclin K as the typical canonical cyclin binding partner of CDK13 (Dai *et al.*, 2012; Greifenberg *et al.*, 2016; Liang *et al.*, 2015) in melanoma. Instead, they found CDK13 bound to Cyclin T1, the typical cyclin binding partner of CDK9 (Wei *et al.*, 1998). This result implicated Cyclin T1 as the cyclin binding partner of CDK13 in human melanoma, which might be required for the kinase activity of CDK13. Several attempts were performed to generate a recombinant protein complex of CDK13/CycT1, but they were not successful so far, as the subunits do not stabilize each other. As a result, we repeated the *in vitro* kinase assay experiment using the recombinant protein complex of CDK9/CycT1 in order to investigate whether Cyclin T1 is required to activate CDK13 in order to phosphorylate ZC3H14 as a specific substrate. As shown in figure 44, ZC3H14 (591-736), the protein construct containing the zinc finger domain only, was not phosphorylated over the reviewed time period. In comparison to the control measurement, the obtained [^{32}P] counts per minute can be considered as equivalent. However, in contrast to the previous result obtained with CDK13/CycK, ZC3H14 (312-736) was phosphorylated in a time dependent manner by CDK9/CycT1, showing significant discrepancy as compared to the control measurement at indicated time points. Thus, on the one hand, this result can be considered as an evidence that Cyclin T1 is required for the kinase activity of CDK13 in melanoma. On the other hand, this result has shown that only ZC3H14 (312-736), but not ZC3H14 (591-736) was phosphorylated by CDK9/CycT1, which supports the presence of the predicted phosphorylation site at S475 of the ZC3H14 protein.

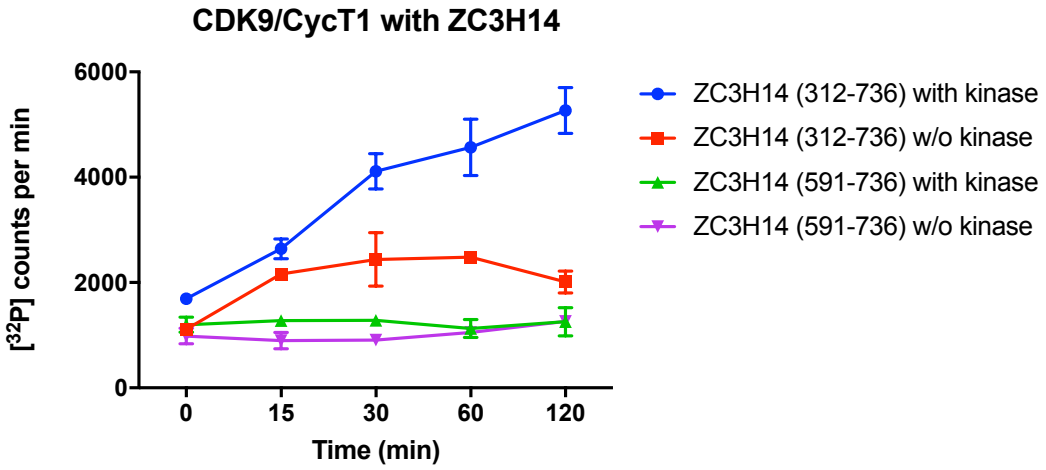


Figure 44: *In vitro* kinase time course measurement of CDK9/CycT1-mediated phosphorylation of ZC3H14 (312-736) and MBP-ZC3H14 (591-736). Kinase activity was only indicated for ZC3H14 (312-736) (blue) over a time period of 2 h. Negative control without kinase is shown in red for ZC3H14 (312-736) and in purple for ZC3H14 (591-736). Data are shown as mean ±SD; n=3.

Cell cycle-related kinase – CDK14

4.4 Structural and biochemical analysis of CDK14/Cyclin Y

4.4.1 Crystallization of the CDK14/Cyclin Y complex

Until to date, the protein structures of both CDK14 and Cyclin Y remain unknown. To design best suitable protein constructs for crystallization approaches, Swiss-Model, a structural bioinformatics tool, was applied for homology modeling of CDK14 and Cyclin Y protein structures (Biasini *et al.*, 2014; Schwede, 2003). The program makes use of pre-existing protein structures of evolutionarily related proteins as templates to build models for proteins of interest with unknown structure. For CDK14, the best alignment result was received with CDK16 with a sequence identity of 60.1% over a sequence length of 320 residues (Figure 45A). Like CDK14, CDK16 is a member of the CDK5 subfamily and is also activated by Cyclin Y (P. Mikolcevic *et al.*, 2012; Shehata *et al.*, 2012). Based on the sequence alignment and existing structure of CDK16 (PDB: 5G6V; (Dixon-Clarke *et al.*, 2016), a model of CDK14 protein structure was illustrated involving predicted secondary structures (purple: alpha-helix; green: beta-sheet) (Figure 45B). On the basis of this modeling result, a series of different crystallization constructs of the kinase domain of CDK14 was generated intending to generate a stable and minimal construct in favour of protein crystallization. Totally, two different N-terminal starting boundaries in combination with three different C-terminal endings were set for molecular evaluation (Figure 45A and Figure 46A). Expression of the CDK14 constructs in *Sf9* cells have shown that both starting boundaries 125 and 133 worked well with position 447 or 469 as ending, whereas 430 seems to be too short to receive high yield and pure proteins (Figure 46B).

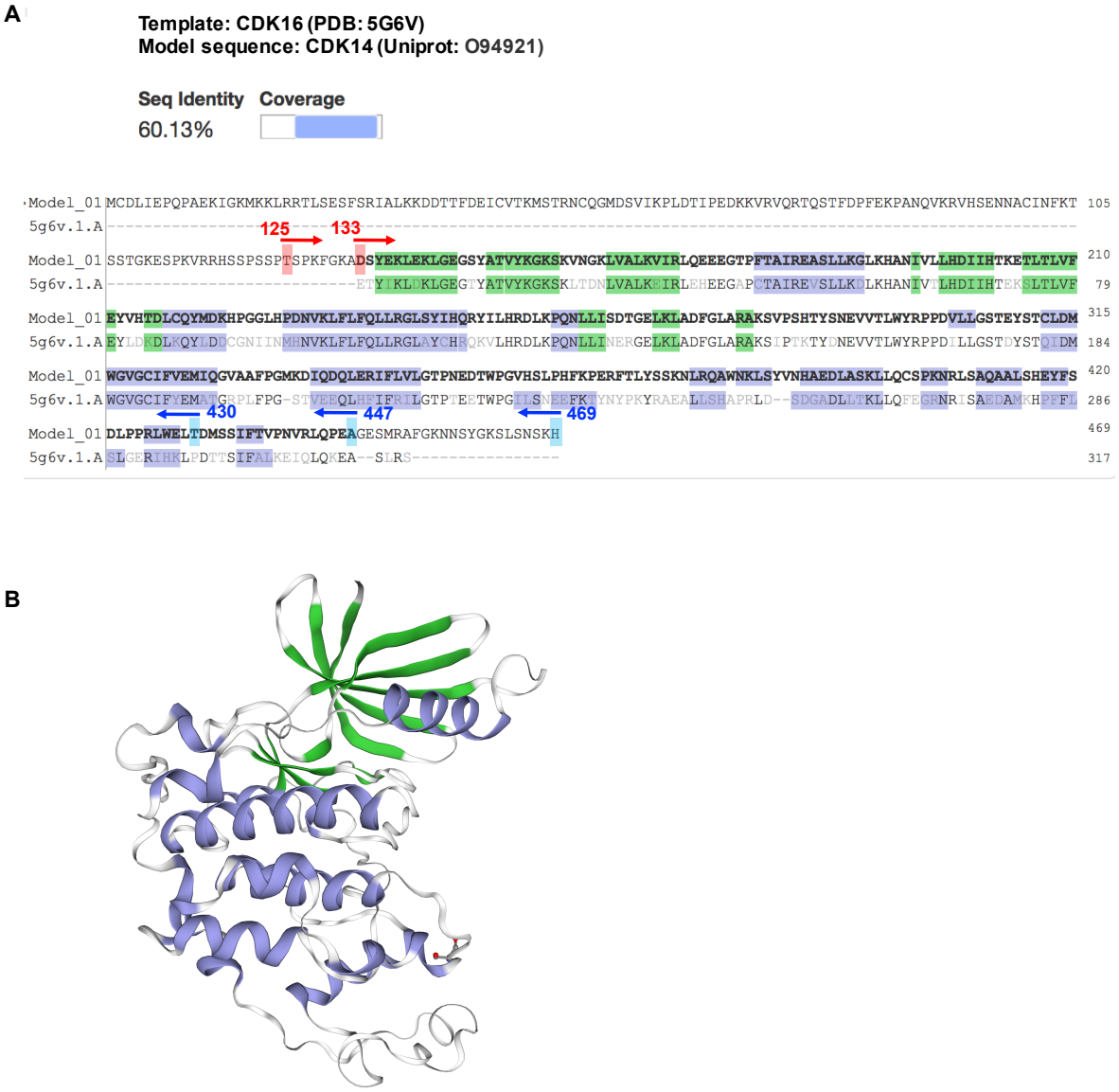


Figure 45: Homology modeling of CDK14 using SWISS-MODEL. (A) Sequence alignment of CDK14 with CDK16. CDK16 (PDB: 5G6V; (Dixon-Clarke *et al.*, 2016) was identified as the best template for the modeling of CDK14. N-terminal starting boundaries (red) and C-terminal terminating boundaries (blue) of different crystallization constructs are illustrated in the sequence. Amino acid sequences of predicted secondary structures of CDK14 are highlighted in purple for alpha-helix and green for beta-sheet. **(B)** Model of CDK14 protein structure.

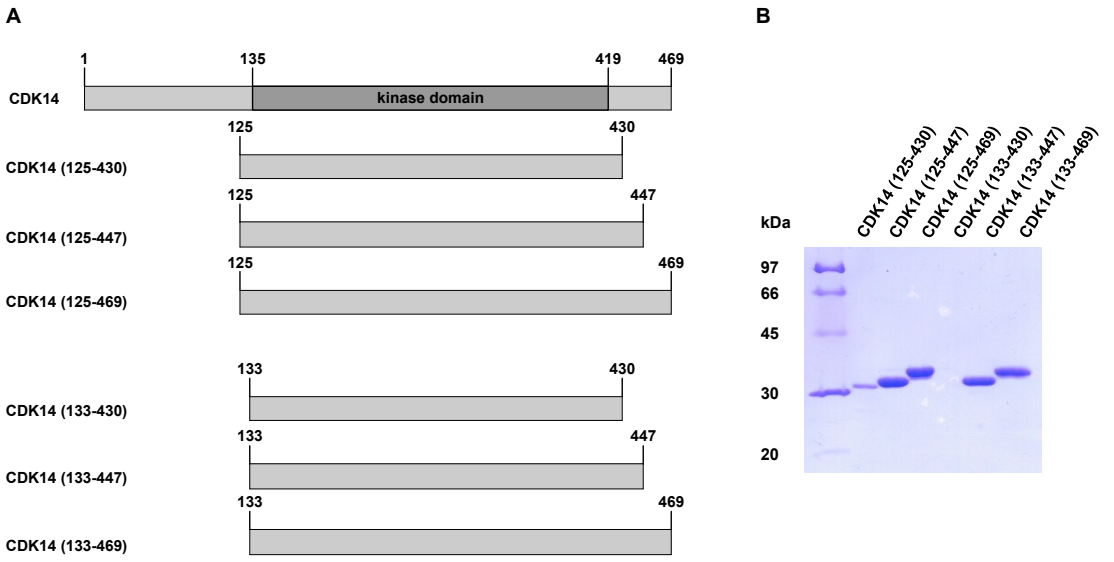
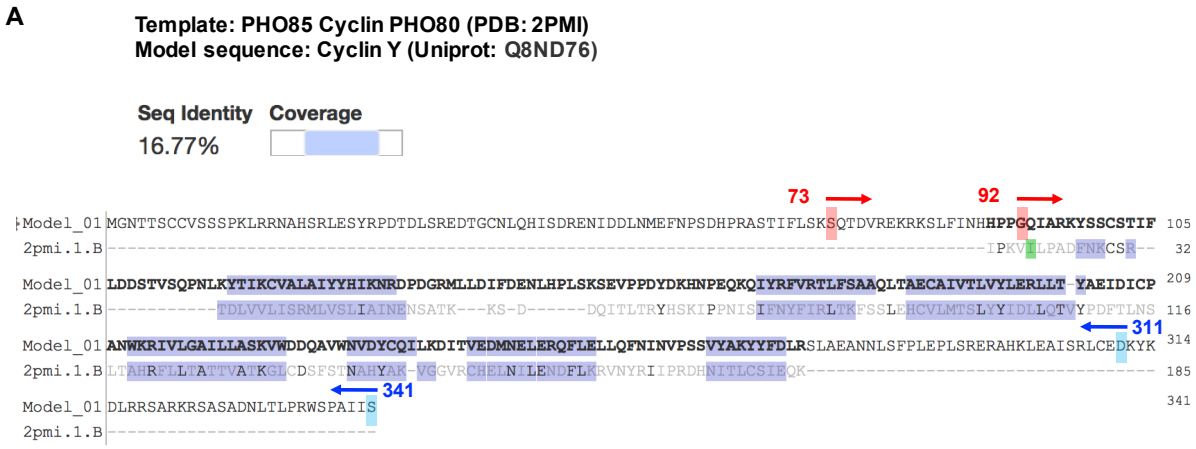


Figure 46: (A) Overview of CDK14 protein expression constructs intended for crystallization trials. **(B)** SDS-PAGE of purified recombinant CDK14 proteins after removing of the GST tag by TEV-digestion. C-terminal boundary at 430 appears to be too short for successful protein expression.

The same procedure of homology modeling was repeated for Cyclin Y. The best alignment result was received with PHO85 Cyclin PHO80 with a sequence identity of 16.8% over a sequence length of 190 residues (Figure 47A). PHO85 Cyclin PHO80 is the the Cyclin of the Pho85-Pho80 CDK-Cyclin complex (PDB: 2PMI; (Huang *et al.*, 2007). A model of Cyclin Y protein structure was illustrated involving predicted secondary structures (purple: alpha-helix; green: beta-sheet) (Figure 47B).



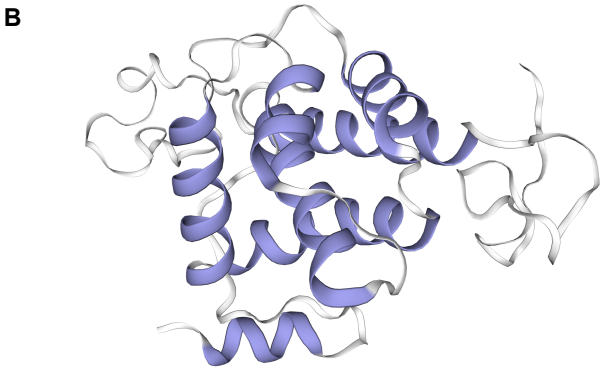


Figure 47: Homology modeling of Cyclin Y using SWISS-MODEL. (A) Sequence alignment of Cyclin Y with PHO85 Cyclin PHO80. PHO85 Cyclin PHO80 (PDB: 2PMI; (Huang *et al.*, 2007) was defined as the best template for the modeling of Cyclin Y. N-terminal starting boundaries (red) and C-terminal terminating boundaries (blue) of different crystallization constructs are illustrated in the sequence. Amino acid sequences of predicted alpha-helix structures of Cyclin Y are highlighted in purple. **(B)** Model of Cyclin Y protein structure.

For the Cyclin again, two boundaries of each N- and C-terminus were defined, intended for protein production and crystallization trials (Figure 47A and Figure 48A). In addition of considering the alignment result and secondary structure prediction based on PHO85 Cyclin PHO80, boundaries setting was also determined based on the knowledge that not only the cyclin box but also the regions flanking the cyclin box, especially aa 85-136, are required for interaction with CDK14 (Jiang *et al.*, 2009). Successful protein expression was received in *E. coli* when starting with 73 or 92 in combination with 341, but not with 311 as ending point (Figure 48B).

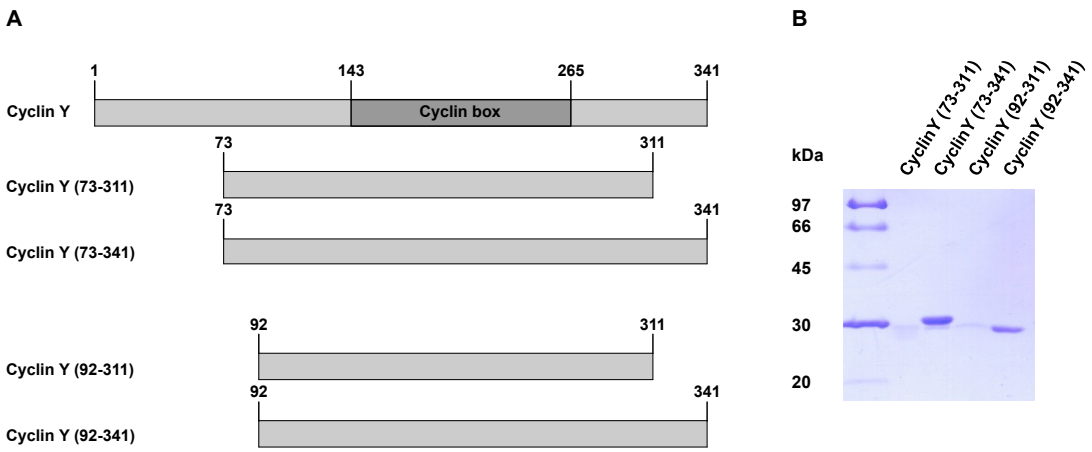


Figure 48: (A) Overview of Cyclin Y protein expression constructs intended for crystallization trials. **(B)** SDS-PAGE of purified recombinant Cyclin Y proteins after removing of the GST-tag by TEV-digestion. C-terminal boundary of position 311 appears to be too short for successful protein expression.

Based on the protein expression results of truncated CDK14 and Cyclin Y, the next step was to generate a short kinase complex by co-purification of CDK14 expressed in *Sf9* with Cyclin Y expressed in *E. coli*. Here, two combinations of CDK14 with Cyclin Y were determined as starting complexes for crystallization trials: CDK14 (125-469) with Cyclin Y (73-341) and CDK14 (133-447) with Cyclin Y (92-341), combining the two longest and the two shortest constructs of CDK14 and Cyclin Y (Figure 49A). As shown in figure 49B, the selected combinations of CDK14 and Cyclin Y have been successfully co-purified as protein complex with a ratio slightly differ from 1:1. Both co-purified CDK14/CycY complexes were further applied to crystallization trials. As the mass spectrometry analysis of CDK14 kinase domain (125-469) confirmed efficient labeling of CDK14 by potential small molecule inhibitor compound (Figure 50), crystallization trials were also started with CDK14/CycY complexes pre-incubated with the compound (kindly provided by the laboratory of Nathanael Gray, Boston, USA) intending to stabilize the ATP binding pocket thereby improving the protein crystallizability.

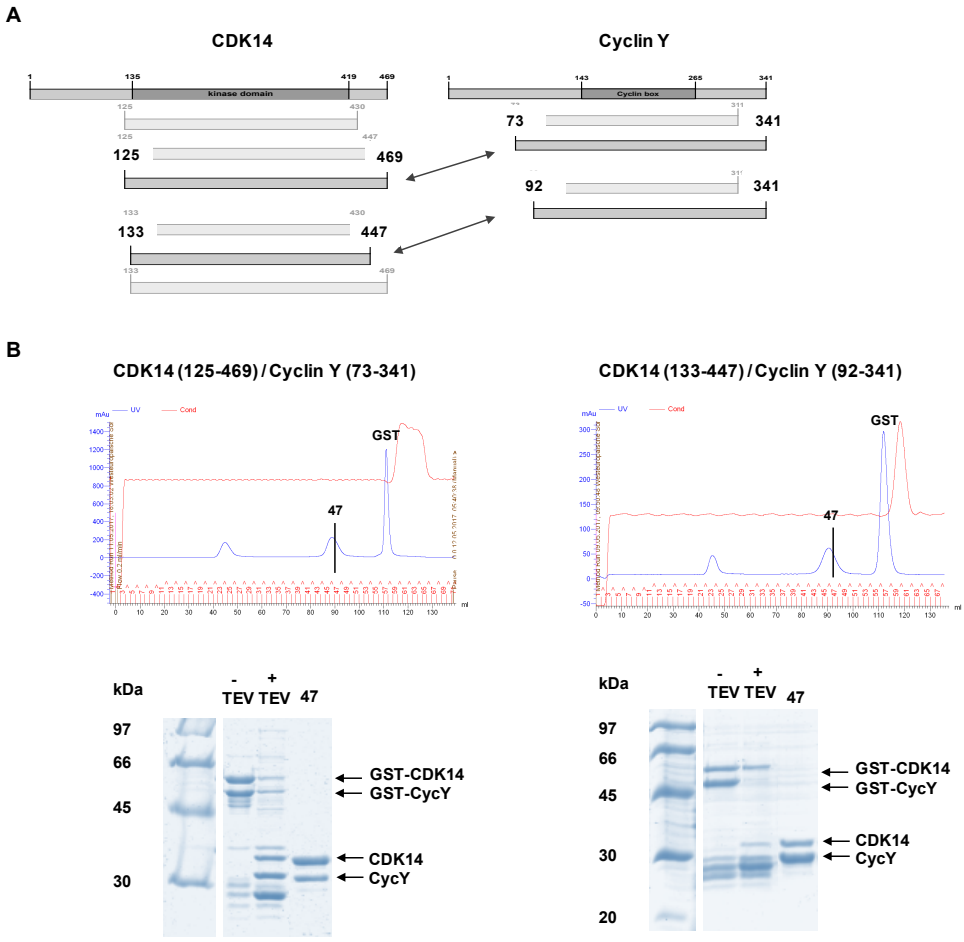


Figure 49: (A) Overview of selected combinations of truncated CDK14 and Cyclin Y constructs intended for crystallization trials. **(B)** Gel filtration chromatogram and SDS-PAGE of co-purified CDK14/CycY complexes of the determined combinations including status before and after TEV-digestion.

