Biochemical characterization of the human Cyclin-dependent kinases Cdk7 and Cdk10

Dissertation

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

vorgelegt von
Robert Salim Düster
aus
Damaskus, Syrien

Bonn, 2019

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

1. Gutachter: Prof. Dr. Matthias Geyer, Institut für Strukturbiologie, Uniklinik Bonn

2. Gutachter: Prof. Dr. Oliver Gruss, Institut für Genetik, Universität Bonn

Tag der Promotionsprüfung: 19.08.2020

Erschienen im Jahr 2021

Parts of this thesis are published in:

Structural and Functional Analysis of the Cdk13/Cyclin K Complex.
Greifenberg AK, Hönig D, Pilarova K, Düster R, Bartholomeeusen K, Bösken CA, Anand K, Blazek D, Geyer M. Cell Rep. 2016 Jan 12;14(2):320-31.
doi: 10.1016/j.celrep.2015.12.025

P-TEFb Activation by RBM7 Shapes a Pro-survival Transcriptional Response to Genotoxic Stress.
Bugai A, Quaresma AJC, Friedel CC, Lenasi T, Düster R, Sibley CR, Fujinaga K, Kukanja P, Hennig T, Blasius M, Geyer M, Ule J, Dölken L, Barborič M. Mol Cell. 2019 Apr 18;74(2):254-267.
doi: 10.1016/j.molcel.2019.01.033

MYC Recruits SPT5 to RNA Polymerase II to Promote Processive Transcription Elongation.
Baluapuri A, Hofstetter J, Dudvarski Stankovic N, Endres T, Bhandare P, Vos SM, Adhikari B, Schwarz JD, Narain A, Vogt M, Wang SY, Düster R, Jung LA, Vanselow JT, Wiegering A, Geyer M, Maric HM, Gallant P, Walz S, Schlosser A, Cramer P, Eilers M, Wolf E. Mol Cell. 2019 May 16;74(4):674-687.
doi: 10.1016/j.molcel.2019.02.031

Development of a Selective CDK7 Covalent Inhibitor Reveals Predominant Cell-Cycle Phenotype.
Olson CM, Liang Y, Leggett A, Park WD, Li L, Mills CE, Elsarrag SZ, Ficarro SB, Zhang T, Düster R, Geyer M, Sim T, Marto JA, Sorger PK, Westover KD, Lin CY, Kwiatkowski N, Gray NS. Cell Chem Biol. 2019 Jun 20;26(6):792-803.
doi: 10.1016/j.chembiol.2019.02.012

Table of content

I List of Abbreviations

aa	amino acid
ADP	adenosine diphosphate
APS	ammonium persulfate
ATP	adenosine triphosphate
ATP-γ-S	adenosine-5'-(γ-thio)-triphosphate
CAK	Cdk-activating kinase
Cdk	Cyclin-dependent kinase
Cdk10^{as}	analogue-sensitive Cdk10 variant
cpm	counts per minute
CTD	C-terminal domain
Cyc	Cyclin
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide
DSF	differential scanning fluorimetry
DYRK	dual-specificity tyrosine regulated kinase
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
FPLC	fast performance liquid chromatography
GSH	reduced glutathione
GST	glutathion-S-transferase
HEPES	4-(2-hydroxyethyl)-1-piperauineethalonamine
IC_{50}	concentration at which 50% is inhibited
IPTG	isopropyl β-D-1-thiogalactopyrase
K_D	binding affinity
kd	dissociation constant
LB	lysogeny broth
LMW	low molecular weight
MBP	maltose binding protein
mRNA	messenger ribonucleic acid
OD	optical density
PBS	phosphate buffered saline
PDB	protein data bank
PMSF	phenylmethylsulfonyl fluoride
PNBM	para-nitrobenzylmesylate
pol	polymerase
RNA	ribonucleic acid
rpm	rotations per minute
RT	room temperature
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharomyces cerevisiae*
SD	standard deviation
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate - polyacrylamide gel electrophoresis
Sf	*Spodoptera frugiperda*
SPR	surface plasmon resonance

TBS	tris buffered saline
TCEP	tris(2-carboxyethyl)phosphine
TEMED	tetramethylethylendiamine
TEV	Tobacco etch virus
Tm	melting temperature
Tris	tris(hydroxymethyl)aminomethane
v/v	volume per volume
w/v	weight per volume
w/w	weight per weight
Wt	wild type
β-ME	β-mercaptoethanol

II List of Figures

1 Abstract

Protein kinases comprise a large superfamily of enzymes regulating a plethora of cellular processes. Dysregulation of protein kinases, such as loss of activity, hyperactivity or mislocalisation, is often accompanied with severe malignancies. The family of Cyclin-dependent kinases (CDKs) has been first described for their regulation of the mitotic cell cycle. However, of the 21 CDK family members found in man only the CDKs 1, 2, 4, and 6 are directly involved in cell cycle regulation. Another subset regulates gene expression at the level of transcription and transcription associated processes, whereas several other CDKs have no established function yet. Common to all Cyclin-dependent kinases is the requirement of the association with a regulatory cyclin subunit for activation. In addition to cyclin binding, phosphorylation of the kinase within an activation segment, called T-loop, is required for kinase activation.

Cdk7 forms a ternary complex with Cyclin H and the protein Mat1. Cdk7 is involved in both cell cycle regulation and transcription. Cdk7 phosphorylates the T-loop of the cell cycle CDKs 1, 2, 4, and 6 leading to their activation. Additionally, Cdk7 is involved in transcription initiation as it associates with the general transcription factor TFIIH. Within recent years, pharmacologic inhibition of Cdk7 has emerged as a promising therapeutic target in cancer treatment. Cdk7 has been studied for almost three decades, but nonetheless regulation of Cdk7 kinase activity at a molecular level is not understood in detail. In contrast to Cdk7, only very few studies address the functional role of Cdk10 to date. Concomitantly, the regulatory cyclin subunit of Cdk10, Cyclin M, has not been identified until 2013. In cancer, Cdk10 is mostly described as a negative regulator of cell cycle progression and reduced Cdk10 levels were found to confer tamoxifen resistance in breast cancer. In this thesis, the human cyclin-dependent kinases Cdk7 and Cdk10 in complex with their regulatory subunits have been analysed by biochemical and biophysical techniques.

In contrast to other CDKs, Cdk7 is phosphorylated at two sites within its T-loop, Ser164 and Thr170. Of these two sites, Thr170 resembles the canonical activation residue which is conserved among CDKs. The second phosphorylation site at Ser164 has been implicated in stability and activity regulation of Cdk7 but no apparent function has been attributed to this phosphorylation mark so far. A focus of the first part of this thesis was to study the effect of Cdk7 Ser164 phosphorylation on kinase activity and complex formation using recombinant protein. By mutational analyses, in which Cdk7 Ser164 was replaced by a non-phosphorylatable alanine, it could be shown that phosphorylation of Cdk7 is required for full activity of the ternary Cdk7/CycH/Mat1 complex. Importantly, the Cdk7 S164E mutation, which was introduced in order to mimic the Cdk7 Ser164 phosphorylation, did not recapitulate the effect of Ser164 phosphorylation in wild type Cdk7 preparations. However, the effect of Ser164 phosphorylation within ternary complexes could be restored in the Cdk7 S164E mutant by truncation of the Cyclin H C-terminus to amino acid 291 (CycH 1-291). Surprisingly, the Cdk7 S164A mutant failed to bind the truncated CycH (1-291) protein in co-purification experiments. This result was unexpected as the Cyclin H C-terminus is supposed to be unstructured and does not contain any canonical feature for Cdk/Cyclin interaction. Together, these data suggest an inhibitory regulatory element residing in the C-terminus of Cyclin H which is reallocated upon Cdk7 Ser164 phosphorylation.

Despite its discovery 25 years ago, only little is known about the Cyclin-dependent kinase 10 (Cdk10) and its cellular functions. In the second part of this thesis, human recombinant Cdk10/CycM from *Sf9* insect cells was analysed for activity and substrate specificity. Moreover, new substrates of Cdk10/CycM were identified by mass spectrometry analyses. Expression of wild type Cdk10/CycM complexes in *Sf9* cells resulted in active kinase complexes without special needs to activate the kinase recombinantly. Initial activity measurements in radioactive kinase assays using recombinant substrates revealed *in vitro* kinase activity towards the typical CDK substrates c-MYC and RNA polymerase II C-terminal domain (CTD).

In order to identify new Cdk10 substrates by mass spectrometry, an *in vitro* chemical genetic screen was performed. To this end, the Cdk10 gatekeeper residue methionine 117 was mutated to glycine resulting in an analogue-sensitive Cdk10 M117G mutant. The Cdk10 M117G mutation enlarges the ATP binding pocket which allows Cdk10 to tolerate bulky ATP-analogues as substrates. Due to their size, these ATP-analogues are not used by native, cellular kinases which confers specificity of the labelling reaction to the mutant Cdk10^{M117G}/CycM complex. To discriminate and purify Cdk10^{M117G}/CycM phosphorylated proteins from other phosphoproteins in the lysate, ATP-γ-S analogues were used which allows for the specific enrichment by a covalent capture and release protocol. Mass spectrometric identification of thio-phosphorylated proteins from cell lysates and nuclear extracts was performed in collaboration with Prof. Dr. Henning Urlaub at the Max Planck Institute for Biophysical Chemistry in Göttingen. In total, 66 putative Cdk10/CycM substrates were identified. Analysis of the biological function by GO-term analysis shows an enrichment of RNA regulatory and translation related proteins. Six of the putative substrates were selected and validated as Cdk10/CycM substrates *in vitro* using recombinant proteins in kinase activity measurements.

In this thesis a new regulatory mechanism for Cdk7 activity regulation was discovered. Given the central role of Cdk7 in cell cycle and transcription regulation as well as its pharmacological relevance as a target in cancer, mechanisms that alter Cdk7 activity are of major importance to understand the physiological and pathological roles of Cdk7. Future studies shall be performed to address the regulation by Cdk7 Ser164 phosphorylation *in vivo*. In case of Cdk10, this study will facilitate future investigations by providing a protocol for the recombinant expression and purification of Cdk10/CycM complexes. Moreover, the thesis contains a rich dataset of Cdk10 substrates, which will help to elucidate the biological function of this kinase complex.

2 Zusammenfassung

Die Familie der Proteinkinasen umfasst im Menschen 518 verschiedene Mitglieder welche eine Vielzahl zellulärer Prozesse und Signalwege regulieren. Fehlregulation von Kinasen wie beispielsweise Mutationen, die zu Verlust oder Hyperaktivität der enzymatischen Aktivität führen sind häufig mit Krankheiten assoziiert. Die Familie der Zyklin-abhängigen Kinasen (engl. Cyclin-dependent kinase, CDK) ist bekannt für die Regulation der zellulären Proliferation. Allerdings haben von den 21 CDKs, die im Menschen vorhanden sind nur vier, nämlich Cdk1, 2, 4 und 6 einen direkten Einfluss auf die Regulation des Zellzyklus. Andere Vertreter dieser Gruppe regulieren die Transkription der DNA. Gemeinsam ist diesen Kinasen, dass sie erst durch Bindung einer regulatorischen Untereinheit, dem Zyklin, in eine Aktive Konformation überführt werden. Zudem ist für die Aktivität eine Phosphorylierung im Aktivierungssegment, dem T-loop, nötig.

In Zellen assoziiert die Kinase Cdk7 mit ihrer regulatorischen Untereinheit Zyklin H (engl. Cyclin H, CycH) sowie einem weiteren, stabilisierend wirkendem Protein, Mat1. Innerhalb der CDKs kommt Cdk7 eine Sonderstellung zu, da es einerseits die Zellzyklus CDKs 1, 2, 4 und 6 durch Phosphorylierung im T-loop aktiviert und andererseits als Teil des allgemeinen Transkriptionsfaktors TFIIH unmittelbar an der Regulation der Transkription beteiligt ist. Zudem kommt Cdk7 eine in den letzten Jahren wachsende Bedeutung für die Therapie verschiedener Krebserkrankungen zu. Trotz des großen Interesses sind die molekularen Grundlagen für die Regulation von Cdk7 jedoch noch immer nur unzureichend verstanden. Im Gegensatz zu Cdk7, welches in den vergangenen Jahrzehnten Gegenstand intensiver Forschung war ist die Zyklin-abhängige Kinase 10 (Cdk10) bisher kaum erforscht. Die identifizierung der regulatorischen Zyklin-Untereinheit erfolgte zudem erst 2013. In Krebsmodellen fungiert Cdk10 meist als ein negativer Regulator des Zellzyklus und verminderte Cdk10 Expressionslevel wurden als mögliche Ursache für Tamoxifenresistenz bei Brustkrebs identifiziert. In dieser Doktorarbeit wurden die humanen Zyklin-abhängigen Kinasen Cdk7 und Cdk10, sowie deren regulatorische Untereinheiten heterolog exprimiert und mittels biochemischer und biophysikalischer Methoden untersucht.

Im Gegensatz zu anderen CDKs wird Cdk7 an zwei Stellen, Ser164 und Thr170, innerhalb seines Aktivierungssegments phosphoryliert. Dabei handelt es sich bei Thr170 um die kanonische Phosphorylierungsstelle für die Aktivierung der Kinase, welche auch in den anderen CDKs konserviert ist. Der zusätzlichen Phosphorylierung von Cdk7 an Ser164 wurde eine Rolle für die Stabilität, sowie die Aktivität der Kinase zugeschrieben. Allerdings konnte in bisherigen Studien keine Eindeutige molekulare Funktion dieser post-translationalen Modifikation gefunden werden. Ein Fokus dieser Arbeit war der Einfluss der Cdk7 Ser164 Phosphorylierung auf die Kinaseaktivität, sowie die Komplexbildung von Cdk7 mit seinen regulatorischen Untereinheiten.

Cdk7 Serin164 wurde dazu durch die Aminosäure Alanin, welche nicht phosphoryliert werden kann oder die Aminosäure Glutamat, welche eine konstitutive Phosphorylierung simuliert, ersetzt. Beide Punktmutanten wurden dann mit an Ser164/Thr170 doppelt phosphoryliertem wildtyp Cdk7 Präparationen verglichen. Es zeigte sich, dass die zusätzliche Phosphorylierung an Cdk7 Ser164 notwendig für die volle Aktivität von heterotrimeren Cdk7/CycH/Mat1Komplexen ist. Auf die Aktivität heterodimerer Cdk7/CycH Komplexe hatte die phosphorylierung hingegen keinen Einfluss. Die Ser164 Mutationen hatten keinen Einfluss auf die Bindungsaffinität des Cdk7/CycH Komplexes

an Mat1. Interessanterweise, lies sich die Phosphorylierung von Cdk7 Ser164 nicht durch die Punktmutation von Serin164 zu Glutamat adäquat imitieren. Allerdings führte eine C-terminale Trunkierung von Cyclin H an Aminosäure 291 (CycH 1-291) zu einer gesteigerten Aktivität in trimeren Cdk7^{S164E}/CycH (1-291)/Mat1 Komplexen, die mit der von wildtyp Cdk7 vergleichbar war. Überaschenderweise führte die Entfernung des C-Terminus von Cyclin H zu einem kompletten Verlust der Bindung an die Cdk7 S164A Mutante. Dies ist insofern überraschend als das sich im C-terminale Abschnitt von Cyclin H keine Elemente befinden, die für die Cdk/Zyklin Interaktion beschrieben sind. Zusammen weisen diese Daten darauf hin, dass der C-terminus von Cyclin H, in Verbindung mit Cdk7 Ser164 Phosphorylierung, eine unerwartete Rolle bei der Cdk7/CycH Bindung und der Aktivitätsregulation von ternären Cdk7/CycH/Mat1 Komplexen einnimmt.

Trotz der ersten Beschreibung von Cdk10 vor bereits 25 Jahren ist wenig über die zelluläre Funktion der Zyklin-abhängigen Kinase Cdk10 bekannt. Im zweiten Teil dieser Arbeit wurde humane recombinant hergestellte Cdk10/CycM Komplexe auf ihre Aktivität und Substratspezifität hin untersucht. Zudem wurden neue Substrate von Cdk10/CycM mittels Massenspektrometrie.

Initiale Aktivitätsmessungen mit wildtyp Cdk10/CycM zeigten eine *in vitro* Aktivität gegenüber der RNA polymerase II C-terminalen Domäne, sowie des Transkriptionfaktors c-MYC. Bisher sind lediglich die Proteine ETS-2, sowie PKN2 als Cdk10 Substrate beschrieben. Zur Erweiterung des Substratspektrums wurden in dieser Arbeit neue Cdk10 Substrate *in vitro* mittels Massenspektrometrie bestimmt. Dazu wurde per Punktmutation von Cdk10 Methionin zu Glycin eine Analog-sensitive Cdk10/CycM Mutante generiert. Die Punktmutation vergrößert die ATP Bindungstasche und erlaubt so die Verwendung von N^6-modifizierten ATP-Analogen, welche nicht von nativen Kinasen verwendet werden können. Darüber hinaus ermöglicht die Verwendung von N^6-modifiziertem ATP-γ-S zu einer bio-orthogonalen thio-phosphorylierung, die die spätere Diskriminierung und Anreicherung der Cdk10^{M117G}/CycM Substrate ermöglicht. In Aktivitätstests konnte gezeigt werden, dass die Punktmutation Cdk10 M117G nicht zu einem Verlust der Kinaseaktivität führt. Zudem konnte die sensitivität gegenüber N^6-modifizierter ATP-Analoga bestätigt werden. Die Anreicherung der thio-phosphorylierten Substrate, sowie die Massenspektrometrische identifizierung etwaiger Cdk10/CycM substrate erfolgte in Kollaboration mit Prof. Dr. Henning Urlaub vom Max Planck Institut für Biophysikalische Chemie in Göttingen. Insgesamt konnten 66 neue Cdk10/CycM Substrate identifiziert werden. Die Analyse der biologischen Funktion mittels GO-term zeigte eine Anreicherung für RNA modifizierende und translatorische Prozesse.

In dieser Arbeit konnte gezeigt werden, dass die Phosphorylierung von Cdk7 an Ser164 für die volle Aktivität des trimeren Cdk7/CycH/Mat1 Komplexes notwendig ist. In Anbetracht der zentralen Rolle von Cdk7 für Zellzyklus und Transkription sollten zukünftige Studien den phosphorylierungsstatus von Cdk7 in ihre Analyse einbeziehen. Im Falle von Cdk10/CycM liefert diese Arbeit erste Einblicke in das Substratspektrum dieser Kinase, sowie in die damit verbundenen zellulären Prozesse. Zudem ermöglicht das entwickelte Protokoll für die rekombinante Darstellung des Cdk10/CycM Komplexes eine wichtige Basis für weitere Studien.

3 Introduction

3.1 The cell cycle

Cell proliferation is a fundamental biological process in living organisms. Dysregulation of cell division is associated with severe malignancies and cancer (Hanahan und Weinberg 2011). The cycle in which cells grow and divide is known as the cell cycle or cell division cycle. The cell cycle consists of two distinct phases: Mitosis (M-Phase) in which the chromosomes are separated and the new daughter cell is departed from its mother cell and an interphase in which the cells grow and duplicate their DNA for subsequent cell division. Mitosis is characterized by a prophase, pro-metaphase, metaphase, anaphase, and telophase. The prophase is characterized by condensation of the chromosomes. In the pro-methaphase the nuclear envelope is disintegrated and the chromosomes are attached to microtubules at their kinetochores and are aligned equatorial in the metaphase. During anaphase the sisterchromatids are divided until they reach the spindle poles in telophase. The interphase is further subdivided in the G1, the S, and the G2 phase. In mammalian cells the M phase lasts about 1 hour, whereas S-phase, in which the DNA is duplicated takes approximately 10-12 hours. The Gap (G1 and G2) phases are characterized by growth of the cell, expression of proteins required for proliferation, and control mechanisms which assure completion of previous steps. In continuously dividing cells G1, phase lasts about 3 hours, but can last longer in absence of growth signals. Most cells in multicellular organisms leave the cell cycle and rest in a quiescent state, called G0 (Alberts et al. 2015) (Figure 3.1.1).

3.1.1 Checkpoints govern the cell cycle progression

Failures in cell division are detrimental in development. Therefore, the cell cycle is highly regulated and contains several checkpoints at which completion of the previous processes is assured. Two main checkpoints of the cell cycle are found at G1/S transition and at G2/M transition. The G1/S checkpoint, also termed restriction point, is mainly affected by external growth stimuli. The checkpoint at G2/M is mainly driven by assurance of complete DNA duplication and a sensing mechanism for DNA damage (El-Aouar Filho et al. 2017).

In eukaryotic cells a specific subset of kinases, cyclin-dependent kinases (CDKs), is associated with cell cycle progression and regulation. Cyclin-dependent kinases are heterodimers of a catalytic subunit (kinase) and a regulatory subunit (cyclin). Association of the kinase with the cyclin is required for kinase activation. The cyclin subunits were found to be expressed and degraded in relation to the cell cycle (Evans et al. 1983). The mammalian CDKs best studied for their involvement in cell cycle regulation are Cdk1, Cdk2, Cdk4, and Cdk6 and have distinct roles during the cell cycle. Whereas Cdk4 and Cdk6 associate with Cyclin D and are specifically activated at the G1/S transition, Cdk2 is active in complex with CycE at G1/S transition and in complex with CycA during S-phase. Cdk1 also associates with CycA in S-phase. In G2/M phase Cdk1 association with Cyclin B is favoured. Active Cdk1/CycB complexes are important for G2/M transition. Cdk1/CycB moreover orchestrates several mitotic events. CDKs are additionally regulated by activating and inhibiting phosphorylations as well as by inhibitory binding partners.

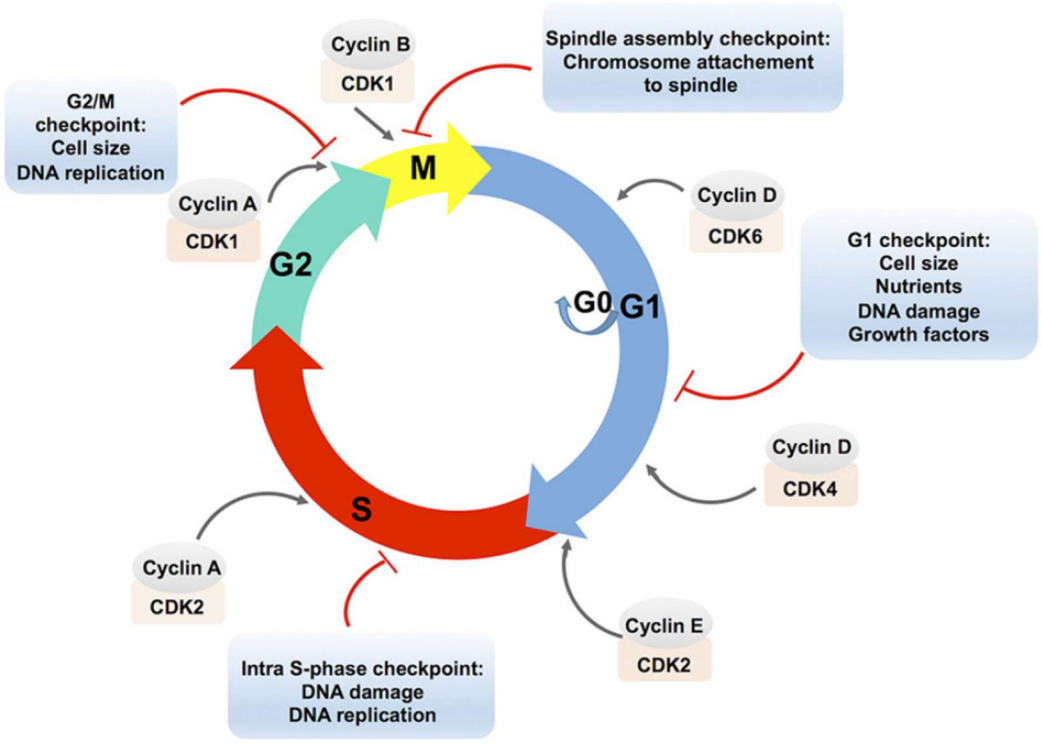


Figure 3.1.1: The eukaryotic cell cycle.
The eukaryotic cell cycle consists of two Gap phases (G1 and G2), the DNA-Synthesis phase (S), and the mitosis phase (M). Cells can also leave the cell cycle and enter a quiescent state (G0). The cell cycle is regulated by timely controlled action of several Cyclin-dependent kinases (CDK). The cycle harbours several checkpoints. The G1- and the G2/M checkpoint are described in the main text (El-Aouar Filho et al. 2017).

3.1.2 Molecular principles at G1/S and G2/M checkpoint

In G1 phase the cell integrates growth stimuli which either promote cell division or result in a quiescent state. Inside the cell, external stimuli are converted into cell-autonomous signal cascades. At a certain threshold cells commit to divide and enter the S-phase. If cells pass this restriction point and committed to cell division they do not rely on growth factors anymore. The restriction point hence demarcates the border between mitogen dependent and mitogen independent cell division. A well understood signalling cascade at the restriction point is the Rb – E2F axis. In absence of growth stimuli, the transcription factor E2F is inactivated by association with retinoblastoma protein (Rb). Upon mitogenic signals the Cyclin dependent kinases Cdk4 and Cdk6 are activated. Cdk4 and Cdk6 in complex with D-type cyclins phosphorylate the Rb protein which in turn liberates E2F from the RB-E2F complex. E2F then activates several genes associated with proliferation. Among the first genes activated is Cyclin E. Expression of Cyclin E results in the formation of active Cdk2/CycE complexes which further hyper-phosphorylates the Rb protein. On a second axis Cdk2/CycE inactivates the ubiquitin ligase (APC^{Cdh1}) which otherwise blocks entry into S-phase. Based on this, several molecular key components of cell division at the restriction point can be described. As such the overexpression of either c-MYC, Cyclin E or the transcription factor E2F were shown to overcome cell cycle arrest at the G1/S restriction point in absence of external growth stimuli (for review see: Fisher 2016).

The checkpoint at G2/M shall reassure that no DNA damaged cells enter mitosis. The kinase Cdk1/CycB is a key component for G2/M transition. Cdk1 is kept inactive by phosphorylation of

Thr14 and Tyr15 by the kinases Myt1 and Wee1. Activation of the phosphatase Cdc25 results in dephosphorylation and activation of the Cdk1/CycB complex. Upon detection of DNA damage two main regulatory pathways are activated by the action of the kinases DNA-PK, ATM and ATR which both result in inactivation of Cdk1/CycB. In a first cascade, the action of Chk-Kinases inactivates Cdc25. In a second cascade, the tumour suppressor protein p53 is activated. Part of the p53 transcripitonal response is the expression of 14-3-3 proteins which bind to Cdk1/CycB and translocate it to the cyctoplasm (Taylor und Stark 2001). Moreover, upregulation by GADD45 leads to dissociation of Cdk1 and Cyclin B and the expression of the Cyclin-dependent kinase inhibitor p21 Cip/KIP further inhibits Cdk1 activity. Hence, cell cycle progression is subject to a vivid cross-talk between DNA damage response factors, transcriptional regulation and kinase dependent signal cascades (Bertoli et al. 2013).

3.2 Gene expression

All living organisms store their genetic information in the form of deoxyribonucleic acid (DNA). In the early 20th century it was known, that genes encode building plans for proteins but it was still a debate in which form information is stored. It was not until 1944 when Avery and his colleagues could demonstrate for the first time that deoxyribonucleic acid (DNA) is the molecule harbouring the genetic information (Avery et al. 1944). About a decade later the structure of DNA was inferred as an α-helix with a deoxyribose backbone and the nucleobases laying inside interacting by hydrogen bonds (Watson and Crick 1953). Within the nucleus DNA is bound to proteins and organized in chromosomes. These proteins are called histones. There are four different histones H2a, H2b, H3, and H4 forming an octamer around which the DNA is wound (Kornberg and Thomas 1974). This complex associated with DNA is called nucleosome.

Still in 1953, Crick formulated what became the dogma of molecular biology. This says that genetic information is an unidirectional flow from DNA to RNA to protein. The identification of retroviral RNA viruses, in which genetic information is stored in RNA and reversely transcribed into DNA in the host cell, should be regarded as an extension rather than a disprove of the dogma. For synthesis of proteins DNA is transcribed into pre-mRNA. This pre-mRNA is processed to mRNA which is translocated from the nucleus to the cytoplasm. The mRNA is then used as template for protein synthesis (Figure 3.2.1).

3.2.1 Transcription in eukaryotic cells

In eukaryotes, genes are transcribed into mRNA by RNA polymerase II (RNA pol II). A major difference to bacteria is the organization of DNA in nucleosomes. This leads to the requirement of several factors making the DNA accessible for transcription. Many transcription factors are in fact chromatin modifying proteins. The process of transcription is grossly divided into three phases: 1. Initiation, 2. Elongation, and 3. Termination. RNA transcripts are modified co- and post-transcriptionally by adding a 7-methyl-guanosine cap at the 5'-end, a poly A tail at the 3'-end, and by a genetic re-organization called splicing. This processing of mRNA is regulated to great extent by a repetitive structure at the C-terminal part of the largest RNA polymerase II subunit Rpb1 (CTD).

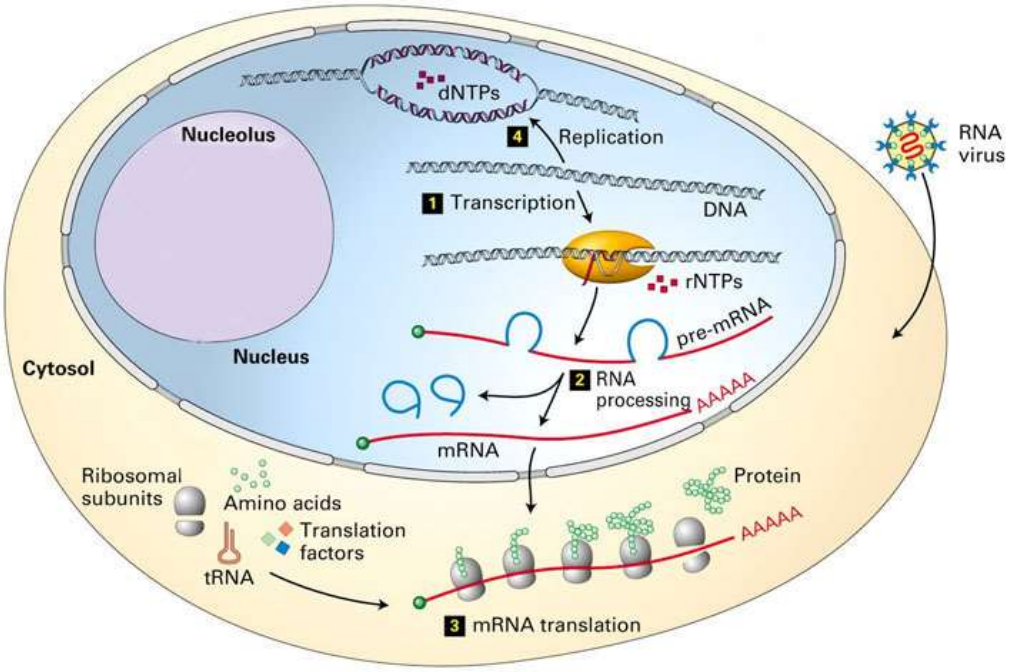


Figure 3.2.1: Gene expression in eukaryotic cells.
Gene expression is a multistep process. (1) Transcription: The gene is transcribed into a pre-mRNA copy. (2) pre-mRNA processing: The pre-mRNA is subjected to splicing and other modifications. (3) mRNA translation: Processed mRNA is translocated from the nucleus to ribosomes in the cytosol. Here, the protein encoded by the gene transcribed is translated from the mRNA. (4) Replication: DNA serves as storing molecule for genetic information and is replicated prior to cell division. Modified from: (Lodish 2008) .