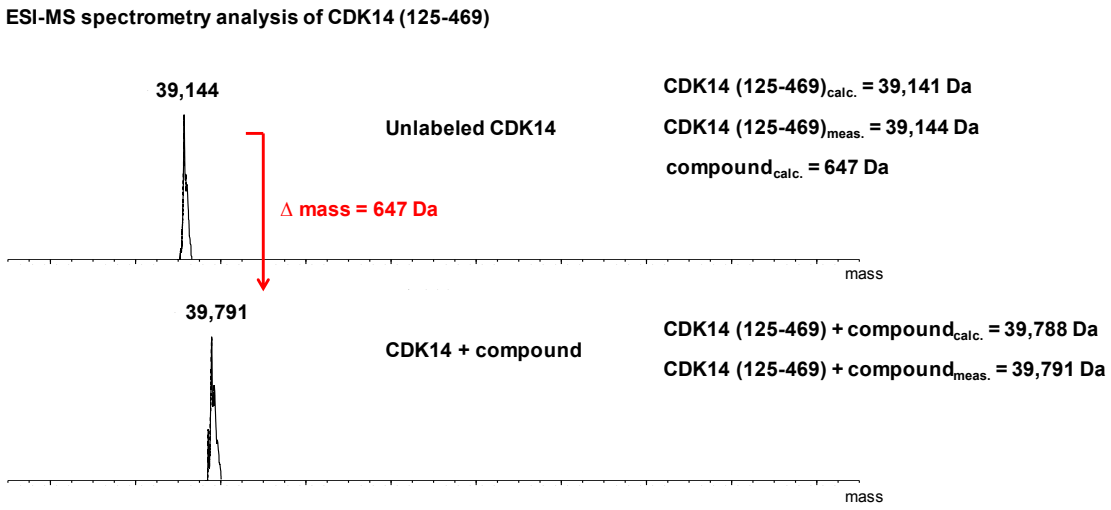


Figure 50: ESI-Mass spectrometry analysis of unlabeled CDK14 and CDK14 labeled with compound (provided by the laboratory of Nathanael Gray). The measured raw mass (39,144 Da) fits well with the calculated mass of CDK14 (125-469) (39,141 Da), which contains a phosphorylation group at S289. The mass shift of 647 Da occurred after compound addition is corresponding to the mass of the compound.

The purified recombinant CDK14/CycY complexes was incubated over night at 4°C with the potential small molecule inhibitor compound in a ratio of 1:1.5. Excess free compound was removed by size-exclusion chromatography using a disposable PD10 desalting column (Merck, Darmstadt, Germany). CDK14/CycY complexes without and with addition of compound were concentrated up to 8-9 mg/mL for the crystallization trials. Both, untreated CDK14/CycY and CDK14/CycY complexes pre-incubated with compound, led to protein crystallization (Figure 51). Some of the crystals were measured at the SLS synchrotron and showed diffraction spots, with the combination of CDK14 (125-469)/CycY (73-341) labeled with compound performed best. Stabilization of the ATP pocket by compound appeared to be more effective in generation of higher qualitative protein crystals. However, the resolution was not yet sufficient for a complete data set recording. Thus, application of compound and protein crystallization need further optimization to improve the quality of the protein crystals.

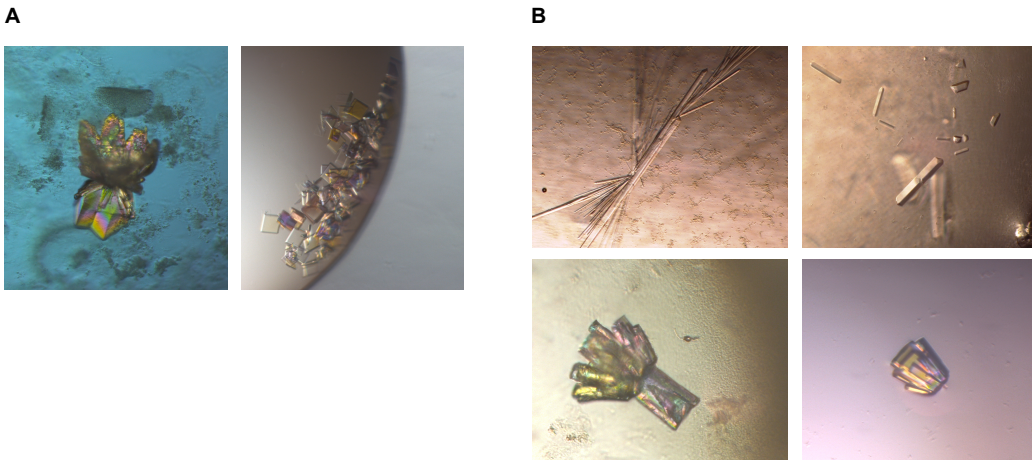


Figure 51: Examples of generated protein crystals of the two CDK14/CycY complexes. (A) Pictures of CDK14/CycY crystals with addition of compound. CDK14 (133-447)/CycY (92-341) (left) and CDK14 (125-469)/CycY (73-341) (right). **(B)** Pictures of protein crystals of CDK14 (125-469)/CycY (73-341) without compound addition.

4.4.2 Identification of CDK14/Cyclin Y binding proteins required for kinase activity

To test if a series of small molecules may act as potent and putatively covalent CDK14 inhibitors, the protein complex of full length kinase with corresponding full length cyclin is the most suitable construct to ensure activity, specificity and regulation of the kinase. Thus, we expressed the kinase complex of full length CDK14 and full length Cyclin Y each tagged N-terminal with a GST-anchor for affinity chromatography as fusion construct in *Sf9* cells (Figure 52A).

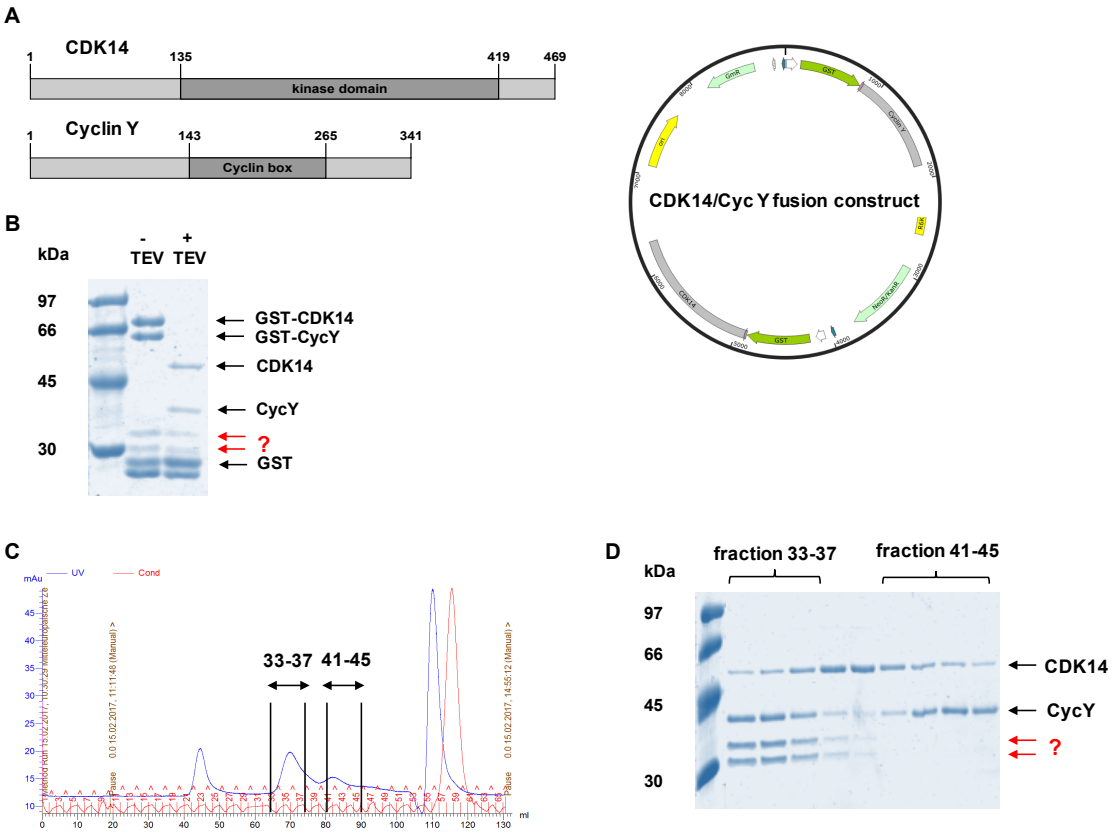


Figure 52: Overview of generated fusion complex of full length CDK14/Cyclin Y. (A) Constructs of full length CDK14 and full length Cyclin Y represented as fusion construct. **(B)** SDS-PAGE of purified recombinant CDK14/CycY before and after TEV digestion. **(C)** Gel filtration chromatogram of purified recombinant CDK14/CycY. **(D)** SDS-PAGE of purified recombinant CDK14/CycY after gel filtration showing two additional bands in fraction 33-37.

After removing of the respective GST-tags by TEV digestion, the protein complex was purified to homogeneity before applying to the analysis in kinase assays. Although purification of the full length kinase and cyclin worked well without the appearance of protein aggregation or degradation, two

unknown bands at ~30 kDa in association with the kinase and cyclin were additionally exposed by gel filtration at an early fractionation stage (Figure 52C and D). It can be excluded that these two bands have been resulted from GST cleavage, as they were already present before TEV digestion (Figure 52B). The SDS-PAGE of fraction 33 to 37 in figure 52D is more likely to show that the appearance of the doublet band is strongly correlated with the presence of Cyclin Y. During the early fractionation, intensity of the doublet band is equivalent to the protein band of Cyclin Y, but quite the contrary to the course of CDK14.

In search of an active CDK14 kinase complex, the kinase activity of differently formed CDK14/CycY complexes were tested in *in vitro* kinase activity assays using radioactive labelled [^{32}P]-γ-ATP (Figure 53). Following three constructs were used here: CDK14 (125-469)/CycY (73-341), the short complex co-purified for crystallization approach. Full length CDK14/CycY fusion complex containing the unknown doublet band (fractions 33-37) and the protein sample only containing CDK14 and Cyclin Y (fractions 41-45). IBER-IRStide protein, a recombinant artificial protein consisting the C-terminal fragment of the human retinoblastoma (RB) protein typically applied as *in vitro* CDK14 substrate (Shu *et al.*, 2007), was used as CDK14 substrate in the kinase assays performed here. Additional, Histone H3.1 protein, not known as a phosphorylation target of CDK14, was used as negative substrate control.

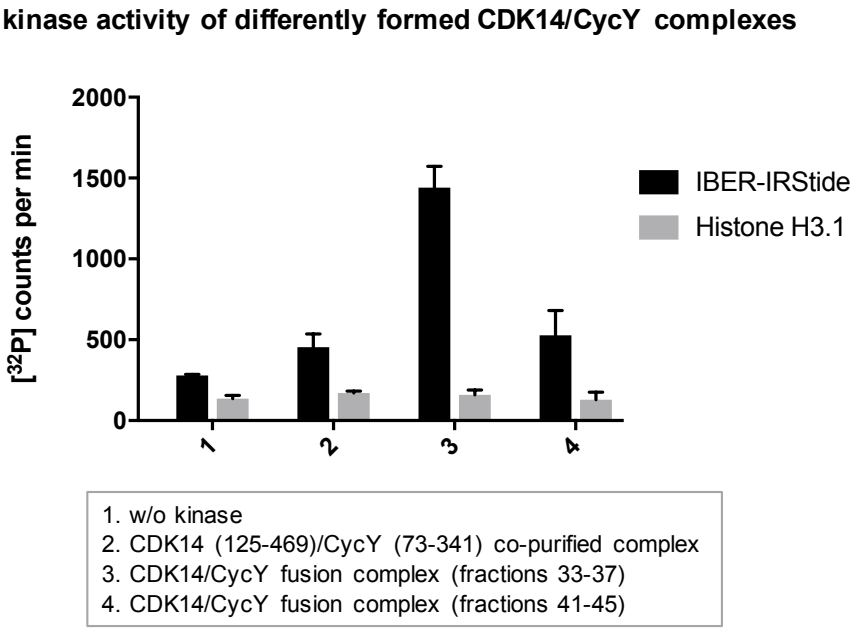


Figure 53: *In vitro* kinase activity assay to test the activity of differently formed CDK14/CycY complexes. 0.2 µM of purified recombinant CDK14/CycY complex and 50 µM IBER-IRStide as substrate were applied to each approach with 1 mM ATP for 30 min. Histone H3.1 was used as negative substrate control. Data are shown as mean ±SD; n=2.

Interestingly, the protein sample containing the CDK14/CycY fusion complex with the additional doublet band (fractions 33-37) was significantly more active than the other two samples and the control without kinase. The protein samples containing only the CDK14/CycY fusion complex as well as the co-purified short CDK14/CycY complex displayed a higher kinase activity compared to the negative control. However, in comparison to the fusion complex containing the doublet band, the kinase activity of these two samples was reduced approximately by factor 3. To exclude a potential correlation between this result and the phosphorylation level of CDK14, which is crucial for the kinase activity, ESI-MS analysis was performed to confirm the proper phosphorylation of the purified recombinant CDK14 kinase domain (125-469). As shown in Figure 54, almost half of the measured CDK14 protein was significantly phosphorylated (+ 79 Da), indicating the presence of active kinase. Thus, the missing kinase activity was not resulted from a potential absence of active CDK14. As Histone H3.1 was not phosphorylated in all three approaches, phosphorylation of IBER-IRStide protein by CDK14/CycY fusion complex (fraction 33-37) can be considered as specific. This active CDK14/CycY kinase complex can be further used to test potential CDK14 small molecule inhibitors for their action.

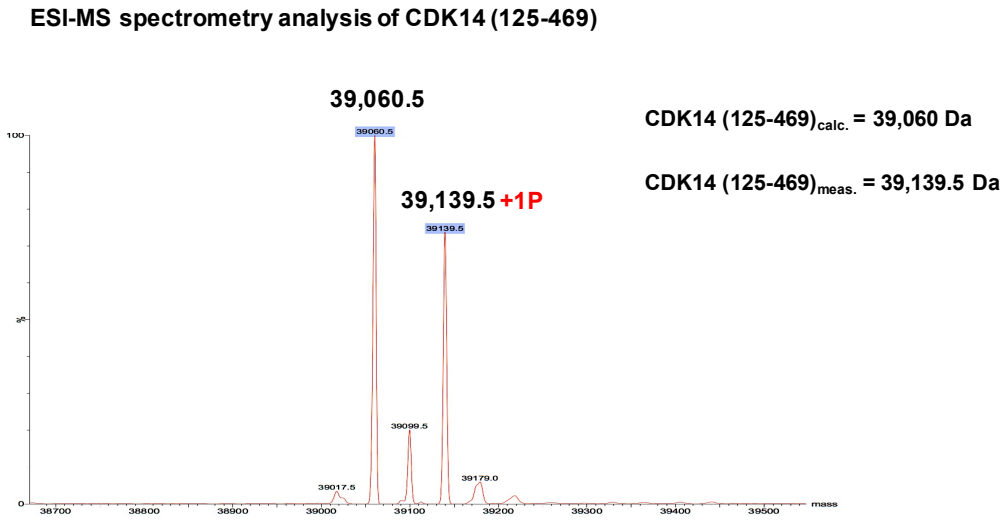


Figure 54: ESI-Mass spectrometry analysis of CDK14 (125-469). The spectrum was generated using Xcalibur (Thermo Fisher Scientific, Dreieich, Germany). The determined mass shift of 79 Da indicates clearly the single phosphorylation of CDK14.

To additionally confirm the activity of the generated CDK14/Cyclin Y construct as well as to examine if the compounds have an inhibitory effect on CDK14 by covalent targeting, *in vitro* kinase activity assays of pre-incubation time course series with different concentrations of the lead compound were conducted. Following the respective pre-incubation time, the kinase reaction was

carried out over a period of 30 min. As shown in figure 55, the kinase activity at 0.5 µM CDK14 concentration was almost completely blocked after 5 min pre-incubation time for 100, 10 and 1 µM inhibitor concentrations. In case of lower inhibitor concentrations (0.1 and 0.01 µM), the kinase activity decreased less dramatically with increasing pre-incubation time but decreased stronger than the DMSO control without inhibitor. In general, the lower the inhibitor concentration, the more pre-incubation time of kinase with inhibitor was required to impair the kinase activity at 0.5 µM concentration. Here, it has been shown on the one hand that the purified recombinant CDK14/CycY complex is suitable for inhibitor compound testing. On the other hand, it demonstrated the effect of covalent binding, as increasing pre-incubation time augmented the inhibitor potency (comparable with the results obtained from the CDK12 measurements previously described in chapter 4.1) thus confirming the covalent targeting of CDK14 by the tested compound.

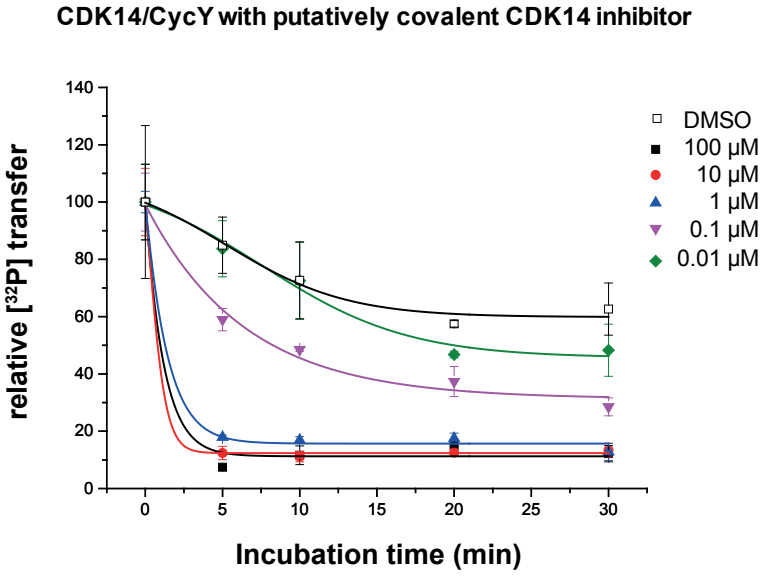


Figure 55: *In vitro* kinase activity assay of CDK14/Cyclin Y with various inhibitor concentrations in association with increasing pre-incubation times. The assay was performed using 0.5 µM kinase, 50 µM IBER-IRStide as substrate and 1 mM ATP. Following the respective pre-incubation time, the kinase reaction was carried out over a period of 30 min. The counts per minute of the kinase activity measurements were normalized to the relative ^{32}P transfer. Data are shown as mean ±SD; n=3.

As the additional doublet band at ~30 kDa appeared to be crucial for the kinase activity of CDK14, peptide mass finger print analysis was performed to identify the unknown doublet band. All four protein bands of the protein sample originating from fractions 33-37 were submitted to in-depth MS analysis. As expected, the two protein bands at ~50 kDa (band 1) and ~40 kDa (band 2) were identified as CDK14 and Cyclin Y. Interestingly, the unknown doublet band at ~30 kDa (band 3 and

4) was identified as 14-3-3 proteins, a group of multifunctional proteins that bind to specific phosphorylated motifs in protein targets thereby modulating the function of a broad range of various cellular proteins (Muslin *et al.*, 1996). Especially three of the seven 14-3-3 isoforms were identified by peptide mass finger print analysis: epsilon, beta and eta (Figure 56). This result provided the finding that 14-3-3 proteins are involved in CDK14 kinase complex formation, which were required for the proper kinase activity.

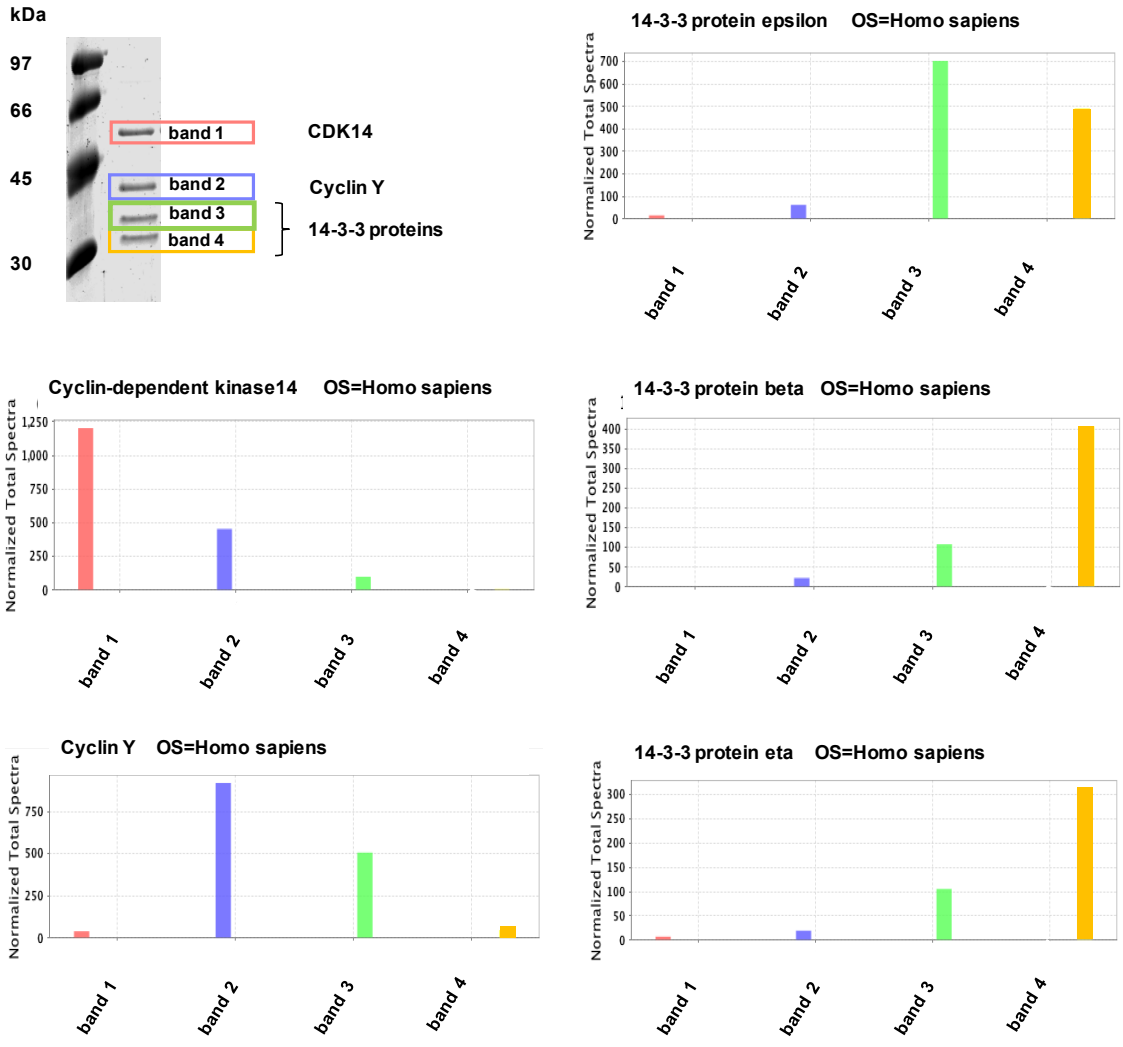


Figure 56: Illustration of the identified proteins by peptide mass finger print analysis. 14-3-3 proteins (band 3 and 4) were additional identified as binding partner of CDK14 kinase complex.

4.4.3 Study of CDK14–Cyclin Y–14-3-3 interactions in kinase complex formation

To gain further insights into the associations between CDK14, Cyclin Y, and 14-3-3 proteins, interaction studies were performed to analyze the kinase complex formation. Although 14-3-3 proteins were identified as specific binding proteins of the CDK14/CycY fusion complex required for the kinase activity, the direct interaction partner of 14-3-3 remained unclear in this context. To address this question, full length CDK14 and Cyclin Y were separately expressed in *Sf9* cells in order to determine whether 14-3-3 proteins could be co-purified by either of them. In terms of CDK14, only the single protein band of the full length GST-tagged CDK14 was detectable without binding of 14-3-3 proteins (Figure 57A). In contrast, the doublet band of 14-3-3 proteins at ~30 kDa was clearly co-purified in association with the full length Cyclin Y (Figure 57B), confirming the previously made assumption based on the gel filtration outcome of CDK14/CycY fusion complex. The GST-tag of the cyclin was cleaved off by TEV-digestion. To examine whether 14-3-3 binding only occur in *Sf9* expression system, the full length Cyclin Y construct was expressed in *E. coli* cells as well. In contrast to Cyclin Y expressed in *Sf9* cells, the doublet band of 14-3-3 proteins disappeared (Figure 57C). Only the protein band of the full length Cyclin Y was detectable. Together, it can be noted that 14-3-3 proteins do not interact with CDK14 only and the binding of 14-3-3 proteins to Cyclin Y is only present if Cyclin Y was expressed in *Sf9* but not in *E. coli* cells.

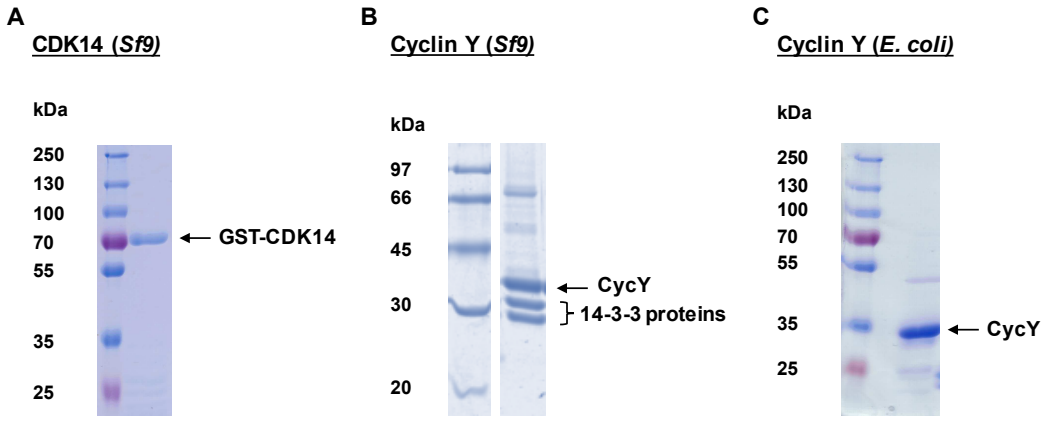


Figure 57: Overview of full length CDK14 and Cyclin Y separately expressed in *Sf9* and *E. coli* cells. **(A)** SDS-PAGE of GST-CDK14 purified from *Sf9* cells. **(B)** SDS-PAGE of Cyclin Y purified from *Sf9* cells in association with 14-3-3 proteins. **(C)** SDS-PAGE of Cyclin Y purified from *E. Coli* cells.

As 14-3-3 binding only occur with Cyclin Y expressed in *Sf9* cells, the protein sample of Cyclin Y expressed in *E. coli* cells can be perfectly used to perform pull-down assays in order to confirm the

missing binding of 14-3-3 proteins to Cyclin Y expressed in *E. coli*. GST-pull-down assays were conducted using full length GST-tagged Cyclin Y expressed in *E. coli* as bait protein immobilized to GSH affinity beads and five 14-3-3 isoforms (beta, gamma, epsilon, ceta and eta) as prey protein. As shown in figure 58, only GST-Cyclin Y was 'pulled down' by elution (E) in each case, while unbound 14-3-3 proteins remained in the flow through (FT). Thus, this result is in accordance with the previous finding that binding of 14-3-3 proteins to Cyclin Y was not available when Cyclin Y was purified from *E. coli* cells.

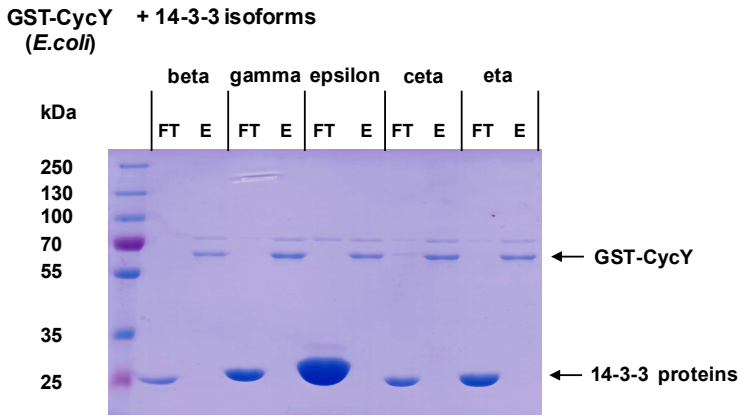


Figure 58: GST-pull-down assay of GST-Cyclin Y (*E. coli*) with different 14-3-3 isoforms. Flow through (FT) and eluate (E) of each 14-3-3 lane are shown by SDS-PAGE. In none of the cases interact 14-3-3 protein with Cyclin Y.

After examination of the respective interaction between 14-3-3 proteins with CDK14 and Cyclin Y, the next step was to investigate if formation of CDK14 kinase complex requires 14-3-3 proteins bound to Cyclin Y. Here, GST-pull-down assay was performed using full length GST-tagged CDK14 purified from *Sf9* cells as bait protein and full length Cyclin Y as prey protein expressed both in *E. coli* and *Sf9* cells without and with bound 14-3-3 proteins. Interestingly, neither Cyclin Y without nor with bound 14-3-3 were pulled down by CDK14 (Figure 59). The existence of beads bound GST-CDK14 was clearly demonstrated by the positive control as GST-CDK14 was only present in the eluate. This result implies that 14-3-3 proteins do not represent the limiting factor for the formation of the native CDK14 kinase complex.

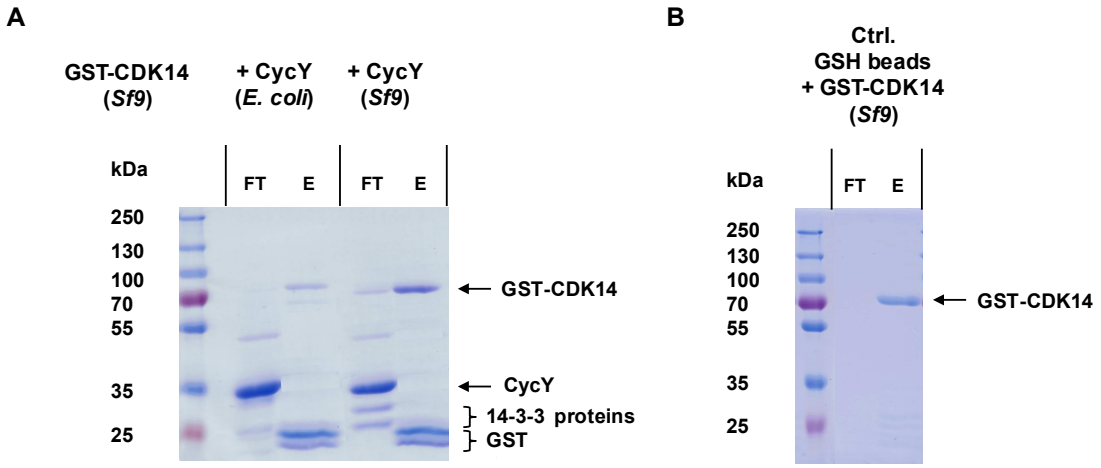


Figure 59: (A) GST-pull-down assay of GST-CDK14 (*Sf9*) with Cyclin Y without (*E. coli*) and with binding of 14-3-3 proteins (*Sf9*). Flow through (FT) and eluate (E) of each Cyclin Y lane are shown by SDS-PAGE. **(B)** GST-pull-down assay of beads bound GST-CDK14 as control.

It seems reasonable to suppose that kinase complex formation requires an interplay between CDK14 and Cyclin Y enabled by co-expression in *Sf9* cells. It is very likely that additional phosphorylations of Cyclin Y by CDK14 facilitate the kinase-cyclin interaction. To identify if specific phosphorylation sites of Cyclin Y only occur in the presence of co-expressed CDK14, peptide mass fingerprint analyses were performed with following full length Cyclin Y proteins: Cyclin Y without 14-3-3 binding (*E. coli*), Cyclin Y with bound 14-3-3 proteins (*Sf9*) and Cyclin Y co-expressed with CDK14 as fusion complex with bound 14-3-3 proteins (*Sf9*) (Figure 60).

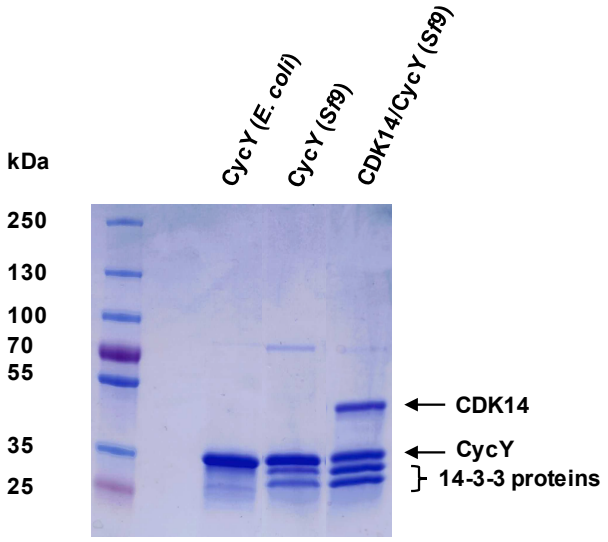


Figure 60: Different full length Cyclin Y proteins were sent to peptide mass finger print analysis to identify specific phosphorylation sites.

In case of Cyclin Y expressed in *E. coli* cells, no phosphorylation site was detected by peptide mass fingerprint analysis as expected for the bacterial expression system. In contrast, several identical phosphorylation sites were identified both in Cyclin Y expressed in *Sf9* cells and in Cyclin Y co-expressed with CDK14 as fusion complex (Figure 61). Especially phosphorylation at S100 and S326 (green box), previously described as essential sites for 14-3-3 binding (Li *et al.*, 2014), were significantly phosphorylated in both Cyclin Y constructs with probabilities close to 100%. Interestingly, two phosphorylation sites: pS33 and pT331 (red), which were significantly identified in Cyclin Y co-expressed with CDK14, were missing in the single expressed Cyclin Y protein (*Sf9*). Thus, it can be assumed that phosphorylation of Cyclin Y at position S33 and T331 is depending on the presence of CDK14 and might be crucial for the direct CDK14-Cyclin Y interaction. Mass spectrometry spectra derived from peptide mass fingerprint analyses of the Cyclin Y constructs are included in the appendix.

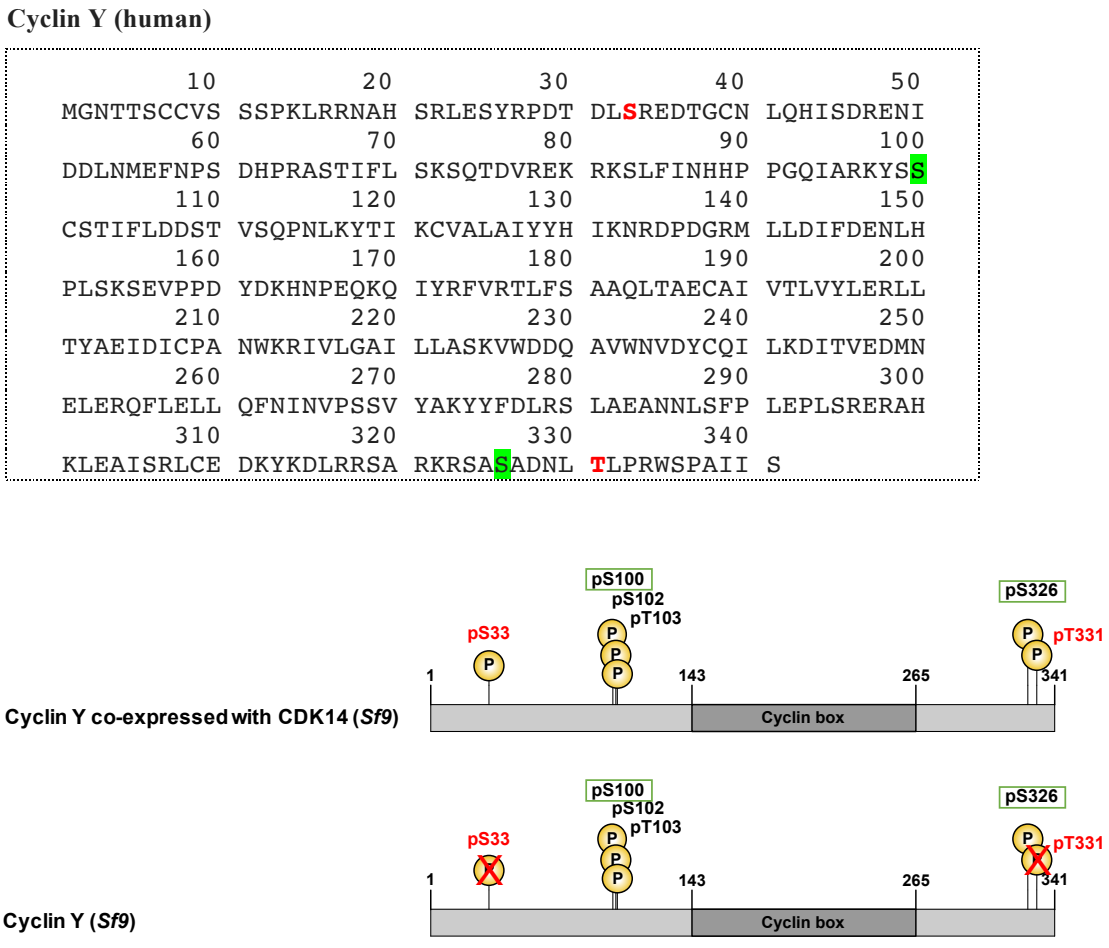


Figure 61: Overview of identified phosphorylation sites in different Cyclin Y constructs by peptide mass fingerprint analyses. Four phosphorylation sites are identical in both constructs, while two sites (red) are missing in the single expressed Cyclin Y protein (*Sf9*). Essential sites for 14-3-3 binding are highlighted in green box.

To determine whether the previously defined 14-3-3 binding sites at S100 and S326 of Cyclin Y provide docking sites for 14-3-3 proteins, we mutated theses serine residues to non-phosphorylatable alanine. Therefore, full length Cyclin Y mutant proteins containing the single mutation S100A, S326A, respectively and the double mutation S100A/S326A were generated. The Cyclin Y mutants were further co-expressed with the full length CDK14 as fusion construct in *Sf9* insect cells. Each of the proteins was tagged N-terminal with a GST-anchor for affinity chromatography. The recombinant putative kinase complex was then purified with gelfiltration after TEV-digestion. As shown in figure 62, mutation of S100 to alanine completely abolished 14-3-3 binding with Cyclin Y, as only the single protein band of Cyclin Y could be detected without the presence of the doublet band of 14-3-3 proteins. Furthermore, the association of Cyclin Y with CDK14 was greatly reduced as well. Although the GST-tag of CDK14 was not cleaved by TEV-digestion in this case, it made even more obvious that the binding of Cyclin Y with CDK14 was abolished as they were exposed by gel filtration at different fractionation stages significantly far apart from each other. Thus, the obtained result from this mutagenesis experiment suggested not only that the phosphorylation of Cyclin Y at S100 is crucial for the interaction between Cyclin Y and 14-3-3 proteins but also indicated that 14-3-3 binding is required for the association of Cyclin Y with CDK14. However, recombinant kinase complex of full length CDK14 with Cyclin Y(S326A) as well as with Cyclin Y(S100A/S326A) mutant could not be successfully expressed after several attempts.

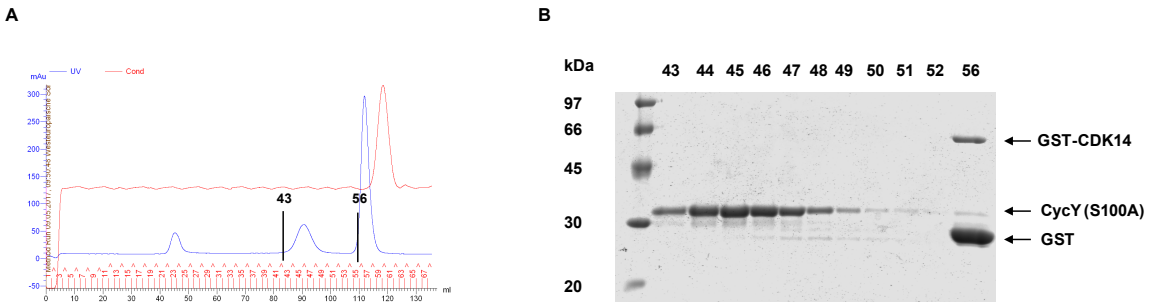


Figure 62: Co-expression of full length Cyclin Y(S100A) mutant with full length CDK14 as fusion construct in *Sf9* cells. (A) Gel filtration chromatogram of CycY(S100A) co-expressed with CDK14. **(B)** SDS-PAGE after TEV-digestion and gel filtration indicated clearly that both the association of Cyclin Y with 14-3-3 and the association of Cyclin Y with CDK14 were abolished. GST-tag of CDK14 was not cleaved by TEV-digestion.

5 Discussion

Transcriptional kinases – CDK12 and CDK13

5.1 Protein structure-based approach assisting in understanding of protein-ligand interaction

Identification of potential small molecule inhibitors to target disease-related proteins is considered to be a time-consuming and expensive process (Morgan *et al.*, 2011; Paul *et al.*, 2010). Various techniques have been developed over the years, such as combinatorial chemistry and high-throughput screening, which have been considered as common methods in traditional drug development (Haggarty *et al.*, 2000; Young *et al.*, 1998). With the advancement of technology and the ongoing development of scientific knowledge, computer-based drug discovery methods become more and more the leading technique over time, applied whenever searching for target proteins or to identify hit compounds (Sliwoski *et al.*, 2013). Since protein–ligand interactions play a fundamental role in drug discovery, a number of computational approaches have been developed to characterize and use the knowledge of such interactions to identify potential drug candidates. Computer-based methods can be mainly divided into protein structure-based and ligand-based approaches (Jorgensen, 2004; Ou-Yang *et al.*, 2012).