Initiation of transcription in eukaryotes requires several factors

Initiation of transcription requires the action of the mediator complex and several general transcription factors. These factors are termed general transcription factors (GTFs) since they are required for expression of all genes in contrast to specific transcription factors, which act on selective genes. Mediator and GTFs act in a sequential manner and mediate changes in chromatin structure, recruit RNA pol II and unwind the DNA for transcription. The mediator complex opens the chromatin structure and interacts with RNA pol II. The sequential action of the GTFs starts with TFIID (transcription factor D of RNA pol II), a multimeric complex which contains a TATA binding protein (TBP) interacting with a specific sequence motif upstream of the gene (TATA box). TBP modifies the DNA structure by inserting a β-strand into the DNA's minor groove. TFIIB interacts with the major groove upstream of the preformed complex. The asymmetric binding of TFIIB to TBP is important to determine directionality of transcription. TFIIF is associated with RNA pol II. TFIIE is required for binding of TFIIH. TFIIH is a multiprotein complex including helicase activity and Cyclin-dependent kinase 7 (Sainsbury et al. 2015).

The action of RNA polymerase II is organized in a transcription cycle

Transcriptional activity can be observed to follow a cyclic process (Figure 3.2.2)(Fuda et al. 2009). Regulation of this cycle is mediated by a subset of transcriptional CDKs (Cdk7, Cdk8, Cdk9, Cdk11, Cdk12, Cdk13). As a first step the nucleosome structure has to unwind to allow binding of further factors. Then a pre-initiation complex is formed by the mediator and general transcription factors.

These act sequentially as described and finally lead to the clearance of RNA pol II from the promoter. This promoter clearance requires Cdk7 kinase activity. RNA pol II starts transcription but is paused after an initial synthesis of 10-50 nucleotides. To overcome this state of promoter proximal pausing and to engage into the productive elongation phase, it is necessary that both the factors inhibiting elongation and the RNA pol II C-terminal domain (CTD) become phosphorylated by Cdk9/CycT1, the positive transcription elongation factor b (P-TEFb). P-TEFb phosphorylates the negative elongation factor (NELF) and the DRB sensitivity inducing factor (DSIF) as well as RNA pol II. After entering the productive transcription elongation phase RNA pol II runs through the entire gene body until polymerase reaction is terminated. Upon termination, RNA pol II is released and can be utilized for another transcription cycle. In addition to this simplistic view it has been shown that RNA pol II stalls at several points throughout the gene (Core und Adelman 2019). Moreover, several co-transcripitonal events as mRNA capping, or cleavage and polyadenylation are regulated by CDKs.

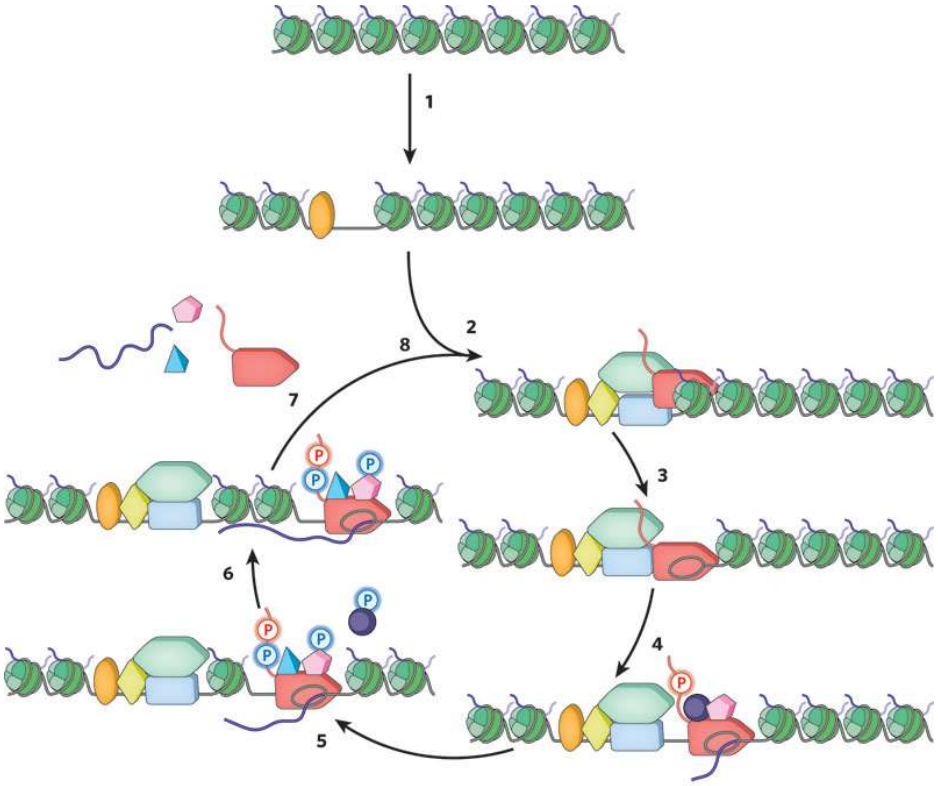


Figure 3.2.2: The transcription cycle in eukaryotic cells.
Transcription is a multistep process. Steps are described in detail in the text. In keywords: 1. Chromatin opening. 2. Formation of the pre-initiation complex. 3. Initiation. 4. RNA pol II release from promoter. 5. Escape from promoter proximal pausing. 6. Productive elongation. 7. Termination and release of RNA pol II. 8. Recycling of RNA pol II (Fuda et al. 2009).

3.2.2 The C-terminal domain of the RNA polymerase II subunit Rpb1

Many processes of transcription are mediated by the C-terminal domain of RNA polymerase II (CTD). The CTD is a repetitive structure of the seven amino acids Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 at the Rpb1 subunit. The length of the CTD differs among organisms. Human CTD consists of 52 heptad repeats, whereas in yeast only 26 repeats are found (Figure 3.2.3). In human, the N-

terminal part the consensus sequence YSPTSPS is highly conserved, while the C-terminal part shows more alterations from the consensus sequence. Most prominent is a change at position Ser7 to lysine where half of the serines are substituted by other amino acids. The last repeat is extended to 17 amino acids and was shown to be important for CTD stability *in vivo (Chapman et al. 2008)*.

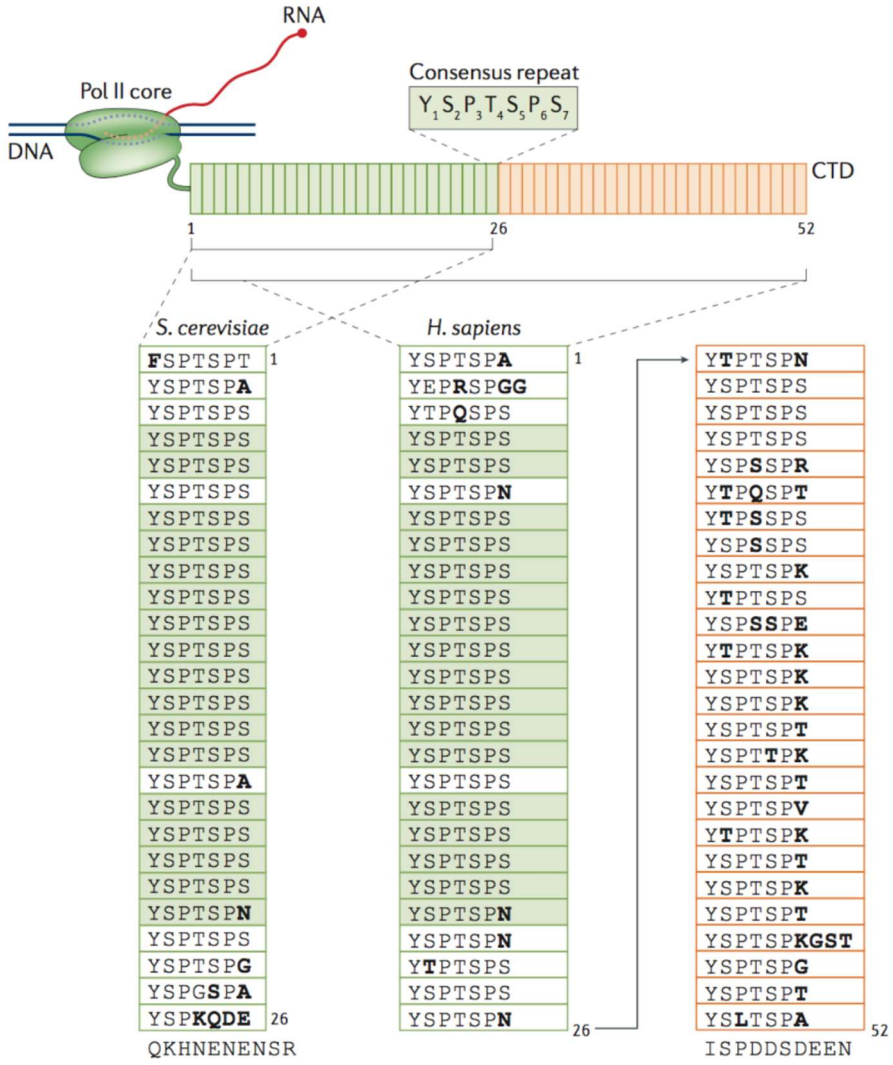


Figure 3.2.3: CTD sequence of human and yeast RNA polymerase II C-terminal domain.
Human CTD consists of 52 repeats while yeast CTD harbors 26 repeats. The first 26 CTD repeats are conserved in humans. Variations of the consensus sequence YSPTSPS are most frequently found in the distal part of the CTD (Harlen and Churchman 2017).

Studies in yeast showed, that severe disruption, but not deletion of single repeats lead to growth defects. It was showed, that substitution of Ser2, Ser5, and Ser7 to glutamate is lethal, while substitution of Tyr1 has different effects dependent on the yeast strain (West and Corden 1995; Zhang et al. 2012; Schwer and Shuman 2011). The CTD can be modified in many ways including phosphorylation, methylation, acetylation, and glycosylation, with phosphorylation being the probably most frequent and best studied modification. Specific phosphorylation of the CTD can be mapped to specific states of the transcription cycle by ChIP experiments (Figure 3.2.4).

Several transcriptional CDKs have been found to phosphorylate the CTD but the significance of this is unclear and many studies suggest a redundancy among these kinases for RNA pol II CTD phosphorylation. In recent years, several studies identified other substrates than the RNA pol II of the transcriptional kinases Cdk9 and Cdk12. These substrates comprise many transcription related factors, which highlights the requirement of these kinases at multiple steps of the transcription cycle (Vos et al. 2018; Sansó et al. 2016; Krajewska et al. 2019).

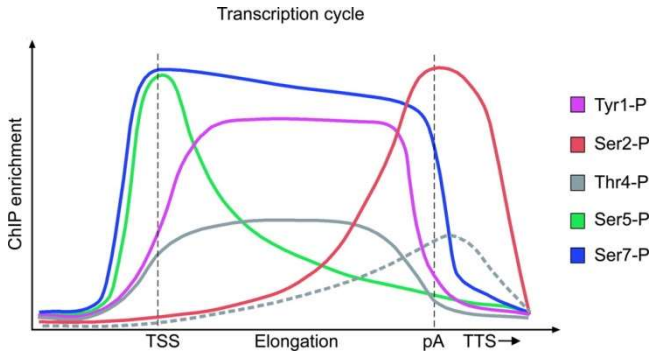


Figure 3.2.4: Phosphorylation status of the CTD during transcription.
Phosphorylation of specific residues within a CTD repeat were mapped to the position of RNA pol II at the gene. TSS – transcription start site. pA – poly A. TTS – Transcription termination site (Eick and Geyer, 2013).

3.3 Kinases

Eukaryotic proteins are subject to post-translational modifications. These modifications often affect protein function or localisation expanding the mechanistic toolbox of the human proteome. A common and widespread modification is phosphorylation which can affect threonine, serine or tyrosine residues. The reversible nature of a phosphorylation mark makes it an ideal modification for the regulation of transient processes. In fact, phosphorylation and dephopshorylation is the principle of many signalling cascades. All protein kinases catalyse the transfer of a gamma-phosphoryl group of ATP to a hydroxyl group of serine, threonine or tyrosine. Despite utilizing the same catalytic mechanism and sharing of several conserved elements protein kinases differ in their ability to either phosphorylate serine and threonine or tyrosine residues and are thus mechanistically subdivided into serine/threonine and tyrosine directed kinases.

Based on sequence analysis, the human kinome of protein kinases consists of 518 kinases which makes up 1.7% of the human genome (Figure 3.3.1). Among these 518 protein kinases, 478 can be identified by sequence similarity of a common eukaryotic protein kinase domain. The remainder 40 kinases are regarded as atypical kinases as they differ in sequence but were shown to possess kinase activity experimentally. Eukaryotic protein kinases are subdivided into nine different groups eight groups are built by the conserved kinase families tyrosine kinases (TK), tyrosine kinase like (TKL), Homologues of yeast sterile 7, 11, and 20 (STE), receptor guanylate cyclases (RGC), Cdk, MAPK, GSK3, and Clk (CMGC), calcium/calmodulin-dependent kinases (CAMK), protein kinase A, G, and C (AGC), casein kinase 1(CK1). The ninth group is composed of atypical protein kinases (not shown). The group of RGC kinases does not represent active kinases. Moreover, ATP binding by the kinase domain might assist guanylate-cyclase activity (Manning et al. 2002).

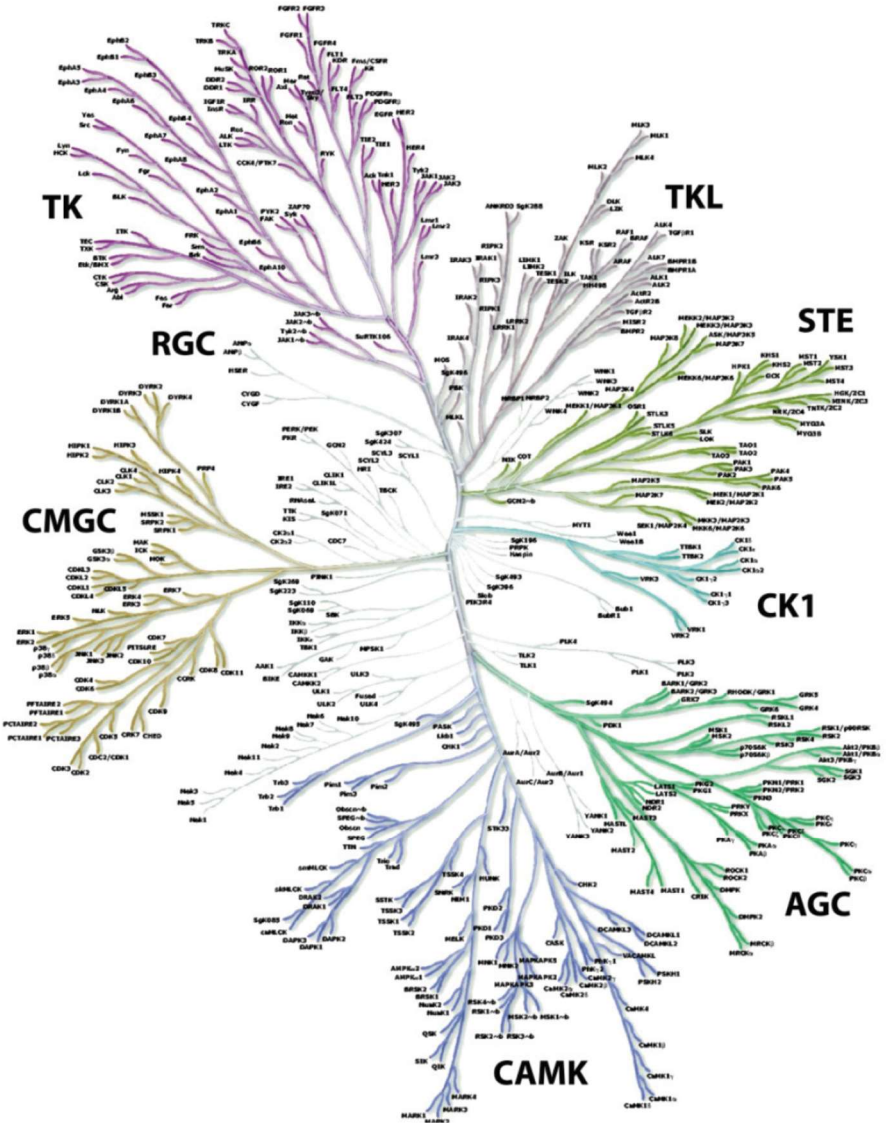


Figure 3.3.1: The human kinome.
Kinases were grouped based on sequence similarities. TK – tyrosine kinases; RGC – receptor guanylate cyclases; CMGC – Cdk/MAPK/GSK3/Clk; CAMK – Calcium/Calmodulin dependent kinases; AGC – protein kinase A, G, and C; CK1 – casein kinase1, STE – sterile of yeast homolog 7; TKL – tyrosine kinase like. The atypical protein kinases which cannot be identified by sequence similarity are not depicted (modified from: Manning et al. 2002).

3.3.1 Cyclin-dependent kinases (CDKs)

Cyclin-dependent kinases (CDKs) belong to the class of kinases termed CMGC kinases comprising CDKs, MAPK, GSK and CLKs and have initially been described as regulators of the cell cycle. Studies in yeast identified kinases which are activated by oscillating regulatory subunits – the cyclins – expression of which is related to certain cell cycle stages. The same concept can be found in mammalian cells, yet the number of the cyclin-dependent kinases and cyclins has increased. Central for all cyclin-dependent kinases is the requirement of a regulatory subunit, the cyclin, for activation of the kinase. According to sequence homology the group of CDKs comprises 21 family members (Figure 3.3.2). Among the cyclin-dependent kinases Cdk1, Cdk2, Cdk4, and Cdk6 have well established roles during the cell cycle. Another subgroup of CDKs (Cdk7-13, 19, 20) is emerging as regulators of transcription (Malumbres 2014). A well described role in transcription is established for Cdk7 as part of TFIIH, Cdk8/Cdk19 as kinase module of the mediator complex, and Cdk9 as P-TEFb. Cdk12 and Cdk13 are emerging as transcriptional kinases with roles in transcription elongation and mRNA processing. Cdk11 is less well studied but was implicated in mRNA splicing (Trembley et al. 2002). The role of other CDKs is poorly understood and has not been subject to intensive studies in the past. CDKs are serine/threonine directed kinases with a strong preference for proline in the +1 position. The empirically determined consensus sequence of cell cycle CDKs is (S/T)Px(R/K). However, recent findings show that CDKs involved in transcription do not strictly depend on this sequence (Sansó et al. 2016; Krajewska et al. 2019).

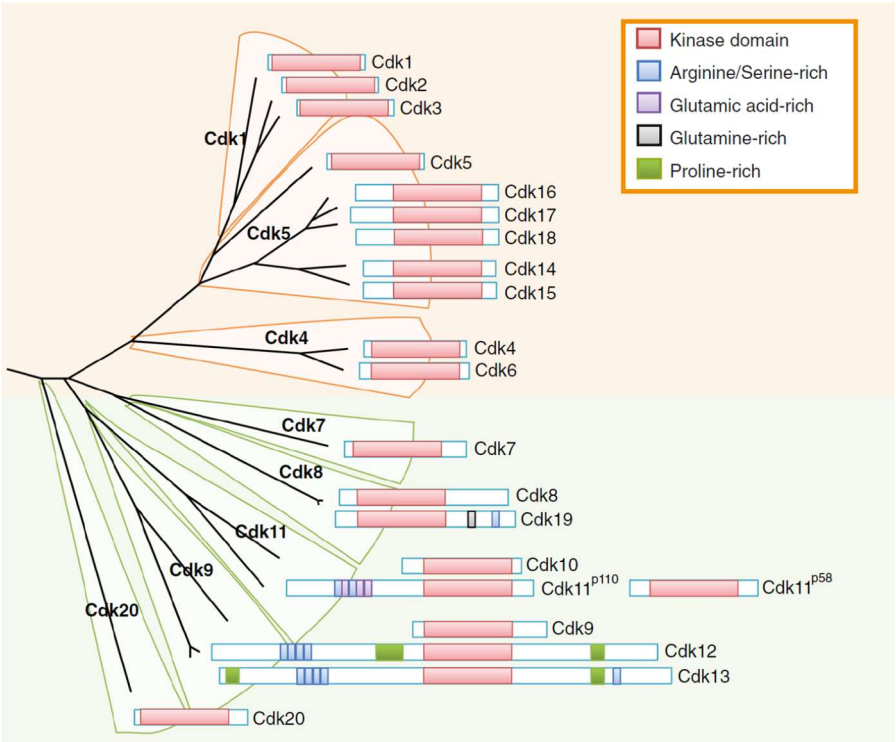


Figure 3.3.2: Cyclin-dependent kinases.
Cyclin-dependent kinases were grouped based on sequence homology and grossly divided into cell cycle related CDKs (CDKs 1-6 and CDKs 14-18) and transcription related CDKs (CDKs 7-13 and CDKs 19, 20). The twenty different CDKs are encoded by twenty-one genes. Two genes exist for Cdk11 (Cdc2L2 and Cdc2L1) which encode two nearly identical Cdk11 proteins Cdk11A and Cdk11B, respectively (not shown in the figure). The truncated $Cdk11^{p58}$ isoform originates from an internal ribosomal entry site of the $Cdk11^{p110}$ isoform (Malumbres 2014).

3.3.2 Structural features of CDKs

All protein kinases share a common architecture comprising an N-terminal and an C-terminal lobe which are connected by a flexible linker (Endicott et al. 2012). In between these two lobes resides the active centre of the kinase. Upon activation a distinct α-helix from the N-lobe is repositioned allowing the N-lobe and the C-lobe to move in relation to each other providing space for substrate binding. In CDKs, the αC-helix is interacting with the cyclin. Upon cyclin binding, the αC-helix is rearranged promoting an open conformation of the kinase active site. Cyclins are classified by their cyclin boxes which form five helices each and mediate the interaction with the kinase. The DFG motif chelates the Mg^{2+} ions required for stabilization of the negatively charged ATP. In addition, a glycine rich loop is involved in stabilization of the ATP (Wood und Endicott 2018). Cyclin-dependent kinases have an activation segment which needs to be phosphorylated for full activity. As the phosphorylated residue is a threonine in most cases this segment is called T-loop. Upon phosphorylation it coordinates three arginine residues from different regions (Arg50, Arg126 and Arg150 in Cdk2) (Russo et al. 1996b). This structural rearrangement is required for full activity. In addition to the active site, which structurally favours a proline in +1 position, allosteric substrate recognition contributes to efficient and site directed phosphorylation. Structures of transcriptional CDKs revealed the existence of an extension helix, and basic cluster which contribute to substrate specificity and activity.

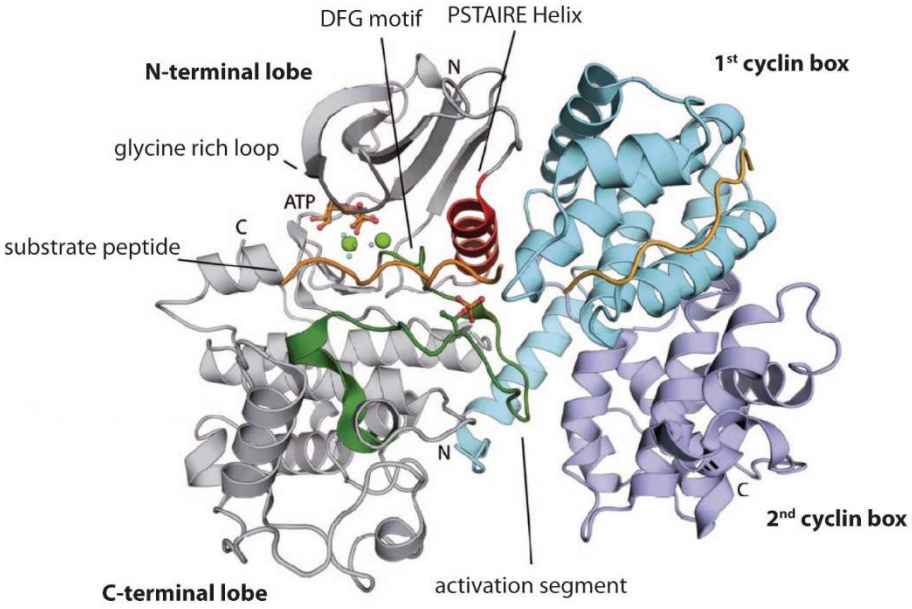


Figure 3.3.3: Structure of the Cdk2/CycA2 complex.
Structure of the Cdk2/CycA2 complex with substrate peptide (PDB: 3QHR, Bao et al. 2011). Conserved structural features are highlighted and labelled (modified from: Bösken 2013).

3.3.3 Regulation of Cyclin-dependent kinases

Cell cycle CDKs are known to be regulated by several means. A general mode of regulation is represented by the timely controlled expression and degradation of the cyclin subunit. In addition to cyclin binding, the phosphorylation of the Cdk activation segment (T-loop) is commonly required for full activity. The order of events in which the kinase is activated differs among the cyclin-dependent kinases. Whereas for Cdk2 it was described that T-loop phosphorylation improves cyclin binding, Cdk4 and Cdk6 are only efficiently phosphorylated when already in complex with their regulatory Cyclin D. In addition to the activating phosphorylation in the T-loop, inhibiting phosphorylations at allosteric sites have been described. Phosphorylation of tyrosine residues within the glycine rich loops of Cdk1 and Cdk2 by Wee1 kinase inhibit Cdk activity. Apart from post-translational modifications Cdk/Cyclin complexes are regulated by interaction with proteins of the CIP/KIP and INK family. INK4 binds Cdk6 and reduces its affinity to ATP, moreover it prevents binding to the cyclin subunit (Brotherton et al. 1998). Inhibitors of the CIP/KIP family bind to the cyclin and Cdk subunit and insert a small helix into the ATP binding pocket preventing ATP binding and imposing conformational changes inhibiting the kinase (Russo et al. 1996a).

In contrast to cell cycle CDKs 1, 2, 4, and 6 of which the regulatory cyclin subunit is known to be differentially expressed and degraded during progression of the cell cycle, transcriptional CDKs are thought to stably associate with their cyclins. Moreover, the classical Cdk inhibitory proteins of the CIP/KIP and INK family do not bind to these kinases and inhibiting phosphorylation of allosteric sites has not been described yet. A way for regulation of these kinases is the aforementioned phosphorylation status of the activation loop. Actually, regulation of transcriptional CDKs by alterations of the T-loop phosphorylation *in vivo* has been found for Cdk9 and Cdk7 (Larochelle et al. 2012a; Schachter et al. 2013). It is likely that also other CDKs are regulated at the level of T-loop phosphorylation but has not experimentally established yet. The only transcriptional Cdk with a well described negative regulation mechanism is Cdk9. Inactivation of Cdk9/CycT1 requires association with Hexim which tethers Cdk9/CycT1 to 7SK RNA and sequesters Cdk9 in an inactive state (Quaresma et al. 2016). However, regulatory factors for the other transcription associated kinases Cdk8, 10, 11, 12, and 13 are less well studied or have not even been identified yet.

3.4 Cdk7/CycH/Mat1

Cdk7 research dates back until 1990 with the description of MO15 (Cdk7) as a cdc2-related kinase (Shuttleworth et al. 1990). It was soon recognized, that MO15 regulates cell cycle activity by activating other cyclin-dependent kinases and thus represents a human Cdk-activating kinase (CAK) (Fesquet et al. 1993; Poon et al. 1993; Solomon et al. 1993). In 1994, Cyclin H was identified to associate with MO15 leading to re-naming of MO15 in Cdk7 (Fisher und Morgan 1994). In the following years it was established that Cdk7/CycH associates with a third protein, *ménage à trois* protein 1 (Mat1), and that this ternary complex is part of the general transcription factor TFIIH (Serizawa et al. 1995; Devault et al. 1995; Fisher et al. 1995; Shiekhattar et al. 1995; Tassan et al. 1995; Yee et al. 1995). Follow up studies in yeast and human cells confirmed its roles in cell cycle and transcription. Recently, Cdk7 emerged as a therapeutic target in cancer therapies (Kwiatkowski et al. 2014). Despite the longstanding interest in Cdk7 biology, no structure of either the binary Cdk7/CycH complex or the ternary Cdk7/CycH/Mat1 complex has been determined yet. Structural

information is available for individual parts as monomeric Cdk7 (PDB: 1UA2) (Lolli et al. 2004) and monomeric Cyclin H (PDB: 1JKW; 1KXU) (Andersen et al. 1997; Kim et al. 1996). Moreover, an NMR structure of the Mat1 N-terminus (amino acids 1-65) is available (PDB: 1G25) (Gervais et al. 2001). Recent studies elucidating the structure of the full TFIIH complex of both human and *S. cerevisiae* revealed that Mat1 forms a long helix which contacts the ARCH domain of XPD. However, these studies were not able to resolve the Cdk7/CycH/Mat1 kinase module of the complex (PDBs: 5OQM; 6NMI) (Schilbach et al. 2017; Greber et al. 2017).

3.4.1 Cdk7 as Cdk-activating kinase (CAK)

Full activation of cyclin-dependent kinases requires phosphorylation within the activation loop of the kinase. A Cdk-activating kinase (CAK) activity was purified from starfish oocytes and identified to be highly similar to Xenopus cDNA clone MO15 (Cdk7) (Fesquet et al. 1993). Subsequent studies confirmed Cdk7 CAK activity *in vitro* and *in vivo*. Interestingly, CAK activity is differentially executed in budding and fission yeast. Kin28, the *S. cerevisiae* Cdk7 homologue does not possess any CAK activity. Instead, CAK function is carried out by a separate kinase, Cak1, which has no homologous counterpart in mammals. In contrast, the fission yeast Cdk7 counterpart Mcs6 has CAK activity, but is complemented by Csk1 which is capable to activate Mcs6 and Cdk1/CycB (Figure 3.4.1) (Reviewed in: Fisher 2005).

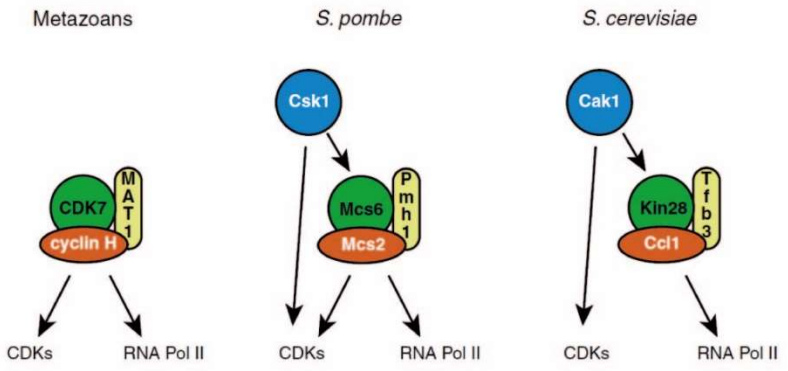


Figure 3.4.1: Dual functions of human Cdk7 as CAK and in transcription.
In metazoans Cdk7 fulfils tasks as Cdk-activating kinase and in transcription of which its functions in transcription are conserved in *S. pombe* and *S. cerevisiae*. In *S. pombe* CAK function is partly executed by Csk1 and Cdk7. In *S. cerevisiae*, the Cdk7 homologue Kin28 does not possess any CAK activity but contains a separate enzyme, Cak1, for CAK function (Fisher 2005).

In human cells, Cdk7 is the only well confirmed cdk-activating kinase but a recent study suggests Cdk4 activation by JNK1 (Colleoni et al. 2017). Cdk7 was shown to activate Cdk1, Cdk2, Cdk4, Cdk6, and Cdk9 *in vitro* and *in vivo* (Larochelle et al. 2012a; Merrick et al. 2008; Larochelle et al. 2007; Schachter und Fisher 2013; Larochelle et al. 1998; Garrett et al. 2001). That other cyclin-dependent kinases are activated by Cdk7 seems likely but lacks experimental confirmation. The fact that Cdk7 cannot phosphorylate itself within its activation segment immediately prompted to the question which kinase holds responsible for Cdk7 activation. *In vitro*, Cdk1 and Cdk2 have been shown to possess Cdk7 pThr170 directed activity but a kinase phosphorylating Cdk7 *in vivo* has not been identified yet (Martinez et al. 1997).

The fact, that Cdk7 is unable to phosphorylate itself moreover highlights, that recognition of the Cdk activation segment is not primarily dependent on the underlying sequence, since Cdk T-loops are similar. This idea was further substantiated experimentally by using Cdk T-loop chimera. If the Cdk2 T-loop was placed at Cdk7, the Cdk2 T-loop sequence was no longer phosphorylated by wild-type Cdk7. In the reverse situation in which a Cdk7 T-loop was inserted into Cdk2, the previously not recognized Cdk7 T-loop sequence was phosphorylated by Cdk7 (Garrett et al. 2001). It is also worth mentioning, that most Cdk T-loops do not contain a proline in +1 position of the activatory serine or threonine site and thus do not represent the minimal consensus motif (S/T)P. Mutagenesis and molecular docking studies suggest a head to tail interaction of Cdk7 to Cdk2 and identified Cdk2 Leu 166 as important for efficient Thr161 phosphorylation by Cdk7. This leucine is conserved among Cdk1, 2, 4, 5, and 6 but not in Cdk7. Other Cdk2 residues which are distant from the activation loop but which affect T-loop phosphorylation are K9, K88, and K89 (Lolli und Johnson 2007).

3.4.2 Cdk7 in transcriptional regulation

Cdk7 is part of the general transcription factor TFIIH and Cdk7 activity has shown to be critical for promoter clearance. Cdk7 was found to phosphorylate the RNA pol II CTD at serine 5 and serine 7 (Akhtar et al. 2009). Cdk7 mediated CTD phosphorylation has also shown to positively affect mRNA capping (Nilson et al. 2015). Moreover, Cdk7 activates Cdk9 resulting in indirect regulation of transcription elongation (Larochelle et al. 2012b). Recently it was shown that TFIIE can recruit Cdk7 independent of TFIIH adding another layer of complexity to Cdk7 involvement in transcription (Compe et al. 2019).

Despite the association with TFIIH and the requirement of Cdk7 activity for transcription in *in vitro* reconstituted systems it is currently not clear to what extend Cdk7 affects global transcription levels. Studies in yeast revealed different Cdk7 dependencies. Analysis of transcription in *S. cerevisiae* after thermal inactivation of a temperature sensitive Kin28 mutant shows a global reduction of RNA pol II transcripts (Holstege et al. 1998). In contrast, inactivation of Mcs6 in *S. pombe* affected only a small portion of all transcripts but had a large impact on cell cycle dependent genes (Lee et al. 2005; Rustici et al. 2004). These data were reconciled recently upon specific inhibition of Cdk7 in HAP1 cells, showing that the loss of Cdk7 activity in dividing cells mainly affects transcription of cell-cycle regulated genes (Olson et al. 2019). Importantly, Cdk7 inhibition results in cell cycle arrest at G1 in these cells. It is therefore not clear if these genes are directly affected by loss of Cdk7 activity or by secondary effects induced by the cell cycle arrest (Olson et al. 2019). It remains an interesting question how transcription in terminally differentiated cells, like neurons, responds to Cdk7 inhibition.

TFIIH is not only involved in transcription initiation, but also in nucleotide excision repair (NER). Cdk7 activity is dispensable for this function and the CAK module dissociates from NER-active complexes upon XPA binding (Coin et al. 2008). The cryo-EM structure of the NER complex shows, that the NER factor XPA alters the conformation of the XPD ARCH domain and thereby frees CAK from the complex. In the reverse, excess of Cdk7/CycH/Mat1 reverted the association of TFIIH with XPA and reduced NER activity (Kokic et al. 2019).

3.4.3 Regulation of Cdk7 activity

Cdk7 in complex with TFIIH was found to exhibit increased Cdk7 activity towards RNA pol II CTD (Yankulov und Bentley 1997; Rossignol et al. 1997) and subsequent studies demonstrated, that association of the C-terminal part of Mat1 with Cdk7/CycH is sufficient for this effect (Busso et al. 2000). However, there are no data indicating a context dependent, reversible association of Mat1 with Cdk7/CycH. Mat1 appears to be stably associated and thus rather represents a scaffolding protein than a true regulatory subunit (Figure 3.4.2). The TFIIH subunit XPD is known to form a quaternary complex together with Cdk7/CycH/Mat1. Early studies identified XPD interaction as a way to regulate Cdk7 CAK activity (Chen et al. 2003). Follow up experiments suggest that this effect is primarily depending on shuttling Cdk7 away from its cell cycle targets and not by directly affecting Cdk7 activity (Stettler et al. 2015; Li et al. 2010).

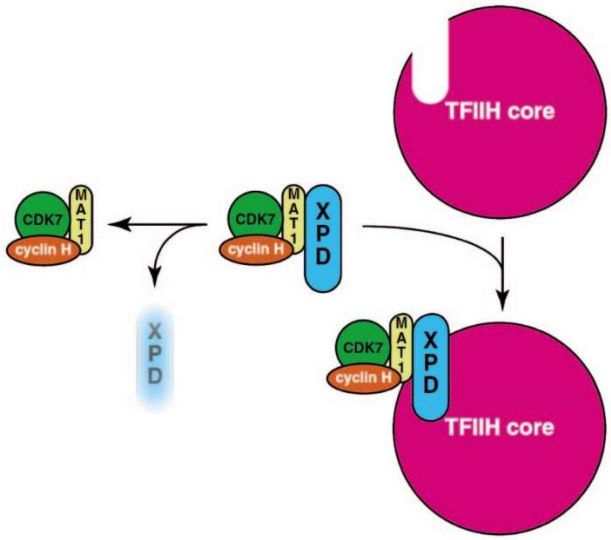


Figure 3.4.2: Cdk7 containing complexes in human cells.
Cdk7 is found in three different complexes in human cells. A trimeric Cdk7/CycH/Mat1 complex, a quaternary complex of Cdk7/CycH/Mat1 with XPD, and associated with TFIIH.

A common way to regulate enzymatic activities are post-translational modifications. In fact, cyclin dependent kinases are commonly regulated by inhibitory and activatory phosphorylations. Similar to other Cyclin-dependent kinases, except Cdk8 and Cdk19, Cdk7 is phosphorylated in its activation segment at Thr170. Cdk7 Thr170 phosphorylation is crucial for Cdk7 binding to Cyclin H. Interestingly, Mat1 association circumvents the requirement of Cdk7 Thr170 phosphorylation for Cyclin H binding and activity (Fisher et al. 1995). Even if Thr170 phosphorylation is dispensable for activity in presence of Mat1 it could be demonstrated that Thr170 phosphorylation increases Cdk7/CycH/Mat1 activity in a substrate dependent manner. The Cdk7 CAK activity remained constant regardless of the Cdk7 Thr170 phosphorylation status, whereas phosphorylation of RNA pol II CTD is drastically increased. This has led to the idea, that Cdk7 Thr170 phosphorylation serves as a regulatory switch to specifically increase Cdk7 activity towards its transcriptional substrates. This idea was further supported by descriptions, that TFIIE and SPT5 phosphorylation by Cdk7 is also affected by Mat1 and Cdk7 Thr170 phosphorylation whereas Cdk11 activation is not (Larochelle

et al. 2006). However, the strict view of increasing transcriptional activity over CAK activity is challenged by the observation, that Cdk4 phosphorylation and thus activation is similarly regulated by Cdk7 Thr170 phosphorylation (Schachter et al. 2013). In addition, old studies suggested an equal increase of Cdk2 phosphorylation and RNA pol II CTD peptides by Mat1 (Busso et al. 2000). In the past 25 years several Mat1 dependent Cdk7 substrates have been described including retinoic acid receptor and p53 (Rochette-Egly et al. 1997; Ko et al. 1997). It remains an open question, if regulation of Cdk7 activity against all Mat1 dependent substrates requires phosphorylation of Thr170. However, the significance of Thr170 phosphorylation in cells is still debated. Cdk7 p170 levels are not globally fluctuating in response to cell cycle progression or external stimuli. Only a small fraction of probably chromatin bound Cdk7 was found to be dephosphorylated upon serum starvation (Schachter et al. 2013).

In addition to threonine 170, Cdk7 contains another phosphorylation site at serine 164 in its activation loop. All Cdk activation segments contain serines N-terminal of the activatory residue, but serine 164 in Cdk7 has emerged interest for two reasons: First, it is located in the sequence SPNR (single letter code) reflecting a perfect CDK/MAPK consensus site and can indeed be phosphorylated by Cdk2 *in vitro* which results in improved binding to Cyclin H. (Martinez et al. 1997). Second, phosphorylation at Ser164 has been shown to occur in cells and is highly enriched in nocodazole arrested cells (Akoulitchev und Reinberg 1998). A detailed study in *Drosophila melanogaster* revealed, that Cdk7 Ser164 phosphorylation levels changes during embryonal development (Larochelle et al. 2001). Additionally, Cdk7 S164A mutant flies exhibited temperature sensitivity. However, data interpretation was difficult since Cdk7 S164A mutation also resulted in reduced levels of Thr170 phosphorylation, supporting the idea that Cdk7 Ser164 phosphorylation is stabilising dimeric and trimeric Cdk7 complexes. In 1998, two groups independently discovered that transcription associated Cdk7 activity is reduced in mitosis (Akoulitchev und Reinberg 1998; Long et al. 1998). The laboratory of Danny Reinberg assigned this effect to phosphorylation of Cdk7 at Ser164 which correlates with high Ser164 phosphorylation levels in nocodazole treated cells which are arrested in mitosis. Even though this model is appealing as it provides a regulatory option to specifically inhibit TFIIH associated Cdk7 activity, it does not fit with biochemical and genetic observations. Another proposed mechanism to regulate Cdk7 activity is phosphorylation of its cyclin subunit. Cdk7/CycH autophosphorylates Cyclin H *in vitro* (Lolli et al. 2004). Also the kinases Cdk8 and casein kinase 2 have been found to phosphorylate Cyclin H *in vitro* (Akoulitchev et al. 2000; Schneider et al. 2002) but the significance of these observations lacks confirmation *in vivo*.

3.4.4 Implications for Cdk7 in cancer

Because of its central involvement in cell cycle regulation and transcription, inhibition of Cdk7 as a treatment for cancer has been discussed early on but has always been dealt with care as cdk7 inhibition would also affect healthy cells. Initial ideas hypothesized the development of inhibitors which might affect only Cdk7 CAK function but not interfere with its role in transcription. These could have impacted cell cycle progression in malignant cells without affecting transcription in quiescent or terminally differentiated cells. Even if genetics enables to generate Cdk7 mutants which are devoid of CAK function in *Drosophila* (Cdk7 P140S) but remain transcriptionally active (Larochelle et al. 1998), no small molecule inhibitors have been developed that specifically target Cdk7 CAK function.

Instead, direct inhibition of cell cycle CDKs Cdk1, Cdk2, Cdk4, and Cdk6 has evolved as a therapeutic option over the past decades.

Interest in Cdk7 inhibition has re-emerged recently by the concept of transcriptional addiction in cancer (Bradner et al. 2017). It was found that several cancers highly depend on the expression of growth promoting factors and are thus vulnerable to inhibition of pathways that are required for transcription. This concept immediately raised interest in transcriptional cyclin-dependent kinases. In 2014, the group of Nathanael Gray at the Dana-Farber Cancer Center in Boston described the development of a covalent inhibitor, THZ-1, which potently and selectively inhibits Cdk7, Cdk12, and Cdk13 (Kwiatkowski et al. 2014). In this and in subsequent studies it was establish, that interference with the transcriptional machinery affects super-enhancer driven tumors via RUNX1 in T cell acute lymphoblastic leukemia and via MYCN in neuroblastoma (Chipumuro et al. 2014). Further development of THZ-1 resulted in THZ-531 which is highly selective for Cdk12 and Cdk13 but lacks inhibitory potential towards Cdk7 (Zhang et al. 2016). Studies using THZ-531 to dissect the roles of Cdk7 and Cdk12/13 in cancer revealed, that combinatorial inhibition of Cdk12/13 and Cdk7, but not either of the kinases alone, is required to target MYC dependent ovarian cancer (Zeng et al. 2018). During the course of this thesis the Gray lab developed the potent covalent Cdk7 inhibitor YKL-5-124, which was highly specific for Cdk7 compared to Cdk12 or Cdk13 in biochemical assays with recombinant proteins (Olson et al. 2019). Recently, the first clinical trial of a Cdk7 inhibitor (SY-1365) was launched (Hu et al. 2019).

3.5 Cdk10/Cyclin M

Even though Cdk10 was identified 25 years ago little is known about its function. A reason for this is probably the lack of a Cdk10 homologue in yeast which prevented easy genetic ablation or modification in early years. In 2013 the gene product of FAM58A was identified as the cyclin partner of Cdk10 and subsequently termed Cyclin M. During the course of this thesis the recommended name for Cyclin M was changed to Cyclin Q (uniprot accession code: Q8N1B3). Throughout this thesis Cyclin M will be used as it is still more common in the literature.

CDK10 (PISSLRE) is located on chromosome 16 (Bullrich et al. 1995). The Cdk10 gene contains 10 exons and was identified in 1994 by two groups upon PCR based screens trying to identify cdc2 related protein kinases in human cells (Brambilla und Draetta 1994; Graña et al. 1994). Two isoforms can be found according to their studies, a full length Cdk10 isoform of 360 amino acids and another N-terminal truncated and C-terminal modified of 316 amino acids. Despite detection of these isoforms on mRNA level in PCR based studies evidence for the existence of the latter isoform on protein level is scant. Moreover, data from a yeast two hybrid screen suggested that the truncated isoform 2 is not able to interact with Cyclin M and thus is very likely to not form a functional kinase-cyclin complex (Guen et al. 2013). Early studies suggested cell cycle specific expression in G2/M phase, but these data were purely PCR based and require a detailed re-investigation on protein level (Li et al. 1995).

Only few studies addressed the cellular roles of Cdk10 until today but mutations in Cdk10 were found in a small group of individuals with developmental deficits (Windpassinger et al. 2017). Genetic analysis revealed heterogenous mutations in the Cdk10 gene which affected splicing of the transcribed mRNA. It is hypothesized that these alterations result in non-functional or less functional Cdk10 variants. Subsequent genetic ablation in mouse was found to be embryonal lethal, but mouse embryonal fibroblasts cultured from these animals showed no apparent growth defects. Additionally, knock out of Cdk10 in adult mice using an inducible Cre-deleter system was not associated with severe defects. However, embryonal fibroblasts responded differently to growth factor suppression by serum deprivation.

Morpholino based studies in zebrafish revealed deficits in brain development undermining a developmental function of Cdk10 (Yeh et al. 2013). The best studied mechanism is the promotion of the proteasomal degradation of the transcription factor ETS-2 by Cdk10 mediated phosphorylation (Liu et al. 2012). Accordingly, most studies identify an increase in ETS-2 levels after interference with Cdk10 activity. Cdk10 was found to be downregulated in breast cancer resistant to tamoxifen treatment. It has been suggested that resistance is a consequence of increased ETS-2 levels and subsequent activation of the MAPK pathway (Iorns et al. 2008). Apart from ETS-2, Cdk10 was found to phosphorylate and thereby downregulate the activity of the kinase PKN2. The resulting alterations in actin cytoskeleton organization are found to affect the formation of the primary cilia. Cdk10 deficient cells possess prolonged cilia upon serum starvation. Moreover, Cdk10 was found to be translocated to the cilia body after serum starvation (Guen et al. 2016). As alterations in ETS-2 levels as well as dysfunction of primary cilia have widespread functional consequences the exact role of Cdk10 in the observed phenotypes is still elusive.

The FAM58A gene (Cyclin M) is located at the X-chromosome and mutations of this gene lead to a rare X-linked gonosomal syndrome called STAR, an acronym of the phenotype of the disease which is characterized by toe **s**yndactyly, **t**elecanthus, **a**nogenital, and **r**enal malformations. It is anticipated that the described mutations result in lost or impaired function of the Cdk10/CycM kinase complex, but it is possible that other, Cdk10 independent mechanisms contribute to the observed phenotype. An interesting observation in this regard is that CycM knock-down results in growth deficits while genetic knock-out of Cdk10 has no apparent effect on cellular proliferation (Windpassinger et al. 2017; Unger et al. 2008).

3.6 Aims of the thesis

Despite a longstanding interest in Cdk7 biology and function its regulation remains incompletely understood. The first part of this thesis shall biochemically analyse binary Cdk7/CycH and ternary Cdk7/CycH/Mat1 complexes. In particular, the effect of Cdk7 Ser164 phosphorylation on complex formation, complex stability and kinase activity shall be analysed. Therefore, human Cdk7 is expressed in *Sf9* cells either alone or in complex with its regulatory binding partner Cyclin H and the Cdk7/CycH scaffolding protein Mat1. The effect of Cdk7 serine 164 phosphorylation will be addressed by mutational studies. To this end, Cdk7 serine 164 is mutated into the non-phosphorylatable amino acid alanine as well as into the phosphorylation mimicking amino acid glutamate. Wild type Cdk7 and Cdk7 mutants are then analysed for kinase activity in *in vitro* kinase activity assays as well as for their ability to form functional complexes with Cyclin H and Mat1. Due to the emerging interest in Cdk7 inhibition for cancer therapy, the pharmacologic potential of novel small molecule inhibitors shall be characterized.

The second part of this thesis will address functional and structural aspects of the Cdk10/Cyclin M complex. A pre-requisite for these investigations is the development of an expression and purification strategy which is applicable at laboratory scale and allows the generation of milligram amounts of pure and homogenous Cdk10/CycM complexes. Of note, Cdk10/CycM is not commercially available and thus establishing a protocol for generation of this complexes represents an important requirement for future research. The Cdk10/CycM complexes shall be functionally characterized and used for structure determination by X-ray crystallography.

Since only little is known about the biological pathways to which Cdk10/CycM contributes, an *in vitro* chemical genetic screen shall be applied to identify new substrates of Cdk10/CycM. To this end the Cdk10 gatekeeper residue methionine 117 of the Cdk10 ATP binding pocket is mutated to the small amino acid glycine. The $Cdk10^{M117G}$ mutant should be able to utilize bulky ATP-analogues as substrates which are not tolerated by native kinases. Cdk10 substrates are then identified by mass spectrometry after incubation of cell lysate with $Cdk10^{M117G}$/CycM and an N^6-modified ATP-γ-S analogue. The generated dataset will be analysed to elucidate Cdk10/CycM substrate recognition properties and Cdk10/CycM involving pathways. Putative substrates identified by the screen shall be validated in *in vitro* kinase activity assays using recombinant proteins.

These investigations will broaden our understanding of Cdk biology in general but will also provide detailed insights into the regulation of Cdk7 activity. Given the central role of cyclin-dependent kinases in many cellular processes and diseases, the information gathered from this thesis will be of interest to a broad scientific community as well as to clinically related researchers.

4 Materials and Methods

4.1 Materials

4.1.1 Chemicals

All chemicals used were of analytical grade. Chemicals of the following suppliers were used: AppliChem (Darmstadt, Germany), Becton Dickinson (Franklin Lakes, USA), Roth (Karlsruhe, Germany), and Sigma Aldrich (München, Germany).

4.1.2 Nucleotides

Nucleotide	Supplier
ATP, solid	Roth, Karlsruhe
[γ-^{32}P]-ATP,5µCi/µl, 3000 mCi/mmol	Perkin Elmer, Boston, USA
ATP-γ-S	Biolog Life Science Institute, Bremen
N^6-methyl-butyl-ATP-γ-S	Biolog Life Science Institute, Bremen
N^6-benzyl-ATP-γ-S	Biolog Life Science Institute, Bremen
N^6-phenyl-ATP-γ-S	Biolog Life Science Institute, Bremen
N^6-phenylethyl-ATP-γ-S	Biolog Life Science Institute, Bremen
N^6-fufuryl-ATP-γ-S	Biolog Life Science Institute, Bremen
N^6-cyclohexyl-ATP-γ-S	Biolog Life Science Institute, Bremen
N^6-cyclopentyl-ATP-γ-S	Biolog Life Science Institute, Bremen
dNTPs	Roche, Mannheim

4.1.3 Consumables

Consumable	Supplier
Ultrafiltration device (3k, 10k, 30k)	Millipore, Amicon, Witten
Optitran BA-S85 reinforced Membrane	Whatman, Maidstone, UK
6-well-plates	BD Falcon, Franklin Lakes, USA
Cuvette, PE (1 cm path length)	Sarstedt, Nümbrecht
Bottle Top Filter (0.2 µm)	Thermo Fisher Scientific, Waltham, USA
Bottle Top Filter (0.45 µm)	Thermo Fisher Scientific, Waltham, USA
Cuvette (electroporation)	Eppendorf, Hamburg
Reaction tubes (0.5 ml, 1 ml und 2 ml)	Eppendorf, Hamburg
Syringe filter (0.22 µm)	Roth, Karlsruhe
Syringe filter (0.45 µm)	Roth, Karlsruhe

Kits	Supplier
Qiaquick Gel Extraction Kit	Qiagen, Hilden
Qiaprep Spin miniprep Kit	Qiagen, Hilden
HiSpeed Plasmid midi Kit	Qiagen, Hilden
Qiaquick PCR Purification Kit	Qiagen, Hilden
ExtractMe, DNA purification kit	Blrt, Gdansk, Poland
Enhanced chemiluminescence substrate	Sigma-Aldrich, Steinheim

4.1.4 Chromatography Media and columns

Column	Chromatography Medium	Supplier
HisTrap crude FF	Ni^{2+} Sepharose 4 fast flow	GE Healthcare
GSTrap FF	Glutathione-Sepharose 4 fast flow	GE Healthcare
MBP-Trap HP	Dextrin sepharose	GE Healthcare
HiLoad Superdex 75 16/600 pg	Dextran crosslinked agarose	GE Healthcare
HiLoad Superdex 200 16/600 pg	Dextran crosslinked agarose	GE Healthcare

4.1.5 Marker

Marker	Supplier
PageRuler plus prestained protein ladder	Thermo Fisher Scientific, Waltham (USA)
LMW Marker for SDS gel electrophoresis	GE Healthcare, Freiburg
100 bp DNA ladder	Roth, Karlsruhe
1 kbp DNA ladder	Roth, Karlsruhe

4.1.6 Antibodies

Antibody target	Host species	Dilution	Supplier
pSer2-CTD (3E8)	rat	1:100	Prof. Dirk Eick
pSer5-CTD (3E10)	rat	1:100	Prof. Dirk Eick
pSer7-CTD (4E12)	rat	1:100	Prof. Dirk Eick
Thiophosphate ester (51-8)	rabbit	1:5000	Abcam, Cambridge, USA
Cdk7	mouse	1:1000	Invitrogen
Cdk7 pThr170	mouse	1:1000	Affinity Biosciences
Cdk7 pSer164	mouse	1:1000	Affbiotech
GST	n. a.	1:1000	Invitrogen
Anti-Rat IgG, HRP-coupled	chicken	1:5000	Santa Cruz Biotechnology
Anti-Rabbit IgG, HRP-coupled	goat	1:10000	Invitrogen
Anti-mouse IgG, HRP-coupled	goat	1:10000	Invitrogen
Anti-mouse IgG, IRdye 800cw	goat	1:10000	Licor
Anti-mouse IgG, IRdye 680cw	goat	1:10000	Licor

4.1.7 Microorganisms and Cell lines

E. coli strain	Genotype/Origin
BL21 (DE3)	*E. coli B F⁻ dcm ompT hsdS($r_B^-m_B^-$) gal λ(DE3)*
BL21 (DE3) pLysS	*E. coli B F⁻ dcm ompT hsdS($r_B^-m_B^-$) gal λ(DE3) [pLysS Camr]*
TOP10	*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ$^-$*
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::Amp^r chiA::LoxP */ pMON7124*
pirHC	F- Δlac169 rpoS(Am) robA1 creC510 hsdR514 endA recA1 uidA(ΔMluI)::pir-116

Cell line	Genotype
Sf9	*Cell line of Spodoptera frugiperda (Vaughn et al., 1977[a])*
HEK293	*Regensburg cell culture*

4.1.8 Enzymes

Enzyme	Supplier
Restriction endonucleases	New England Biolabs, Ipswich, USA
T4 DNA ligase	New England Biolabs, Ipswich, USA
Q5 Polymerase	New England Biolabs, Ipswich, USA
OptiTaq Polymerase	EurX, Gdansk, Poland
Cre Recombinase	New England Biolabs, Ipswich, USA
DNase	Applichem, Darmstadt
TEV protease	Geyer lab

4.1.9 Nucleic Acids / Vectors

Nucleic acids		Supplier
Primer:		All primers were ordered at Metabion, Planegg/Steinkirchen
Vector:		
	pET28a	GE Healthcare, Freiburg
	pET28a-MBP-TEV-site modified	Geyer lab, modified from pET28a
	pGEX-4T1-TEV-site modified	GE Healthcare, Freiburg
	pGEX-6P1	GE Healthcare, Freiburg
	pACEBac1	ATG Biosynthetics, Merzhausen
	pACEBac1-GST	Geyer lab, modified from pACEBac1
	pACEBac1-MBP-tev	Geyer lab, modified from pACEBac1
	pACEBac1-His-tev	Geyer lab, modified from pACEBac1
	pIDC	ATG Biosynthetics, Merzhausen
	pIDK	ATG Biosynthetics, Merzhausen
	pIDK-GST	Geyer lab, modified from pIDK

4.1.10 Synthetic peptides

Name	Sequence	Supplier
cons. $CTD_{[3]}$	YSPTSPSYSPTSPSYSPTSPS-PEG2-RR-amid	Biosynthan, Berlin
pY1 $CTD_{[3]}$	pYSPTSPSpYSPTSPSpYSPTSPS-PEG2-RR-amid	Biosynthan, Berlin
pS2 $CTD_{[3]}$	SYpSPTSPSYpSPTSPSYpSPTSPS-PEG2-RR-amid	Biosynthan, Berlin
pT4 $CTD_{[3]}$	YSPpTSPSYSPpTSPSYSPpTSPS-PEG2-RR-amid	Biosynthan, Berlin
pS5 $CTD_{[3]}$	YSPTpSPSYSPTpSPSYSPTpSPS-PEG2-RR-amid	Biosynthan, Berlin
pS7 $CTD_{[3]}$	YSPTSPpSYSPTSPpSYSPTSPpSY-PEG2-RR-amid	Biosynthan, Berlin
K7 $CTD_{[3]}$	YSPTSPKYSPTSPKYSPTSPK-PEG2-RR-amid	Biosynthan, Berlin
A5 $CTD_{[3]}$	YSPTAPSYSPTAPSYSPTAPS-PEG2-RR-amid	Sascha Gentz, MPI Dortmund

p indicates that the following amino acid is phosphorylated

4.1.11 Buffer and media

LB medium (1l)
10 g Bacto™ tryptone
5 g Bacto™ yeast extract
10 g NaCl

For agar plates add 1.5% (w/v) Agar

TB medium (1l)
12 g Bacto™ tryptone
24 g Bacto™ yeast extract
4 ml Glycerin
2.31 g KH_2PO_4
12.54 g K_2PO_4

Insect cell medium
Sf-900™ SFM III, Thermo Fisher

PBS
150 mM NaCl
20 mM Na_2HPO_4
4.6 mM NaH_2PO_4
pH 7.5

TBS-(T)
150 mM NaCl
50 mM Tris-HCl
(0.05% Tween-20)
pH 8

Buffers for the purification of Cdk10/CycM and Cdk7/CycH/(Mat1) complexes are found at in chapter 4.4.6 at page 51.

4.1.12 Devices

Device	Name/Type	Manufacturer/Supplier
Agarose gel chamber	Mini-Sub® Cell GT Cell	Biorad, Hercules, USA
Autoclave	5075 EL	Systec, Linden
Balance	CPA 3245	Sartorius, Göttingen
Balance	PCD	Kern, Balingen-Frommern
Cell culture bench	HeraSafe	Heraeus Instruments, Hanau
Centrifuge	Avanti J-26S XP	Beckman Coulter
Centrifuge	Cenrifuge 5804	Eppendorf, Hamburg
Centrifuge, table top	Centrifuge 5424	Eppendorf, Hamburg
Cooling chamber	Unichromat 1500	UniEquip, Martinsried
Electroporator	Eporator	Eppendorf, Hamburg
FPLC system	Äkta prime plus	GE Healthcare
Gel and Western Blot Imager	Odyssey	LiCor, Bad Homburg
Gel documentation system	ChemiDocXRS+	Biorad, Hercules, USA
Heat steriliser	Typ ST6120	Heraeus Instruments, Hanau
Incubator (cell culture)	Multitron cell	Infors HT, Bottmingen, Switzerland
Incubator (*E. coli*, plates)	Heratherm Incubator	Thermo Fisher Scientific, Waltham, USA
Incubator (*E. coli*, shaking)	Multitron pro	Infors HT, Bottmingen, Switzerland
Incubator (*E. coli*, shaking)	Innowa40	New Brunswick Scientific, Eppendorf, Hamburg
Liquid scintillation counter	LS6500	Beckman Coulter, Brea, USA
Magnetic stirrer	MR 3002	Heidolph, Schwabach
Microscope (inverted, cell culture)	Eclipse TS100	Nikon, Tokio, Japan
Microwave	Privileg 8018	Privileg
PCR machine	Mastercycler Nexus SX1	Eppendorf, Hamburg
pH meter	Lab 850	Xylem Analytics, Mainz
Power supply	Power Supply Basic	Biorad, Hercules, USA
SDS-PAGE electrophoresis unit	MiniProtean III System	Biorad, Hercules, USA
SDS-PAGE pouring station	MiniProtean	Biorad, Hercules, USA
Semi-dry blotter	V20 SDB	SciePlas, Cambridge, UK
Shaker, rolling	RM5-30V CAT	NeoLab, Heidelberg
Sonifier	Sonics	Vibra Cell, Newtown, USA
Sonifier waterbath	Bandelin Sonorex	Sonorex Digitec, Berlin
Spectrophotometer	Biophotometer	Eppendorf, Hamburg
Surface plasmon resonance spectroscope	Biacore 8K	GE Healthcare
Thermal stability	Prometheus Nt.48	Nanotemper
Thermo mixer	Thermomixer comfort	Eppendorf, Hamburg
UV-Spectrometer	Nanodrop 2000C	Thermo Scientific, Wilmington, USA
Vortexer	Vortex Genie	Bender & Hobein, Bruchsal
Water bath	Julabo 5	Julabo, Seelbach

4.2 Methods – Molecular Biology

4.2.1 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used to amplify or modify DNA fragments. Polymerase chain reaction was usually performed in a 50 µL reaction volume in 0.2 ml PCR tubes in an Eppendorf Nexus SX1 thermoblock. For amplification, Q5 polymerase (New England Biolabs) or OptiTaq polymerase (EurX) were used with the respective buffers according to the manufacturer's instructions. Primers were used at a concentration of 1 µM and dNTPs at a concentration of 100 µM.
DNA was amplified in a three step PCR protocol of denaturing, annealing and elongation. Usually, DNA was amplified in 30 PCR cycles. DNA from PCR was purified from the PCR reaction mixture directly or after agarose gel electrophoresis using the EXTRACTME DNA CLEAN-UP KIT (BLIRT, Gdansk, Poland) according to the manufacturer's instructions.