In search of potential CDK12 and CDK13 small molecule inhibitors, we chose a protein structure-based approach using molecular docking by structure superimposition of X-ray structures of CDK12 and CDK13 as target protein with X-ray structures of kinases containing the desired compound. Molecular docking of flavopiridol into the ATP binding site of CDK12 not only provided valuable information for the modest activity of flavopiridol against CDK12 by indicating the side chain of H1040 as a major contributor of steric hindrance for flavopiridol but also defined a general exclusion criterion which can be further used to identify hit compounds. Through additional mutagenesis experiments, the side chain of H1040 has been successfully confirmed as a major contributor of steric hindrance for flavopiridol.

Using a protein structure-based approach, dinaciclib was predicted not to have a steric clash but rather to form a favorable face-to-face stacking interaction with CDK12 H1040. Kinase activity assays indeed confirmed dinaciclib as a potent CDK12 and CDK13 inhibitor. It has been assumed that the predicted favorable base stacking interaction could be the one factor determining for the potent inhibitory activity of dinaciclib against CDK12. As demonstrated in earlier works of aromatic interactions, this type of non-covalent interactions can provide both stability and selectivity within

the interaction partners (Waters, 2002). Furthermore, attractive non-bonded interactions between aromatic rings are often seen in medical and pharmaceutical sectors. The vast majority of medicinal products contain aromatic substituents and their differential recognition by proteins is likely dominated by aromatic interactions (Gilman *et al.*, 1993).

As docking of dinaciclib into CDK12 active site was resulting from superimposition of known crystal structures, it is very likely that the supposed π-π stacking interaction is just a visual impression created by small rearrangements of the individual moieties that are beyond the accuracy of the superimposition procedure. This assumption can be supported by following evidences: When calculating the distance between atoms of the histidine imidazole ring of H1040 and the pyridine-N-oxide ring of dinaciclib in the binding pose, the shortest measured distance was 1.0 Angstrom. This distance, however, is too close for a reliable π-π stacking interaction as 3.3-3.8 Å. is the general accepted centroid-centroid distance (Alvarez, 2013; Janiak, 2000). It is far more probable that this distance would favor a displacement mode of binding. Additional, it has been shown that a near face-to-face alignment with overlapping of most of the ring-plane area, which has been proposed by the docking model of dinaciclib, is extremely rare. The usual face-to-face stacking interaction is an offset or slipped stacking, whereby the rings are parallel displaced forming an angle of about 20° between the ring centroids (Janiak, 2000). To address this question, additional site directed mutagenesis experiments of H1040 and its surrounding residues in CDK12 C-terminal extension helix were performed. Based on the obtained results from these experiments we are convinced that the aromatic ring of H1040 will be simply displaced by the pyridine-N-oxide ring of dinaciclib rather than contributing significant stabilization of the bound state through a face-to-face stacking interaction. Thus, in contrast to flavopiridol, the histidine residue of H1040 may have no impact on the interaction between CDK12 and dinaciclib. The determined IC_{50} value of dinaciclib for CDK12 H1040G (103 nM) does not show substantial deviation towards the determined IC_{50} value of dinaciclib for CDK12 wild type (58 nM). The created impression of a conformation that favors a π-π stacking interaction and thus leading to a highly stable protein-ligand complex is misleading. There must be alternative stable binding poses resulting in specific mutual recognition based on non-covalent interactions that are present at the interface between the kinase active site and dinaciclib.

Any investigation for this requires the three-dimensional X-ray structure of dinaciclib in complex with CDK12. For further proceeding, more efforts have to be put into protein crystallization trials which is definitely the most straight forward approach to elucidate the real binding conformation of the bound compound. By this way, non-covalent interactions can be accurately visualized and analyzed by detecting amino acids essential for proper protein-ligand interaction. Afterwards, further predictions can be made for mutations that may disrupt dinaciclib binding while preserving kinase

function. Moreover, based on the fact that dinaciclib also has potent inhibitory activity against other kinases (CDK1, 2, 5, 9 and 13), close examination of the ligand binding site in the X-ray structure of CDK12-dinaciclib may provide suggestions how the compound can be further chemically modified in order to prevent off-target effects by improving target selectivity.

5.2 Interpretation of the potency of non-covalent and covalent kinase inhibitors

Non-covalent reversible small-molecule kinase inhibitors are predominately affected by ATP concentration, as they interact with their biological targets under equilibrium binding conditions, where the desired drug–protein interaction is a rapid and reversible process. The reversible kinase inhibitor has to compete constantly against intracellular ATP for the kinase active binding site. It has been demonstrated that high intracellular ATP concentrations affect the cellular potency of ATP competitive kinase inhibitors under physiological conditions (Knight and Shokat, 2005).

In contrast to reversible inhibitors, covalent irreversible kinase inhibitors interact with their biological targets by forming a covalent attachment. This irreversible mechanism of action helps to mitigate against competition by high intracellular concentrations of ATP (Yun *et al.*, 2008). As previously indicated, the competitive influence of the ATP concentration becomes less important after pre-incubation time of 30 min was applied, resulting in lower IC_{50} values. This result was also shown by Barf and Kaptein, who additionally demonstrated that the selection of the length of pre-incubation time determines the inhibitory potency of the irreversible inhibitor, unless maximal inhibition is achieved when all available target protein is occupied by the inhibitor (Barf and Kaptein, 2012). At this point, the competitive influence of ATP was abolished completely. If they are given sufficient time to react, irreversible inhibitors will provide complete and permanent blockade of their target protein. Thus, in this context it should be noted that irreversible inhibitors interact in a time-dependent manner and the reaction proceeds to completion rather than to equilibrium. Since targeting under equilibrium condition, the competitive status with ATP remains unchanged for reversible kinase inhibitors regardless of the pre-incubation time.

Based on the above mentioned facts that irreversible inhibitors are able to achieve complete inactivation of their target protein given enough time, caution should be taken with the interpretation of the observed IC_{50} values, which will be influenced by the selection of the length of the pre-incubation time. In contrast to reversible inhibitors, inactivation of the kinase by covalent inhibitors involves one further parameter in addition to the affinity of initial non-covalent binding (K_i), which

is depending on the ATP concentration. The potency of irreversible inhibitors is further governed by the rate of target inactivation (K_{inact}) through the subsequent bond formation reaction in order to form the final covalent complex (Figure 63). Thus, even without a pre-incubation period, the efficacy of covalent kinase inhibitor is significantly higher. Once the covalent attachment has taken place, return to unbound inhibitor and free target protein (k_{-2}) is very unlikely, if the bond formation is effectively irreversible. The better the juxtaposition of the two interacting partners, which is primarily influenced by the inherent reactivity of the electrophilic group as well as the distance to the nucleophilic trap, the better is the kinase inactivation step (Krippendorff *et al.*, 2009). As the kinase is effectively removed from the equilibrium with ATP and substrate over time, incubation for a different period of time would give a different IC_{50} value. Depending on the values for K_i and k_2, the IC_{50} value can change significantly when measured at different pre-incubation times, which has been shown in the previous results of the covalent CDK12 inhibitor.

Consequently, the potency of irreversible inhibitors, which usually become more potent over time, cannot be easily characterized by using conventional IC_{50} measurements that can be time dependent. The measured IC_{50} values can be misleading and cause incorrect interpretations. Instead, the K_{inact}:K_i ratio represents a more credible approach to characterize the inhibitory potency of covalent irreversible inhibitors, which is normally preferred over conventional IC_{50} values (Krippendorff *et al.*, 2009; Noe and Gilbert, 2012).

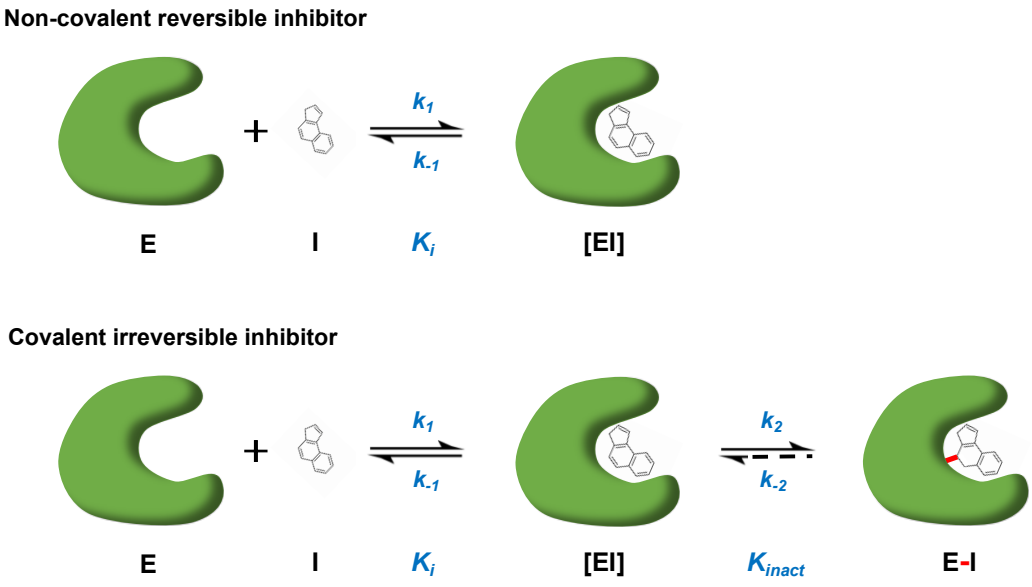


Figure 63: Comparison of the action of a target-specific inhibitor under non-covalent reversible and covalent irreversible binding condition. E: enzyme. I: inhibitor. Modified figure based on (Bauer, 2015).

5.3 Phosphorylation activities of CDK12 and CDK13 contribute to co-transcriptional RNA processing

In the validation study of distinct pre-mRNA processing factors as new CDK12 and CDK13 substrates, our findings derived from *in vitro* kinase activity assays and peptide mass finger print analyses strongly support that both transcription kinases, CDK12 and CDK13 in complex with Cyclin K, directly phosphorylate the substrates CDC5L, SF3B1, CSTF64 and SPT6H, determined by the *in vivo* phosphoproteomic analysis.

In vitro kinase time course measurements have shown that kinase-mediated phosphorylation resulted in a higher rate of [^{32}P]-incorporation for CDK12 than for CDK13 with respect to the substrates CDC5L, SF3B1 and CSTF64. It can be suggested that CDK12 has a stronger effect against these substrates than CDK13 by phosphorylating more sites. However, the stronger effect of CDK12 may not only based on its specific phosphorylation potency against these substrates, but it is also possible that there is a substantial discrepancy between the kinase activity of the purified recombinant complexes of CDK12 and CDK13, depending on the rate of the T-loop phosphorylation. Thus, it is possible that the process of substrate phosphorylation was additional influenced by a kinase activity-dependent manner, as the concentration of active CDK12 and CDK13 was different in the applied reaction mixture. Nevertheless, the measured radioactive counts in presence of CDK13 was significantly beyond the values measured in the control experiments, confirming CDK13-mediated phosphorylation of CDC5L, SF3B1 and CSTF64.
The measurement with SPT6H as CDK12 and CDK13 substrate, however, showed no increase in radioactive counts by kinases CDK12 or CDK13 if compared to the control measurement without kinase. In general, it has been noted that the values of the radioactive counts obtained in this time course experiment were more than 10-fold reduced after 6 h compared to CDC5L substrate phosphorylation by CDK12. Thus, SPT6H cannot be confirmed as a phosphorylation target of CDK12 and CDK13 by performing *in vitro* kinase activity assay. One explanation of this discrepancy with the results from *in vivo* phosphoproteomic analysis could be that full length SPT6H protein in its native cellular protein complex environment may be required to allow specific phosphorylation by the kinases. However, the truncated SPT6H protein was phosphorylated by CDK9/CycT1, indicating in particular that the recombinant truncated SPT6H protein is not an artificial product, but was able to be phosphorylated. This finding led to the assumption that the performed *in vitro* kinase activity assay may not be sufficiently sensitive for reliable detection if only one possible phosphorylation site for CDK12 and CDK13 is present as predicted by the *in vivo* phosphoproteomic analysis. *In vitro* time course measurements of CDK12 with CDC5L and SF3B1, where numerous

phosphorylation sites were detected by peptide mass finger print analysis for each of them, resulted in 3- to 6-fold higher radioactive counts after 6 h compared to CSTF64, which was also predicted to have only one phosphorylation site for CDK12 and CDK13. Thus, the number of the phosphorylation sites might be a quantitative measure for the level of radioactive counts measured in the *in vitro* kinase activity assay. By performing peptide mass finger print analysis of SPT6H, this assumption has been confirmed by detection of the one supposed phosphorylation site. Based on this result, SPT6H can be considered as a direct phosphorylation target of CDK12 and CDK13.

Phosphorylation of SPT6H by the kinase CDK9 and Cyclin T1 as part of the positive transcription elongation factor b (P-TEFb), which is required for the release of paused RNAPII II into elongation activation (Marshall and Price, 1995; Wei *et al.*, 1998), was also determined by *in vitro* phosphorylation site mapping (Vos *et al.*, 8/2018b). Although the formation of an activated RNAPII II elongation complex *in vitro* requires the presence of SPT6H next to the kinase function of P-TEFb and the elongation factors PAF1 complex, activation of SPT6H is not initiated directly via phosphorylation by P-TEFb, but depending on the phosphorylation of the RNAPII II CTD linker by P-TEFb. After phosphorylation of the CTD linker at specific sites by P-TEFb, SPT6H binds directly to the phosphorylated CTD linker and subsequently exercises its function to active elongation (Vos *et al.*, 8/2018b). Thus, it is unlikely that phosphorylation of SPT6H by CDK9 is associated with cellular importance. Although the findings reported in the current study of Krajewska et al. have shown that inhibition of CDK12 results in a transcriptional elongation defect corresponding to those reported with inhibition of CDK9, where increased RNAPII II pausing leads to a defect in elongation (Gressel *et al.*, 2017; Shao and Zeitlinger, 2017), there might be no specific relationship between phosphorylation of SPT6H by CDK12 and CDK9.

Despite the demonstration that CSTF64 was a target protein of CDK12 and CDK13 by significant phosphorylation rates measured over time, subsequent peptide mass fingerprint analysis did not confirm this result by identifying a phosphorylation site of the recombinant CSTF64 protein construct. This result could be due to various circumstances like unfavorable proteolytic fragment digest, poor performance in the ESI-MS detection analytics or low protein stability occurred under the applied conditions. All of these factors may impair the quality of the identification of protein modifications. Because of failure in detection of certain parts of the protein sequence due to inadequate size or unfavorable ionization properties of the digested peptide fragments, it is not unusual that digestion of proteins results in a loss of information, such as the presence and connectivity of PTMs (Lanucara and Eyers, 2013). Another issue relates to the large size, poor solubility and poor ionization efficiency of intact proteins, which often lead to insufficient separation resolution and low MS detection sensitivity, thereby represent a big challenge for the ESI-MS analysis (Gregorich and Ge, 2014).

In vivo, CDK12 was identified as the predominant kinase mediating the transcriptional effects of THZ531 in treated cells, as premature cleavage and polyadenylation (PCPA) was determined as the main defect resulting from THZ531 treatment, an outcome that is mediated primarily by the targeting of CDK12. In contrast, less profound effects were observed for CDK13 (Krajewska *et al.*, 2019). However, our findings derived from the *in vitro* experiments have shown that both CDK12 and CDK13 induced the phosphorylation of pre-mRNA processing proteins, which have established roles in co-transcriptional RNA processing by RNAPII. Based on the findings that both kinases appear to have the same phosphorylation target and additional inhibition by THZ351 indicated the same outcome, the distinction among the phosphorylation targets was not as clear-cut by using previous *in vitro* methods. Thus, further studies are required in order to resolve this overlap between CDK12 and CDK13.

5.4 Implication and substrate specificity of CDK13 in melanoma

Based on our data obtained from structural and biochemical analyses of melanoma-associated CDK13 mutations we confirmed that the identified mutations near the ATP-binding site of CDK13 are not nonsense-mutations as they inactivate the kinase activity dramatically *in vitro*. Together with the findings that CDK13 downregulation or mutation correlated with decreased overall survival derived by analysis of human TCGA data (Cerami *et al.*, 2012; Gao *et al.*, 2013) as well as expression of these CDK13 mutations in zebrafish led to accelerated melanoma onset (Insco *et al.*, submitted), it can be supposed that the identified CDK13 patient mutations may represent the leading cause for the development of cutaneous melanoma by inactivating the kinase activity of CDK13.

Using *in vivo* experiments, Insco and co-workers further identified ZC3H14 as a CDK13 binding protein and suggested that ZC3H14 is also a CDK13 phosphorylation substrate in human melanoma cells. Moreover, they identified Cyclin T1 but not Cyclin K as the cyclin binding partner for CDK13 in human melanoma, which was also demonstrated to be required for the CDK13-mediated melanomagenesis (Insco *et al.*, submitted). In contrast, our data from *in vitro* kinase activity assay did not indicate phosphorylation of purified recombinant truncated ZC3H14 proteins by CDK13 in complex with Cyclin K. In the first instance this outcome can be considered as not sufficiently powerful, as truncated proteins may not adequately reflect the actual biological state. The full length protein construct may be required to achieve a proper phosphorylation by CDK13/CycK. However, by using CDK9 in complex with Cyclin T1, phosphorylation activity was able to be measured with

ZC3H14 (312-736) but it was still missing in case of the zinc finger domain only construct, ZC3H14 (591-736). On the one hand this result reinforced the predicted phosphorylation site at S475 of ZC3H14, since only ZC3H14 (312-736), but not ZC3H14 (591-736) was phosphorylated by CDK9/CycT1. On the other hand, based on this finding, supposed phosphorylation of the SP-motif at position 620 of the zinc finger domain could not be confirmed.

Furthermore, this result additional supported the hypothesis that in melanoma cells Cyclin T1 may be required to generate an active CDK13 complex in order to phosphorylate ZC3H14, which was also determined as a CDK13 phosphorylation substrate in human melanoma. It is assumed that the cyclin subunit determines the specificity for substrates and regulation factors recognition by contributing interaction surfaces, mediating substrate binding across a shallow, hydrophobic peptide binding groove on the cyclin subunit (Lee *et al.*, 2007). In association with CDKs, the different cyclin-CDK complexes display differences in substrate specificities in order to regulate cell cycle progression (Adams *et al.*, 1996; Peeper *et al.*, 1993) and gene transcription, as in case of CDK9/CycT1 where Cyclin T1 binds directly to the HIV-1 TAT protein (Wei *et al.*, 1998) and Hexim1 (Dames *et al.*, 2007; Michels *et al.*, 2003; Schulte *et al.*, 2005). Apart from this, based on the result derived from the *in vivo* IP experiment, it can be assumed that Cyclin T1 is more highly expressed in melanoma cells than Cyclin K. As the expression levels of the various interacting partners have a direct influence on the formation of protein complexes (Jansen *et al.*, 2002), the supposed overexpression of Cyclin T1 may significantly enhance the formation of the CDK13/CycT1 complex in melanoma cells. In this context, the possibility that CDK13 uses different cyclin binding partners to enable distinct cellular processes by phosphorylating a range of key protein substrates in a site and time-specific manner is rather high. To assess the global importance of the cyclin specificity in CDK13-mediated melanomagenesis, the next logical step is to confirm the phosphorylation activity of a stable recombinant CDK13/CycT1 complex against ZC3H14 protein.

Together, our studies support the existence of a new layer of regulation of CDK13 by which both the presence and the intrinsic biochemical properties of different cyclins facilitate the correct kinase substrate phosphorylation. Beyond that, this finding might contribute to an in-depth understanding of the specificity and regulation of the entire family of the transcription kinases involving in various cellular processes. It is well accepted in the cell cycle field that the cyclin contributes the interaction surfaces for substrates and regulators (Jonathon Pines, 1995; Satyanarayana and Kaldis, 2009). It is possible that different cyclins bind to the transcription kinases as well and thus determine the specificity for both substrates and regulation factors, by this way allowing the kinases to adopt various regulation functions in the process of transcription.

Cell cycle-related kinase – CDK14

5.5 Regulatory interplay between CDK14, Cyclin Y, and 14-3-3 in mediating CDK14/CycY association

In the current study, 14-3-3 proteins were accidentally co-purified as binding proteins of the CDK14/CycY kinase complex during the search of an active CDK14 kinase. It has been demonstrated subsequently that binding of 14-3-3 not only significantly enhanced the association between Cyclin Y and CDK14 as described previously (Li *et al.*, 2014), but was also required for the full kinase activity of the CDK14 kinase. Exactly the same CDK14/CycY complex but without 14-3-3 association exhibited only one third of the kinase activity toward the IBER-IRStide substrate. Here, a doublet band of 14-3-3 has been detected in association with CDK14/CycY. It is often believed that the one band represents exogenously expressed 14-3-3 whereas the other band represents co-purified endogenous 14-3-3 proteins (Al-Hakim, 2005; Gannon-Murakami and Murakami, 2002). This point of view, however, does not apply in our case, since 14-3-3 was not expressed as recombinant protein. Thus, both protein bands represent co-purified endogenous 14-3-3 proteins. It has also been suggested that each of the two bands might represent a modified form or a distinct isoform of 14-3-3 as the molecular weights of the 14-3-3 isoforms are very similar (Kumagai *et al.*, 1998). As self-dimerization is important for the ability of 14-3-3 proteins to bind phosphoprotein targets *in vivo*, whereas the seven 14-3-3 isoforms exist in a dynamic balance of homo- or heterodimers (Benzinger *et al.*, 2005; Messaritou *et al.*, 2010; Shen *et al.*, 2003; Verdoodt *et al.*, 2006), it could therefore be reasonably assumed that the doublet band represents a 14-3-3 heterodimer formed to target the kinase complex (Figure 64).