4.2.2 Site directed mutagenesis

Point mutations were introduced into plasmids by primer directed mutagenesis in a PCR reaction and subsequent digestion of the non-mutated parental plasmid. The non-mutated parental plasmid was specifically digested by the methylation sensitive restriction endonuclease DpnI. DpnI digests DNA of the sequence GATC in which adenine nucleotides are methylated. The DNA methylation is generated naturally by amplification of the plasmid in *E. coli* cells. The newly PCR generated, mutated DNA lacks these methylations and is therefore not digested by DpnI.

Primers for site directed mutagenesis consisted of approximately 45 nucleotides carrying the mutation in the centre. Parental plasmids were digested by adding 5 µL 10x cutsmart buffer (NEB), and 1 µL DpnI (NEB) for 1h at 37°C directly to the PCR reaction mixture after PCR. After restriction digest 5 µL to 10 µL of the reaction mixture were used to transform NEBβ10 cells.

4.2.3 Agarose gel electrophoresis

DNA was separated and analysed by gel electrophoresis in agarose gels. Agarose gel electrophoresis was performed in 1xTAE buffer at 100 V, constant using Mini-Sub® or Midi-Sub® Cell GT Cell (BioRad). DNA was stained by peqgreen (peqlab) added 1:20,000 to the agarose. Gels were visualized and documented under UV-light with a ChemDocXRS+ (BioRad). For size determination of DNA, the 100 bp DNA ladder (Roth) and the 1 kbp ladder (Roth) were used.

<u>TAE buffer</u>
40 mM Tris
20 mM acetic acid
1 mM EDTA
pH 8.5

<u>Gel Loading Dye, Purple (6X) (NEB)</u>
15% Ficoll®-400
60 mM EDTA
20 mM Tris-HCl
0.48% SDS
0.12% Dye 1
0.006% Dye 2
pH 8

4.2.4 Isolation of DNA from Agarose Gels (Gel Extraction)

For extraction of DNA from agarose gels with a clean scalpel, stained DNA was excised with a clean scalpell. DNA was extracted using the EXTRACTME DNA CLEAN-UP KIT (BLIRT, Gdańsk, Poland) according to the manufacturer's instructions.

4.2.5 Restriction digest

For cloning or analysis of plasmid DNA, DNA was digested using specific endonucleases. All endonucleases used in this thesis were purchased at New England Biolabs and the reactions were performed in the buffers supplied by the manufacturer according to the manufacturer´s instructions.

4.2.6 Ligation

Ligation of DNA was performed using T4 DNA Ligase (NEB). For generation of recombinant plasmid DNA, vector and insert were mixed in a 1:3 to 1:10 molar ratio in T4 Ligase buffer (NEB) and incubated for 2 h at room temperature with T4 DNA Ligase.

4.2.7 Transformation of *E. coli* cells

For expression of protein or vector amplification, competent *E. coli* strains were transformed with plasmid DNA. After transformation, bacteria recovered in antibiotics-free LB medium at 37°C in a shaking incubator. Then, cells were plated on agar plates. Positive clones were selected by antibiotics. For transformation, electro competent and heat shock competent cells were used. Cells were made competent in house using common Glycerol or $CaCl_2$-Protocols.

For transformation by electroporation cells were thawed on ice. When thawed, 1 µl of plasmid DNA was added to 60 µl of competent bacteria. Cells and DNA were mixed by pipetting up and down and then transferred to an electroporation cuvette. The cuvette was placed in the electroporator (Eppendorf) and cells were electroporated at 1.5 kV – 2.0 kV. Immediately after electroporation 1 ml LB was added to the cells and cells were transferred to a tube. For recovery, cells were incubated in a shaking incubator at 37°C for 1 h. Then, bacteria were plated on agar-plates containing antibiotics for selection of transformed cells.

For transformation by heat shock, cells were thawed on ice. When thawed 1 µl DNA was added to the cells and mixed by pipetting. Cells were incubated with DNA on ice for 15 minutes. After incubation, cells were heat shocked for 60 seconds at 42°C. After the heat shock, cells were immediately placed on ice for 2 minutes. Then, 1 ml LB medium was added to the cells. For recovery, cells were incubated in a shaking incubator at 37°C for 1 h. Afterwards, bacteria were plated on agar-plates containing antibiotics for selection of transformed cells.

4.2.8 Plasmid preparation from *E. coli* cells

Competent TOP10 or NEBβ10 cells were transformed and grown on agar-plates as described. For small scale DNA amplification, single cell clones were picked and grown in 2-5 ml LB containing the respective antibiotics overnight at 37°C in a shaking incubator. Plasmid DNA was purified with spin

columns using the EXTRACTME DNA CLEAN-UP KIT (BLIRT, Gdansk, Poland) according to the manufacturer's instruction. In brief, cells were collected by centrifugation, resuspended and lysed by alkaline lysis. After neutralization, lysate was cleared by centrifugation. Cleared lysate was applied to columns carrying silica membranes to bind DNA. DNA was washed with a 70% ethanol containing wash buffer and finally eluted with either distilled water (prior to ligation) or with the elution buffer supplied with the kits.

4.2.9 DNA sequencing

All generated expression constructs were analysed by DNA sequencing to confirm correct sequence and insertion. DNA samples were sequenced by Sanger sequencing at GATC Biotech using adequate primer.

4.2.10 Generation of multi gene expression vectors

For expression of proteins in *Sf*9 insect cells the MultiBac^{Turbo} System (ATG Biosynthetics, Merzhausen) was used (Bieniossek et al. 2008). The system offers the possibility to generate multi-gene expression vectors by Cre-LoxP fusion of plasmids. This is achieved by a set of vectors which act as donor and recipient vectors. Genes to be expressed together are inserted separately into their respective vector. Donor and recipient vectors are fused by Cre-recombinase reaction of LoxP sites to form a multigene fusion vector. Fusion of vectors was performed according to the protocol in the MultiBac^{Turbo} User Manual. Vectors were mixed in equal concentrations and Cre recombinase reaction was performed for 1 h at 37°C. After recombination reaction mixture was used to transform competent TOP10 or NEBβ10 *E. coli* cells. After transformation, cells recovered in 1 ml LB medium for 1 h up to 16 h. Successfully recombined plasmids were selected by antibiotics. To assure correct assembly of fusion vectors, vectors were analysed by restriction digest and compared to a restriction pattern generated *in silico* using CreACEMBLER software.

4.3 Methods – Cell Biology

4.3.1 Maintenance of *Sf9* insect cell culture

Sf9 cells were grown as suspension culture in SF900 III SFM Medium (Thermo Fisher Scientific) at 27°C, 80 rpm in a shaking incubator. Cells were kept at 0.5 x10^6 – 4x10^6 cells/mL and cells were splitted two to three times per week. *Sf9* cells were cultured using sterilized glassware of appropriate size to allow good aeration.

4.3.2 Generation of recombinant baculovirus for protein expression in *Sf9* cells

Some proteins require post-translational modifications or have specific folding requirements, which are not met by expression in bacteria. In this study, *Sf9* insect cells from lepidoptera were used as eukaryotic expression system (Vaughn et al. 1977).

Bacmid preparation

The baculoviral genome in the MultiBac system is present as a bacterial artificial chromosome (BAC) in the DH10 MultiBac^{Turbo} cells. The BAC contains a recognition site for the bacterial Tn7 transposase within a *lacZα* gene. The acceptor vectors pACEBac1 and pACEBac2 also carry a Tn7 transposase site which facilitates their integration into the baculoviral genome. Insertion of the vector into the baculoviral genome results in disruption of the *lacZα* gene. Successful transposition can therefore be monitored by a blue/white screening using galactopyranoside chromogenic substrates such as 5-Bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal).

Competent DH10 MultiBac^{Turbo} cells were transformed with 100 ng plasmid DNA. DH10 MultiBac^{Turbo} cells were electroporated at 2 kV. For recovery, cells were grown in 1 ml LB for at least 2 h at 37°C in a shaking incubator. Cells were plated on agar plates containing kanamycin (50 µg/ml), gentamycin (7 µg/ml), ampicillin (100 µg/ml), tetracyclin (10 µg/ml), X-Gal (100 µg/ml), and IPTG (40 µg/ml). Agar plates were incubated for about 48 h until blue/white selection was possible. Then, 4 white, and 1 blue colonies were streaked out on a fresh agar plate containing kanamycin (50 µg/ml), gentamycin (7 µg/ml), ampicillin (100 µg/ml), tetracyclin (10 µg/ml), X-Gal (100 µg/ml), and IPTG (40 µg/ml) and bacteria were grown for another 24 h to assure correct selection.

White colonies were picked from agar plate and grown overnight at 37°C in 2 ml LB medium containing kanamycin (50 µg/ml), gentamycin (7 µg/ml), ampicillin (100 µg/ml), and tetracyclin (10 µg/ml). Cells were harvested by centrifugation, resuspended, and lysed using EXTRACTME DNA CLEAN-UP KIT buffers P1, and P2. After neutralization of the lysate with P3 buffer, lysate was cleared from precipitations by centrifugation at 14,000 rpm (5424 rotor, Eppendorf) for 10 minutes. Supernatant was transferred to a fresh tube and centrifuged again for 5 minutes at 14,000 rpm. Afterwards, supernatant was transferred to a fresh tube and 800 µl ice-cold isopropanol was added to precipitate the Bacmid DNA. Precipitated Bacmid DNA was collected by centrifugation at 14,000 rpm for 30 minutes at 4°C. Isopropanol was removed and DNA was washed two times with 70% ice-cold ethanol and centrifuged for 15 minutes at 14,000 rpm at 4°C. After the second wash, the tube was kept closed and immediately transferred to a sterile tissue culture hood. The tube was opened in the hood and ethanol was removed. Then, tube was left open for about 10 minutes to evaporate residual ethanol, which could interfere with DNA solubilisation. Afterwards, Bacmid DNA was solubilised in H_2O or in *Sf9* cell medium.

Transfection of *Sf9* cells and virus propagation

Sf9 cells were transfected with Bacmid DNA by lipofection using Cellfectin® (Invitrogen). Transfection was performed in 6-well format with 2 ml ($0.35x10^6$ cells/ml) *Sf9* cells per well. For transfection, 10 µl of Bacmid DNA was mixed with 100 µl serum-free medium. In a different tube 8 µl Cellfectin was mixed with 100 µl medium. Both solutions were mixed and incubated for 15-30 minutes to allow formation of DNA-lipid complexes. After incubation, 200 µl were added to the respective well. Cells were incubated for 4 h at 27°C. After 4 h, medium was replaced with 3 ml fresh medium and cells were subsequently incubated for 72 h at 27°C. After 72 h the virus containing supernatant was collected (V_0) and sterile filtered. V_0 was then used for virus amplification.

V_0 was used to infect 50 ml of *Sf9* cells ($0.5x10^6$ cells/ml). Successful infection of the *Sf9* cells was monitored by a stop of cell division of transfected cells due to viral infection. Therefore, cells were

counted every adjusted to $0.5x10^6$ cells/ml (50 ml) until the cells stopped dividing. After replication has stopped, cells were incubated for another 48 h and then centrifuged at 500 rpm for 20 minutes. The supernatant containing the virus was collected (V_1). To obtain higher titer Virus (V_2), 1 ml of V_1 was used to infect 100 ml *Sf9* cells ($1x10^6$ cells/ml). Cells were incubated for four days and afterwards centrifuged at 500 rpm for 20 minutes and supernatant was collected (V_2). All viral stocks were sterile filtered and stored at 4°C.

4.4 Methods – Protein Biochemistry

4.4.1 Protein Expression in *E. coli* cells

For expression of proteins in bacteria, *E. coli* cells were grown in LB medium containing appropriate antibiotics at 37°C (pre-culture). The next day optical density at 600 nm (OD_{600nm}) was determined and the pre-culture was diluted into larger volumes of LB or TB medium to an OD_{600nm} of 0.1. Cultures were grown to optical densities of 0.8 to 1.2 at 37°C for induction of expression. Protein expression was induced by adding IPTG to a final concentration of 0.1-0.5 mM, and temperature was set to 18°C – 37°C.

To harvest bacteria, cells were collected in 1 L buckets by centrifugation at 5000 rpm (JLA.8.1 rotor, Beckmann coulter) for 20 minutes. Cell pellets were resuspended in PBS and pelleted again by centrifugation. Bacterial pellets were subjected to cell lysis or snap frozen in liquid nitrogen and stored at -20°C for later use.

List of proteins expressed in *E. coli* and their expression parameters

Protein	Vector	Temperature (°C)	Time (h)	IPTG (mM)
GST-$CTD_{[52]}$	pGex6P1	18	16	0.1
GST-$CTD_{[9]KKK}$	pGex6P1	30	4	0.3
GST-$CTD_{[16]}$A2	pGex4T1-tev	30	4	0.3
GST-$CTD_{[9]}$K7	pGEX 4T1-tev	30	4	0.3
His-NPM1 (1-265), Isoform2	pProExHta	18	16	0.5
GST-LIMK2 (149-242)	pET28a-GST-tev	18	16	0.5
GST-Ubap2L (355-652)	pGEX 4T1-tev	18	16	0.3
GST-SRRT (495-602)	pET28a GST-tev	18	16	0.3
GST-SRSF7 (1-238 f.l.)	pGEX4T1-tev	30	4	0.5
MBP-HSF1 (271-384)	pET28a MBP-tev	30	4	0.3
GST-Cdk2 (1-298, f.l.)	pGEX6P1	22	4	0.5
His-c-MYC (17-167)	pProExHTa	30	4	0.5

4.4.2 Protein expression in *Sf9* cells

For expression, *Sf9* cells were grown to a density of $1.5x10^6$ cells/ml. 2% (v/v) of baculovirus V_2 preparation was added to the cells and cell density was checked the next day. Cells were harvested after 72 h by centrifugation at 2000 rpm (JLA8.1 rotor, Beckmann Coulter), washed carefully with PBS, snap frozen in liquid nitrogen, and stored at -20°C or -80°C.

4.4.3 Cell lysis of bacterial cells

Cell pellets were resuspended in lysis buffer supplemented with PMSF (1 mM) and DNase (NEB, 1 µg/ml). Cells were resuspended using a magnetic stirrer, vortexing, or pipetting. Cells were lysed by sonication (Vibra Cell). For volumes up to 15 ml a tapered Microtip was used. Volumes larger than 15 ml were sonified using the standard probe. The device was always tuned after changing the tip. Cell lysis was performed by sonication for 1 minute (5 second on/off interval) with an amplitude of 40% (Microtip) or 80% to 100% (Standard Probe). Crude bacterial cell lysate was centrifuged for 20 minutes at 20000 rpm (JA 25.50 rotor, Beckmann Coulter). Supernatant was subsequently filtered through 0.45 µm filters.

4.4.4 Cell lysis of *Sf9* insect cells

Sf9 cell pellets were resuspended in ice cold lysis buffer (25 mL Lysis buffer/Liter of *Sf9* cell culture) containing 1 mM PMSF and DNAse. Cells were lysed as described for bacteria. After cell lysis, cell lysate was centrifuged for 1 h at 45000 (45Ti Rotor, Beckmann Coulter) or at 25000 rpm (JA 25.50 Rotor, Beckmann Coulter) to remove cellular debris. Supernatant was subsequently filtered through 0.45 µm filters.

4.4.5 Purification of Proteins

Proteins were purified from filtered cell lysates by affinity chromatography and size exclusion chromatography using an FPLC system (ÄktaPrime Plus/ÄktaStart, GE Healthcare)

Affinity chromatography

Affinity chromatography is used to separate compounds in solution (e.g nucleic acids, proteins) by their binding towards a ligand, which is immobilized to a stationary phase. Binding is non-covalent and allows recovery of the extracted compound. Non-specifically caught compounds are removed by washing. Often a high salt buffer is used to disrupt compounds bound by charge. Elution is usually done by a applying a competitor with higher binding efficiency to the immobilized ligand.

For affinity chromatography GSTrap, MBPTrap, and HisTrap (GE Healthcare) columns were used for purification of GST-tagged, MBP-tagged, and His-tagged proteins respectively. GST, MBP, or a hexahistidine motif are introduced into the protein sequence via the plasmid vector. In case of GST, affinity purification is done with a glutathione covered matrix. GST-tagged proteins are caught from lysate via GST-glutathione interactions, MBP is affinity purified via dextrin-sepharose matrix. Columns are then washed in a high salt buffer to remove non-specifically bound proteins. After washing, proteins are eluted from the column with a buffer containing reduced glutathione (GSH). In case of His-tagged proteins, affinity purification makes use of the binding of histidine residues to nickel ions. Fractions were analysed by SDS-PAGE. Fractions containing the protein of interest in high amount and purity were pooled and further subjected to sample concentration, TEV protease digest, and size exclusion chromatography.

Size exclusion chromatography

In size exclusion chromatography proteins are applied to a chromatography gel matrix. Proteins of higher molecular mass pass through the matrix faster than smaller proteins. In this work Superdex75pg and Superdex200pg (GE Healthcare) were used for size exclusion chromatography. Columns were equilibrated with one column volume gel filtration buffer, then, sample was applied via an injection loop of the FPLC to the column. Column flow through was collected in fractions of 2 ml or 3 ml and samples were analysed by SDS-PAGE. Fractions containing the protein of interest were pooled, concentrated as described (4.4.8). For storage, proteins were aliquoted, snap frozen in liquid nitrogen, and stored at -80°C.

4.4.6 Protocols of newly established purifications

Purification of MBP-tev-Cdk10/His-tev-CycM

For purification of Cdk10/Cyclin M complexes at high purity and milligram amounts a combination of immobilized metal affinity chromatography and MBP affinity chromatography was used. Cdk10/Cyclin M was co-expressed as MBP-Cdk10/His_6-CycM in *Sf9* insect cells.

The following protocol is applicable to all MBP-tev-Cdk10/His-tev-CycM constructs and mutants.

Lysis buffer			**1 Liter**	**500 ml**
	50 mM HEPES	238.31 g/mol	11.92 g	5.96 g
	300 mM NaCl	58.44 g/mol	17.53 g	8.76 g
	20 mM Imidazol	68.08 g	1.36 g	0.68 g
	5 mM bME	78.13 g/mol	385 µl	195 µl
	pH 7.6			

His Elution buffer				**100 ml**
	prepare from Gefi buffer by adding imidazole to 250 mM final concentration			
	250 mM Imidazol			1.7 g
	adjust to pH 7.6			

MBP Elution buffer				**100 ml**
	prepare from Gefi buffer by adding Maltose to 10 mM final concentration			
	10 mM Maltose	342.3 g/mol		0.34 g
	adjust to pH 7.6			

SEC Running buffer			**1 Liter**	**500 ml**
	20 mM HEPES		4.76 g	2.38 g
	300 mM NaCl		17.53 g	8.77 g
	1 mM TCEP	286.65 g/mol	0.286 g	0.143 g
	pH 7.6			

300 mM NaCl is important; 150 mM leads to strong precipitation

- Lyse *Sf9* cells in 25 ml lysis buffer / L of culture
- add PMSF and DNase to the sample
- Lyse cells by sonication
- spin at 25000 rpm for approx. 1 h
- filter lysate through 0.45 µM filter
- equilibrate a HisTrap and an MBPTrap column with lysis buffer
- remove MBPTrap and apply sample to HisTrap (flow rate 1 ml/min)
- wash with 10% of His elution buffer until baseline
- connect HisTrap to MBPTrap
- elute with His elution buffer directly onto MBPTrap at 0.5 ml/min (impurities from His-Purification will mainly run through the MBPTrap)
- wash briefly with His elution buffer (~25 ml)
- remove HisTrap and wash MBPTrap with Gefibuffer
- Elute with MBP elution buffer
- Digest with TEV-protease o/n at 4°C
- Equilibrate S75 pg column with SEC running buffer
- Apply sample to S75 pg column
- Elute with SEC running buffer in 1 ml fractions

Purification of Cdk7 containing complexes

The following protocol is applicable to all GST-tev-Cdk7/CycH, GST-tev-Cdk7/CycH/GST-Mat1, and GST-tev-Cdk7/CycH/Mat1 constructs and mutants.

Lysis/wash buffer			**1 Liter**	**500 ml**
	50 mM HEPES	238.31 g/mol	11.92 g	5.96 g
	150 mM NaCl	58.44 g/mol	8.76 g	4.38 g
	5 mM bME	78.13 g/mol	385 µl	195 µl
	pH 7.6			

GST elution buffer			**100 ml**
	prepare from Gefi buffer by adding GSH to 10 mM final concentration		
	10 mM GSH	307.33 g/mol	0.307 g
	adjust to pH 7.6		

SEC running buffer			**1 Liter**	**500 ml**
	20 mM HEPES		4.76 g	2.38 g
	150 mM NaCl		8.76 g	4.38 g
	1 mM TCEP	286.65 g/mol	0.286 g	0.143 g
	pH 7.6			

- Lyse *Sf9* cells in 25 ml lysis buffer / L of culture
- add PMSF and DNase to the sample
- Lyse cells by sonication
- spin at 25000 rpm for approx. 1 h
- filter lysate through 0.45 µM filter
- apply GSTrap columns (1x5 ml column for smaller batches) with a flow rate of 1.5 ml/min
- **important**: binding capacity of a 5ml GSTrapFF column for GST-Cdk7/CycH is approx. 3 - 4 mg only
  for large expressions do multiple binding/elution steps
- wash with lysis buffer
- elute with elution buffer at a flow rate of 1.5 ml/min
- digest with TEV protease o/n at 4°C (1:50 (mg TEV:mg Protein)
- run SEC (S200) connected to 5ml GSTrap column at 0.7 ml/min

The following protocol is applicable to GST-Cdk7/CycH/MBP-Mat1 complexes from separately expressed GST-Cdk7/CycH and MBP-Mat1.

Lysis/wash buffer			**1 Liter**	**500 ml**
	50 mM HEPES	238.31 g/mol	11.92 g	5.96 g
	150 mM NaCl	58.44 g/mol	8.76 g	4.38 g
	5 mM bME	78.13 g/mol	385 µl	195 µl
	pH 7.6			

MBP Elution buffer			**100 ml**
	prepare from Gefi buffer by adding Maltose to 10 mM final concentration		
	10 mM Maltose	342.3 g/mol	0.34 g
	adjust to pH 7.6		

SEC Running buffer		**1 Liter**	**500 ml**
	20 mM HEPES	4.76 g	2.38 g
	150 mM NaCl	8.76 g	4.38 g
	1 mM TCEP	286.65 g/mol	0.286 g	0.143 g
	pH 7.6		

- Lyse *Sf9* cells in 25 ml lysis buffer / L of culture
- add PMSF (final 1 mM) and DNase (final 1 µg/ml) to the sample
- Lyse cells by sonication (4 min, 40%, 5sec on, 10 sec off, on ice)
- spin at 25000 rpm for approx. 1 h (if possible, pre-clear lysate by centrifugation at 5000 rpm for 10 min)
- filter lysate through 0.45 µM filter
- equilibrate Tube B with MBP elution buffer
- equilibrate Tube A and the MBPTrap columns
- apply sample at 1 ml/min
- for large volumes use a higher flow-rate and do multiple binding/elution steps
- wash with lysis buffer (2.5 ml/min) until baseline
- elute with elution buffer at a flow rate of 1 ml/min in 2.5 ml Fractions
- pool samples and concentrate to 2 ml
- digest with TEV protease o/n at 4°C
- equilibrate SEC with SEC running buffer
- run SEC (Superdex 200 pg) connected to 5ml GSTrap and 5 ml MBPTrap column at 0.7 ml/min
- analyse fractions by SDS-PAGE and pool relevant fractions

4.4.7 Determination of protein concentration

Concentration of purified protein was determined by measuring absorption at 280 nm with a Nanodrop 2000c (Thermo Scientific). In parallel, absorption at 260 nm was measured to monitor possible RNA contamination. Protein concentration was calculated from absorbance using the theoretical absorption coefficient and the molecular mass of the protein calculated by EXPASY ProtParam tool.

4.4.8 Concentration of proteins

Proteins were concentrated by ultrafiltration in Amicon Ultra Centrifugal Filter Units (Amicon). In this application protein solution is filtered through a cellulose membrane. Dependent on pore size proteins of specific sizes are retained, while smaller ones pass through the filter membrane. Ultrafiltration devices with a molecular cut off of 10 kDa and 30 kDa were used.

4.4.9 Removal of affinity tag by TEV-protease digestion

For removal of affinity tags after affinity purification, tags were cleaved off by Tobacco Etch Virus (TEV) protease. TEV protease is the 27 kDa catalytic domain of the Nuclear Inclusion a (NIa) protein encoded by the tobacco etch virus (TEV). TEV protease is a highly specific protease which cleaves after the recognition motif ENLYFQ. Cleavage of the peptide sequence occurs after the glutamine. For TEV-protease digestion, TEV-protease was added in a 1:50 ratio (mg TEV / mg Protein) and incubated overnight at 4°C. TEV-protease was afterwards removed from the sample by size exclusion chromatography.

4.4.10Sodium dodecyl sulfate polyacrylamide gel electrophoresis

For analysis, proteins were separated using SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and subsequently analysed by coomassie-staining or transferred to a nitrocellulose membrane for immunoblot analysis. In SDS-PAGE proteins are separated in a polyacrylamide gel according to their relative molecular weight. Samples were reduced and denatured by boiling in SDS-sample buffer. Throughout this study discontinuous SDS-PAGE with a Tris-HCl/Tris-Glycine buffer system was performed (Laemmli, 1970). Stacking Gels contained 5% acrylamide, separation Gels contained 10% - 20% acrylamide. Cross-linking of the acrylamide was induced with ammonium persulfate and catalysed by TEMED. SDS-Gels were poured using the Biorad miniPROTEAN system. Electrophoresis was performed in Biorad miniPROTEAN chambers.

Dissolving Gel Buffer	Stacking Gel Buffer	SDS Running Buffer	4x SDS Sample buffer
1.5 M Tris pH 8.8	0.5 M Tris pH 6.8	25 mM Tris	62.5 mM Tris pH6.8
0.4 % (w/v) SDS	0.4% (w/v) SDS	194 mM Glycine	40% Glycerol
		0.1% (w/v) SDS	1.6% SDS
			0.04% bromophenolblue
			20% β-Mercaptoethanol

Coomassie staining

Proteins in SDS-gels were stained with coomassie brilliant blue R250.

staining solution	destaining solution
0.1 % (w/v) coomassie R250	10 % (v/v) ethanol
40 % (v/v) ethanol	5 % (v/v) Acetic acid
10 % (v/v) acetic acid	

4.4.11Immunoblot (Western Blot)

For western blotting, proteins were separated in a SDS-gel as described and then transferred to a nitrocellulose membrane (Optitran BA-S 85, pore size 0.45 µM, GE Healthcare). Proteins were transferred to a nitrocellulose membrane in a semi-dry blotting chamber at a constant current of 140 mA/gel for 60 minutes. After transfer, membrane was blocked in 5% milk-powder in TBS-T for 1 h at room temperature or at 4°C overnight. Then, membrane was incubated with primary antibody. Afterwards the blot was washed 3 x 5 minutes in TBS-T and was then incubated with secondary antibody for 1 h. Membrane was again washed 3 x 5 minutes in TBS-T. Fluorophore-coupled secondary antibodies were incubated and washed in the dark. For visualization, two different systems depending on the secondary antibody were used. For HRP-coupled secondary antibodies, membranes were incubated with ECL-solution for 1 minute and then analysed with a CCD camera in a Biorad XRSChemDoc system. Fluorophore-coupled secondary antibodies were analysed using a Licor Odyssey system.

4.4.12 Kinase assays

The activity of kinases was analysed *in vitro* using purified recombinant kinases and different substrate peptides or proteins. For detection of phosphorylation, a filter-binding assay using radioactive labelled [γ-^{32}P]-ATP, western blot, or SDS-PAGE were used.

[γ-^{32}P]-ATP Kinase assay

Kinase assays using [γ-32P]-ATP allow quantitative, substrate independent measurements of kinase activities. In this thesis a radioactive filter binding assay was used to analyse kinase activity *in vitro* (Hastie et al. 2006). Kinase assays were performed in kinase assay buffer (see table xy) at 30°C. The reaction was started by adding ATP to a final concentration of 1 mM – 2 mM and radioactive ATP at 3 µCi. The reactions were stopped with EDTA (50 mM final concentration) and the sample transferred onto nitrocellulose filter membrane. The membrane was washed 3x5 min in 0.75% phosphoric acid (5 ml wash buffer/filter) to remove free ATP and subsequently measured in a liquid scintillation counter for 1 minute. Typically, two technical replicates/sample were measured.

Detection of phosphorylation by SDS-PAGE or Western Blot

Protein phosphorylation was analysed by SDS-PAGE and western blot. Some proteins display a different migration behaviour in SDS-PAGE upon phosphorylation, and can thereby identified in SDS-PAGE. Phosphorylation can also be monitored with western blot using phosphorylation specific antibodies. For analysis of phosphorylation with SDS-PAGE or western blot, kinase reactions were performed in a reaction volume of 10 µL. The reaction was stopped with 10 µL of 2xSDS-sample buffer. For analysis by coomassie staining between 1 µg and 2 µg of substrate protein were loaded. If phosphorylation was analysed by western blot 50 ng to 200 ng substrate were used.

Kinase assay buffer
50 mM HEPES
34 mM KCl
7 mM $MgCl_2$
2.5 mM DTE
5 mM β-Glycerophosphate

4.4.13Mass spectrometry

Peptide Mass Fingerprint analysis

Peptide mass fingerprint can be used to analyse the identity of a protein or to detect and localize posttranslational modifications. In peptide mass fingerprint analysis proteins are digested by proteases (usually Trypsin or Chymotrypsin). Protein fragments are then analysed by mass spectrometry and compared with theoretical peptides in a database. For peptide mass fingerprint analysis, proteins in solution were separated by SDS-PAGE and stained with coomassie. Protein bands of interest were excised and analysed by mass spectrometry at the proteomics facility of the Max-Planck-Institute for Biophysical Chemistry in Göttingen (Prof. Henning Urlaub).

4.4.14 Identification of Cdk10 substrates from cell lysates

Cdk10 substrates were identified by a chemical genetic screen in HeLa full cell lysate and nuclear extracts. The chemical genetic screen makes use of a genetically modified kinase which is able to tolerate bulky-ATP analogues which are not tolerated by any native kinase (Bishop et al. 2000). The use of this bulky-ATP analogue ensures kinase-specific labelling of substrates. To discriminate newly phosphorylated substrates from those already phosphorylated by endogenous kinases in the lysate, bulky ATP-γ-S analogues were used, in which an oxygen atom in the γ-phosphate is replaced by sulphur. Moreover, this allows a covalent capture and release purification of the thio-phosphorylated peptides (Blethrow et al. 2008).

For specific labelling of proteins in cell extracts, N^6-Phenyl-ethyl-ATP-γ-S was used as bulky ATP analogue. For each replicate 2 mg protein from nuclear extracts were incubated with wild-type Cdk10/Cyclin M or analogue sensitive mutant $Cdk10^{M117G}$/Cyclin M complexes, 1 mM ATP, 0.1 mM PhEt-ATP-γ-S, 10 mM $MgCl_2$, 0.2 mg/ml Creatine phospho-kinase, and 40 mM Creatine phosphate in a volume of 100 µL The addition of ATP prevents non-specific use of the PhEt-ATP-γ-S. Creatine phosphate and creatine phospho kinase function as an ATP regenerating system which

was shown to further suppress non-specific labelling (Larochelle et al. 2006). The mixture was incubated for 60 min at 25°C at 350 rpm in a shaking incubator. The reaction was stopped by addition of 20 µL 0.5 M EDTA per replicate. Samples for mass spectrometry were snap-frozen in liquid nitrogen and stored at -80°C. For analysis of the reaction by western blot a small aliquot was alkylated with 2.5 mM para-nitrobenzylmesylate (PNBM) for 30 min at room temperature and then mixed with 2xSDS sample buffer prior to western blot analysis. Membranes were probed with a alkylation specific antibody, generated to detect the PNBM alkylated thio-phosphorylation (Allen et al. 2007).