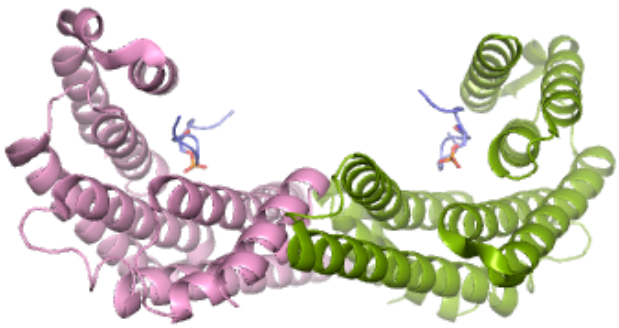


Figure 64: Structure of a 14-3-3 dimer in complex with phosphorylated 14-3-3 binding motif (PDB: 6A5S;(Xu *et al.*, 2019)**.** Each monomer of 14-3-3 is shown in a different colour. The phosphorylated 14-3-3 binding motif (VTSSpSCPADLT; shown in purple with the phosphate group highlighted in orange) binds in an extended conformation to 14-3-3 within the conserved groove of each monomer.

This assumption can be supported by the finding that if 14-3-3 was alone without the existence of any interaction partners, only a single protein band of 14-3-3 should be visualized by the gel electrophoresis. However, in conjunction with the presence of Cyclin Y–14-3-3 association the doublet band of 14-3-3 appeared. In addition, the doublet band was allocated to different 14-3-3 isoforms [epsilon (ϵ) , beta (β) , eta (η)], exactly the same Cyclin Y targeting 14-3-3 isoforms also identified by Li et al. *in vitro* and *in vivo* (Li *et al.*, 2014). Besides, unlike some isoforms preferring to adopt strictly homodimeric interactions (Benzinger *et al.*, 2005; Wilker *et al.*, 2005), 14-3-3ϵ has been implicated in preferential heterodimer formation among others with 14-3-3β and 14-3-3η (Chaudhri *et al.*, 2003). Heterodimer of 14-3-3β and 14-3-3ϵ is also a well known regulator, required for example to regulate aldosterone of the epithelial sodium channel (Liang *et al.*, 2008).

Our data further provided the finding that 14-3-3 proteins could not be co-purified with CDK14 (*Sf9)* but only with Cyclin Y expressed in *Sf9*. One might think at first that as 14-3-3 are not naturally occurring proteins in *E. coli* cells (based on UniProt research), co-purification of endogenous 14-3-3 with Cyclin Y from *E. coli* cannot be achieved at all. However, pull-down assays of Cyclin Y (*E. coli*) with different recombinant 14-3-3 isoforms clearly confirmed the missing interaction between 14-3-3 and Cyclin Y purified from *E. coli*. These findings indicated that Cyclin Y is the predominant interaction partner of 14-3-3, whereby posttranslational modifications of the cyclin are required for proper targeting. The latter is supported on the one hand by the fact that the vast majority of cellular-binding partners interact with 14-3-3 proteins in a phosphorylation-dependent manner, as 14-3-3 binding is typically primarily phosphorylation-dependent (Pozuelo Rubio *et al.*, 2004). On the other hand, the previously published finding that S100 and S326 of Cyclin Y are crucial for the interaction between Cyclin Y and 14-3-3 (Li *et al.*, 2014) further indicates the need of posttranslational phosphorylation which can only be provided by the insect cell expression system. This finding was confirmed by performing mutagenesis experiments. We have shown, that mutation of S100 to non-phosphorylatable alanine completely abolished 14-3-3 binding with Cyclin Y as well as the association of Cyclin Y with CDK14, additionally indicating the requirement of 14-3-3 binding for the proper association of Cyclin Y with CDK14. Phosphorylations of Cyclin Y at S100 and S326 were also detected by our peptide mass finger print analysis *in vitro*, but only when Cyclin Y was expressed in *Sf9* cells, according to the above mentioned findings.

To understand how 14-3-3 proteins enhance the kinase-cyclin association and thus facilitate the kinase activity by forming a heterodimer, the modes of action of 14-3-3 interaction with target proteins have to be first taken into consideration. In general, 14-3-3 interaction with target proteins can be classified into three major modes of action: (i) directed conformational changes, (ii) masking of a specific region on the target, and (iii) scaffolding of two proteins (Braselmann and McCormick,

1995; Bridges and Moorhead, 2005; Chiang *et al.*, 2001; Tzivion *et al.*, 2001). As 14-3-3 dimer was only present in association with Cyclin Y, it can be supposed in the first instance that 14-3-3 binds to Cyclin Y and thereby induces conformational changes in the target protein, which facilitates the association between kinase and cyclin (Figure 65A). In addition, it is also likely that 14-3-3 serves as a scaffold protein which stabilizes the CDK14 and Cyclin Y interaction after targeting to Cyclin Y first (Figure 65B).

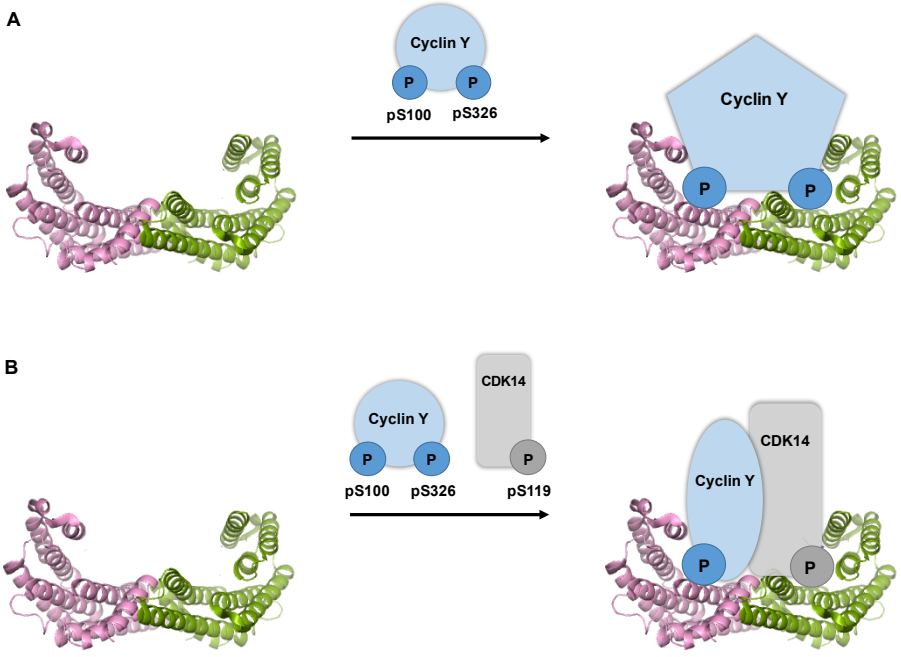


Figure 65: Possible regulatory role of 14-3-3 dimer contributing to CDK14/CycY association. (**A)** Directed conformational changes in Cyclin Y. (**B)** Scaffolding of CDK14 and Cyclin Y. Structure of 14-3-3 dimer is based on 14-3-3 gamma (PDB: 6A5S; (Xu *et al.*, 2019).

However, several points speak against the role of 14-3-3 as a phosphorylation-dependent scaffold protein by targeting both CDK14 and CyclinY. Firstly, based on our data, 14-3-3 proteins were not able to be co-purified with CDK14 alone and secondly, Cyclin Y in association with 14-3-3 did not interact with CDK14. Although particular 14-3-3 isoforms have been previously proven to bind to CDK14 as well via pS119 of the kinase (Gao et al., 2006), abolishment of S119 on CDK14 did not impair the function of 14-3-3 proteins in facilitating CDK14/CycY association. Only mutation of S100 and S326 of Cyclin Y affected the association between kinase and cyclin adversely (Li et al., 2014). Thus, potential interaction between 14-3-3 and CDK14 does not contribute to CDK14/CycY association, which in turn means that 14-3-3 may not act as a scaffold protein in this context.

In addition, the presence of two 14-3-3 phosphorylation binding sites on Cyclin Y, S100 and S326, represents the possibility for 14-3-3 dimer to directly interact only with Cyclin Y as each monomer is able to bind one discrete phosphoserine or -threonine peptide independently within its conserved groove. (Liu *et al.*, 1995; Rittinger *et al.*, 1999). Here, 14-3-3 binding to Serotonin–N-acetyltransferase protein provides a bona fide example of the binding mechanism supposed above in which each of the separate subunit of the 14-3-3 dimer bind to a phosphorylated site at two separate regions on one protein (Obsil *et al.*, 2001). By the use of the homology modeled Cyclin Y structure, illustration of S100 indicates its potential location on the protein surface, allowing the access for 14-3-3 binding (Figure 66). The second phospho-site at S326 cannot be displayed due to the limited length of the amino acid sequence of the template protein (PHO85 Cyclin PHO80). Nevertheless, based on the position of the C-terminal ending of the template protein displayed in the model (shown in lines), it can be assumed that S326 is located at a separate region not in close proximity to S100, thus encouraging targeting by each of the subunit of the 14-3-3 dimer.

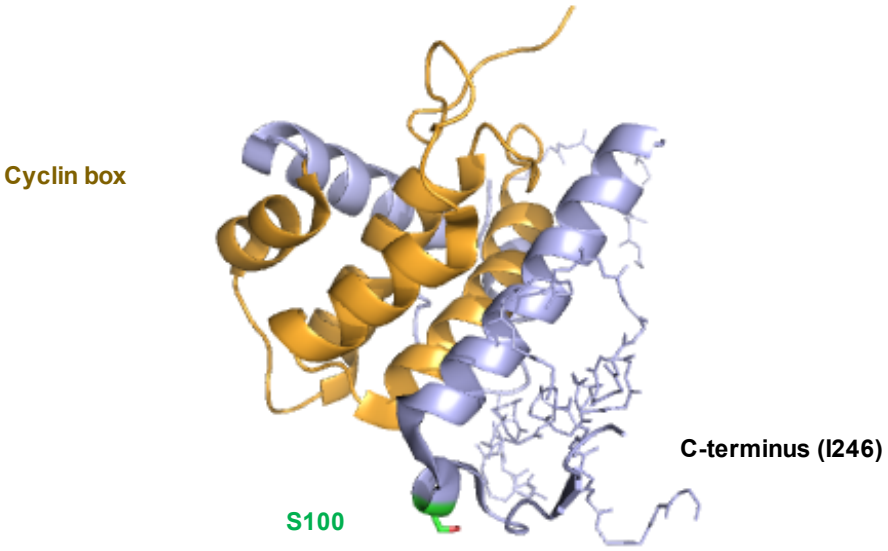


Figure 66: Illustration of the potential location of S100 and S326 on the homology model of Cyclin Y. S100 is located on the protein surface (shown in green). The cyclin box is shown in gold and the C-terminal part of the template protein (PHO85 Cyclin PHO80) is displayed in lines.

Taking all of the described indications together, it is highly possible that 14-3-3 contributes to CDK14/CycY association by only bind to Cyclin Y as heterodimer. Due to the highly α-helical nature of the 14-3-3 proteins, the global rigidity of this protein leads to deformation of the target protein upon binding with little or no change in the structure of the 14-3-3 dimer (Yaffe, 2002). It can be suggested that binding of 14-3-3 dimer at the phospho-sites enclosing the cyclin box (S100 is located to the N-terminal and S326 is located to the C-terminal of the cyclin box) may allow the cyclin box

to be exposed more outward (shown in gold in figure 66), resulting in substantial rearrangement of the active site. By this way, 14-3-3 binding could alter the protein surface of Cyclin Y involved in protein-protein interactions and thus increases the mutual access between the cyclin box of Cyclin Y and the 'PFTAIRE' motif of CDK14. This in turn might result in a higher binding affinity of Cyclin Y to CDK14, leading to a full kinase activity at the end. This 14-3-3 driven binding mechanism might provide an explanation of how Cyclin Y is able to interact and activate CDK14 with only a single cyclin box as conventional cyclins usually require two such boxes for activity. One is typically required for CDK binding and activation and the other is necessary for the proper folding of the cyclin protein itself (Noble *et al.*, 1997).

Our data also suggested that besides the requirement of 14-3-3 binding, CDK14/CycY association is additionally regulated through phosphorylation of Cyclin Y by CDK14 as only the presence of 14-3-3 in association with Cyclin Y did not seem to be sufficient to form a CDK14/CycY complex. This assumption can be supported by the previously published data that Cyclin Y exhibited enhanced phosphorylation *in vivo* through its association with CDK14, indicating that CDK14 might be required for the phosphorylation of Cyclin Y (Jiang *et al.*, 2009). Additionally, a CDK-dependent phosphorylation site on Cyclin Y was also identified for CDK16 (Shehata *et al.*, 2015), which is closely related to CDK14 (Lazzaro *et al.*, 1997). Although our *in vitro* peptide analysis favored S33 and T331 as CDK14-dependent phosphorylation sites, early study has demonstrated that exogenously expressed Cyclin Y only displayed a hybridization band with anti-phospho-serine antibody, but not with anti-phospho-threonine (Jiang *et al.*, 2009). This finding suggested that Cyclin Y was only phosphorylated at serine residues *in vivo*. Therefore, it is likely that *in vitro* phosphorylation at T331 belongs to unspecific nature, suggesting S33 as potential CDK14-dependet phosphorylation site, which is required for CDK14/CycY association simultaneously with the binding of the 14-3-3 dimer. To confirm this assumption, the next step will be the generation of a phospho-deficient S33A mutant of Cyclin Y in order to examine if this phospho-site is crucial for the CDK14/CycY association.

Together, the present work has emphasized the regulatory interplay between CDK14, Cyclin Y and 14-3-3 proteins in a phosphorylation-dependent manner. It provides a possible new mechanistic insight into activation of CDK14, which is mediated through interaction with the specific phosphorylated form of Cyclin Y in complex with 14-3-3 dimer. A model illustrating this supposed regulatory interplay between CDK14, Cyclin Y and 14-3-3 in mediating CDK14/CycY association is shown in figure 67. As CDK14 is involved in a broad spectrum of diseases by promoting tumor cell proliferation, migration and invasion in different types of cancer such as breast cancer (Gu *et al.*, 2015; Wang *et al.*, 2017), lung cancer (Liu *et al.*, 2016), liver cancer (Leung *et al.*, 2011), gastric

cancer (Yang *et al.*, 2015), pancreatic cancer (Zheng *et al.*, 2015) and ovarian cancer (Ou-Yang *et al.*, 2017), a better understanding of the regulatory interplay might open new paths to target CDK14 in the field of cancer treatments.

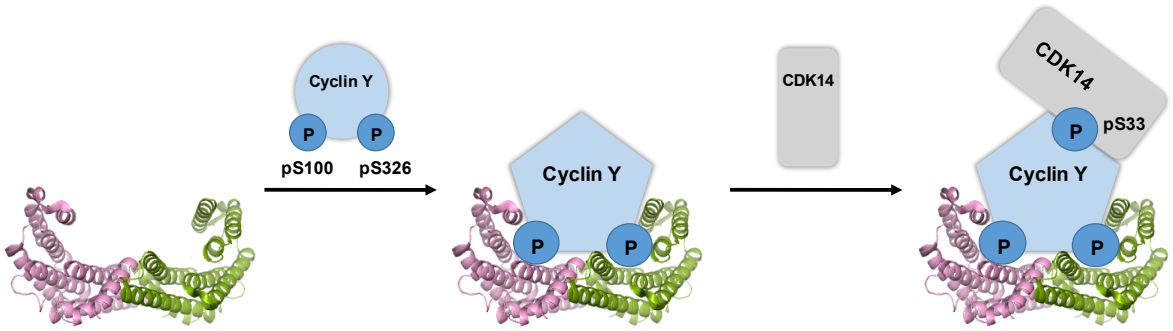


Figure 67: Proposed model of the regulatory interplay between CDK14, Cyclin Y and 14-3-3 in mediating CDK14/CycY association. Structure of 14-3-3 dimer is based on 14-3-3 gamma (PDB: 6A5S; (Xu *et al.*, 2019).

6 References

Adams, P.D., Sellers, W.R., Sharma, S.K., Wu, A.D., Nalin, C.M. and Kaelin, W.G. 1996. Identification of a cyclin-cdk2 recognition motif present in substrates and p21-like cyclin-dependent kinase inhibitors. *Mol. Cell. Biol.*, **16**: 6623–6633.

Adelman, K. and Lis, J.T. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.*, **13**: 720–731.

Ahmad, K. and Henikoff, S. 2002. Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **99 Suppl 4**: 16477–16484.

Ahn, S.H., Kim, M. and Buratowski, S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell*, **13**: 67–76.

Aitken, A. 1995. 14-3-3 proteins on the MAP. *Trends Biochem. Sci.*, **20**: 95–97.

Akhtar, Md.S., Heidemann, M., Tietjen, J.R., Zhang, D.W., Chapman, R.D., Eick, D. and Ansari, A.Z. 2009. TFIIH Kinase Places Bivalent Marks on the Carboxy-Terminal Domain of RNA Polymerase II. *Molecular Cell*, **34**: 387–393.

Akoulitchev, S., Chuikov, S. and Reinberg, D. 2000. TFIIH is negatively regulated by cdk8-containing mediator complexes. *Nature*, **407**: 102–106.

Al-Hakim, A.K. 2005. 14-3-3 cooperates with LKB1 to regulate the activity and localization of QSK and SIK. *Journal of Cell Science*, **118**: 5661–5673.

Allison, L.A., Wong, J.K., Fitzpatrick, V.D., Moyle, M. and Ingles, C.J. 1988. The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. *Mol. Cell. Biol.*, **8**: 321–329.

Alvarez, S. 2013. A cartography of the van der Waals territories. *Dalton Transactions*, **42**: 8617.

Baek, K., Brown, R.S., Birrane, G. and Ladias, J.A.A. 2007. Crystal Structure of Human Cyclin K, a Positive Regulator of Cyclin-dependent Kinase 9. *Journal of Molecular Biology*, **366**: 563–573.

Ball, L.J., Kühne, R., Schneider-Mergener, J. and Oschkinat, H. 2005. Recognition of Proline-Rich Motifs by Protein-Protein-Interaction Domains. *Angewandte Chemie International Edition*, **44**: 2852–2869.

Barf, T. and Kaptein, A. 2012. Irreversible Protein Kinase Inhibitors: Balancing the Benefits and Risks. *Journal of Medicinal Chemistry*, **55**: 6243–6262.

Bartel, D.P. 2009. MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, **136**: 215–233.

Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T. and Greenleaf, A.L. 2010. CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes & Development*, **24**: 2303–2316.

Baskaran, R., Chiang, G.G. and Wang, J.Y. 1996. Identification of a binding site in c-Ab1 tyrosine kinase for the C-terminal repeated domain of RNA polymerase II. *Molecular and Cellular Biology*, **16**: 3361–3369.

Baskaran, R., Dahmus, M.E. and Wang, J.Y. 1993. Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proceedings of the National Academy of Sciences*, **90**: 11167–11171.

Bataille, A.R., Jeronimo, C., Jacques, P.-É., Laramée, L., Fortin, M.-È., Forest, A., Bergeron, M., Hanes, S.D. and Robert, F. 2012. A Universal RNA Polymerase II CTD Cycle Is Orchestrated by Complex Interplays between Kinase, Phosphatase, and Isomerase Enzymes along Genes. *Molecular Cell*, **45**: 158–170.

Bauer, R.A. 2015. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discovery Today*, **20**: 1061–1073.

Baumann, M., Pontiller, J. and Ernst, W. 2010. Structure and basal transcription complex of RNA polymerase II core promoters in the mammalian genome: an overview. *Mol. Biotechnol.*, **45**: 241–247.

Baumli, S., Lolli, G., Lowe, E.D., Troiani, S., Rusconi, L., Bullock, A.N., Debreczeni, J.É., Knapp, S. and Johnson, L.N. 2008. The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. *The EMBO Journal*, **27**: 1907–1918.

Becker, P.B. and Hörz, W. 2002. ATP-dependent nucleosome remodeling. *Annu. Rev. Biochem.*, **71**: 247–273.

Bellier, S., Dubois, M.F., Nishida, E., Almouzni, G. and Bensaude, O. 1997. Phosphorylation of the RNA polymerase II largest subunit during Xenopus laevis oocyte maturation. *Molecular and Cellular Biology*, **17**: 1434–1440.

Bentley, D.L. 2014. Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*, **15**: 163–175.

Benzinger, A., Popowicz, G.M., Joy, J.K., Majumdar, S., Holak, T.A. and Hermeking, H. 2005. The crystal structure of the non-liganded 14-3-3sigma protein: insights into determinants of isoform specific ligand binding and dimerization. *Cell Res.*, **15**: 219–227.

Bergfors, T.M. (ed). 1999. *Protein crystallization: techniques, strategies, and tips: a laboratory manual*.

Berro, R., Pedati, C., Kehn-Hall, K., Wu, W., Klase, Z., Even, Y., Geneviere, A.-M., Ammosova, T., Nekhai, S. and Kashanchi, F. 2008. CDK13, a New Potential Human Immunodeficiency Virus Type 1 Inhibitory Factor Regulating Viral mRNA Splicing. *Journal of Virology*, **82**: 7155–7166.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L. and Schwede, T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **42**: W252–W258.