Proteins in cell lysates are then proteolytically digested by trypsin or chymotrypsin. The resulting thio-phosphorylated peptides as well as cysteine containing peptides are then coupled to iodacetyl agarose. Thio-phosphorylated peptides are then released by oxidation with the peroxide reagent oxone which converts the thio-phosphorylated peptides into conventional phosphopeptides. Phosphopeptides are then analysed by mass spectrometry.

Proteolysis of proteins and enrichment of thio-phosphorylated peptides and mass spectrometric analysis was performed in collaboration with Yin Yanlong, PhD student in the laboratory of Prof. Dr Henning Urlaub at the Max-Planck Institute for Biophysical Chemistry, Göttingen.

<u>Reaction mixture for thio-phosphorylation of proteins in cell lysate/extracts</u>

2 mg	total protein per replicate
0.5% – 1% (w/w)	$Cdk10^{M117G}$/CycM
0.2 mg/ml	Creatine phospho kinase
40 mM	Creatine phosphate
1 mM	ATP
0.1 mM	PhEt-ATP-γ-S
10 mM	$MgCl_2$
Ad 100 µL	Buffer

4.5 Biophysical Methods

4.5.1 Determination of thermal protein stability

The thermal stability of proteins and protein complexes in solution was determined by nano differential scanning fluorometry (nanoDSF) in a Prometheus device (Nanotemper). Protein stability is monitored by a change of the intrinsic absorbance of tryptophan residues which changes as tryptophans become solvent accessible by unfolding of the protein. Fluorescence is monitored at 350 nm and 330 nm wavelengths. For determination of thermal stability, the transition point (T_m) of the first derivation of the 350 nm/330 nm ratio was determined automatically by the software provided by the manufacturer. For generation of figures, the data were exported and processed with graph pad prism and adobe illustrator. The data were not re-evaluated in graph pad prism.

For thermal shift assay 10 µL of protein sample were filled into glass capillaries by capillary forces. Protein unfolding was monitored from 20°C to 95°C at constant heating of 1.5°C per minute. All proteins and protein complexes analysed in this thesis were kept in the same buffer formulation at 20 mM HEPES pH7.6, 150 mM NaCl, and 1 mM TCEP. All samples were assayed in technical triplicates.

4.5.2 Surface plasmon resonance spectroscopy

Surface plasmon resonance (SPR) spectroscopy is an optical technique to characterize noncovalent binding events between an immobilized molecule (ligand) and a solvent binding partner (analyte). Surface plasmon resonance spectroscopy measures differences in the generation of surface plasmons when activated by a light beam. The generation of the surface plasmons is dependent on the surface. Binding is hence monitored by a change in the surface upon association of the analyte to the immobilized ligand resulting in an altered plasmon resonance.

The experiments in this thesis were performed using a BIACORE 8 K (GE Healthcare) equipped with a research-grade CM5 sensor chip. The ligand (MBP-Mat1 (230-309) was immobilized using an anti-MBP antibody which was coupled to the chip. Another uncoated flow cell served as a reference surface. The surfaces were blocked for 7 min with 1 M ethanolamine, pH 8.0. The analyte in 20 mM HEPES, 150 mM NaCl, 0.05 % Tween20, 1 mM TCEP pH 7.4, was injected over the flow cells at different concentrations to obtain a multi cycle kinetics at a temperature of 20 °C. Data were collected at a rate of 1 Hz. Binding analysis was performed by Dr. Karl Gatterdam at the Institute of Structural Biology, Bonn.

5 Results

5.1 Characterisation of the Cdk7/CycH/Mat1 complex

Cdk7 fulfils versatile functions in the cell. With its unique position as part of the general transcription factor TFIIH and its function as the cdk-activating kinase (CAK) it is directly involved in regulation of transcription and cell cycle progression. In cells, three different Cdk7 containing complexes have been described. A trimeric complex composed of Cdk7/CycH/Mat1. A quaternary assembly of Cdk7/CycH/Mat1 in complex with XPD and a multisubunit complex of Cdk7/CycH/Mat1 associated with the core TFIIH (Drapkin et al. 1996). Importantly, dimeric Cdk7/CycH complexes have not been described so far. The trimeric Cdk7/CycH/Mat1 complex is often referred to as CAK or free-CAK, to highlight that this form is not associated with TFIIH. Despite the longstanding interest in transcriptional and cell cycle regulation by Cdk7, the regulation of Cdk7 remains enigmatic. It is described, that association with Mat1 increases the activity of Cdk7/CycH towards several substrates (Rossignol et al. 1997), but Mat1 appears to be stably associated with Cdk7/CycH and hence rather represents a scaffold to allow activity than acting as a regulatory subunit (Fisher 2005). During the last two decades two modes of Cdk7 regulation have been described: association with XPD and Cdk7 Thr170 phosphorylation (Chen et al. 2003; Larochelle et al. 2001). Association with XPD seems to alter the cellular localisation and thereby directing Cdk7 to or away from its substrate (Li et al. 2010). Due to its prominent role in cell cycle and transcriptional regulation, Cdk7 has been implicated as a drug target in cancer early on, but druggability has been hampered by its various essential functions (Baumli et al. 2008; Lolli und Johnson 2005). Interest in Cdk7 inhibition has increased in recent years by the concept of transcriptional addiction in cancer. It is assumed, that viability of cancer cells is dependent on the transcription of oncogenes whose expression is driven by so-called super-enhancers. These dependencies make cancer cells vulnerable to inhibition of transcription (Bradner et al. 2017).

This part of the thesis provides detailed information of Cdk7 activity and stability. The activity of different recombinant Cdk7 complexes and the binding affinity of Mat1 to dimeric Cdk7/CycH has been determined. Functionally, Cdk7 Serine164 phosphorylation was found as a pre-requisite for full activity of the trimeric Cdk7/CycH/Mat1 complex. Moreover, it is shown, that the C-terminal part of Cyclin H contributes to Cdk7 binding and activity regulation in a Ser164 dependent manner and contributes to complex stability. Finally, the inhibitory potential and cross-reactivity of YKL-5-124, a new covalent Cdk7 inhibitor, is investigated.

5.1.1 Domain Architecture of the human Cdk7/CycH/Mat1 complex

The human cyclin-dependent kinase 7 (Cdk7; uniprot accession code: P50613) is a 346 amino acid protein. In cells, the initiator methionine is removed. Cdk7 contains a kinase domain which spans most of the protein. At the N- and the C-terminus basic sequences can be found which might serve as nuclear localisation signals but no experimental evaluation regarding the functionality is available. Within the kinase domain Cdk7 contains a variation of the PSTAIRE sequence, a common feature of cyclin-dependent kinases. In Cdk7 the sequence is NRTALRE and hence less conserved compared to other CDKs. It has been suggested that sequence variation in this sequence plays a role of DNA bound RNA pol II recognition by Cdk7 (Lolli 2009). The Cdk7 activation loop contains two phosphorylation sites at Ser164 and Thr170. Phosphorylation at Thr170 is required for complex formation and activity of the dimeric Cdk7/CycH complexes. Mat1 can overcome the requirement of Thr170 phosphorylation resulting in active Cdk7/CycH/Mat1 complexes even after mutation of Thr170 to alanine (Fisher et al. 1995). Phosphorylation at Ser164 has been implicated in stabilisation of Cdk7/CycH complexes and in regulation of activity (Martinez et al. 1997). Moreover, Cdk7 Ser164 is located in a consensus sequence for CDKs/MAPKs which is SPXK/R and has been found to be phosphorylated by Cdk2 *in vitro* (Garrett et al. 2001). A comparable motif within the activation loop of CDKs is only found in Cdk11A, Cdk11B, Cdk8, and Cdk19.

The Cdk7 cyclin partner Cyclin H (CycH, uniprot accession code: P51946) is a 323 amino acid protein containing two central cyclin boxes. At the C-terminus it contains a functional nuclear localisation signal (Krempler et al. 2005). The CycH C-terminus is mainly unstructured and contains many charged amino acids.

The Cdk7/CycH assembly factor ménage a trois (Mat1; uniprot accession code: P51948) comprises 309 amino acids to which three functional parts can be assigned: an N-terminal Zinc-Finger of the RING type family, a central domain interacting with TFIIH, and a C-terminal hydrophobic part which interacts with Cdk7 and CycH. The N-terminal RING-Finger is supposed to assist in transcription by contacting the DNA when the holo-TFIIH complex is formed. The central helical domain is contacting the TFIIH subunits XPB and XPD and thereby regulates the assembly of the core TFIIH with the Cdk7 kinase module. The C-terminal part of Mat1 has been shown to be both required and sufficient for binding and stabilising Cdk7/CycH (Busso et al. 2000; Greber et al. 2017). The domain structures of Cdk7, Cyclin H, and Mat1 are shown in Figure 5.1.1.

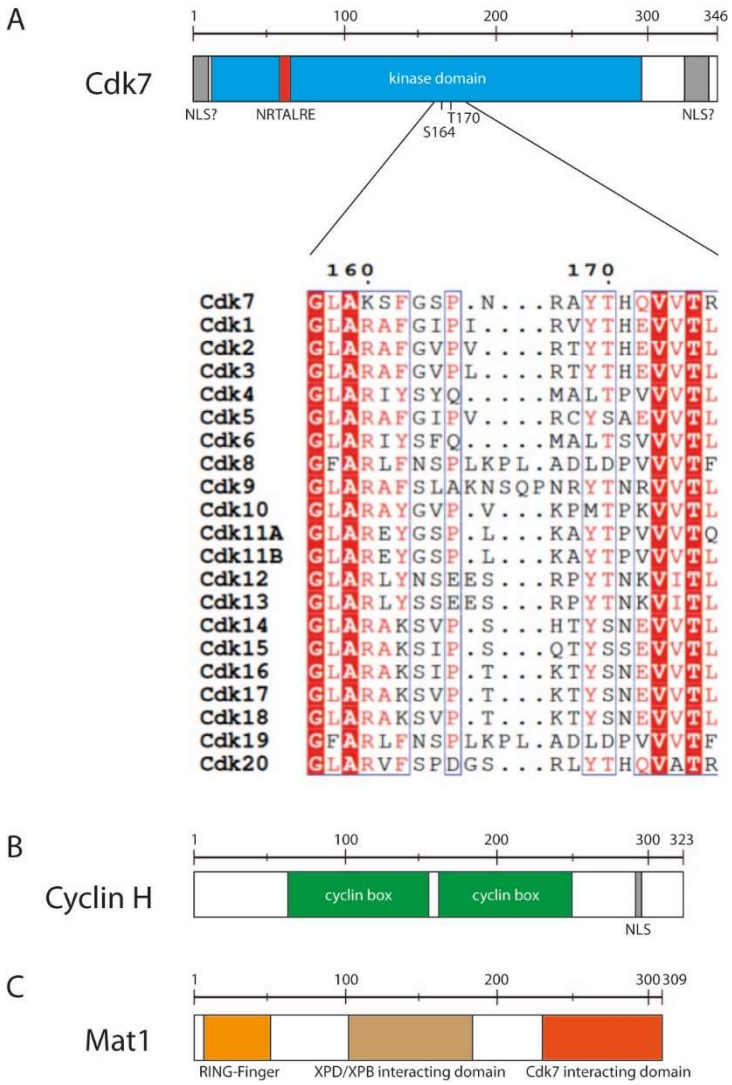


Figure 5.1.1: Domain architecture of Cdk7, Cyclin H and Mat1
A) Domain architecture of Cdk7. The kinase domain spans nearly the entire protein. The NRTALRE sequence for cyclin interaction is depicted in red. The Cdk7 activation domain harbours two sites of phosphorylation, Ser164 and Thr170. Thr170 is required for activation of the kinase. At the N- and the C-terminus are basic regions which might act as nuclear localisation signals. A sequence alignment of all Cdk activation segments is shown below. **B**) Domain structure of Cyclin H. Cyclin H harbours two cyclin boxes in the middle part of the protein. At the C-terminus, Cyclin H contains a nuclear localisation signal. **C**) Domain structure of Mat1. Mat1 contains a RING-finger at the N-terminus. The central region mediates interactions to the TFIIH subunits XPD and XPB. A hydrophobic domain at the C-terminus binds to Cdk7 and Cyclin H.

5.1.2 Cdk7 preparations differ in T-loop phosphorylation

Cdk7 was affinity purified from *Sf9* insect cells and analysed by SDS-PAGE before and after the GST-tag was cleaved off by TEV protease. Upon TEV digest, Cdk7 appears as two distinct bands in SDS-PAGE (Figure 5.1.2). This phenomenon has been described and assigned to Cdk7 T-loop phosphorylation at two sites serine 164 and threonine 170 (Larochelle et al. 2001). To validate, if the observed Cdk7 migration behaviour in SDS-PAGE is indeed dependent on phosphorylation, serine 164 was mutated to alanine. The Cdk7 S164A mutant appears as a single band upon removal of the GST-tag. The running behaviour resembles the one of the upper migrating Cdk7 band (Figure 5.1.2,

B). This result is in accordance with published mass spectrometry data which show that the upper migrating Cdk7 is phosphorylated at Thr170 and the lower migrating form is bis-phosphorylated at Ser164 and Thr170 (Lolli et al. 2004). Moreover it has been described, that the phosphorylation status of Cdk7 at Ser164 and Thr170 differs when expressed in either dimeric Cdk7/CycH or trimeric Cdk7/CycH/Mat1 complexes (Larochelle et al. 2001). Indeed, co-expression of Cdk7/CycH resulted in Cdk7 which was nearly exclusively running in the lower migrating, bis-phosphorylated form, while co-expression of the trimeric Cdk7/CycH/Mat1 resulted in the upper migrating non- or singly phosphorylated form (Figure 5.1.2, C).

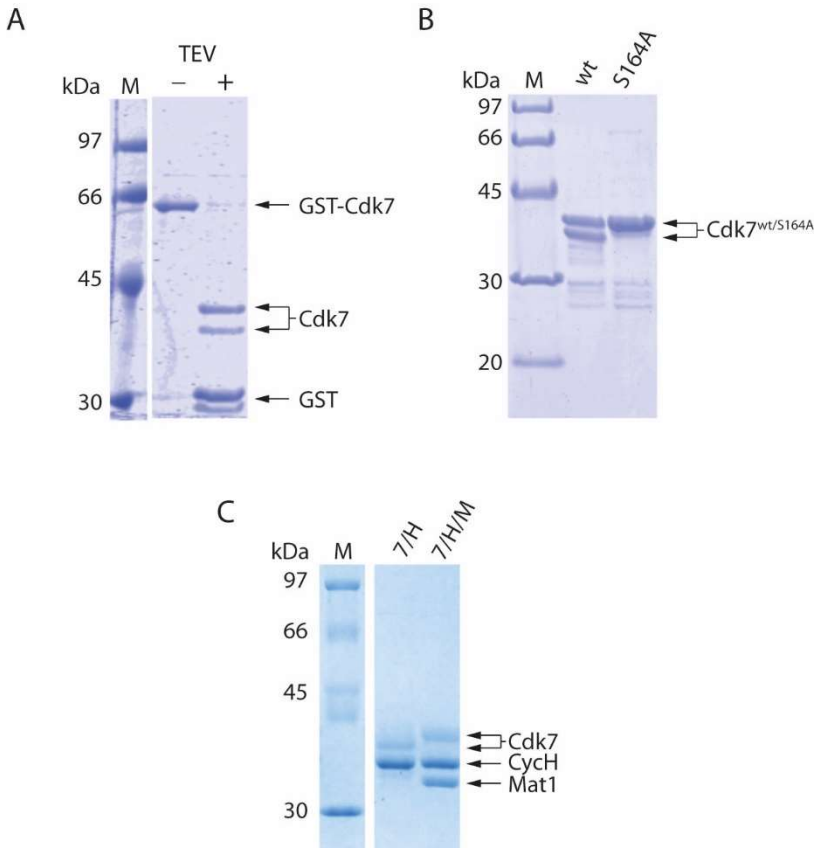


Figure 5.1.2: Cdk7 SDS-PAGE running behaviour is dependent on Ser164 phosphorylation.
A) Coomassie staining of 1 µg affinity purified GST-Cdk7 before and after TEV cleavage. **B**) Coomassie staining of purified Cdk7 and Cdk7^{S164A} (1 µg each). **C**) Coomassie staining of purified dimeric Cdk7/CycH and trimeric Cdk7/CycH/Mat1.

To analyse T-loop phosphorylation in more detail, different Cdk7 preparations either with or without Mat1 were analysed for Thr170 and Ser164 phosphorylation by Western Blot using phosphorylation specific antibodies. Samples were analysed as both GST-tagged Cdk7 preparations and after tag-removal. The preparations tested comprise dimeric Cdk7/CycH preparations including Cdk7^{S164A}/CycH and Cdk7^{S164E}/CycH mutants, trimeric complexes which were co-expressed Cdk7/CycH/Mat (1-309), Cdk7/CycH/Mat1 (230-309), as well as trimeric Cdk7/CycH + MBP-Mat1, in which Cdk7/CycH and Mat1 were expressed separately and then pooled for purification.

SDS PAGE analysis of the samples revealed that all complexes showed good purity after affinity chromatography (Figure 5.1.3). Only the GST-Cdk7/CycH/GST-Mat1 (230-309) sample contained

considerable amounts of leaky GST. Upon tag removal Cdk7 migrates at several heights in SDS PAGE. Cdk7/CycH and Cdk7^{S164E}/CycH run mostly in the lower migrating form. Co-expression of Cdk7/CycH with either full length Mat1 or N-terminal truncated Mat1 (230-309) resulted in the upper migrating form. Cdk7 in the trimeric Cdk7 complex in which Cdk7/CycH and Mat1 (230-309) were expressed separately and pooled prior to affinity chromatography appears as in dimeric Cdk7/CycH. The Cdk7^{S164A} mutant in complex with CycH migrates between Cdk7 in the dimeric and trimeric complex.

Western blot analysis of the GST-tagged samples revealed that dimeric Cdk7/CycH complexes were all phosphorylated at Cdk7 Thr170 at comparable levels regardless if Cdk7 Ser164 was mutated or not. Cdk7/CycH also depicted a strong signal for Ser164 phosphorylation, which was absent in Ser164 mutants. Co-expressed trimeric Cdk7/CycH/Mat1 complexes displayed lower levels of Cdk7 pThr170, and Cdk7 pSer164 phosphorylation compared to dimeric Cdk7 preparations. Instead, the separately expressed Cdk7/CycH + Mat1 complex showed high levels of Thr170 and Ser164 phosphorylation as dimeric Cdk7/CycH (Figure 5.1.3, C).

Albeit, the co-expressed trimeric complexes contained less Thr170 and Ser164 phosphorylation compared to dimeric Cdk7/CycH, a considerable amount of T-loop phosphorylated Cdk7 was detected. The samples were purified via the GST-tag at the kinase, T-loop phosphorylation could hence originate from free non-complexed Cdk7. To evaluate T-loop phosphorylation further, samples after tag-removal by TEV protease digestion and SEC were analysed. CDk7 Thr170 was phosphorylated in all dimeric Cdk7/CycH preparations and in the Cdk7/CycH + Mat1 (230-309) preparation. In contrast, no pThr170 and no pSer164 signal was detected in the co-expressed trimeric samples tested. Probing of with an anti Cdk7 antibody confirmed that Cdk7 was present in all lanes.

Co-expression of trimeric complexes results in a lack of T-loop phosphorylation at both Ser164 and Thr170. Truncation of Mat1 to the C-terminal part required for interaction with Cdk7/CycH does not affect the phosphorylation status of co-expressed Cdk7/CycH/Mat1. Mutation of Cdk7 serine 164 does not affect phosphorylation at threonine 170. Co-purification of separately expressed GST-Cdk7/CycH and Mat1 (230-309) preserves high Cdk7 Ser164/Thr170 double phosphorylation and thus serves as a purification strategy to obtain large quantities of phosphorylated trimeric complexes.

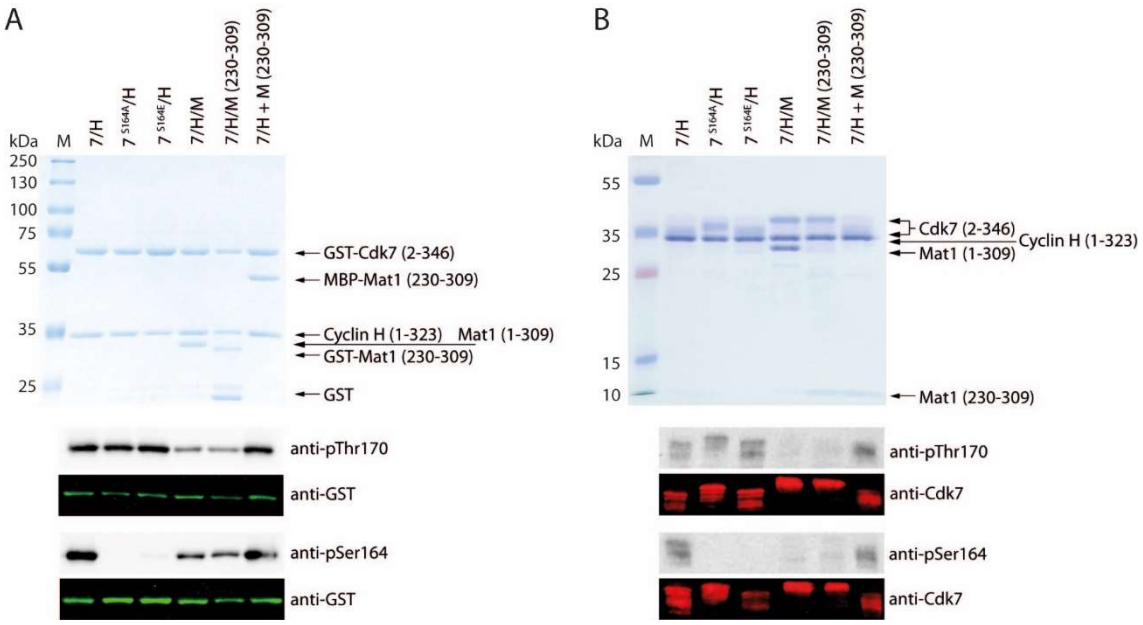


Figure 5.1.3: Analysis of the T-loop phosphorylation in different Cdk7 preparations.
A) Analysis of the Cdk7 T-loop phosphorylation in GST-Cdk7 samples. Coomassie staining of 10 µL of different Cdk7 complexes at a concentration of 0.97 µM. This corresponds to 1 µg GST-Cdk7/CycH complex. For WB 50 ng of GST-Cdk7/CycH were loaded, and all other proteins were adjusted accordingly. Analysis of the Cdk7 T-loop phosphorylation at Thr170 and Ser164 were done using Cdk7 phospho-specific primary antibodies and an HRP-coupled secondary antibody. As loading control, the amount of total GST-Cdk7 on the blot was detected by an anti GST-antibody and visualized using a fluorophore coupled secondary antibody. 7/H + M (230-309) indicates, that Cdk7/CycH and Mat1 were expressed separately. **B**) Analysis of Cdk7 T-loop phosphorylation after removal of the affinity tags and further purification by SEC. Coomassie staining of 10 µL of different Cdk7 complexes at a concentration of 1.3 µM. This corresponds to 1 µG Cdk7/CycH complex. Cdk7 T-loop phosphorylation at Thr170 and Ser164 was analysed as in A. As loading control, blots were stripped and probed with a pan-Cdk7 antibody. Cdk7 signal was detected by a fluorophore coupled secondary antibody. 50 ng of GST-Cdk7/CycH were loaded, all other proteins were adjusted accordingly.

5.1.3 T-loop phosphorylation is correlated with activity

Mat1 is generally accepted as a factor that stabilises and increases the kinase activity of the Cdk7/CycH complex. However, this view is oversimplified as early studies demonstrated that co-expression of trimeric Cdk7 complexes in *Sf9* cells show only a mild increase in stability and activity compared to dimeric Cdk7/CycH. Interestingly, *in vitro* reconstitution of trimeric Cdk7 complexes from Cdk7/CycH and Mat1, which were expressed and purified separately lead to CAK complexes with greatly increased stability and activity. It has been shown, that this effect is caused by alterations in Cdk7 phosphorylation levels that are a consequence of different expression strategies (Larochelle et al. 2001). This phenomenon was analysed by testing the Cdk7 preparations shown in Figure 5.1.3 for activity in a radioactive kinase assay. Mutation of Cdk7 serine 164 to either alanine or glutamate has no effect on the activity of dimeric Cdk7/CycH. The non-phosphorylated Cdk7/CycH/Mat1 complex displayed similar activity to Cdk7/CycH. Truncation of Mat1 in non-phosphorylated trimeric Cdk7/CycH/Mat1 (230-309) moderately increases the activity. Instead, the activity of the pSer164/pThr170 double phosphorylated Cdk7/CycH + Mat1 (230-309) sample was increased by 4.5-fold compared to the dimeric Cdk7/CycH complex (Figure 5.1.4).

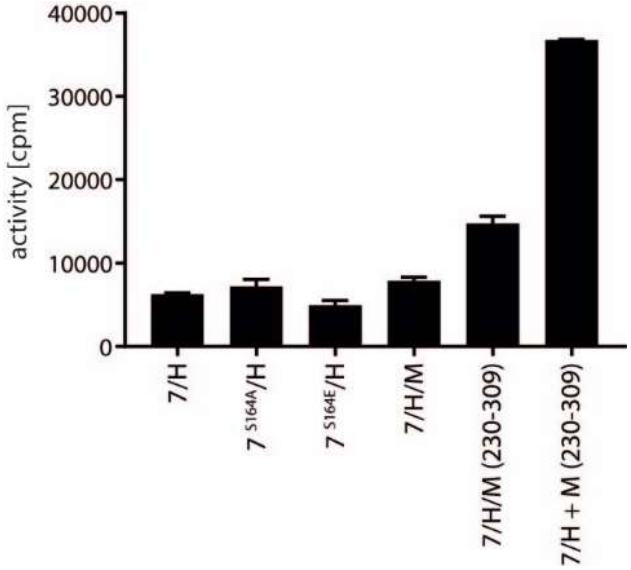


Figure 5.1.4: Activity of different Cdk7 preparations.
The samples shown in Figure 5.1.3 were probed for their activity in a radioactive kinase assay. 0.1 µM of each kinase complex was incubated with 10 µM GST-$CTD_{[52]}$ and 1 mM ATP for 15 minutes at 30°C. Data represent mean +SD from a duplicate measurement.

5.1.4 Substrate specificity of Cdk7/CycH/Mat1 towards RNA pol II CTD

Cdk7 phosphorylates the RNA pol II CTD at position Ser5 and Ser7. It was analysed if Mat1 changes substrate specificity towards different serines within a CTD heptad repeat (Figure 5.1.5). Analysis of the dimeric Cdk7/CycH and trimeric Cdk7/CycH/Mat1 preparations in a radioactive kinase assay confirmed equal activity of the complexes towards GST-$CTD_{[52]}$. For western blot, GST-$CTD_{[52]}$ was incubated in a time course experiment for 20 minutes and samples were taken after 0 min, 5 min, 10 min, and 20 min and specific phosphorylation of RNA pol II CTD was analysed using antibodies specific for Ser2, Ser5, or Ser7 phosphorylation. In accordance with published data, Cdk7 phosphorylated Ser5 and Ser7 positions of the hepta repeats. Only a weak signal for Serine 2 phosphorylation was detected. Signals for Serine5 and Serine7 phosphorylation were prominent. The dimeric Cdk7/CycH and the trimeric Cdk7/CycH/Mat1 show equal pSer5 and pSer7 phosphorylation patterns. Thus, Mat1 does not alter the substrate specificity of the Cdk7/CycH complex towards specific residues within the RNA pol II CTD.

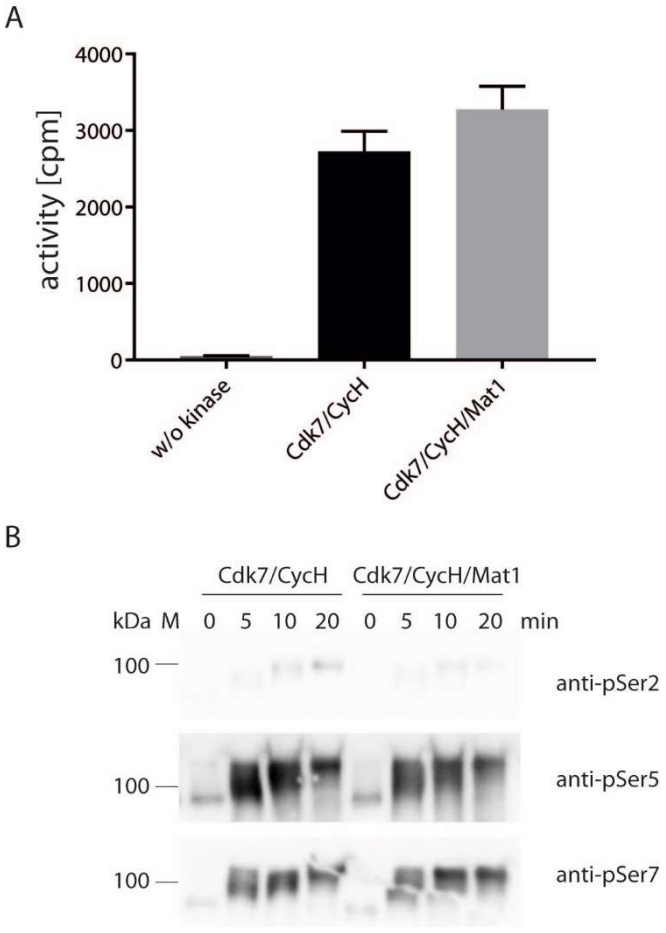


Figure 5.1.5: Activity and specificity of Cdk7/CycH and Cdk7/CycH/Mat1.
A) Radioactive kinase assay of 0.2 µM of Cdk7/CycH and Cdk7/CycH/Mat1 with 10 µM GST-$CTD_{[52]}$ and 2 mM ATP at 30°C for 15 minutes. **B**) Analysis of the GST-$CTD_{[52]}$ phosphorylation with phospho-specific antibodies. Phosphorylation reaction was performed as in A. Samples were taken at the indicated time points. 1 µg GST-$CTD_{[52]}$ was loaded in each lane for WB analysis.

5.1.5 Cdk7 activity and substrate specificity upon *in vitro* addition of Mat1

In vitro reconstitution of trimeric Cdk7/CycH/Mat1 from dimeric Cdk7/CycH and singly expressed Mat1 from *Sf9* insect cells results in a profound increase in Cdk7 kinase activity (Larochelle et al. 2001). It was therefore intended to analyse effects on activity and substrate specificity after *in vitro* addition of Mat1. *In vitro* reconstitution prior to the kinase assay was given preference over co-purified Cdk7/CycH + Mat1 (230-309) as it makes use of the same kinase preparation and hence eliminates batch to batch variations in kinase activity arising from differences in purification.
For *in vitro* reconstitution, the C-terminal part of Mat1 which is sufficient for Cdk7/CycH binding and activity regulation was expressed as MBP-fusion protein in *Sf9* insect cells. MBP-Mat1 (230-309) was purified via affinity chromatography followed by SEC resulting in pure MBP-Mat1 (230-309) (Figure 5.1.6, A). For analysis, 0.1 µM dimeric Cdk7/CycH was incubated with 0.4 µM MBP-Mat1 (230-309) in presence of 10 µM GST-$CTD_{[52]}$ for 10 minutes at room temperature prior to the kinase assay. Kinase assays were started by addition of ATP. Pre-incubation of Cdk7/CycH with Mat1 resulted in a nearly 10-fold increase in kinase activity compared to Cdk7/CycH alone (Figure 5.1.6, B). Pre-incubation of Cdk7/CycH with MBP had no effect on Cdk7 activity. Samples without GST-$CTD_{[52]}$ substrate gave no signal ruling out that the effect was due to MBP-Mat1 (230-309) being a substrate of Cdk7. Moreover, a sample lacking Cdk7/CycH was also negative in GST-$CTD_{[52]}$ phosphorylation demonstrating, that there is no contamination with a CTD kinase in the Mat1 preparation.