Bieniossek, C., Imasaki, T., Takagi, Y. and Berger, I. 2012. MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem. Sci.*, **37**: 49–57.

BioMol.net. 2019.

Blazek, D. 2012. The cyclin K/Cdk12 complex: An emerging new player in the maintenance of genome stability. *Cell Cycle*, **11**: 1049–1050.

Blazek, D., Kohoutek, J., Bartholomeeusen, K., Johansen, E., Hulinkova, P., Luo, Z., Cimermancic, P., Ule, J. and Peterlin, B.M. 2011. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes & Development*, **25**: 2158–2172.

Bonnet, F., Vigneron, M., Bensaude, O. and Dubois, M.-F. 1999. Transcription-independent phosphorylation of the RNA polymerase II C-terminal domain (CTD) involves ERK kinases (MEK1/2). *Nucleic Acids Research*, **27**: 4399–4404.

Bösken, C.A., Farnung, L., Hintermair, C., Merzel Schachter, M., Vogel-Bachmayr, K., Blazek, D., Anand, K., Fisher, R.P., Eick, D. and Geyer, M. 2014. The structure and substrate specificity of human Cdk12/Cyclin K. *Nature Communications*, **5**.

Bostwick, B.L., McLean, S., Posey, J.E., Streff, H.E., Gripp, K.W., Blesson, A., Powell-Hamilton, N., Tusi, J., Stevenson, D.A., Farrelly, E., Hudgins, L., Yang, Y., Xia, F., Wang, X., Liu, P., Walkiewicz, M., McGuire, M., Grange, D.K., Andrews, M.V., *et al.* 2017. Phenotypic and molecular characterisation of CDK13-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders. *Genome Medicine*, **9**.

Bouwman, P. and Jonkers, J. 2014. Molecular Pathways: How Can BRCA-Mutated Tumors Become Resistant to PARP Inhibitors? *Clinical Cancer Research*, **20**: 540–547.

Braselmann, S. and McCormick, F. 1995. Bcr and Raf form a complex in vivo via 14-3-3 proteins. *EMBO J.*, **14**: 4839–4848.

Brenner, S. and Crick, F. 1957. CSHL Archives Repository | On Protein Synthesis.

Bridges, D. and Moorhead, G.B.G. 2005. 14-3-3 proteins: a number of functions for a numbered protein. *Sci. STKE*, **2005**: re10.

Brookes, E. and Pombo, A. 2009. Modifications of RNA polymerase II are pivotal in regulating gene expression states. *EMBO reports*, **10**: 1213–1219.

Brown, N., Noble, M., Endicott, J., Garman, E., Wakatsuki, S., Mitchell, E., Rasmussen, B., Hunt, T. and Johnson, L. 1995. The crystal structure of cyclin A. *Structure*, **3**: 1235–1247.

Buratowski, S. 2009. Progression through the RNA Polymerase II CTD Cycle. *Molecular Cell*, **36**: 541–546.

Buratowski, S. 2003. The CTD code. *Nature Structural & Molecular Biology*, **10**: 679–680.

Cao, L., Chen, F., Yang, X., Xu, W., Xie, J. and Yu, L. 2014. Phylogenetic analysis of CDK and cyclin proteins in premetazoan lineages. *BMC Evolutionary Biology*, **14**: 10.

Carneiro, T., Krepischi, A., Souza da Costa, S., Tojal da Silva, I., Vianna-Morgante, A., Valieris, R., Ezquina, S., Bertola, D., Otto, P. and Rosenberg, C. 2018. Utility of trio-based exome sequencing in the elucidation of the genetic basis of isolated syndromic intellectual disability: illustrative cases. *The Application of Clinical Genetics*, **Volume 11**: 93–98.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A.P., Sander, C. and Schultz, N. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*, **2**: 401–404.

Chao, D.M., Gadbois, E.L., Murray, P.J., Anderson, S.F., Sonu, M.S., Parvin, J.D. and Young, R.A. 1996. A mammalian SRB protein associated with an RNA polymerase II holoenzyme. *Nature*, **380**: 82–85.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E. and Eick, D. 2007. Transcribing RNA Polymerase II Is Phosphorylated at CTD Residue Serine-7. *Science*, **318**: 1780–1782.

Chaudhri, M., Scarabel, M. and Aitken, A. 2003. Mammalian and yeast 14-3-3 isoforms form distinct patterns of dimers in vivo. *Biochem. Biophys. Res. Commun.*, **300**: 679–685.

Chen, H.-H., Wang, Y.-C. and Fann, M.-J. 2006. Identification and Characterization of the CDK12/Cyclin L1 Complex Involved in Alternative Splicing Regulation. *Molecular and Cellular Biology*, **26**: 2736–2745.

Chen, H.-H., Wong, Y.-H., Geneviere, A.-M. and Fann, M.-J. 2007. CDK13/CDC2L5 interacts with L-type cyclins and regulates alternative splicing. *Biochemical and Biophysical Research Communications*, **354**: 735–740.

Chiang, C.W., Harris, G., Ellig, C., Masters, S.C., Subramanian, R., Shenolikar, S., Wadzinski, B.E. and Yang, E. 2001. Protein phosphatase 2A activates the proapoptotic function of BAD in interleukin- 3-dependent lymphoid cells by a mechanism requiring 14-3-3 dissociation. *Blood*, **97**: 1289–1297.

Chilà, R., Guffanti, F. and Damia, G. 2016. Role and therapeutic potential of CDK12 in human cancers. *Cancer Treatment Reviews*, **50**: 83–88.

Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev.*, **11**: 3319–3326.

Cobb, M. 2017. 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol.*, **15**: e2003243.

Colgan, D.F. and Manley, J.L. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.*, **11**: 2755–2766.

Corden, J.L. 1990. Tails of RNA polymerase II. *Trends Biochem. Sci.*, **15**: 383–387.

Corden, J.L., Cadena, D.L., Ahearn, J.M. and Dahmus, M.E. 1985. A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proceedings of the National Academy of Sciences*, **82**: 7934–7938.

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J.F. and Vannini, A. 2008. Structure of eukaryotic RNA polymerases. *Annu Rev Biophys*, **37**: 337–352.

Czudnochowski, N., Bösken, C.A. and Geyer, M. 2012. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nature Communications*, **3**.

Dai, Q., Lei, T., Zhao, C., Zhong, J., Tang, Y., Chen, B., Yang, J., Li, C., Wang, S., Song, X., Li, L. and Li, Q. 2012. Cyclin K-containing Kinase Complexes Maintain Self-renewal in Murine Embryonic Stem Cells. *Journal of Biological Chemistry*, **287**: 25344–25352.

Dames, S.A., Schonichen, A., Schulte, A., Barboric, M., Peterlin, B.M., Grzesiek, S. and Geyer, M. 2007. Structure of the Cyclin T binding domain of Hexim1 and molecular basis for its recognition of P-TEFb. *Proceedings of the National Academy of Sciences*, **104**: 14312–14317.

Datta, S.R., Katsov, A., Hu, L., Petros, A., Fesik, S.W., Yaffe, M.B. and Greenberg, M.E. 2000. 14-3-3 proteins and survival kinases cooperate to inactivate BAD by BH3 domain phosphorylation. *Mol. Cell*, **6**: 41–51.

Davidson, G. and Niehrs, C. 2010. Emerging links between CDK cell cycle regulators and Wnt signaling. *Trends in Cell Biology*, **20**: 453–460.

Davidson, G., Shen, J., Huang, Y.-L., Su, Y., Karaulanov, E., Bartscherer, K., Hassler, C., Stannek, P., Boutros, M. and Niehrs, C. 2009. Cell cycle control of wnt receptor activation. *Dev. Cell*, **17**: 788–799.

Day, P.J., Cleasby, A., Tickle, I.J., O'Reilly, M., Coyle, J.E., Holding, F.P., McMenamin, R.L., Yon, J., Chopra, R., Lengauer, C. and Jhoti, H. 2009. Crystal structure of human CDK4 in complex with a D-type cyclin. *Proceedings of the National Academy of Sciences*, **106**: 4166–4170.

De Bondt, H.L., Rosenblatt, J., Jancarik, J., Jones, H.D., Morgant, D.O. and Kim, S.-H. 1993. Crystal structure of cyclin-dependent kinase 2. *Nature*, **363**: 595–602.

Dessau, M.A. and Modis, Y. 2011. Protein Crystallization for X-ray Crystallography. *Journal of Visualized Experiments*, doi: 10.3791/2285.

Devaiah, B.N., Lewis, B.A., Cherman, N., Hewitt, M.C., Albrecht, B.K., Robey, P.G., Ozato, K., Sims, R.J. and Singer, D.S. 2012. BRD4 is an atypical kinase that phosphorylates Serine2 of the RNA Polymerase II carboxy-terminal domain. *Proceedings of the National Academy of Sciences*, **109**: 6927–6932.

Dhanasekaran, N. and Reddy, E.P. 1998. Signaling by dual specificity kinases. *Oncogene*, **17**: 1447–1455.

Dixon-Clarke, S.E., Galan Bartual, S., Elkins, J., Savitsky, P., Kopec, J., Mackenzie, A., Tallant, C., Heroven, C., Burgess-Brown, N., von Delft, F., Arrowsmith, C.H., Edwards, A.M., Bountra, C. and Bullock, A. 2016. Crystal structure of the PCTAIRE1 kinase in complex with inhibitor. *(:unav)*, doi: 10.2210/pdb5g6v/pdb.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder,

M., Kokocinski, F., Abdelhamid, R.F., *et al.* 2012. Landscape of transcription in human cells. *Nature*, **489**: 101–108.

Echalier, A., Endicott, J.A. and Noble, M.E.M. 2010. Recent developments in cyclin-dependent kinase biochemical and structural studies. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1804**: 511–519.

Edwards, M.C., Wong, C. and Elledge, S.J. 1998. Human Cyclin K, a Novel RNA Polymerase II-Associated Cyclin Possessing Both Carboxy-Terminal Domain Kinase and Cdk-Activating Kinase Activity. *Molecular and Cellular Biology*, **18**: 4291–4300.

Egloff, S. and Murphy, S. 2008. Cracking the RNA polymerase II CTD code. *Trends Genet.*, **24**: 280–288.

Eick, D. and Geyer, M. 2013. The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chemical Reviews*, **113**: 8456–8490.

Ekumi, K.M., Paculova, H., Lenasi, T., Pospichalova, V., Bösken, C.A., Rybarikova, J., Bryja, V., Geyer, M., Blazek, D. and Barboric, M. 2015. Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Research*, **43**: 2575–2589.

Errico, A., Deshmukh, K., Tanaka, Y., Pozniakovsky, A. and Hunt, T. 2010. Identification of substrates for cyclin dependent kinases. *Advances in Enzyme Regulation*, **50**: 375–399.

Even, Y., Durieux, S., Escande, M.-L., Lozano, J.C., Peaucellier, G., Weil, D. and Geneviève, A.-M. 2006. CDC2L5, a Cdk-like kinase with RS domain, interacts with the ASF/SF2-associated protein p32 and affects splicing in vivo. *Journal of Cellular Biochemistry*, **99**: 890–904.

Fantl, W.J., Muslin, A.J., Kikuchi, A., Martin, J.A., MacNicol, A.M., Grosst, R.W. and Williams, L.T. 1994. Activation of Raf-1 by 14-3-3 proteins. *Nature*, **371**: 612–614.

Ferguson, F.M., Doctor, Z.M., Ficarro, S.B., Browne, C.M., Marto, J.A., Johnson, J.L., Yaron, T.M., Cantley, L.C., Kim, N.D., Sim, T., Berberich, M.J., Kalocsay, M., Sorger, P.K. and Gray, N.S. 2019. Discovery of Covalent CDK14 Inhibitors with Pan-TAIRE Family Specificity. *Cell Chemical Biology*, **26**: 804-817.e12.

Finch, J.T. and Klug, A. 1976. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U.S.A.*, **73**: 1897–1901.

Fire, A., Samuels, M. and Sharp, P.A. 1984. Interactions between RNA polymerase II, factors, and template leading to accurate transcription. *J. Biol. Chem.*, **259**: 2509–2516.

Flanagan, P.M., Kelleher, R.J., Sayre, M.H., Tschochner, H. and Kornberg, R.D. 1991. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature*, **350**: 436–438.

Flores, O., Lu, H. and Reinberg, D. 1992. Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIH. *J. Biol. Chem.*, **267**: 2786–2793.

Flores, O., Maldonado, E. and Reinberg, D. 1989. Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *J. Biol. Chem.*, **264**: 8913–8921.

Fu, H., Subramanian, R.R. and Masters, S.C. 2000. 14-3-3 Proteins: Structure, Function, and Regulation. *Annual Review of Pharmacology and Toxicology*, **40**: 617–647.

Fuda, N.J., Ardehali, M.B. and Lis, J.T. 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, **461**: 186–192.

Furukawa, Y., Ikuta, N., Omata, S., Yamauchi, T., Isobe, T. and Ichimura, T. 1993. Demonstration of the Phosphorylation-Dependent Interaction of Tryptophan Hydroxylase with the 14-3-3 Protein. *Biochemical and Biophysical Research Communications*, **194**: 144–149.

Fussner, E., Ching, R.W. and Bazett-Jones, D.P. 2011. Living without 30nm chromatin fibers. *Trends Biochem. Sci.*, **36**: 1–6.

Galbraith, M.D., Donner, A.J. and Espinosa, J.M. 2010. CDK8: A positive regulator of transcription. *Transcription*, **1**: 4–12.

Gannon-Murakami, L. and Murakami, K. 2002. Selective Association of Protein Kinase C with 14-3-3 ζ in Neuronally Differentiated PC12 Cells: STIMULATORY AND INHIBITORY EFFECT OF 14-3-3 ζ *IN VIVO*. *Journal of Biological Chemistry*, **277**: 23116–23122.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C. and Schultz, N. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*, **6**: pl1.

Ghamari, A., van de Corput, M.P.C., Thongjuea, S., van Cappellen, W.A., van IJcken, W., van Haren, J., Soler, E., Eick, D., Lenhard, B. and Grosveld, F.G. 2013. In vivo live imaging of RNA polymerase II transcription factories in primary cells. *Genes & Development*, **27**: 767–777.

Gilman, A.G., Rall, T.W., Mies, A.S. and Taylor, P. 1993. *The Pharmaceutical Basis of Therapeutics*, 8th ed.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P. and Bentley, D.L. 2009. TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II. *Molecular and Cellular Biology*, **29**: 5455–5464.

Gregorich, Z.R. and Ge, Y. 2014. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics*, **14**: 1195–1210.

Greifenberg, A.K., Hönig, D., Pilarova, K., Düster, R., Bartholomeeusen, K., Bösken, C.A., Anand, K., Blazek, D. and Geyer, M. 2016. Structural and Functional Analysis of the Cdk13/Cyclin K Complex. *Cell Reports*, **14**: 320–331.

Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D. and Cramer, P. 2017. CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife*, **6**.

Grote, M., Wolf, E., Will, C.L., Lemm, I., Agafonov, D.E., Schomburg, A., Fischle, W., Urlaub, H. and Luhrmann, R. 2010. Molecular Architecture of the Human Prp19/CDC5L Complex. *Molecular and Cellular Biology*, **30**: 2105–2119.

Gu, X., Wang, Y., Wang, H., Ni, Q., Zhang, C., Zhu, J., Huang, W., Xu, P., Mao, G. and Yang, S. 2015. Upregulated PFTK1 promotes tumor cell proliferation, migration, and invasion in breast cancer. *Medical Oncology*, **32**.

Haggarty, S.J., Mayer, T.U., Miyamoto, D.T., Fathi, R., King, R.W., Mitchison, T.J. and Schreiber, S.L. 2000. Dissecting cellular processes using small molecules: identification of colchicine-like, taxol-like and other small molecules that perturb mitosis. *Chemistry & Biology*, **7**: 275–286.

Hamilton, M.J., Caswell, R.C., Canham, N., Cole, T., Firth, H.V., Foulds, N., Heimdal, K., Hobson, E., Houge, G., Joss, S., Kumar, D., Lampe, A.K., Maystadt, I., McKay, V., Metcalfe, K., Newbury-Ecob, R., Park, S.-M., Robert, L., Rustad, C.F., *et al.* 2018. Heterozygous mutations affecting the protein kinase domain of *CDK13* cause a syndromic form of developmental delay and intellectual disability. *Journal of Medical Genetics*, **55**: 28–38.

Hamilton, M.J. and Suri, M. 2019. CDK13-related disorder. In: *Advances in Genetics*, pp. 163–182.

Hawley, D.K. and Roeder, R.G. 1985. Separation and partial characterization of three functional steps in transcription initiation by human RNA polymerase II. *J. Biol. Chem.*, **260**: 8163–8172.

Healio Learn Genomics. 2019. Regulation of Gene Expression in Eukaryotes.

Heidemann, M. and Eick, D. 2012. Tyrosine-1 and threonine-4 phosphorylation marks complete the RNA polymerase II CTD phospho-code. *RNA Biology*, **9**: 1144–1146.

Hengartner, C.J., Myer, V.E., Liao, S.-M., Wilson, C.J., Koh, S.S. and Young, R.A. 1998. Temporal Regulation of RNA Polymerase II by Srb10 and Kin28 Cyclin-Dependent Kinases. *Molecular Cell*, **2**: 43–53.

Hertel, K.J. and Graveley, B.R. 2005. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends in Biochemical Sciences*, **30**: 115–118.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., Chapman, R.D., Andrau, J.-C. and Eick, D. 2012. Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation: CTD Thr4 is required for transcription elongation. *The EMBO Journal*, **31**: 2784–2797.

Hong, S.W., Hong, S.M., Yoo, J.W., Lee, Y.C., Kim, S., Lis, J.T. and Lee, D. -k. 2009. Phosphorylation of the RNA polymerase II C-terminal domain by TFIIH kinase is not essential for transcription of Saccharomyces cerevisiae genome. *Proceedings of the National Academy of Sciences*, **106**: 14276–14280.

Huang, K., Ferrin-O'Connell, I., Zhang, W., Leonard, G.A., O'Shea, E.K. and Quiocho, F.A. 2007. Structure of the Pho85-Pho80 CDK-cyclin complex of the phosphate-responsive signal transduction pathway. *Mol. Cell*, **28**: 614–623.

Insco, M.L., Abraham, B.J., Dubbury, S.J., Dust, S., Chen, K.Y., Liu, D., Wu, C., Ludwig, C.G., Fabo, T., Henriques, T., Adelman, K., Geyer, M., Sharp, P.A., Young, R.A., Boutz, P.L. and Zon, L.I. submitted. CDK13 Mutations Drive Melanoma via Accumulation of Prematurely Terminated Transcripts. *Manuscript submitted for publication*.

International Human Genome Sequencing Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature*, **431**: 931–945.

Iorns, E., Martens-de Kemp, S.R., Lord, C.J. and Ashworth, A. 2009. CRK7 modifies the MAPK pathway and influences the response to endocrine therapy. *Carcinogenesis*, **30**: 1696–1701.

Izban, M.G. and Luse, D.S. 1992. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J. Biol. Chem.*, **267**: 13647–13655.

Jang, M.K., Mochizuki, K., Zhou, M., Jeong, H.-S., Brady, J.N. and Ozato, K. 2005. The Bromodomain Protein Brd4 Is a Positive Regulatory Component of P-TEFb and Stimulates RNA Polymerase II-Dependent Transcription. *Molecular Cell*, **19**: 523–534.

Janiak, C. 2000. A critical account on π–π stacking in metal complexes with aromatic nitrogen-containing ligands †. *Journal of the Chemical Society, Dalton Transactions*, 3885–3896.

Jansen, R., Greenbaum, D. and Gerstein, M. 2002. Relating whole-genome expression data with protein-protein interactions. *Genome Res.*, **12**: 37–46.

Jeffrey, P.D., Russo, A.A., Polyak, K., Gibbs, E., Hurwitz, J., Massagué, J. and Pavletich, N.P. 1995. Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature*, **376**: 313–320.

Jiang, M., Gao, Y., Yang, T., Zhu, X. and Chen, J. 2009. Cyclin Y, a novel membrane-associated cyclin, interacts with PFTK1. *FEBS Letters*, **583**: 2171–2178.

Jones, J.C., Phatnani, H.P., Haystead, T.A., MacDonald, J.A., Alam, S.M. and Greenleaf, A.L. 2004. C-terminal Repeat Domain Kinase I Phosphorylates Ser2 and Ser5 of RNA Polymerase II C-terminal Domain Repeats. *Journal of Biological Chemistry*, **279**: 24957–24964.

Jorgensen, W.L. 2004. The Many Roles of Computation in Drug Discovery. *Science*, **303**: 1813–1818.

Joshi, P.M., Sutor, S.L., Huntoon, C.J. and Karnitz, L.M. 2014. 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. *Journal of Biological Chemistry*, **289**: 9247–9253.

Kanin, E.I., Kipp, R.T., Kung, C., Slattery, M., Viale, A., Hahn, S., Shokat, K.M. and Ansari, A.Z. 2007. Chemical inhibition of the TFIIH-associated kinase Cdk7/Kin28 does not impair global mRNA synthesis. *Proceedings of the National Academy of Sciences*, **104**: 5812–5817.

Kauraniemi, P., Bärlund, M., Monni, O. and Kallioniemi, A. 2001. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. *Cancer Res.*, **61**: 8235–8240.

Kauraniemi, P., Kuukasjärvi, T., Sauter, G. and Kallioniemi, A. 2003. Amplification of a 280-Kilobase Core Region at the ERBB2 Locus Leads to Activation of Two Hypothetical Proteins in Breast Cancer. *The American Journal of Pathology*, **163**: 1979–1984.

Kelly, W.G., Dahmus, M.E. and Hart, G.W. 1993. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J. Biol. Chem.*, **268**: 10416–10424.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C. and Bentley, D.L. 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature Structural & Molecular Biology*, **17**: 1279–1286.

Kim, M., Suh, H., Cho, E.-J. and Buratowski, S. 2009. Phosphorylation of the Yeast Rpb1 C-terminal Domain at Serines 2, 5, and 7. *Journal of Biological Chemistry*, **284**: 26421–26426.

Kim, Y.J., Björklund, S., Li, Y., Sayre, M.H. and Kornberg, R.D. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, **77**: 599–608.

Knight, Z.A. and Shokat, K.M. 2005. Features of Selective Kinase Inhibitors. *Chemistry & Biology*, **12**: 621–637.

Ko, T.K., Kelly, E. and Pines, J. 2001. CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J. Cell. Sci.*, **114**: 2591–2603.

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

Kornberg, R.D. 1974. Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**: 868–871.

Krajewska, M., Dries, R., Grassetti, A.V., Dust, S., Gao, Y., Huang, H., Sharma, B., Day, D.S., Kwiatkowski, N., Pomaville, M., Dodd, O., Chipumuro, E., Zhang, T., Greenleaf, A.L., Yuan, G.-C., Gray, N.S., Young, R.A., Geyer, M., Gerber, S.A., *et al.* 2019. CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. *Nature Communications*, **10**(1):1757.

Krek, W. and Nigg, E.A. 1992. Cell cycle regulation of vertebrate p34cdc2 activity: identification of Thr161 as an essential in vivo phosphorylation site. *New Biol.*, **4**: 323–329.

Krippendorff, B.-F., Neuhaus, R., Lienau, P., Reichel, A. and Huisinga, W. 2009. Mechanism-Based Inhibition: Deriving K_I and k_{inact} Directly from Time-Dependent IC_{50} Values. *Journal of Biomolecular Screening*, **14**: 913–923.

Kuehner, J.N., Pearson, E.L. and Moore, C. 2011. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews Molecular Cell Biology*, **12**: 283–294.

Kumagai, A., Yakowec, P.S. and Dunphy, W.G. 1998. 14-3-3 proteins act as negative regulators of the mitotic inducer Cdc25 in Xenopus egg extracts. *Mol. Biol. Cell*, **9**: 345–354.

Kuras, L. and Struhl, K. 1999. Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenzyme. *Nature*, **399**: 609–613.

Lanucara, F. and Eyers, C.E. 2013. Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrom Rev*, **32**: 27–42.

Lazzaro, M.A., Albert, P.R. and Julien, J.P. 1997. A novel cdc2-related protein kinase expressed in the nervous system. *J. Neurochem.*, **69**: 348–364.

Leclerc, V., Tassan, J.P., O'Farrell, P.H., Nigg, E.A. and Léopold, P. 1996. Drosophila Cdk8, a kinase partner of cyclin C that interacts with the large subunit of RNA polymerase II. *Molecular Biology of the Cell*, **7**: 505–513.

Lee, H.J., Chua, G.H., Krishnan, A., Lane, D.P. and Verma, C.S. 2007. Substrate Specificity of Cyclins Determined by Electrostatics. *Cell Cycle*, **6**: 2219–2226.

Lee, J.T. 2012. Epigenetic Regulation by Long Noncoding RNAs. *Science*, **338**: 1435–1439.

Lelli, K.M., Slattery, M. and Mann, R.S. 2012. Disentangling the many layers of eukaryotic transcriptional regulation. *Annu. Rev. Genet.*, **46**: 43–68.

Leung, W.K.C., Ching, A.K.K., Chan, A.W.H., Poon, T.C.W., Mian, H., Wong, A.S.T., To, K.-F. and Wong, N. 2011. A novel interplay between oncogenic PFTK1 protein kinase and tumor suppressor TAGLN2 in the control of liver cancer cell motility. *Oncogene*, **30**: 4464–4475.

Li, S., Jiang, M., Wang, W. and Chen, J. 2014. 14-3-3 Binding to Cyclin Y contributes to cyclin Y/CDK14 association. *Acta Biochimica et Biophysica Sinica*, **46**: 299–304.

Liang, K., Gao, X., Gilmore, J.M., Florens, L., Washburn, M.P., Smith, E. and Shilatifard, A. 2015. Characterization of Human Cyclin-Dependent Kinase 12 (CDK12) and CDK13 Complexes in C-Terminal Domain Phosphorylation, Gene Transcription, and RNA Processing. *Molecular and Cellular Biology*, **35**: 928–938.

Liang, X., Butterworth, M.B., Peters, K.W., Walker, W.H. and Frizzell, R.A. 2008. An obligatory heterodimer of 14-3-3beta and 14-3-3epsilon is required for aldosterone regulation of the epithelial sodium channel. *J. Biol. Chem.*, **283**: 27418–27425.

Liao, S.-M., Zhang, J., Jeffery, D.A., Koleske, A.J., Thompson, C.M., Chao, D.M., Viljoen, M., van Vuuren, H.J.J. and Young, R.A. 1995. A kinase–cyclin pair in the RNA polymerase II holoenzyme. *Nature*, **374**: 193–196.

Lim, S. and Kaldis, P. 2013. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development*, **140**: 3079–3093.

Lin, X., Taube, R., Fujinaga, K. and Peterlin, B.M. 2002. P-TEFb Containing Cyclin K and Cdk9 Can Activate Transcription via RNA. *Journal of Biological Chemistry*, **277**: 16873–16878.

Liu, D., Bienkowska, J., Petosa, C., Collier, R.J., Fu, H. and Liddington, R. 1995. Crystal structure of the zeta isoform of the 14-3-3 protein. *Nature*, **376**: 191–194.

Liu, D., Guest, S. and Finley, R.L. 2010. Why cyclin Y? A highly conserved cyclin with essential functions. *Fly (Austin)*, **4**: 278–282.

Liu, J. and Kipreos, E.T. 2000. Evolution of Cyclin-Dependent Kinases (CDKs) and CDK-Activating Kinases (CAKs): Differential Conservation of CAKs in Yeast and Metazoa. *Molecular Biology and Evolution*, **17**: 1061–1074.

Liu, M., Shi, S., Li, K. and Chen, E. 2016. Knockdown of PFTK1 Expression by RNAi Inhibits the Proliferation and Invasion of Human Non-Small Lung Adenocarcinoma Cells. *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics*, **24**: 181–187.

Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A.K. and Rosenfeld, M.G. 2013. Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. *Cell*, **155**: 1581–1595.

Long, J.C. and Caceres, J.F. 2009. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal*, **417**: 15–27.

Lorca, T., Labbé, J.C., Devault, A., Fesquet, D., Capony, J.P., Cavadore, J.C., Le Bouffant, F. and Dorée, M. 1992. Dephosphorylation of cdc2 on threonine 161 is required for cdc2 kinase inactivation and normal anaphase. *EMBO J.*, **11**: 2381–2390.

Lui, G.Y.L., Grandori, C. and Kemp, C.J. 2018. CDK12: an emerging therapeutic target for cancer. *Journal of Clinical Pathology*, **71**: 957–962.

Malumbres, M. 2014. Cyclin-dependent kinases. *Genome Biology*, **15**: 122.

Malumbres, M. and Barbacid, M. 2009. Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer*, **9**: 153–166.

Malumbres, M. and Barbacid, M. 2005. Mammalian cyclin-dependent kinases. *Trends in Biochemical Sciences*, **30**: 630–641.

Malumbres, M., Harlow, E., Hunt, T., Hunter, T., Lahti, J.M., Manning, G., Morgan, D.O., Tsai, L.-H. and Wolgemuth, D.J. 2009. Cyclin-dependent kinases: a family portrait. *Nature Cell Biology*, **11**: 1275–1276.

Manley, J.L., Sharp, P.A. and Gefter, M.L. 1982. RNA synthesis in isolated nuclei. *Journal of Molecular Biology*, **159**: 581–599.

Manning, G. 2002. The Protein Kinase Complement of the Human Genome. *Science*, **298**: 1912–1934.

Marqués, F., Moreau, J.-L., Peaucellier, G., Lozano, J.-C., Schatt, P., Picard, A., Callebaut, I., Perret, E. and Geneviève, A.-M. 2000. A New Subfamily of High Molecular Mass CDC2-Related Kinases with PITAI/VRE Motifs. *Biochemical and Biophysical Research Communications*, **279**: 832–837.

Marshall, N.F., Peng, J., Xie, Z. and Price, D.H. 1996. Control of RNA Polymerase II Elongation Potential by a Novel Carboxyl-terminal Domain Kinase. *Journal of Biological Chemistry*, **271**: 27176–27183.

Marshall, N.F. and Price, D.H. 1995. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J. Biol. Chem.*, **270**: 12335–12338.

Martin, M.P., Olesen, S.H., Georg, G.I. and Schönbrunn, E. 2013. Cyclin-Dependent Kinase Inhibitor Dinaciclib Interacts with the Acetyl-Lysine Recognition Site of Bromodomains. *ACS Chemical Biology*, **8**: 2360–2365.

Mathis, D.J. and Chambon, P. 1981. The SV40 early region TATA box is required for accurate in vitro initiation of transcription. *Nature*, **290**: 310–315.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D. and Cramer, P. 2012. CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. *Science*, **336**: 1723–1725.

Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Söding, J. and Cramer, P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nature Structural & Molecular Biology*, **17**: 1272–1278.

McPherson, A. 1999. *Crystallization of biological macromolecules*.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. and Cramer, P. 2005. A structural perspective of CTD function. *Genes Dev.*, **19**: 1401–1415.

Messaritou, G., Grammenoudi, S. and Skoulakis, E.M.C. 2010. Dimerization Is Essential for 14-3-3ζ Stability and Function *in Vivo*. *Journal of Biological Chemistry*, **285**: 1692–1700.

Michels, A.A., Nguyen, V.T., Fraldi, A., Labas, V., Edwards, M., Bonnet, F., Lania, L. and Bensaude, O. 2003. MAQ1 and 7SK RNA interact with CDK9/cyclin T complexes in a transcription-dependent manner. *Mol. Cell. Biol.*, **23**: 4859–4869.

Mikolcevic, Petra, Rainer, J. and Geley, S. 2012. Orphan kinases turn eccentric: A new class of cyclin Y-activated, membrane-targeted CDKs. *Cell Cycle*, **11**: 3758–3768.

Mikolcevic, P., Sigl, R., Rauch, V., Hess, M.W., Pfaller, K., Barisic, M., Pelliniemi, L.J., Boesl, M. and Geley, S. 2012. Cyclin-Dependent Kinase 16/PCTAIRE Kinase 1 Is Activated by Cyclin Y and Is Essential for Spermatogenesis. *Molecular and Cellular Biology*, **32**: 868–879.

Morgan, D.O. 1997. CYCLIN-DEPENDENT KINASES: Engines, Clocks, and Microprocessors. *Annual Review of Cell and Developmental Biology*, **13**: 261–291.

Morgan, S., Grootendorst, P., Lexchin, J., Cunningham, C. and Greyson, D. 2011. The cost of drug development: A systematic review. *Health Policy*, **100**: 4–17.

Mu, R., Wang, Y.-B., Wu, M., Yang, Y., Song, W., Li, T., Zhang, W.-N., Tan, B., Li, A.-L., Wang, N., Xia, Q., Gong, W.-L., Wang, C.-G., Zhou, T., Guo, N., Sang, Z.-H. and Li, H.-Y. 2017. Depletion of pre-mRNA splicing factor Cdc5L inhibits mitotic progression and triggers mitotic catastrophe. *Cell Death & Disease*, **5**: e1151–e1151.

Muslin, A.J., Tanner, J.W., Allen, P.M. and Shaw, A.S. 1996. Interaction of 14-3-3 with signaling proteins is mediated by the recognition of phosphoserine. *Cell*, **84**: 889–897.

Noble, M.E., Endicott, J.A., Brown, N.R. and Johnson, L.N. 1997. The cyclin box fold: protein recognition in cell-cycle and transcription control. *Trends Biochem. Sci.*, **22**: 482–487.

Noe, M.C. and Gilbert, A.M. 2012. Targeted Covalent Enzyme Inhibitors. In: *Annual Reports in Medicinal Chemistry*, pp. 413–439.

Novellasdemunt, L., Antas, P. and Li, V.S.W. 2015. Targeting Wnt signaling in colorectal cancer. A Review in the Theme: Cell Signaling: Proteins, Pathways and Mechanisms. *American Journal of Physiology-Cell Physiology*, **309**: C511–C521.

Obsil, T., Ghirlando, R., Klein, D.C., Ganguly, S. and Dyda, F. 2001. Crystal Structure of the 14-3-3ζ:Serotonin N-Acetyltransferase Complex. *Cell*, **105**: 257–267.

Odawara, J., Harada, A., Yoshimi, T., Maehara, K., Tachibana, T., Okada, S., Akashi, K. and Ohkawa, Y. 2011. The classification of mRNA expression levels by the phosphorylation state of RNAPII CTD based on a combined genome-wide approach. *BMC Genomics*, **12**: 516.

Ong, C.-T. and Corces, V.G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.*, **12**: 283–293.

Orom, U.A. and Shiekhattar, R. 2011. Noncoding RNAs and enhancers: complications of a long-distance relationship. *Trends Genet.*, **27**: 433–439.

Ossipow, V., Tassan, J.P., Nigg, E.A. and Schibler, U. 1995. A mammalian RNA polymerase II holoenzyme containing all components required for promoter-specific transcription initiation. *Cell*, **83**: 137–146.

Ou-Yang, J., Huang, L.-H. and Sun, X.-X. 2017. Cyclin-Dependent Kinase 14 Promotes Cell Proliferation, Migration and Invasion in Ovarian Cancer by Inhibiting Wnt Signaling Pathway. *Gynecologic and Obstetric Investigation*, **82**: 230–239.

Ou-Yang, S., Lu, J., Kong, X., Liang, Z., Luo, C. and Jiang, H. 2012. Computational drug discovery. *Acta Pharmacologica Sinica*, **33**: 1131–1140.

Pak, C., Garshasbi, M., Kahrizi, K., Gross, C., Apponi, L.H., Noto, J.J., Kelly, S.M., Leung, S.W., Tzschach, A., Behjati, F., Abedini, S.S., Mohseni, M., Jensen, L.R., Hu, H., Huang, B., Stahley, S.N., Liu, G., Williams, K.R., Burdick, S., *et al.* 2011. Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. *Proceedings of the National Academy of Sciences*, **108**: 12390–12395.

Pak, V., Eifler, T.T., Jäger, S., Krogan, N.J., Fujinaga, K. and Peterlin, B.M. 2015. CDK11 in TREX/THOC Regulates HIV mRNA 3′ End Processing. *Cell Host & Microbe*, **18**: 560–570.

Parry, D., Guzi, T., Shanahan, F., Davis, N., Prabhavalkar, D., Wiswell, D., Seghezzi, W., Paruch, K., Dwyer, M.P., Doll, R., Nomeir, A., Windsor, W., Fischmann, T., Wang, Y., Oft, M., Chen, T., Kirschmeier, P. and Lees, E.M. 2010. Dinaciclib (SCH 727965), a Novel and Potent Cyclin-Dependent Kinase Inhibitor. *Molecular Cancer Therapeutics*, **9**: 2344–2353.

Paul, S.M., Mytelka, D.S., Dunwiddie, C.T., Persinger, C.C., Munos, B.H., Lindborg, S.R. and Schacht, A.L. 2010. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery*, **9**: 203–214.

Pavletich, N.P. 1999. Mechanisms of cyclin-dependent kinase regulation: structures of cdks, their cyclin activators, and cip and INK4 inhibitors. *Journal of Molecular Biology*, **287**: 821–828.

Peeper, D.S., Parker, L.L., Ewen, M.E., Toebes, M., Hall, F.L., Xu, M., Zantema, A., van der Eb, A.J. and Piwnica-Worms, H. 1993. A- and B-type cyclins differentially modulate substrate specificity of cyclin-cdk complexes. *EMBO J.*, **12**: 1947–1954.

Phatnani, H.P. and Greenleaf, A.L. 2006. Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.*, **20**: 2922–2936.

Pines, J. 1995. Cyclins and cyclin-dependent kinases: a biochemical view. *Biochemical Journal*, **308**: 697–711.

Pines, Jonathon. 1995. Cyclins and Cyclin-Dependent Kinases: Theme and Variations. In: *Advances in Cancer Research*, pp. 181–212.

Pozuelo Rubio, M., Geraghty, K.M., Wong, B.H.C., Wood, N.T., Campbell, D.G., Morrice, N. and Mackintosh, C. 2004. 14-3-3-affinity purification of over 200 human phosphoproteins reveals new links to regulation of cellular metabolism, proliferation and trafficking. *Biochem. J.*, **379**: 395–408.

Ramanathan, Y., Rajpara, S.M., Reza, S.M., Lees, E., Shuman, S., Mathews, M.B. and Pe'ery, T. 2001. Three RNA Polymerase II Carboxyl-terminal Domain Kinases Display Distinct Substrate Preferences. *Journal of Biological Chemistry*, **276**: 10913–10920.

Ramanathan, Y., Reza, S.M., Young, T.M., Mathews, M.B. and Pe'ery, T. 1999. Human and rodent transcription elongation factor P-TEFb: interactions with human immunodeficiency virus type 1 tat and carboxy-terminal domain substrate. *J. Virol.*, **73**: 5448–5458.

Ranuncolo, S.M., Ghosh, S., Hanover, J.A., Hart, G.W. and Lewis, B.A. 2012. Evidence of the Involvement of *O* -GlcNAc-modified Human RNA Polymerase II CTD in Transcription *in Vitro* and *in Vivo*. *Journal of Biological Chemistry*, **287**: 23549–23561.

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K. and Bonner, W. 2002. Histone H2A variants H2AX and H2AZ. *Curr. Opin. Genet. Dev.*, **12**: 162–169.

Reinberg, D. and Roeder, R.G. 1987. Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of initiation factors IIB and IIE. *J. Biol. Chem.*, **262**: 3310–3321.

Rha, J., Jones, S.K., Fidler, J., Banerjee, A., Leung, S.W., Morris, K.J., Wong, J.C., Inglis, G.A.S., Shapiro, L., Deng, Q., Cutler, A.A., Hanif, A.M., Pardue, M.T., Schaffer, A., Seyfried, N.T., Moberg, K.H., Bassell, G.J., Escayg, A., García, P.S., *et al.* 2017. The RNA-binding protein, ZC3H14, is required for proper poly(A) tail length control, expression of synaptic proteins, and brain function in mice. *Human Molecular Genetics*, **26**: 3663–3681.

Rinn, J.L. and Chang, H.Y. 2012. Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.*, **81**: 145–166.

Rittinger, K., Budman, J., Xu, J., Volinia, S., Cantley, L.C., Smerdon, S.J., Gamblin, S.J. and Yaffe, M.B. 1999. Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. *Mol. Cell*, **4**: 153–166.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L. and Buratowski, S. 2000. Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol. Cell. Biol.*, **20**: 104–112.

Rosenquist, M., Sehnke, P., Ferl, R.J., Sommarin, M. and Larsson, C. 2000. Evolution of the 14-3-3 protein family: does the large number of isoforms in multicellular organisms reflect functional specificity? *J. Mol. Evol.*, **51**: 446–458.

Russo, A.A., Jeffrey, P.D. and Pavletich, N.P. 1996. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat. Struct. Biol.*, **3**: 696–700.

Samuels, M., Fire, A. and Sharp, P.A. 1982. Separation and characterization of factors mediating accurate transcription by RNA polymerase II. *J. Biol. Chem.*, **257**: 14419–14427.

Samuels, M. and Sharp, P.A. 1986. Purification and characterization of a specific RNA polymerase II transcription factor. *J. Biol. Chem.*, **261**: 2003–2013.

Satyanarayana, A. and Kaldis, P. 2009. Mammalian cell-cycle regulation: several Cdks, numerous cyclins and diverse compensatory mechanisms. *Oncogene*, **28**: 2925–2939.

Sawadogo, M. and Roeder, R.G. 1985. Factors involved in specific transcription by human RNA polymerase II: analysis by a rapid and quantitative in vitro assay. *Proc. Natl. Acad. Sci. U.S.A.*, **82**: 4394–4398.

Schroeder, S.C., Schwer, B., Shuman, S. and Bentley, D. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev.*, **14**: 2435–2440.

Schulman, B.A., Lindstrom, D.L. and Harlow, E. 1998. Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A. *Proc. Natl. Acad. Sci. U.S.A.*, **95**: 10453–10458.

Schulte, A., Czudnochowski, N., Barboric, M., Schönichen, A., Blazek, D., Peterlin, B.M. and Geyer, M. 2005. Identification of a cyclin T-binding domain in Hexim1 and biochemical analysis of its binding competition with HIV-1 Tat. *J. Biol. Chem.*, **280**: 24968–24977.

Schulze-Gahmen, U., De Bondt, H.L. and Kim, S.-H. 1996. High-Resolution Crystal Structures of Human Cyclin-Dependent Kinase 2 with and without ATP: Bound Waters and Natural Ligand as Guides for Inhibitor Design †. *Journal of Medicinal Chemistry*, **39**: 4540–4546.

Schwede, T. 2003. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, **31**: 3381–3385.