Next it was addressed, if the increase in activity correlates with a change in substrate specificity within a CTD hepta repeat. GST-$CTD_{[52]}$ was phosphorylated by Cdk7/CycH either with or without pre-incubation with MBP-Mat1 (230-309) and subsequently analysed by western blot using antibodies specific for phosphorylation at Ser2, Ser5, or Ser7 (Figure 5.1.6, C). Cdk7 activity towards the GST-$CTD_{[52]}$ in presence of MBP-Mat1 (230-309) was stronger compared to Cdk7/CycH alone. Signals were detected for Ser5 phosphorylation and Ser7 phosphorylation but not Ser2 phosphorylation. Western blot analysis thus confirmed increased activity upon *in vitro* Mat1 binding, but gave no indication for a shift in substrate specificity within the RNA pol II C-terminal domain.

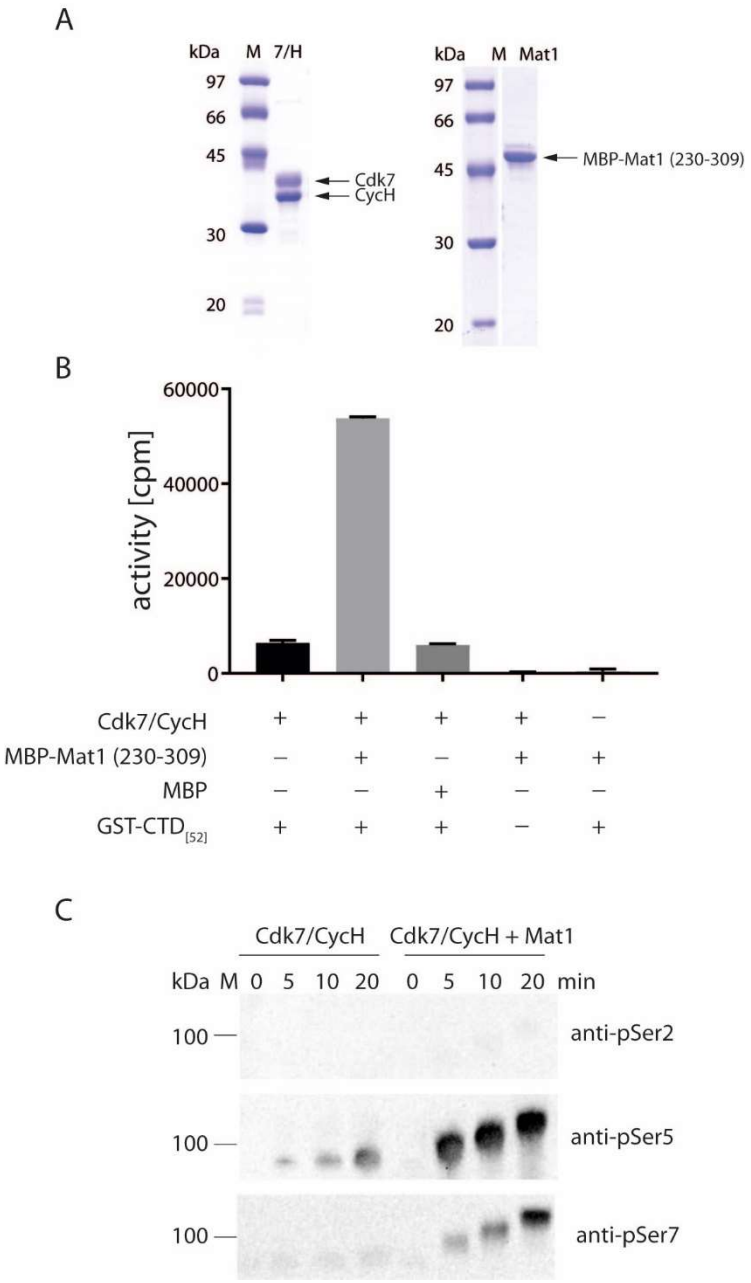


Figure 5.1.6: Activity of substrate specificity of Cdk7/CycH + Mat1 upon *in vitro* reconstitution.
A) Coomassie staining of 2 µg Cdk7/CycH and 2 µg MBP-Mat1 (230-309) used for *in vitro* reconstitution of trimeric Cdk7 complexes. B) Activity of dimeric and *in vitro* reconstituted trimeric Cdk7 complexes. The concentrations of the reactants are: 0.1 µM Cdk7/CycH, 0.4 µM MBP-Mat1 (230-309), 0.4 µM MBP, and 10 µM GST-$CTD_{[52]}$. After mixing of the components, reaction mixture was incubated for 10 minutes at room temperature prior to starting the kinase assay with 1 mM ATP. Reactions were stopped after 15 minutes by addition of EDTA. **B**) Western Blot analysis of GST-$CTD_{[52]}$ phosphorylation by Cdk7/CycH and *in vitro* reconstituted Cdk7/CycH/MBP-Mat1 (230-309). Phosphorylation reaction was performed as in A except for stopping the reaction with SDS sample buffer instead of EDTA. Samples were taken at indicated time points. For analysis 300 ng GST-$CTD_{[52]}$ were loaded in each lane.

It has been reported, that the association of Cdk7/CycH with Mat1 specifically increases the Cdk7 kinase activity towards its transcriptional substrate RNA pol II CTD, while having no effect on activity towards the cell cycle substrate Cdk2 (Yankulov und Bentley 1997). To further characterise Mat1 mediated gain of activity, the activity towards GST-$CTD_{[52]}$ was compared to GST-$CTD_{[9]KKK}$, GST-Cdk2, and towards a set of synthetic $CTD_{[3]}$ peptides harbouring modifications at Ser5, and Ser7. Kinase reactions were analysed in a radioactive kinase assay. Both GST-CTD variants GST-$CTD_{[52]}$ and GST-$CTD_{[9]KKK}$ showed increased phosphorylation upon Mat1 addition with GST-$CTD_{[9]KKK}$ phosphorylation resulting in higher absolute counts compared to GST-$CTD_{[52]}$. Cdk2 was phosphorylated at much lower rates compared to GST-$CTD_{[52]}$. Mat1 also stimulates Cdk7 activity towards GST-Cdk2, but to a lesser extent compared to the effect seen with GST-CTD as substrate (Figure 5.1.7).

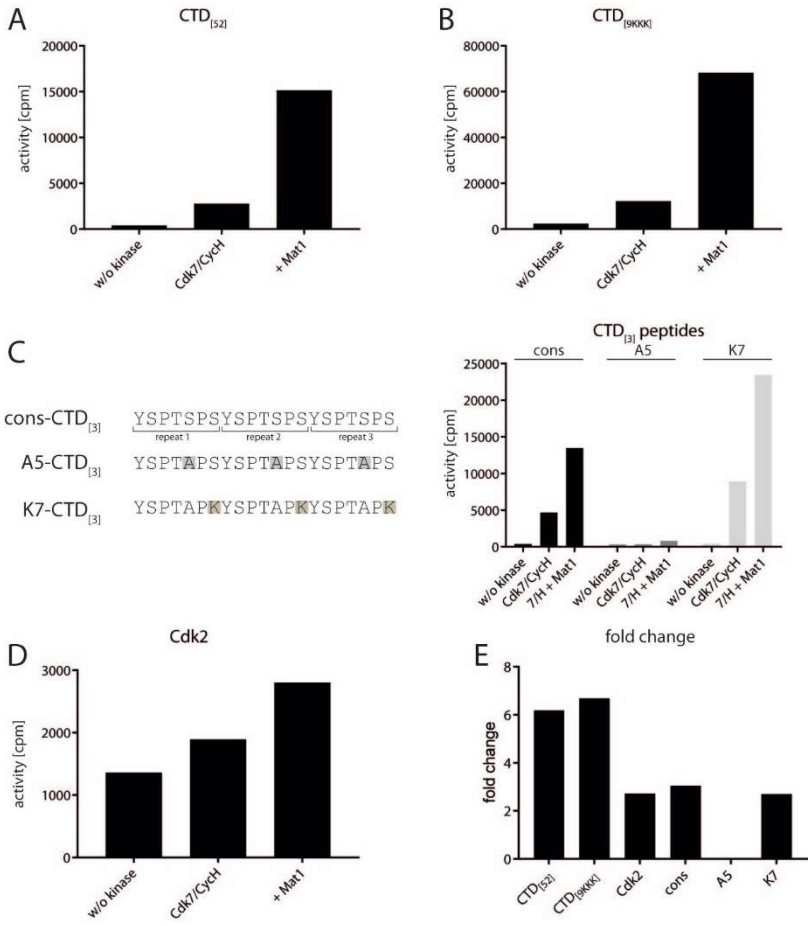


Figure 5.1.7: Mat1 dependent increase in activity towards different substrates.
A-D) Radioactive kinase assays with 0.1 µM Cdk7/CycH in presence or absence of 0.4 µM MBP-Mat1 (230-309) towards **A**) 10 µM GST-$CTD_{[52]}$, **B**) 50 µM GST-$CTD_{[9]KKK}$, **C**) 100 µM cons-$CTD_{[3]}$, 100 µM A5-$CTD_{[3]}$, 100 µM K7-$CTD_{[3]}$. The peptide sequence is shown in the left panel. **D**) 15 µM GST-Cdk2. All samples were incubated for 15 minutes at 30°C at 1 mM ATP. **E**) Fold-change upon Mat1 addition of the data shown in A-D. Data were background corrected for analysis.

Moreover, GST-Cdk2 samples display high background in absence of Cdk7, which might be a consequence of Cdk2 auto-phosphorylation. Phosphorylation of synthetic CTD peptides consisting of three hepta repeats, in which Ser5 was replaced by Ala (A5-$CTD_{[3]}$) or in which Ser7 was replaced

by Lys (K7-CTD$_{[3]}$) were compared to a peptide representing three repetitions of the consensus sequence YSPTSPS (cons-CTD$_{[3]}$). The cons-CTD$_{[3]}$ and the K7-CTD$_{[3]}$ peptide where phosphorylated by Cdk7 with K7 depicting an approximately 2-fold better phosphorylation compared to the consensus peptide. Both, consensus and K7 showed an increased phosphorylation upon Mat1 pre-incubation. The A5 peptide was not phosphorylated in either condition. For analysis, the data described above were background corrected and displayed as fold-change upon Mat1 addition (Figure 5.1.7). Both GST-CTD$_{[52]}$ and GST-CTD$_{[9]KKK}$ phosphorylation increase 6-fold in presence of Mat1. Increase of Cdk2 phosphorylation was less pronounced, but still was 3-fold. For the cons-CTD$_{[3]}$ and the K7-CTD$_{[3]}$ peptide phosphorylation is increased 3-fold and hence similar to GST-Cdk2. Data from the A5-CTD$_{[3]}$ peptide were not analysed for the occurring fold-change due to insufficient phosphorylation.

Cdk2 phosphorylation is hence also affected by Mat1 binding but to a lesser extent than CTD phosphorylation. Interestingly the Mat1 mediated effect towards the short synthetic peptides resembles the increase seen for Cdk2 but not for GST-CTD preparations. Similar results have been obtained in a study using Cdk2 and CTD peptides with four repeats as substrates (Busso et al. 2000). The inability of Cdk7 to phosphorylate the A5-peptide at the non-modified Ser7 position could be due to impaired recognition of the altered sequence or could rely upon a general deficiency to phosphorylate position Ser7 in short peptides. For Cdk9, which also phosphorylates the CTD at position Ser5 and Ser7 *in vitro*, it has been shown that Ser7 phosphorylation is dependent on the number of CTD repeats (Bösken 2013).

5.1.6 Cdk7 Ser164 phosphorylation in complex formation and activity

In the experiments so far, published data could be confirmed, that recombinant Cdk7 appears in different phosphorylation states and that phosphorylation status affects Cdk7 activity and thus represents a way of Cdk7 regulation. Studies in the past focused on regulation by Thr170 phosphorylation but several studies suggest an involvement of Ser164 in positive and negative regulation of Cdk7 activity. As shown in this thesis, Cdk7 Ser164 mutation to non-phosphorylatable alanine or the phosphorylation mimic glutamate do neither affect pThr170 levels nor represents general constraints for Cdk7/CycH complex formation and activity (Figure 5.1.3). These Cdk7 mutants are thus well suited to specifically investigate the effect of Cdk7 Ser164 phosphorylation on complex formation, activity, and stability.

5.1.7 Cdk7^{S164A} fails to bind C-terminally truncated Cyclin H

Cdk7, Cdk7^{S164E} and Cdk7^{S164A} were expressed together with full-length Cyclin H or a C-terminally truncated Cyclin H (1-291). The Cyclin H C-terminus is supposed to be mainly unstructured and thus not resolved in the CycH crystal structures (Andersen et al. 1997; Kim et al. 1996). As previously shown Cdk7^{S164E} and Cdk7^{S164A} mutations can be purified together with full-length Cyclin H. In contrast, Cdk7 and Cdk7^{S164E} formed complexes with CycH (1-291), but the Cdk7^{S164A} mutant did not co-purify the C-terminal truncated CycH (1-291) (Fig. 7.8, A). A radioactive kinase assay with Cdk7 and Cdk7^{S164E} in complex with truncated CycH (1-291) compared to Cdk7 in complex with full-length CycH revealed no alteration in activity towards GST-CTD$_{[52]}$ (Figure 5.1.8).

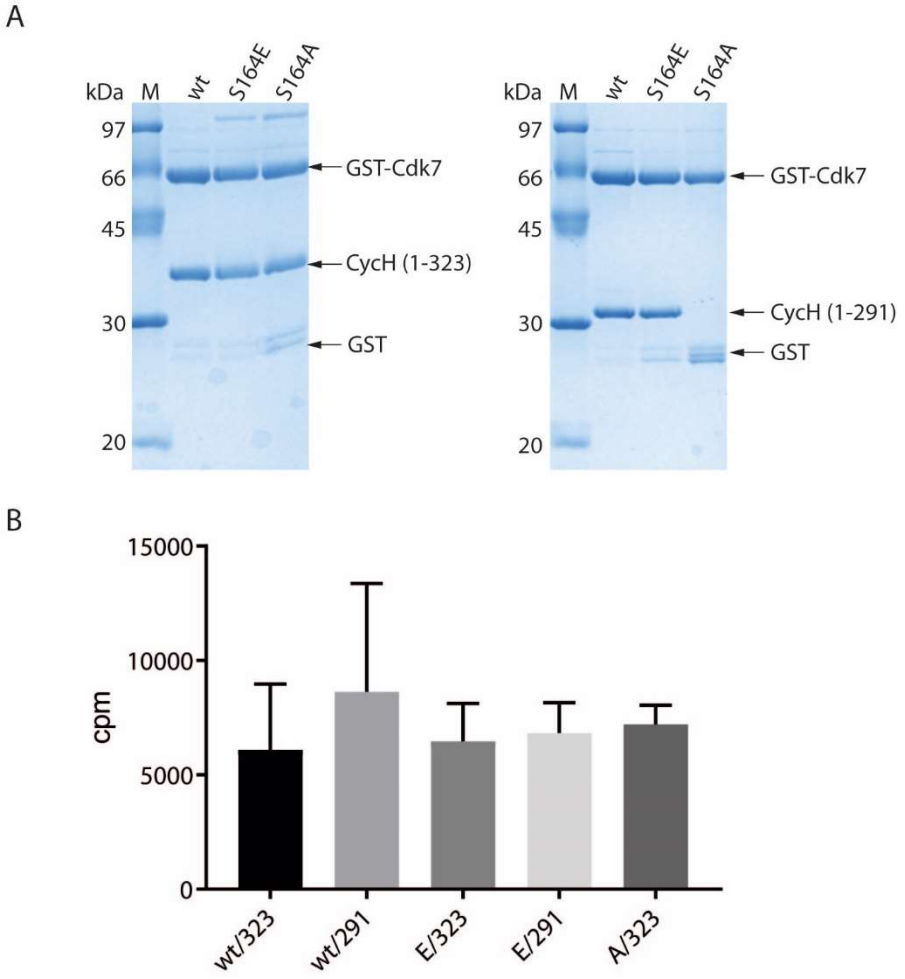


Figure 5.1.8: Cdk^{S164A} fails to bind C-terminally truncated Cyclin H.
A) Co-expression and affinity purification of Cdk7 and Cdk7 serine 164 mutants with full-length CycH (1-323) or C-terminally truncated CycH (1-291). **B**) Radioactive kinase activity assay of the Cdk7/CycH complexes shown in A. 0.1 µM of each complex were incubated with 10 µM GST-CTD$_{[52]}$ and 1 mM ATP for 15 minutes at 30°C. Data represent mean + SD of at least two independent experiments.

5.1.8 Mat1 binding affinity

Despite the long known interaction between Cdk7/CycH and Mat1 no detailed data regarding the binding affinity of the two components are published. The binding affinity of Cdk7/CycH to Mat1 was determined by surface plasmon resonance (SPR) and compared to $Cdk7^{S164A}$/CycH $Cdk7^{S164E}$/CycH and Cdk7/CycH (1-291). For analysis, *Sf9* cell expressed MBP-Mat1 (230-309) was used as ligand and immobilized on the chip with an anti-MBP antibody. As analyte, Cdk7/CycH, $Cdk7^{S164A}$/CycH, $Cdk7^{S164E}$/CycH, and Cdk7/CycH (1-291) were tested in different concentrations. Chips using MBP as ligand were used for control experiments (Figure 5.1.9). SPR measurements were performed in collaboration with Dr. Karl Gatterdam, Institute of Structural Biology, Bonn.

Samples were analysed in multicycle experiments in seven concentrations ranging from 0.2 nM to 150 nM. All Cdk7/CycH complexes tested bound to Mat1 (230-309) in a comparable manner and are therefore described together. Association of the dimeric Cdk7/CycH complexes could be detected at all concentrations tested. Cdk7/CycH binding to Mat1 (230-309) appears to be very stable with rather low dissociation constants. Application of a 1:1 fitting model to determine the K_D did not properly reconcile the curves observed. The K_D of the binding was therefore calculated by analysis of the steady state affinity. Cdk7/CycH complexes were applied as analyte until the association was saturated, followed by a long dissociation phase. Using steady state kinetics of the dissociation constant, the K_D was determined for Cdk7/CycH binding to Mat1 (230-309). The determined binding affinities were 9.8 nM for Cdk7/CycH, 11 nM for $Cdk7^{S164A}$/CycH, and 7.06 nM for $Cdk7^{S164E}$/CycH. Truncation of the Cyclin H C-terminus resulted of an K_D of 3.75 nM. Control measurements with MBP immobilized on the chip and Cdk7/CycH, $Cdk7^{S164A}$/CycH, and $Cdk7^{S164E}$/CycH at a concentration of 50 nM did not show any association to of the Cdk7/CycH complexes to MBP, confirming specificity of the measurement for Mat1.

Taken together, the binding affinity of Cdk7/CycH to Mat1 (230-309) was found to be in the low nanomolar range. Once formed, the dimeric complex is very stable and washout of Cdk7/CycH from the immobilized MBP-Mat1 (230-309) was nearly absent. Neither mutation of Cdk7 serine 164 to alanine or glutamate, nor truncation of the Cyclin H C-terminus had a significant influence on Mat1 binding affinity.

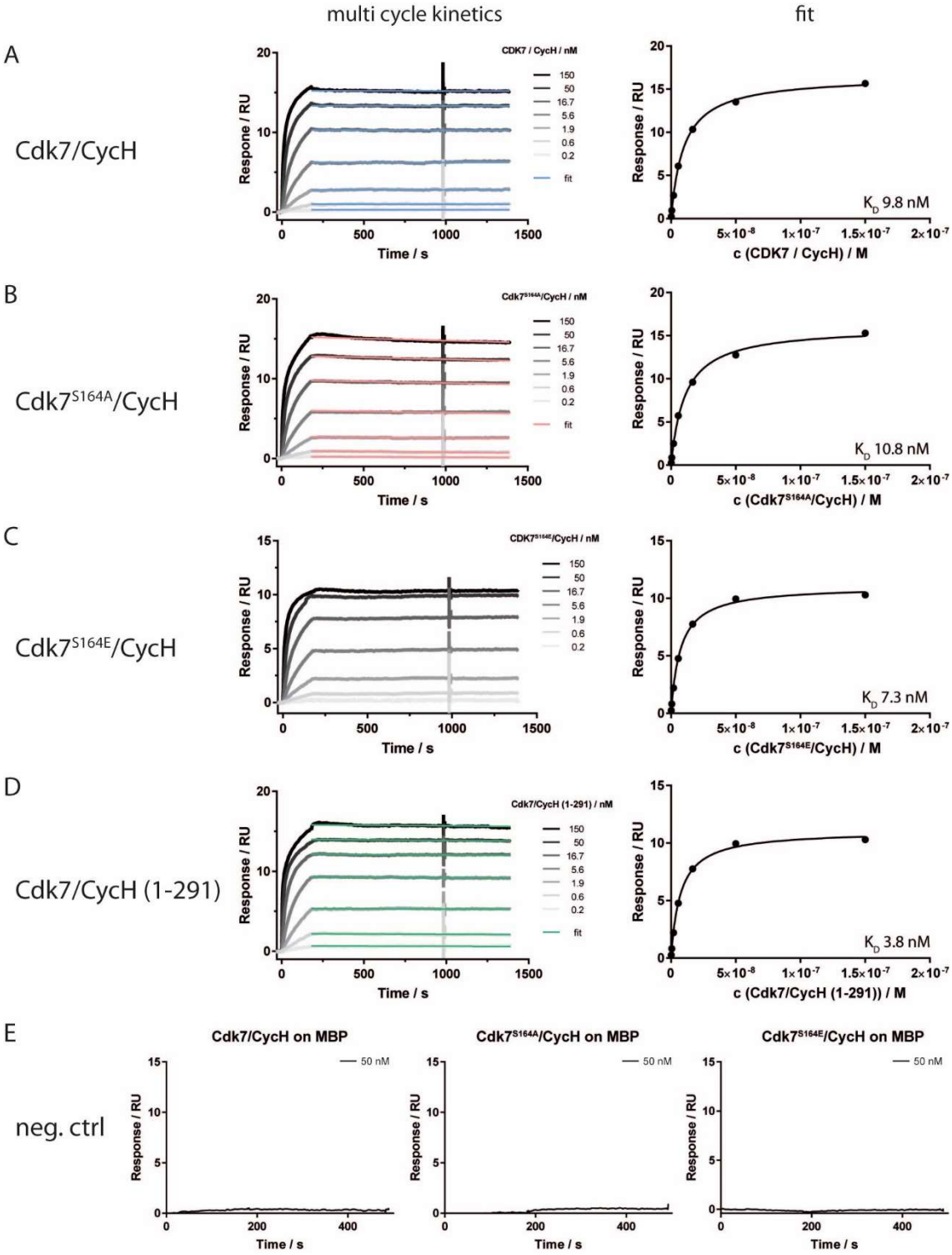


Figure 5.1.9: Determination of the Cdk7/CycH binding affinity to Mat1
Surface plasmon resonance measurements using immobilized MBP-Mat1 (230-309) as ligand. Cdk7/CycH complexes were applied as ligands in multicycle kinetic measurements. Sensorgrams are shown on the left and fit of the multicycle kinetics on the right for **A)** Cdk7/CycH, **B)** Cdk7^{S164A}/CycH, **C)** Cdk7^{S164E}/CycH, **D)** Cdk7/CycH (1-291), and **E)** Cdk7/CycH and S164 mutants on MBP (kinetics sensorgram only).

5.1.9 Cdk7 Ser164 phosphorylation is required for Mat1 dependent gain of activity

To analyse if Cdk7 serine 164 phosphorylation affects Cdk7 activity upon complex formation with Mat1, Cdk7/CycH, Cdk7^{S164A}/CycH, and Cdk7^{S164E}/CycH were incubated with increasing concentrations of MBP-Mat1 (230-309). The concentration of the Cdk7/CycH complex was kept constant at 0.1 µM. Cdk7/CycH, Cdk7^{S164A}/CycH, or Cdk7^{S164E}/CycH were pre-incubated for 10 minutes at room temperature with increasing concentrations (0.1 µM – 6.4 µM) MBP-Mat1 (230-309) (Figure 5.1.10).

Incubation with Cdk7/CycH in absence of Mat1 results in an intermediate shift of GST-CTD$_{[52]}$, which is in accordance with previous activity measurements indicating a moderate GST-CTD$_{[52]}$ phosphorylation at these conditions. Incubation of GST-CTD$_{[52]}$ with MBP-Mat1 (230-309) alone had no effect on GST-CTD$_{[52]}$ migration. Pre-incubation of 0.1 µM Cdk7/CycH with 0.1 µM or 0.2 µM Mat1 (230-309) has no apparent effect on GST-CTD$_{[52]}$ migration behaviour compared to Cdk7/CycH alone. Addition of 0.4 µM Mat1 (230-309) resulted in a profound shift of the GST-CTD$_{[52]}$. Further increase in Mat1 did not alter GST-CTD$_{[52]}$ migration behaviour anymore.

As expected, GST-CTD$_{[52]}$ phosphorylated with Cdk7^{S164A}/CycH or Cdk7^{S164E}/CycH in the absence of Mat1 showed a similar shift in GST-CTD$_{[52]}$ migration as Cdk7/CycH, confirming equal activity of the dimeric complexes. In contrast to Cdk7wt, addition of Mat1, even at a concentration 6.4 µM, did not result in a complete shift of the GST-CTD into the upper migrating form. Only a mild shift in the migration behaviour is observed (Figure 5.1.10, A).

To quantitatively validate the effects observed with SDS-PAGE, the same experimental setup was applied using radioactive γ-[^{32}P]-ATP. For better visualization, data were normalized to the activity of the respective Cdk7/CycH complex and displayed as fold-change in activity (Figure 5.1.10, B). Pre-incubation of Cdk7/CycH and Mat1 (230-309) at equimolar concentrations of 0.1 µM each already resulted in a nearly 4-fold increase in activity. Increase in activity to 7-fold of the Cdk7/CycH control level was observed until a Mat1 concentration of 0.4 µM. A further increase in Mat1 concentration did not further increase Cdk7 kinase activity. Cdk7^{S164A} and Cdk7^{S164E} mutant activity mildly increased in presence of 0.1 µM Mat1. In general, activity of the Cdk7 Ser164 mutants increased by only 1.6-fold (S164A) and 2-fold (S164E) even at high Mat1 concentration (Figure 5.1.10, B). In accordance with these results, co-purification of Cdk7^{S164A}/CycH supplemented with Mat1 (230-309) resembles the activity of dimeric Cdk7/CycH preparations and does not exhibit increased activity as observed with Cdk7/CycH and Mat1 (data not shown).

To exclude that the observed effect is artificial due to Mat1 truncation, recombinant full length MBP-Mat1 (1-309) was expressed and purified from *Sf9* insect cells. 0.1 µM Cdk7/CycH complexes were analysed for their activity after pre-incubation with 0.4 µM MBP-Mat1 (230-309) or 0.4 µM MBP-Mat1 (1-309). As observed before, Cdk7 activity was strongly increased (8.4-fold) by MBP-Mat1 (230-309), whereas Cdk7 Ser164 mutants only increased approximately 2-fold. The data obtained upon addition of full-length Mat1 mirrored the findings of the truncated Mat1 with the only restriction, that the increase in activity was slightly less pronounced with a 6.4-fold increase of Cdk7 activity. In accordance with this reduction, also the Cdk7 Ser164 mutants were activated albeit to a lesser extent (Figure 5.1.10, C)

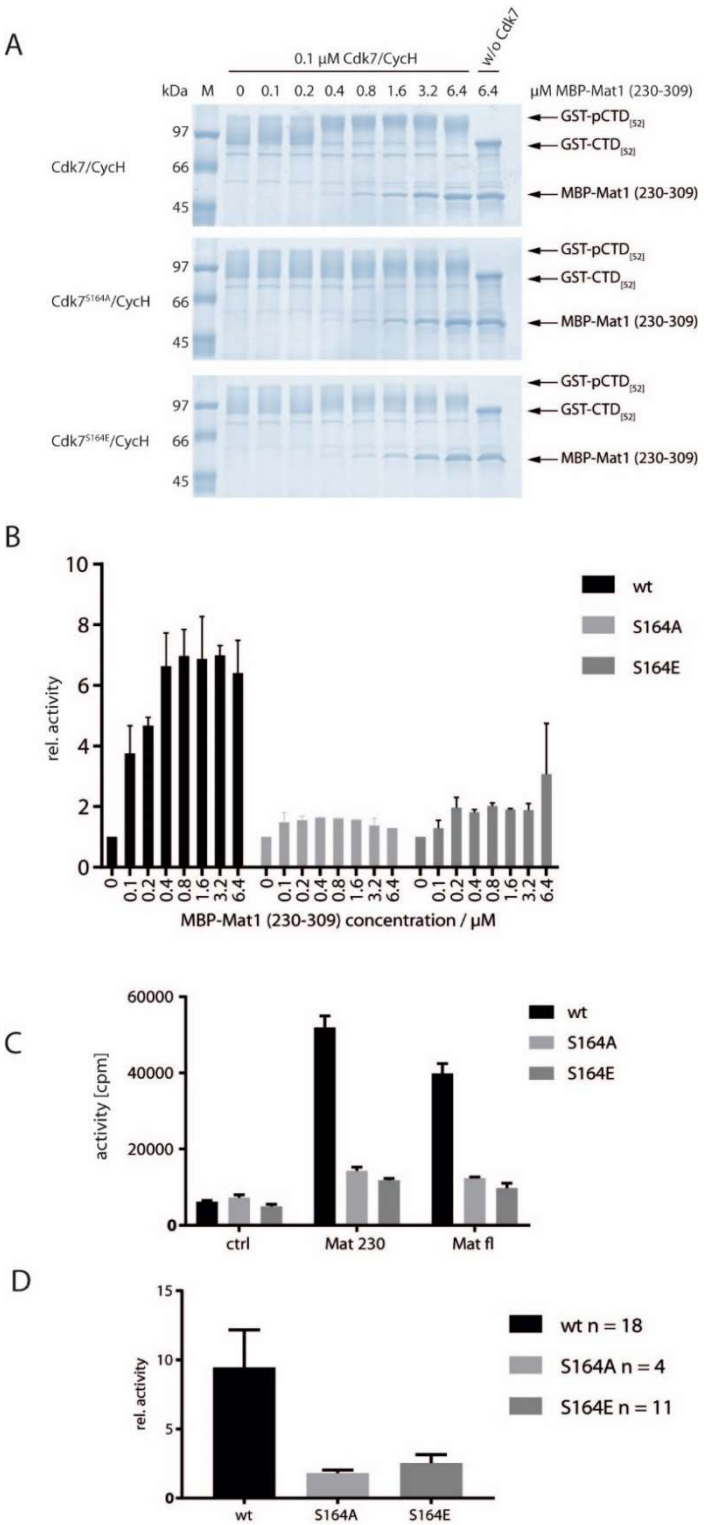


Figure 5.1.10: Cdk7 serine 164 phosphorylation is required for Mat1 dependent increase in activity.
A) SDS-PAGE analysis. 0.1 µM Cdk7/CycH and Cdk7Ser164 mutants were incubated with increasing amounts of MBP-Mat1 (230-309) for 10 minutes prior to kinase assay. Assay was started by addition of 2 mM ATP and samples were incubated for 15 minutes at 30°C. Reaction was stopped with SDS sample buffer and 2 µg GST-CTD$_{[52]}$ was analysed by SDS-PAGE. **B**) Radioactive measurements of the assay shown in A. Samples were treated as in A with the exception for stopping the kinase reaction with EDTA instead of SDS sample buffer. For analysis, data were normalized to the respective kinase activity in absence of Mat1. Data are presented as mean + SD from two independent experiments. **C**) Radioactive kinase assay. 0.1 µM Cdk7/CycH, Cdk7^{S164A}/CycH, and Cdk7^{S164E}/CycH were pre-incubated with 0.4 µM MBP-Mat1 (230-309), or 0.4 µM MBP-Mat1 (1-309) prior to kinase assay. Assay conditions were as in B. **D**) Summary of all experiments performed at the same conditions using a Cdk7/CycH concentration of 0.1 µM and a MBP-Mat1 (230-309) concentration of 0.4 µM. Data are shown as fold-change upon Mat1 addition and represent mean ± SD.

The Mat1 mediated effect on Cdk7 kinase activity varies between experiments. To prevent data misinterpretation due to variation within the experiments, data from all experiments performed at identical conditions are summarized as fold-change upon Mat1 addition in Figure 5.1.10, D. The conditions of these measurements were 0.1 µM Cdk7/CycH, 0.4 µM MBP-Mat1 (230-309), and 10 µM GST-$CTD_{[52]}$, a pre-incubation time of 10 minutes, 1 mM ATP, and an assay incubation time of 15 minutes. This results in a total n = 18 for Cdk7/CycH, n=4 for $Cdk7^{S164A}$/CycH, and n=11 for $Cdk7^{S164E}$/CycH. According to the summarized data, Mat1 increases Cdk7 activity by 9.4-fold. $Cdk7^{S164A}$ activity is increased 1.8-fold, and $Cdk7^{S164E}$ activity 2.5-fold (Figure 5.1.10, D).

5.1.10 The CycH C-terminus regulates Cdk7/CycH/Mat1 activity

In this thesis it was demonstrated that the unstructured C-terminal part of Cyclin H contributes to Cdk7 binding in a Cdk7 Ser164 phosphorylation dependent manner (Figure 5.1.8).

It was therefore examined, if Cyclin H truncation also affects activity upon Mat1 binding. Interestingly, the Cyclin H C-terminus (aa 291-323) is highly polar comprising 19 charged residues within the C-terminal 33 amino acids (Figure 5.1.11, A). The twelve glutamate or aspartate residues are sparsed into three patches of four amino acids each. In addition, there are two sites of basic amino acids of which the amino acids 292-295 (KKRK) function as the Cyclin H nuclear localisation signal (Krempler et al. 2005). For investigation, Cyclin H was truncated in four successive steps at amino acids 313, 301, 295, and 291 and expressed together with $Cdk7^{S164E}$ (Figure 5.1.11, B). The activity of Cdk7/CycH (1-291) increases in presence of MBP-Mat1 (230-309) to a comparable extent as seen when complexed to full length Cyclin H. As observed before, $Cdk7^{S164E}$/CycH fails to show a similar increase in activity. C-terminal truncation of CycH at amino acids 313, 301, and 295 respectively, all resembled the the data obtained with $Cdk7^{S164E}$/CycH (1-323). In contrast, further truncation of Cyclin H to CycH (1-291) and thus removal of the basic cluster KKRK resulted in a Mat1 mediated increase in activity similar to the effect observed with wild type Cdk7/CycH (Figure 5.1.11, C).

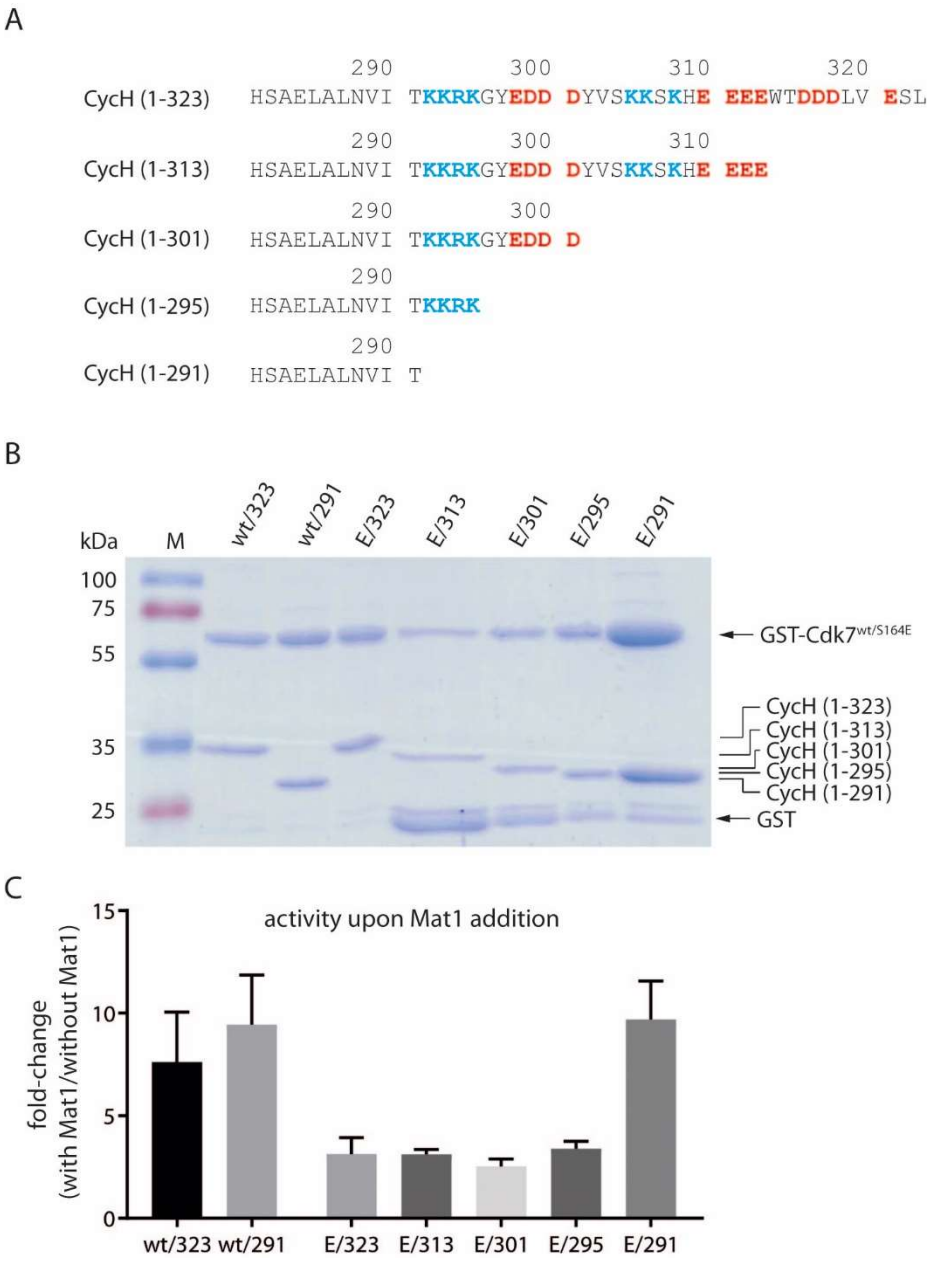


Figure 5.1.11: The Cyclin H C-terminus regulates Cdk7/CycH/Mat1 activity.
A) Amino acid sequence of the Cyclin H C-terminus (amino acids 280-323). Negatively and positively charged amino acids in the C-terminus following aa291 are highlighted by colour. **B**) Coomassie staining of 2 µg of affinity purified GST-Cdk7^{wt} and GST-Cdk7^{S164E} in complex with different C-terminal truncated Cyclin H constructs. **C**) Increase in activity upon MBP-Mat1 (230-309) binding of the Cdk7/CycH preparations shown in B. Kinase activity of 0.1 µM Cdk7/CycH complex towards 10 µM GST-CTD$_{[52]}$ was analysed either in presence or absence of 0.4 µM MBP-Mat1 (230-309). The reaction mixtures were pre-incubated for 10 minutes at room temperature before starting the kinase assay with 1 mM ATP. Upon ATP addition, reaction was incubated for 15 minutes at 30°C. Data are depicted as fold-change of the respective kinase activity upon Mat1 addition. Data represent mean ± SD obtained in three independent experiments.

5.1.11 Stability of binary Cdk7/CycH and ternary Cdk7/CycH/Mat1

To accurately determine the stability of dimeric and trimeric Cdk7 complexes, samples were analysed for thermal stability by nano differential scanning fluorometry (nanoDSF). Protein stability is determined by a change of the intrinsic absorbance of tryptophan residues which changes as tryptophans become solvent accessible by unfolding of the protein. Unfolding was monitored over a temperature from 20°C to 90°C at a constant heating of 1.5°C/minute. All proteins were kept in the same buffer formulation at 20 mM HEPES pH7.6, 150 mM NaCl, and 1 mM TCEP.

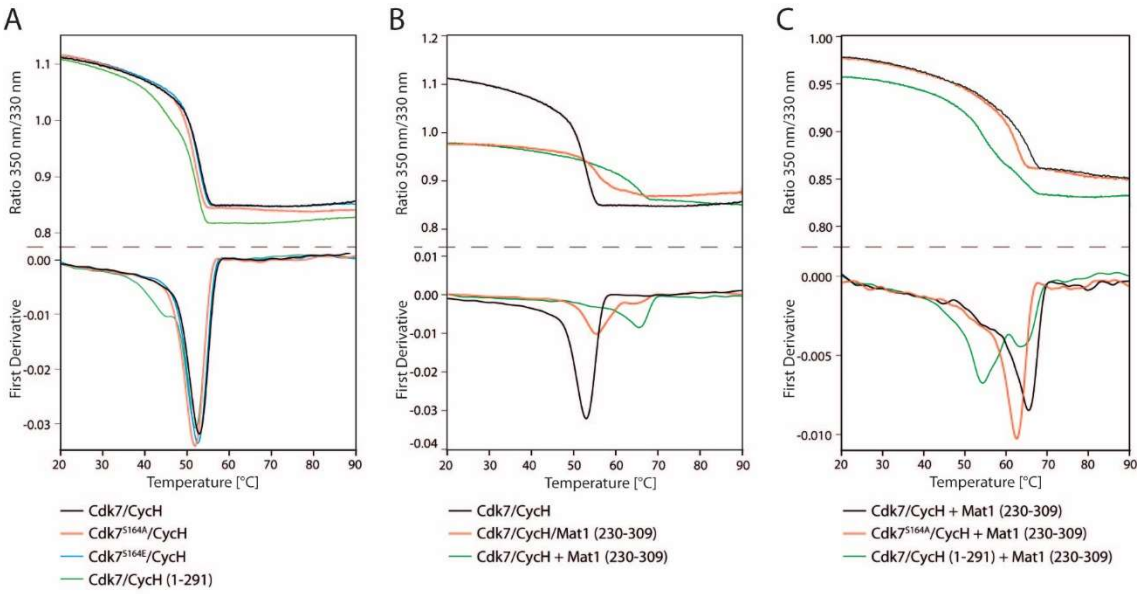
In a first set of experiments the melting point of dimeric Cdk7/CycH preparations was analysed. The stability of Cdk7/CycH was determined and compared to Cdk7^{S164A}/CycH, Cdk7^{S164E}/CycH, and Cdk7/CycH (1-291). Figure 5.1.12 shows the thermal unfolding of the Cdk7/CycH complexes. The upper panel shows the 350 nm/330 nm ratio of the intrinsic tryptophan fluorescence of the complexes. Upon increasing temperature, the ratio changes as a consequence of the unfolding of the complex. The lower panel represents the first derivative of the measured ratio. The peak maximum of the first derivation is defined as the melting temperature (T_m) of the proteins.

Using nanoDSF the Cdk7/CycH melting temperature was at 52.6°C. Cdk7^{S164A}/CycH was less stable by 1°C with a T_m of 51.6°C. Cdk7^{S164E}/CycH unfolded at a temperature of 52.4°C. The complex containing the truncated CycH, Cdk7/CycH (1-291), unfolds at 51.7°C. Hence, Cdk7/CycH and Cdk7^{S164E}/CycH display equal stability whereas Cdk7^{S164A}/CycH is slightly less stable. Interestingly, stability of the cyclin truncated Cdk7/CycH (1-291) resembles the stability of the Cdk7^{S164A}/CycH mutant (Figure 5.1.12, A).

It has been described, that *in vitro* re-constituted trimeric Cdk7/CycH/Mat1 complexes have an increased thermal stability over co-expressed trimeric Cdk7 complexes, but no detailed description has been provided yet (Larochelle et al. 2001). Cdk7/CycH was compared to co-purified Cdk7/CycH/Mat1 (230-309) and Cdk7/CycH + Mat1 (230-309) in which Cdk7/CycH and Mat1 were expressed separately and co-purified for trimeric complex generation. The two trimeric complexes show the same 350/330 ratio in the beginning and the end of the measurement, reflecting, that both complexes do not differ in their subunit composition but only in the way of purification and the associated differences in T-loop phosphorylation. Whereas Cdk7/CycH unfolded at 52.6°C, the co-expressed trimeric Cdk7 complex had a stability of 55.4°C. However, the stability of the *in vitro* reconstituted trimeric complex was increased by almost 10°C to 64.9°C (Figure 5.1.12, B). Similar, the stability of other co-expressed trimeric Cdk7 preparations were analysed. The melting points were determined to be 54°C for Cdk7/CycH/Mat1, 54.6°C for Cdk7/CycH (1-291)/Mat1 (230-309), and 55.9°C for Cdk7^{S164E}/CycH (1-291)/Mat1 (230-309) (Figure 5.1.12, D). Thus, co-purification of trimeric complexes from individual components results in a profound increase in stability compared to co-expression.

Next it was analysed how Cdk7^{S164A} mutation or cyclin truncation affect the stability of separately expressed, co-purified Cdk7/CycH + Mat1 (230-309) complexes. Again, the Cdk7^{S164A} mutation destabilises the complex. In the setup tested, the stability was lowered by 2.6°C compared to Cdk7 wild type complexes to 62.2°C. Nonetheless, the general stabilizing effect was preserved. Complexes containing truncated Cyclin H displayed a bi-phasic unfolding behaviour. A first transition point was observed at 54.6°C resembling the stability of the complex when co-expressed. A second transition point was detected at 64°C reflecting the stability of the co-purified Cdk7/CycH + Mat1 sample (Figure 5.2.12, C). Cyclin H truncation thus results in a dichotomous complex formation with stability

characteristics of dimeric and trimeric Cdk7 complexes. This suggests an involvement of the Cyclin H C-terminus in trimeric Cdk7/CycH/Mat1 complex formation and stability.



D

	complex	Tm
dimeric	Cdk7/CycH	52.6°C
	Cdk7^{S164A}/CycH	51.6°C
	Cdk7^{S164E}/CycH	52.4°C
	Cdk7/CycH (1-291)	51.7°C
trimeric	*Cdk7/CycH/Mat1*	*54.0°C*
	Cdk7/CycH/Mat1 (230-309)	55.4°C
	Cdk7/CycH (1-291)/Mat1 (230-309)	*54.6°C*
	Cdk7^{S164E}/CycH (1-291)/Mat1 (230-309)	*55.9°C*
	Cdk7/CycH + Mat1 (230-309)	64.9°C
	Cdk7^{S164A}/CycH + Mat1 (230-309)	62.4°C
	Cdk7/CycH (1-291) + Mat1 (230-309)	54.6/64.0°C

Figure 5.1.12: Stability of dimeric and trimeric Cdk7 complexes.
Thermal stability of hetero dimeric and trimeric Cdk7 complexes was measured at a concentration of 5 µM. The temperature gradient was set for 20°C - 90°C at a constant heating of 1.5°C/minute. All samples were measured in triplicate. The displayed data represent the mean of all three measurements. **A**) Stability of dimeric Cdk7/CycH complexes. **B**) Stability of dimeric Cdk7/CycH, co-expressed Cdk7/CycH/Mat1 (230-309), and Cdk7/CycH + Mat1 (230-309). **C**) Stability of Cdk7/CycH + Mat1 (230-309), Cdk7^{S164A}/CycH + Mat1 (230-309), and Cdk7/CycH (1-291) + Mat1. **D**) Summary of the determined melting temperatures. Samples which are not represented above are displayed in italics.

The bi-phasic appearance of Cdk7/CycH (1-291) + Mat1 (230-309) could originate from sub-stoichiometric amounts of Mat1. To further address this question, trimeric complexes were re-constituted *in vitro* immediately prior to nanoDSF measurements from purified dimeric Cdk7/CycH and purified full-length Mat1 (1-309) (Figure 5.1.13). When analysed alone, Mat1 does not show a distinct melting point but is gradually unfolded over a long temperature range. This unfolding behaviour might be explained by the structure of Mat1, which was found to contain a long helix between the N-terminal Ring finger, and the C-terminal Cdk7/CycH interaction domain (Greber et al. 2017). To identify the optimal ratio of Cdk7/CycH and Mat1 for nanoDSF measurements, 2.5 µM of Cdk7/CycH complex were incubated with increasing concentrations of Mat1 (Figure 5.1.13, C). Addition of Mat1 results in a decrease in the initial 350 nm/330 nm ratio as seen previously with co-purified trimeric complexes. Upon increasing Mat1 concentration an increased stability of the complex can be observed. For better visualization of the melting point, data of the first derivative shown in Figure 5.1.13, C, lower panel, are displayed for each Mat1 concentration individually compared to Cdk7/CycH (Figure 5.1.13, D-G). Equimolar mixing and a 2.5-fold excess of Mat1 have essentially no effect on Cdk7/CycH stability. At a 5-fold excess of Mat1 a bi-phasic unfolding with maxima at 52.7°C and at 59.3°C can be observed, indicating a mixture of dimeric and trimeric complexes. Further increase of Mat1 to a 10-fold excess completely shifts the observed stability to a uniform melting point at 59.3°C. Thus, *in vitro* reconstitution of the trimeric complexes produces comparable data to co-purification experiments. The lower melting point of 59.3°C compared to 64.9°C of the co-purified Cdk7/CycH + Mat1 (230-309) complex is probably due to the use of full-length Mat1 in the *in vitro* approach.

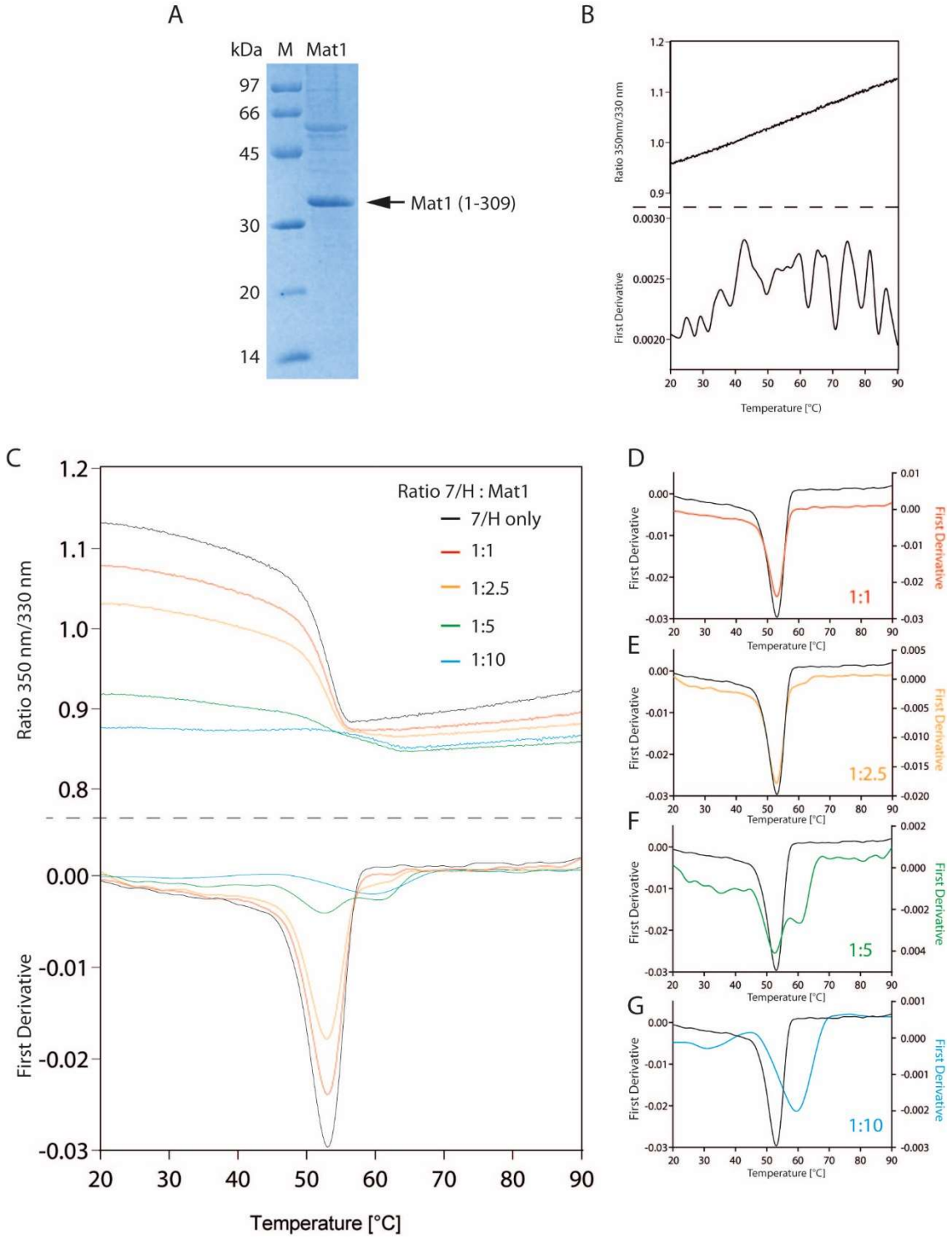


Figure 5.1.13: Thermal stability of Cdk7/CycH/Mat1 complexes upon *in vitro* reconstitution.
A) SDS PAGE analysis of 2 µg purified full length Mat1. **B**) Thermal denaturation of 10 µM Mat1. Upper panel: ratio, lower panel: first derivative of the data above. **C**) The thermal stability of 2.5 µM Cdk7/CycH either alone or with increasing concentrations of Mat1 was monitored. **D-G**) Data of the first derivative shown in C, but displayed for each Cdk7/CycH : Mat1 ratio individually. For better visualization data are displayed on different scale for Cdk7/CycH and Cdk7/CycH + Mat1.

Next, the *in vitro* reconstitution using full-length Mat1 was expanded to Cdk7^{S164A}/CycH, Cdk7^{S164E}/CycH, and Cdk7/CycH (1-291) complexes (Figure 5.1.14). Samples of dimeric complexes at a concentration of 2.5 µM were incubated with a 10-fold excess of Mat1 for 10 minutes at room temperature prior to nanoDSF measurements. Upon Mat1 incubation, stability of Cdk7/CycH was increased from 52.5°C to 59.3°C. Cdk7^{S164A}/CycH stability increased from 51.3°C to 56.8°C, and Cdk7^{S164E}/CycH stability increased from 52.5°C to 58.4°C. Incubation of Cdk7/CycH (1-291) with Mat1 in part increased the stability of the complex from 51.2°C to 59.8°C, but, in contrast to Cdk7 complexes with full-length Cyclin H, no uniform melting point was detected. Hence, data from direct *in vitro* reconstitution from purified full-length Mat1 and different Cdk7/CycH preparations are supportive to the data obtained using co-purified complexes of N-terminally truncated Mat1 (230-309) + Cdk7/CycH. By applying nanoDSF it was possible to get detailed information about Cdk7 complex stability. In general, Cdk7^{S164A} mutation resulted in a decreased stability of both, dimeric and trimeric Cdk7 complexes. Moreover, an involvement of the C-terminus of Cyclin H on complex stability could be assigned.

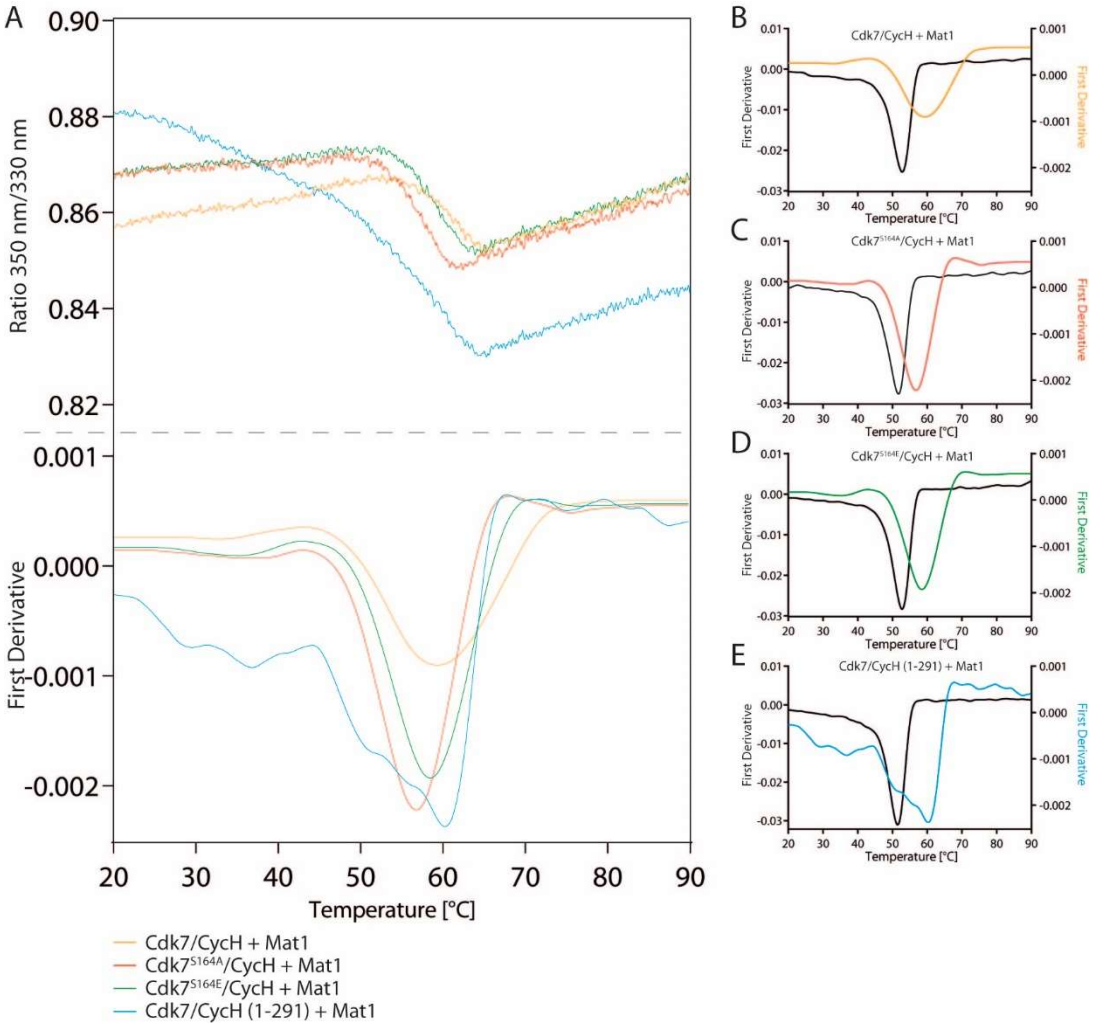


Figure 5.1.14: Stability of *in vitro* reconstituted trimeric Cdk7 complexes.
A) Thermal unfolding of dimeric Cdk7/CycH complexes in presence of Mat1. 2.5 µM of the respective complex was mixed with 25 µM Mat1. After 10 minutes of pre-incubation at room temperature, thermal stability of the samples was analysed. **B-E**) Individual representation of the curves shown in A. The stability of the samples was compared to the respective dimeric complex which was monitored in parallel but not shown in A.

5.1.12 Specific and selective covalent inhibition of Cdk7 by YKL-5-124

Inhibition of transcription regulating kinases is currently debated as a novel treatment option in cancer. Targeting kinases which fulfil general tasks in all but not only malignant cells appears contraintuitive at first sight, but interfering with transcription has been demonstrated to be an effective way for treatment of several cancers. The rationale behind is the finding, that cancer cells highly depend on high expression of certain factors to maintain their malignant state (Bradner et al. 2017). In 2014, the group of Nathanael Gray at the Dana-Farber Cancer Center in Boston described the development of a covalent inhibitor, THZ-1, which potently and selectively inhibits Cdk7, Cdk12, and Cdk13 (Kwiatkowski et al. 2014). In their studies they could establish, that interference with the transcriptional machinery affects super-enhancer driven tumors via RUNX1 in T cell acute lymphoblastic leukemia and via MYCN in neuroblastoma (Kwiatkowski et al. 2014; Chipumuro et al. 2014). Further development of THZ-1 resulted in THZ-531 which is highly selective for Cdk12 and Cdk13 but lacks inhibitory potential towards Cdk7 (Zhang et al. 2016).

During the course of this thesis the laboratory of Dr. Nathanael Gray developed a new inhibitor, YKL-5-124, which was designed to specifically inhibit Cdk7, but not targeting Cdk12 or Cdk13. For generation of the inhibitor, a scaffold originating from the PAK4 inhibitor PF-3758309 (Murray et al. 2010) which was found to inhibit Cdk7 was combined with the covalent THZ-1 warhead.

In this study the inhibitor YKL-5-124 was analysed for its *in vitro* inhibitory potential towards recombinant Cdk7, Cdk12, and Cdk13 complexes. The data were compared to the initially described Cdk7 inhibitor THZ-1. THZ-1 inhibited all three kinases at comparable levels. Strangely, incubation of Cdk7 with high concentration of THZ-1 above 10 µM reversed Cdk7 inhibition in a concentration dependent manner. For determination of the IC_{50} values these data points were excluded. Cdk7 was inhibited by THZ-1 with an IC_{50} of 293 nM and thus was slightly better compared to Cdk12 and Cdk13 with 893 nM and 628 nM, respectively (Figure 5.1.15, A). YKL-5-124 potently inhibited Cdk7 with an IC_{50} of 54 nM. Inhibition of Cdk12 and Cdk13 was only observed at high micromolar concentrations of the inhibitor. IC_{50} for Cdk12 and Cdk13 are calculated as 120 µM and 269 µM, respectively (Figure 5.1.15, B). YKL-5-124 thus represents a new potent inhibitor of Cdk7 with virtually no cross-reactivity towards Cdk12 and Cdk13.

Application of YKL-5-124 in cells results in cell cycle arrest in G1/S phase, probably due to loss of Cdk2 and Cdk4 activation by Cdk7. Consequently, the expression of cell cycle related genes was reduced. However, YKL-5-124 had no effect on global mRNA levels and did not induce apoptosis, as observed with THZ-1. Re-constitution of THZ-1 inhibition using YKL-5-124, and THZ-531 recapitulated the effects seen with THZ-1. Dividing cells thus seem to primarily depend on the CAK function of Cdk7, rather than its functions in transcription (Olson et al. 2019).