Shao, W. and Zeitlinger, J. 2017. Paused RNA polymerase II inhibits new transcriptional initiation. *Nature Genetics*, **49**: 1045–1051.

Shehata, S.N., Deak, M., Morrice, N.A., Ohta, E., Hunter, R.W., Kalscheuer, V.M. and Sakamoto, K. 2015. Cyclin Y phosphorylation- and 14-3-3-binding-dependent activation of PCTAIRE-1/CDK16. *Biochemical Journal*, **469**: 409–420.

Shehata, S.N., Hunter, R.W., Ohta, E., Peggie, M.W., Lou, H.J., Sicheri, F., Zeqiraj, E., Turk, B.E. and Sakamoto, K. 2012. Analysis of substrate specificity and cyclin Y binding of PCTAIRE-1 kinase. *Cellular Signalling*, **24**: 2085–2094.

Shen, Y.H., Godlewski, J., Bronisz, A., Zhu, J., Comb, M.J., Avruch, J. and Tzivion, G. 2003. Significance of 14-3-3 self-dimerization for phosphorylation-dependent target binding. *Mol. Biol. Cell*, **14**: 4721–4733.

Shu, F., Lv, S., Qin, Y., Ma, X., Wang, X., Peng, X., Luo, Y., Xu, B. -e, Sun, X. and Wu, J. 2007. Functional characterization of human PFTK1 as a cyclin-dependent kinase. *Proceedings of the National Academy of Sciences*, **104**: 9248–9253.

Sifrim A., the UK10K Consortium, the Deciphering Developmental Disorders Study, Sifrim, A., Hitz, M.-P., Wilsdon, A., Breckpot, J., Turki, S.H.A., Thienpont, B., McRae, J., Fitzgerald, T.W., Singh, T., Swaminathan, G.J., Prigmore, E., Rajan, D., Abdul-Khaliq, H., Banka, S., Bauer, U.M.M., Bentham, J., *et al.* 2016. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nature Genetics*, **48**: 1060–1065.

Skoulakis, E.M.C. and Davis, R.L. 1998. 14-3-3 proteins in neuronal development and function. *Molecular Neurobiology*, **16**: 269–284.

Sliwoski, G., Kothiwale, S., Meiler, J. and Lowe, E.W. 2013. Computational Methods in Drug Discovery. *Pharmacological Reviews*, **66**: 334–395.

Smith, M.M. 2002. Centromeres and variant histones: what, where, when and why? *Curr. Opin. Cell Biol.*, **14**: 279–285.

Solomon, M.J., Lee, T. and Kirschner, M.W. 1992. Role of phosphorylation in p34cdc2 activation: identification of an activating kinase. *Mol. Biol. Cell*, **3**: 13–27.

Spitz, F. and Furlong, E.E.M. 2012. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**: 613–626.

Sudol, M., Sliwa, K. and Russo, T. 2001. Functions of WW domains in the nucleus. *FEBS Letters*, **490**: 190–195.

Sun, X., Zhang, Y., Cho, H., Rickert, P., Lees, E., Lane, W. and Reinberg, D. 1998. NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell*, **2**: 213–222.

Takagaki, Y. and Manley, J.L. 1997. RNA recognition by the human polyadenylation factor CstF. *Mol. Cell. Biol.*, **17**: 3907–3914.

Takaki, T., Echalier, A., Brown, N.R., Hunt, T., Endicott, J.A. and Noble, M.E.M. 2009. The structure of CDK4/cyclin D3 has implications for models of CDK activation. *Proceedings of the National Academy of Sciences*, **106**: 4171–4176.

The Cancer Genome Atlas Research Network. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature*, **474**: 609–615.

Tietjen, J.R., Zhang, D.W., Rodríguez-Molina, J.B., White, B.E., Akhtar, M.S., Heidemann, M., Li, X., Chapman, R.D., Shokat, K., Keles, S., Eick, D. and Ansari, A.Z. 2010. Chemical-genomic dissection of the CTD code. *Nature Structural & Molecular Biology*, **17**: 1154–1161.

Tjio, J.H. and Levan, A. 2010. THE CHROMOSOME NUMBER OF MAN. *Hereditas*, **42**: 1–6.

Tonna, S., El-Osta, A., Cooper, M.E. and Tikellis, C. 2010. Metabolic memory and diabetic nephropathy: potential role for epigenetic mechanisms. *Nature Reviews Nephrology*, **6**: 332–341.

Tzivion, G. and Avruch, J. 2002. 14-3-3 Proteins: Active Cofactors in Cellular Regulation by Serine/Threonine Phosphorylation. *Journal of Biological Chemistry*, **277**: 3061–3064.

Tzivion, G., Shen, Y.H. and Zhu, J. 2001. 14-3-3 proteins; bringing new definitions to scaffolding. *Oncogene*, **20**: 6331–6338.

Uehara, T., Takenouchi, T., Kosaki, R., Kurosawa, K., Mizuno, S. and Kosaki, K. 2018. Redefining the phenotypic spectrum of de novo heterozygous CDK13 variants: Three patients without cardiac defects. *European Journal of Medical Genetics*, **61**: 243–247.

van den Akker, W.M.R., Brummelman, I., Martis, L.M., Timmermans, R.N., Pfundt, R., Kleefstra, T., Willemsen, M.H., Gerkes, E.H., Herkert, J.C., van Essen, A.J., Rump, P., Vansenne, F., Terhal, P.A., van Haelst, M.M., Cristian, I., Turner, C.E., Cho, M.T., Begtrup, A., Willaert, R., *et al.* 2018. De novo variants in *CDK13* associated with syndromic ID/DD: Molecular and clinical delineation of 15 individuals and a further review. *Clinical Genetics*, **93**: 1000–1007.

van Holde, K. and Zlatanova, J. 1995. Chromatin Higher Order Structure: Chasing a Mirage? *Journal of Biological Chemistry*, **270**: 8373–8376.

Verdoodt, B., Benzinger, A., Popowicz, G.M., Holak, T.A. and Hermeking, H. 2006. Characterization of 14-3-3sigma Dimerization Determinants: Requirement of Homodimerization for Inhibition of Cell Proliferation. *Cell Cycle*, **5**: 2920–2926.

Viladevall, L., St. Amour, C.V., Rosebrock, A., Schneider, S., Zhang, C., Allen, J.J., Shokat, K.M., Schwer, B., Leatherwood, J.K. and Fisher, R.P. 2009. TFIIH and P-TEFb Coordinate Transcription with Capping Enzyme Recruitment at Specific Genes in Fission Yeast. *Molecular Cell*, **33**: 738–751.

Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H. and Cramer, P. 8/2018b. Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature*, **560**: 607–612.

Vos, S.M., Farnung, L., Urlaub, H. and Cramer, P. 08 2018a. Structure of paused transcription complex Pol II-DSIF-NELF. *Nature*, **560**: 601–606.

Wahl, M.C., Will, C.L. and Lührmann, R. 2009. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, **136**: 701–718.

Wang, B., Zou, A., Ma, L., Chen, X., Wang, L., Zeng, X. and Tan, T. 2017. miR-455 inhibits breast cancer cell proliferation through targeting CDK14. *European Journal of Pharmacology*, **807**: 138–143.

Wang, X., Jia, Y., Fei, C., Song, X. and Li, L. 2016. Activation/Proliferation-associated Protein 2 (Caprin-2) Positively Regulates CDK14/Cyclin Y-mediated Lipoprotein Receptor-related Protein 5 and 6 (LRP5/6) Constitutive Phosphorylation. *Journal of Biological Chemistry*, **291**: 26427–26434.

Waters, M.L. 2002. Aromatic interactions in model systems. *Current Opinion in Chemical Biology*, **6**: 736–741.

Wei, P., Garber, M.E., Fang, S.-M., Fischer, W.H. and Jones, K.A. 1998. A Novel CDK9-Associated C-Type Cyclin Interacts Directly with HIV-1 Tat and Mediates Its High-Affinity, Loop-Specific Binding to TAR RNA. *Cell*, **92**: 451–462.

Wilker, E.W., Grant, R.A., Artim, S.C. and Yaffe, M.B. 2005. A structural basis for 14-3-3sigma functional specificity. *J. Biol. Chem.*, **280**: 18891–18898.

Wood, D.J. and Endicott, J.A. 2018. Structural insights into the functional diversity of the CDK–cyclin family. *Open Biology*, **8**: 180112.

Woodcock, C.L.F., Safer, J.P. and Stanchfield, J.E. 1976. Structural repeating units in chromatin. *Experimental Cell Research*, **97**: 101–110.

Xu, Y., Ren, J., He, X., Chen, H., Wei, T. and Feng, W. 2019. YWHA/14-3-3 proteins recognize phosphorylated TFEB by a noncanonical mode for controlling TFEB cytoplasmic localization. *Autophagy*, **15**: 1017–1030.

Yaffe, M.B. 2002. How do 14-3-3 proteins work?-- Gatekeeper phosphorylation and the molecular anvil hypothesis. *FEBS Lett.*, **513**: 53–57.

Yang, L., Zhu, Jia, Huang, H., Yang, Q., Cai, J., Wang, Q., Zhu, Junya, Shao, M., Xiao, J., Cao, J., Gu, X., Zhang, S. and Wang, Y. 2015. PFTK1 Promotes Gastric Cancer Progression by Regulating Proliferation, Migration and Invasion. *PLOS ONE*, **10**: e0140451.

Yang, T. and Chen, J.Y. 2001. Identification and cellular localization of human PFTAIRE1. *Gene*, **267**: 165–172.

Yang, Z., Yik, J.H.N., Chen, R., He, N., Jang, M.K., Ozato, K. and Zhou, Q. 2005. Recruitment of P-TEFb for Stimulation of Transcriptional Elongation by the Bromodomain Protein Brd4. *Molecular Cell*, **19**: 535–545.

Yoh, S.M., Cho, H., Pickle, L., Evans, R.M. and Jones, K.A. 2007. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes & Development*, **21**: 160–174.

Young, K., Lin, S., Sun, L., Lee, E., Modi, M., Hellings, S., Husbands, M., Ozenberger, B. and Franco, R. 1998. Identification of a calcium channel modulator using a high throughput yeast two-hybrid screen. *Nature Biotechnology*, **16**: 946–950.

Yun, C.-H., Mengwasser, K.E., Toms, A.V., Woo, M.S., Greulich, H., Wong, K.-K., Meyerson, M. and Eck, M.J. 2008. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proceedings of the National Academy of Sciences*, **105**: 2070–2075.

Zaret, K.S. and Sherman, F. 1982. DNA sequence required for efficient transcription termination in yeast. *Cell*, **28**: 563–573.

Zha, J., Harada, H., Yang, E., Jockel, J. and Korsmeyer, S.J. 1996. Serine Phosphorylation of Death Agonist BAD in Response to Survival Factor Results in Binding to 14-3-3 Not BCL-XL. *Cell*, **87**: 619–628.

Zhan, T., Rindtorff, N. and Boutros, M. 2017. Wnt signaling in cancer. *Oncogene*, **36**: 1461–1473.

Zhang, T., Kwiatkowski, N., Olson, C.M., Dixon-Clarke, S.E., Abraham, B.J., Greifenberg, A.K., Ficarro, S.B., Elkins, J.M., Liang, Y., Hannett, N.M., Manz, T., Hao, M., Bartkowiak, B., Greenleaf, A.L., Marto, J.A., Geyer, M., Bullock, A.N., Young, R.A. and Gray, N.S. 2016. Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.*, **12**: 876–884.

Zhang, W., Liu, R., Tang, C., Xi, Q., Lu, S., Chen, W., Zhu, L., Cheng, J., Chen, Y., Wang, W., Zhong, J. and Deng, Y. 2016. PFTK1 regulates cell proliferation, migration and invasion in epithelial ovarian cancer. *International Journal of Biological Macromolecules*, **85**: 405–416.

Zhang, W., Prakash, C., Sum, C., Gong, Y., Li, Y., Kwok, J.J.T., Thiessen, N., Pettersson, S., Jones, S.J.M., Knapp, S., Yang, H. and Chin, K.-C. 2012. Bromodomain-containing Protein 4 (BRD4) Regulates RNA Polymerase II Serine 2 Phosphorylation in Human CD4+ T Cells. *Journal of Biological Chemistry*, **287**: 43137–43155.

Zhang, Y. and Reinberg, D. 2001. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.*, **15**: 2343–2360.

Zhao, J., Hyman, L. and Moore, C. 1999. Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**: 405–445.

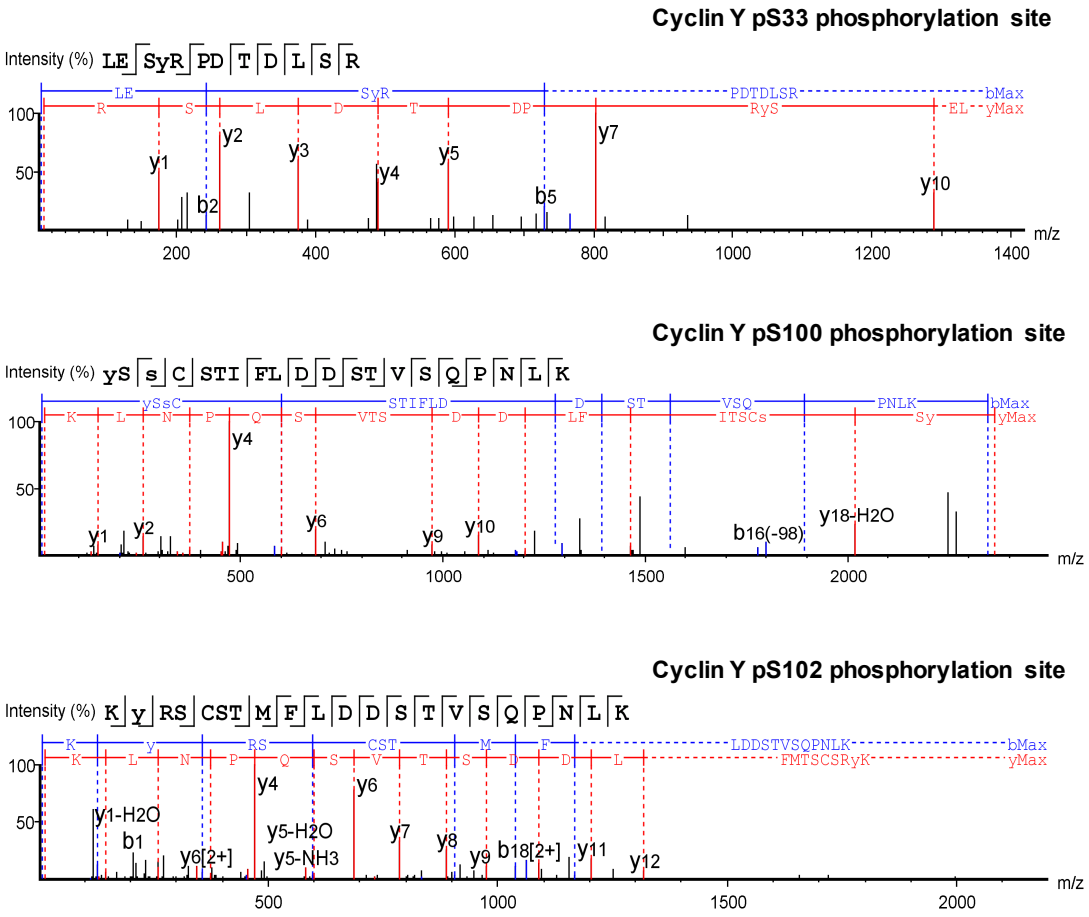
Zheng, L., Zhou, Z. and He, Z. 2015. Knockdown of PFTK1 inhibits tumor cell proliferation, invasion and epithelial-to-mesenchymal transition in pancreatic cancer. *Int J Clin Exp Pathol*, **8**: 14005–14012.

Zhou, Z. and Fu, X.-D. 2013. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma*, **122**: 191–207.

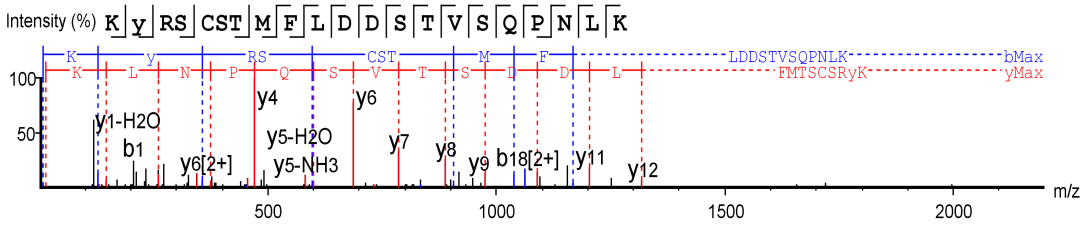
Zi, Z., Zhang, Z., Li, Q., An, W., Zeng, L., Gao, D., Yang, Y., Zhu, X., Zeng, R., Shum, W.W. and Wu, J. 2015. CCNYL1, but Not CCNY, Cooperates with CDK16 to Regulate Spermatogenesis in Mouse. *PLoS Genet.*, **11**: e1005485.

7 Appendix

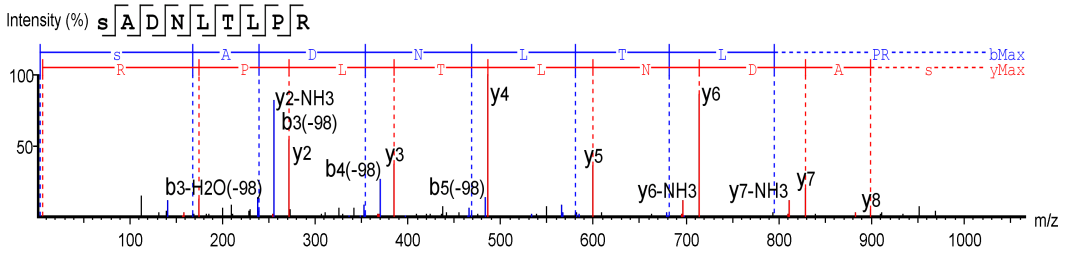
Appendix 1: Mass spectrometry spectra derived from peptide mass fingerprint analyses of the Cyclin Y constructs.



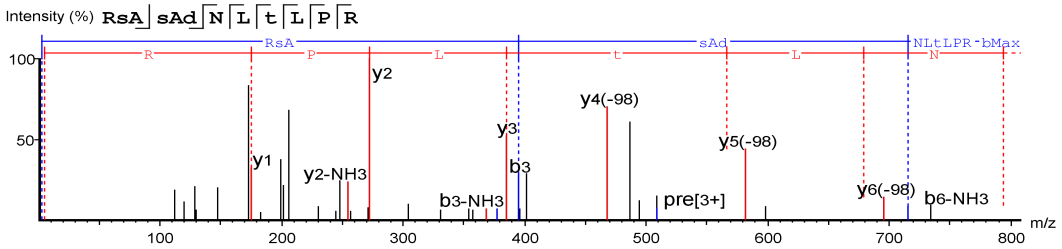
Cyclin Y pT103 phosphorylation site



Cyclin Y pS326 phosphorylation site



Cyclin Y pT331 phosphorylation site



8 Danksagung

Diese Stelle möchte ich nutzen, um allen Mitwirkenden an meiner Doktorarbeit größten Dank auszusprechen.

Zuallererst möchte ich Herrn Prof. Dr. Matthias Geyer danken, für die gegebene Möglichkeit an faszinierenden Projekten mitzuwirken, für die hervorragende Betreuung und Unterstützung während der Bearbeitung meiner Projekte und der Verfassung meiner Doktorarbeit, sowie für die Übernahme des Erstgutachtens.
Frau Prof. Dr. Diana Imhof möchte ich herzlich für die freundliche Übernahme des Zweitgutachtens danken.
Herrn Prof. Dr. Hubert Schorle möchte ich herzlich für die Teilnahme an meiner Prüfungskommission als Drittprüfer danken.
Herrn Prof. Dr. Hasenauer gilt mein herzlicher Dank für die Teilnahme an meiner Prüfungskommission als Viertprüfer.

Ich möchte mich bei allen aktuellen und ehemaligen Arbeitskollegen des Instituts für Strukturbiologie für die angenehme Zusammenarbeit, Diskussions- und Hilfsbereitschaft und tolle Arbeitsatmosphäre bedanken. Danke für die tolle Zeit! Auch möchte ich mich bei den Arbeitskollegen der AG Famulok des caesar Instituts für die Unterstützung und Hilfsbereitschaft bedanken.

Des Weiteren gilt mein Dank Prof. Dr. Nathanael Gray, Dr. Fleur Ferguson und Dr. Baishan Jiang (Dana-Farber Cancer Institute, Boston), Prof. Dr. Geoffrey Shapiro (Dana-Farber Cancer Institute, Boston), Prof. Dr. Rani George und Dr. Malgorzata Krajewska (Dana-Farber Cancer Institute, Boston), Prof. Dr. Leonard Zon und Dr. Megan Insco (Dana-Farber Cancer Institute, Boston) und Dr. Axel Choidas (Lead Discovery Center GmbH, Dortmund) für die freundliche Zusammenarbeit und die Bereitstellung von Proteinen und compounds.

Mein ganz persönlicher Dank gilt im Besonderen meinen Eltern für ihre stetige moralische Unterstützung, die mich mein ganzes Leben lang begleitet haben. Ein besonderer Dank gilt auch meinem Freund Lars. Danke, dass ihr immer für mich da seid. Worte reichen nicht aus um zu beschreiben wie wichtig ihr mir seid!