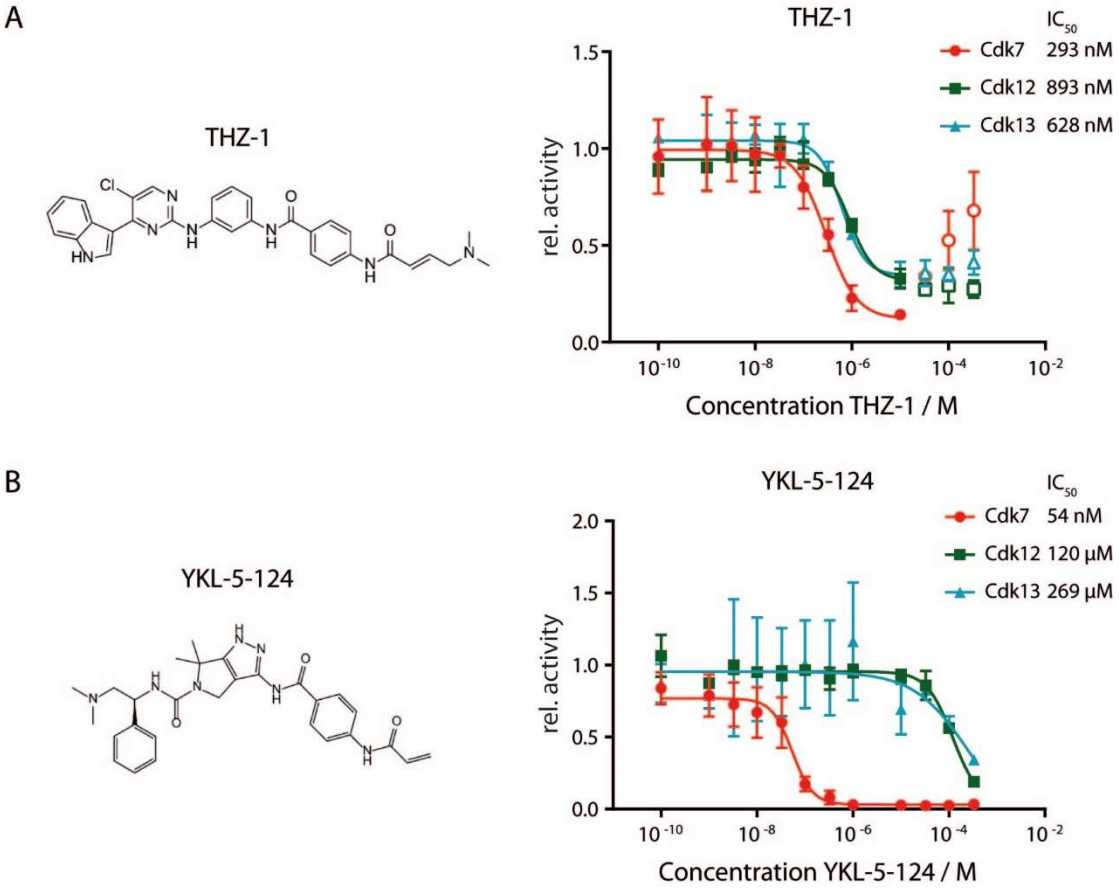


Figure 5.1.15: Pharmacologic inhibition of Cdk7, Cdk12, and Cdk13 by THZ-1 and YKL-5-124.
A) Inhibition of Cdk7, Cdk12, and Cdk13 with THZ-1. Left, chemical structure of THZ-1. Right, dose response measurements in a radioactive kinase assay. Cdk7 (2-346)/CycH (1-323)/Mat1 (1-309), Cdk12 (714-1063)/CycK (1-267), and Cdk13 (694-1093)/CycK (1-267) at a concentration of 0.2 µM were incubated in presence of 1 mM ATP with increasing concentrations of THZ-1 for 5 minutes at 30°C prior to the kinase assay. The kinase reaction was started by addition of substrate and then incubated for 15 minutes at 30°C. Activity was normalized to a DMSO control. Data represent mean ± SD of three independent experiments. Empty symbols represent data which were excluded for fitting and IC_{50} determination. **B**) Inhibition of Cdk7, Cdk12, and Cdk13 with YKL-5-124. Left, chemical structure of YKL-5-124. Right, dose response measurements in a radioactive kinase assay. Reaction conditions were the same as described in A.

5.2 Biochemical characterisation of the Cdk10/CycM complex

Despite its discovery 25 years ago, only little is known about cyclin-dependent kinase 10 (Cdk10) and its cellular functions. The kinase domain of Cdk10 is most closely related to Cdk11 sharing a sequence identity of 53% in humans. Nonetheless, they differ in their cognate cyclins. Cdk11 is found to associate with Cyclin L1, Cyclin L2 and Cyclin D3. The cyclin partner of Cdk10 instead is Cyclin M and has been identified recently (Guen et al. 2013). Cyclin M is the gene product of FAM58A and mutations in this gene have been characterized in a rare X-linked gonosomal disease called STAR syndrome (Guen et al. 2013; Unger et al. 2008). Moreover, knock-out of Cdk10 in mice is lethal during embryogenesis, but Cdk10 is dispensable for maintenance of $Cdk10^{-/-}$ mouse embryonal fibroblasts acquired from these embryos (Windpassinger et al. 2017). In addition, alterations in Cdk10 levels have been found in several cancer types. However, the studies are not conclusive, depicting Cdk10 as a positive or negative regulator in disease progression dependent on the cell types analysed (Iorns et al. 2008). All this illuminates how little is actually understood about the role of Cdk10 in physiologic and pathologic conditions and thus it is not surprising, that Cdk10 was denoted an understudied kinase recently (Rodgers et al. 2018).

No studies have yet been published which provide a detailed structural or functional characterisation of the Cdk10/CycM complex. In this part of the thesis expression and purification strategies for Cdk10/CycM are established which allow for functional and structural investigations. Moreover, functional studies elucidate the kinase activity and pharmacological inhibition of Cdk10. Finally, a chemical genetic approach is used to identify new substrates of this Cdk/Cyclin complex.

5.2.1 Cdk10 and Cyclin M domain architecture

Human Cdk10 (uniprot accession number Q5131) is a protein of 360 amino acids. It does not contain large N-terminal or C-terminal extensions in addition to the kinase domain. Within the kinase-domain it contains the PISSLRE sequence a variation of the Cdk1 PSTAIRE c-Helix, which is the conserved cyclin interaction motif. Aspartate 163 serves as proton acceptor during catalysis, and thus mutation of this residue results in a catalytically inactive enzyme. Cdk10 is activated by phosphorylation at threonine 196 by a yet unidentified kinase. At the C-terminus a bi-partite nuclear localisation signal is found (Figure 5.2.1, A). Cyclin M (uniprot accession number QN81B3) is composed of 248 amino acids and contains two cyclin boxes. Cyclin M does not contain a nuclear localisation signal so that transport to the nucleus is probably dependent on Cdk10 interaction or other binding partners (Figure 5.2.1, B).

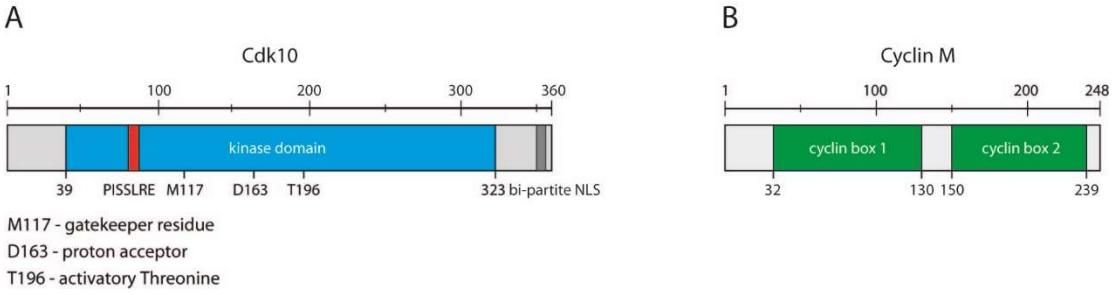


Figure 5.2.1: Domain architecture of Cyclin-dependent kinase 10 and Cyclin M.
A) Domain architecture of Cdk10. The kinase domain is depicted in blue. Within the kinase domain the PISSLRE motif for interaction with the cyclin is shown in red. Important residues are denoted. The bi-partite nuclear localization signal is highlighted in dark grey. **B**) Domain architecture of Cyclin M with the cyclin boxes depicted in green.

5.2.2 Establishing an expression and purification strategy for Cdk10/CycM

To analyse the Cyclin-dependent kinase 10/Cyclin M (Cdk10/CycM) complex, expression and purification strategies were developed that allow to analyse functional and structural aspects of the kinase. To this end, different affinity tags were tested for their effect on yield and solubility of the protein complex. In addition, to increase the expression levels of Cdk10/CycM complexes for crystallisation, a kinase-dead Cdk10^{D163N} mutant was generated.

Cdk10/CycM complexes were co-expressed from a single expression vector in *Sf9* cells using the Multibac^{Turbo} system (Bieniossek et al. 2008). The initial purification strategy utilized GST-tagged Cdk10 and GST-tagged Cyclin M (Figure 5.2.2, A). This strategy has been proven well-working for cyclin-dependent kinases 12 and 13 (Bösken et al. 2014; Greifenberg et al. 2016). For Cdk10/CycM, expression and purification of GST-tagged proteins resulted in stoichiometric Cdk10/CycM complexes. However, the protein complex was expressed at very low levels. In accordance with the low yield, *Sf9* cells infected with baculovirus to express GST-Cdk10/GST-CycM did not grow in size and cell viability decreased considerably during expression. Lowering the virus titer used to infect the cells or shortened expression times did not affect the expression levels (data not shown). Due to the low expression levels it was not possible to generate sufficient material for preparative size exclusion chromatography after TEV protease digestion. In order to increase the yield of the complex, Cdk10

was rendered kinase-dead by mutation of the catalytic aspartic acid 163 to asparagine (Cdk10^{D163N}). Expression of a kinase-dead mutant in fact increased the yield and allowed for tag-removal and further purification by SEC. Nonetheless, expression yield was still not sufficient to allow crystallographic approaches. To improve expression, Cdk10 was fused to Maltose binding protein (MBP) containing a linker for TEV protease digestion. Cyclin M was co-expressed from the same vector with a TEV cleavable N-terminal hexahistidine-tag (His-CycM). Combination of IMAC and affinity chromatography using Ni^{2+} and MBPTrap columns resulted in large quantities of MPB-Cdk10/His-CycM complexes. Subsequent tag removal and SEC with reverse MBPTrap resulted in pure protein. By applying this expression and purification strategy the yield of the complex was increased to 1 mg/L *Sf9* cell culture for Cdk10/Cyclin M and to 2 mg/L *Sf9* cell culture for the kinase-dead Cdk10^{D163N}/Cyclin M variant (Figure 5.2.2, B-E).

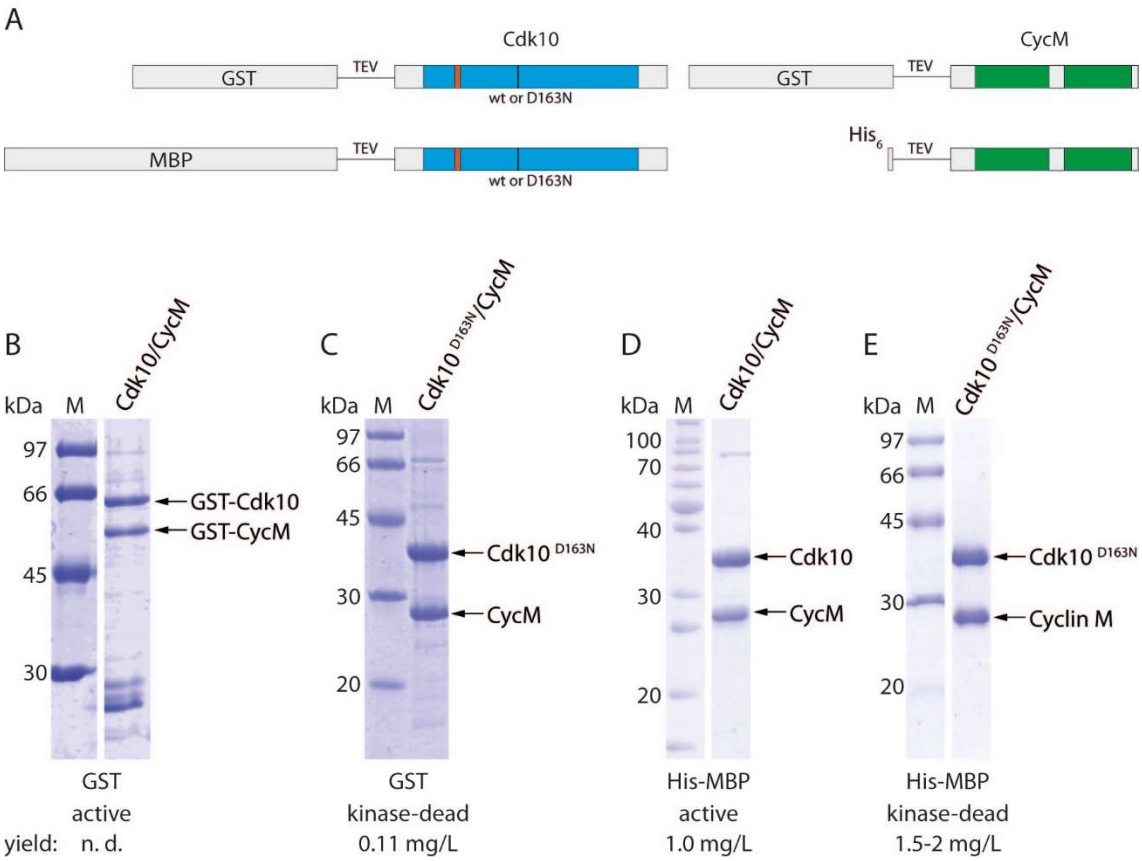


Figure 5.2.2: Optimization of Cdk10/CycM expression and purification.
A) Schematic representation of the constructs used. **B-E**) SDS-PAGE analysis of 4 µg purified Cdk10/CycM complexes after different expression and purification strategies. The indicated yield is shown in mg per liter *Sf9* cell culture below the gels.

5.2.3 Constructs for crystallographic approaches

For protein crystallography different constructs were tested. In addition to full-length Cdk10^{D163N} (1-360), N-terminal and C-terminal truncations of the protein were generated by PCR. These include N-terminal truncation of Cdk10^{D163N} at amino acid position 9, 20, and 27, as well as C-terminal truncation at amino acid position 331. All Cdk10^{D163N} constructs were co-expressed as N-terminal MBP-fusion proteins together with full-length hexahistidine-tagged Cyclin M (1-248) resulting in eight different Cdk10^{D163N}/Cyclin M complexes analysed (Figure 5.2.3).

The yield of the protein complexes decreased with construct size. In addition to decreased yield after affinity chromatography, Cdk10 constructs truncated at amino acid 20 results in insufficient TEV-digest of both MBP-Cdk10^{D163N} (20-360) and MBP-Cdk10^{D163N} (20-331). These constructs were not processed further, due to insufficient amounts for crystallization. N-terminal Cdk10 truncation at amino acid 27 impairs Cyclin M binding and complex formation. When purified via the His_6-CycM via Ni^{2+} coated resin, no protein was obtained. Purification via the MBP-tag at the kinase resulted in co-purification of small, sub-stoichiometric amounts of His-Cyclin M. After removal of the MBP- and the His-tag by TEV digest, proteins eluted as single proteins, but not as a protein complex in size exclusion chromatography using a S75 pg column (data not shown). This observation is consistent with published data in which a comparable Cdk10 construct failed to bind Cyclin M in a yeast two hybrid screen (Guen et al. 2013).

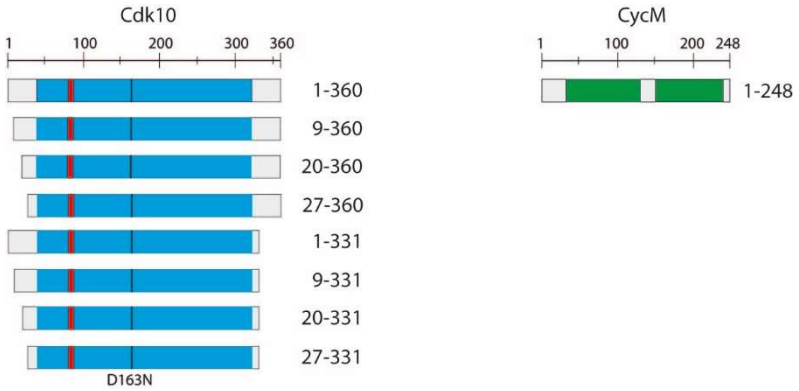


Figure 5.2.3: Cdk10/CycM constructs for crystallography.
N-terminal and C-terminal truncated Cdk10 constructs (left) used for co-expression with full-length CycM (right). The Cdk10 kinase domain is depicted in blue. The Cdk10 constructs used for crystallography all contained the D163N mutation.

Cdk10/Cyclin M complexes of the constructs Cdk10^{D163N} (1-360)/CycM, Cdk10^{D163N} (9-360)/CycM, and Cdk10^{D163N} (1-331)/CycM could be obtained in sufficient amount and homogeneity for crystallography. Crystallography trials were performed by Dr. Kanchan Anand, Institute of Structural Biology, Bonn. Cdk10^{D163N}/CycM constructs reproducibly formed crystals, but crystals mainly grew as thin needles. Until the end of this thesis, no crystals with diffraction spectra which allow for structure determination could be obtained (data not shown).

5.2.4 Cdk10 activity is dependent on Cyclin M co-expression

Cyclin-dependent kinases are activated by two events: first, phosphorylation of the activatory threonine in the activation loop, and second, binding to a cyclin partner. Peptide mass fingerprint analysis of Cdk10, either expressed alone or in combination with Cyclin M, confirmed that Cdk10 is phosphorylated within its activation loop at Thr196. The sample from Cdk10/CycM co-expression was additionally phosphorylated at Ser351 (data not shown). To test whether Cdk10 activity is dependent on Cyclin M, GST-Cdk10 was either expressed alone or co-expressed together with GST-Cyclin M and then purified via glutathione affinity chromatography (Figure 5.2.4, A). Purified GST-Cdk10 and GST-Cdk10/GST-CycM were assayed for activity towards GST-RNA pol II $CTD_{[52]}$ in a radioactive kinase assay. GST-Cdk10 alone does not display any activity, but incubation with GST-Cdk10/GST-CycM resulted in phosphorylation of the substrate (Figure 5.2.4, B). This confirms recent studies that Cdk10 activity is dependent on Cyclin M (Guen et al. 2013).

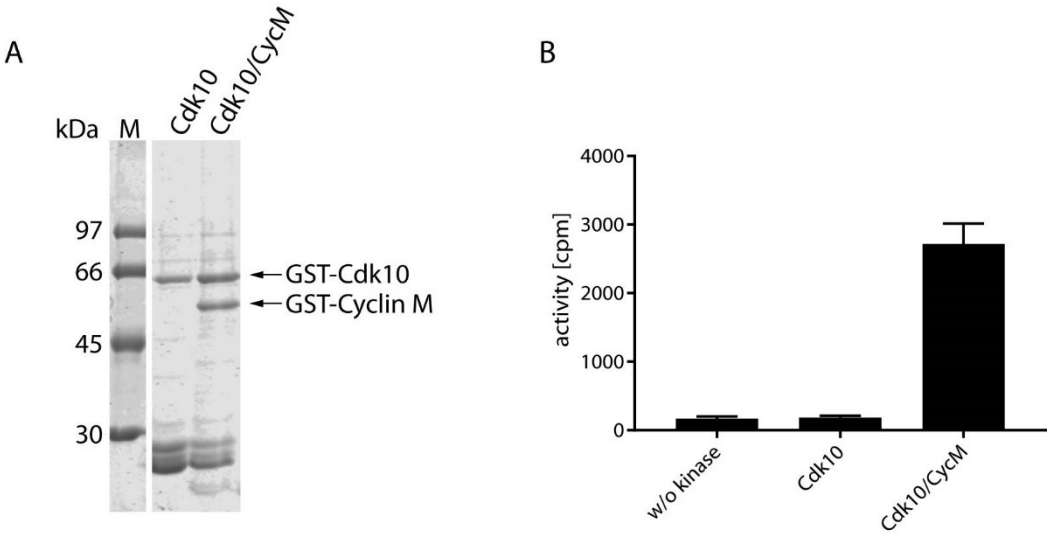


Figure 5.2.4: Cdk10 activity depends on Cyclin M co-expression.
A) SDS-PAGE analysis of 2 µg of affinity purified GST-Cdk10 and GST-Cdk10/GST-CycM. **B**) Radioactive kinase activity assay. For activity determination, 10 µM GST-$CTD_{[52]}$ were incubated for 30 minutes with 2 mM ATP either in absence or presence of 0.2 µM GST-Cdk10 or 0.2 µM GST-Cdk10/GST-CycM.

5.2.5 Cdk10 phosphorylates RNA pol II CTD and c-Myc, but not SRSF7

As a next step the activity of Cdk10/CycM towards RNA pol II C-terminal domain, c-Myc, and SRSF7 was examined. GST-CTD$_{[52]}$ is a common target for several transcriptional cyclin-dependent kinases and c-Myc has been identified as a substrate for Cdk9, Cdk12, and Cdk13 in previous studies (Bösken 2013; Greifenberg 2014). SRSF7 is a known substrate of Cdk11, which shares the highest sequence identity with Cdk10 (Hu et al. 2003). Cdk10/CycM was therefore incubated with GST-CTD$_{[52]}$, His-c-Myc (17-167), and GST-SRSF7 which were expressed in *E. coli*. Cdk10/CycM phosphorylated GST-CTD$_{[52]}$ and c-Myc, but not SRSF7. c-Myc phosphorylation was about 2.5-fold higher compared to GST-CTD$_{[52]}$ phosphorylation (Figure 5.2.5).

Hence, Cdk10 phosphorylates the Cdk substrates RNA pol II CTD, and c-Myc, but substrate specificity differs from Cdk11, which is known to phosphorylate SRSF7.

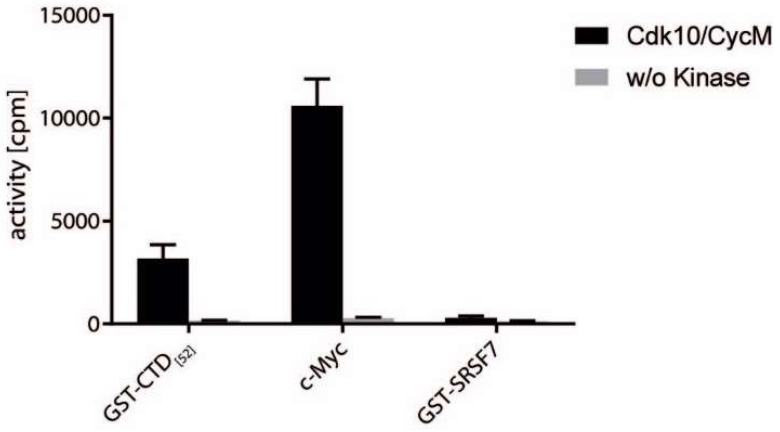


Figure 5.2.5: Cdk10 activity towards different protein substrates.
Radioactive kinase assay. 10 µM GST-CTD$_{[52]}$, 50 µM c-Myc (16-167), or 40 µM full length GST-SRSF7 were incubated with 2 mM ATP in presence or absence of 0.2 µM GST-Cdk10/GST-CycM for 30 minutes at 30°C.

5.2.6 Cdk10/CycM phosphorylation of the RNA pol II CTD

Given the central role of the RNA pol II CTD in transcription regulation, the activity and specificity of Cdk10 was further analysed using antibodies specific for CTD phosphorylation, different CTD constructs, and a set of pre-phosphorylated CTD peptides.

GST-CTD$_{[52]}$ was phosphorylated with 0.2 µM Cdk10/CycM in a time course experiment over a period of 16 h and samples were taken after 0 min, 15 min, 30 min, 60 min, 120 min, 240 min, 480 min, and 16 h. Samples were analysed by coomassie staining of an SDS-Gel and by western blot using antibodies specific for phosphorylation at position Ser2, Ser5, or Ser7 within a CTD hepta repeat. In the coomassie stained Gel GST-CTD$_{[52]}$ phosphorylation is visualised by a change in migration behaviour upon phosphorylation. Increasing phosphorylation turns the GST-CTD$_{[52]}$ into a slower migrating form (GST-pCTD$_{[52]}$). GST-CTD$_{[52]}$ phosphorylation by Cdk10/CycM appears to be rather slow, a difference in migration behaviour is visible after 120 minutes. After 16 h the GST-CTD$_{[52]}$ has completely shifted into its upper migrating form. The corresponding immunoblot analysis shows, that Cdk10 is able to phosphorylate all serine positions (Ser2, Ser5, and Ser7) of the

CTD. Phosphorylation is visible for all serine positions after 15 minutes, and increases over time. Phospho-Ser5 shows the strongest signal, followed by phospho-Ser2 and phospho-Ser7 (Figure 5.2.6, A).

Next, Cdk10 activity towards four different CTD constructs was probed. 1.) GST-$CTD_{[52]}$, containing all repeats of the human RNA pol II CTD. 2.) GST-$CTD_{[9]KKK}$, a construct containing 9 hepta repeats, with N- and C-terminal addition of lysines. 3.) GST-$CTD_{[9]}$K7 a construct containing 9 hepta repeats in which the Ser7 position of every repeat is exchanged by a lysine. Exchange of Ser7 by Lys is the most frequent modification in the distal section of the human CTD. 4.) GST-$CTD_{[16]}$A2, a construct containing 16 heptad repeats in which every serine2 position is replaced by alanine. All CTD constructs were phosphorylated by Cdk10/CycM with human, wild type GST-$CTD_{[52]}$ being the best substrate. The construct containing only 9 CTD repeats was not as well phosphorylated. Interestingly, replacement of Ser7 to Lys in the $CTD_{[9]}$K7 construct resulted in better phosphorylation. $CTD_{[16]}$A2 phosphorylation was comparable to $CTD_{[9]KKK}$ (Figure 5.2.6, B).

The cyclin-dependent kinases Cdk12 and Cdk13 require a pre-phosphorylation on Ser7 of the CTD hepta repeat for full activity, and also for P-TEFb increased phosphorylation rates have been shown if Ser7 is pre-phosphorylated (Czudnochowski et al. 2012; Bösken et al. 2014; Greifenberg et al. 2016). It was therefore analysed if Cdk10 activity requires a priming phosphorylation as well. To this end, Cdk10 activity was tested towards a set of synthetic pre-phosphorylated CTD peptides with 3 hepta repeats, which are pre-phosphorylated at specific sites (Figure 5.2.6, C). Cdk10 exhibited basically no activity towards the synthetic peptides regardless of the phosphorylation state. Only the peptide bearing the consensus CTD sequence and the Ser7 phosphorylated peptide might display signals which are above background levels. Strikingly, the peptide containing the Ser7 to Lys modification was clearly phosphorylated. This resembles the data obtained with recombinant GST-CTD variants, in which Ser7 to Lys substitution had a positive effect on phosphorylation.

Overall, Cdk10 is able to phosphorylate RNA pol II CTD at Ser2, Ser5, and Ser7 position and displays a preference for a Ser7 to lysine modification, the most frequent occurring alteration of the RNA pol II CTD consensus repeat structure in humans. Priming by any pre-phosphorylation, as found for Cdk9, Cdk12, and Cdk13 was not observed.

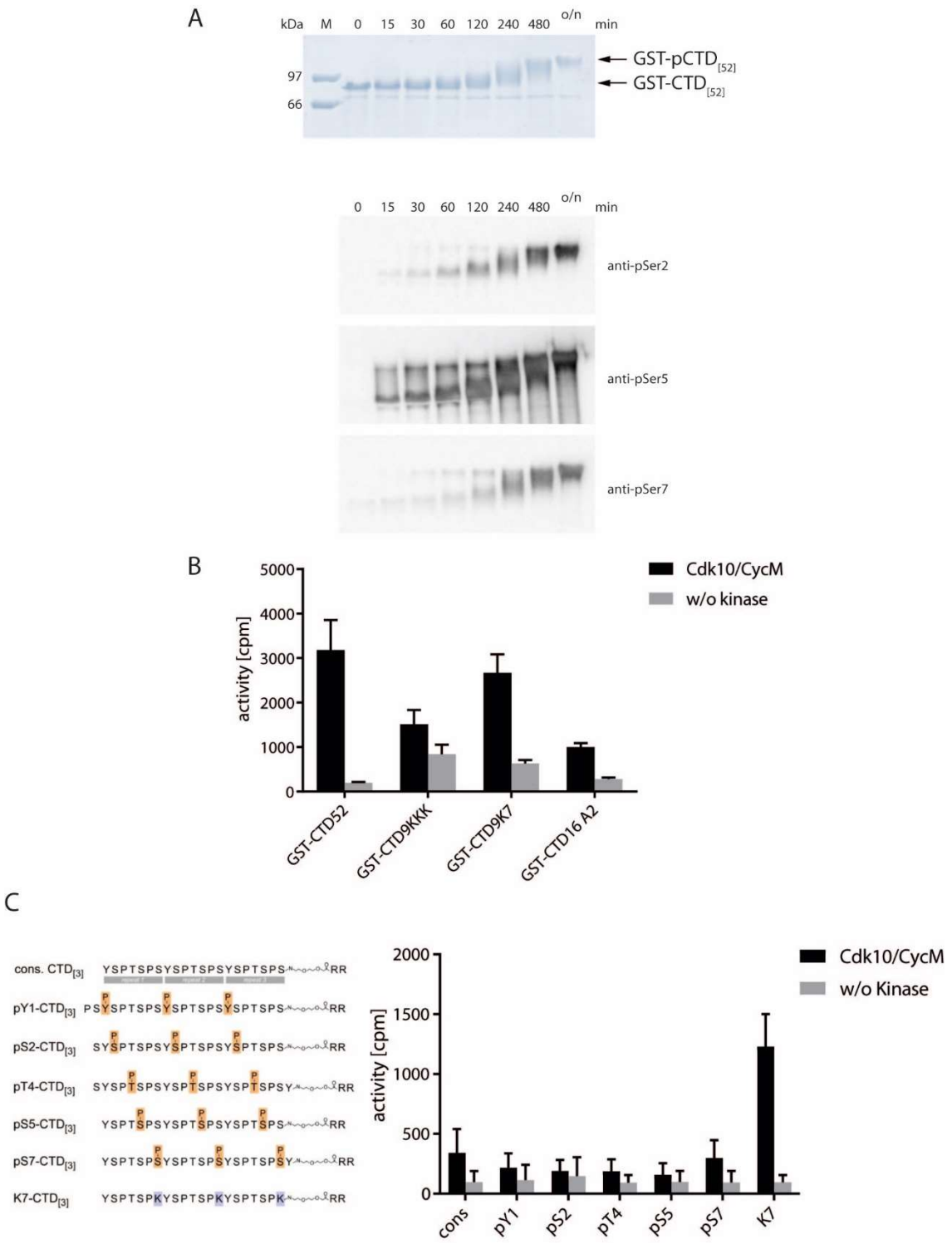


Figure 5.2.6: Substrate specificity towards RNA pol II CTD substrates.
A) Western blot analysis of GST-CTD$_{[52]}$ phosphorylation. After phosphorylation of 10 µM GST-CTD$_{[52]}$, samples were resolved on SDS-PAGE. For analysis by coomassie staining, 2 µg GST-CTD$_{[52]}$ of each sample were loaded. For analysis of the phosphorylation with phosphorylation specific antibodies, 300 ng of each sample were resolved by SDS-PAGE, blotted on nitrocellulose and probed with the respective antibody. **B**) Activity towards different GST-CTD proteins. For analysis, 10 µM GST-CTD$_{[52]}$, 50 µM GST-CTD$_{[9]KKK}$, 50 µM GST-CTD$_{[9]}$K7, or 25 µM GST-CTD$_{[16]}$A2 were incubated for 30 minutes at 30°C in presence or absence of 0.2 µM Cdk10/CycM. C) Activity towards a set of pre-phosphorylated CTD substrates. The pre-phosphorylated CTD peptides used are depicted on the left. Phosphorylation of pre-phosphorylated peptides by Cdk10/CycM was monitored after 1 h of incubation at 30°C with or without 0.2 µM of the kinase complex.

5.2.7 Cdk10/CycM phosphorylates c-Myc at five different sites

The N-terminal c-Myc (17-167) protein construct used as substrate in this thesis contains five (S/T)P motifs at position Thr58, Ser62, Ser71, Thr78, and Ser161. It has been shown, that all these sites are phosphorylated by Cdk9, Cdk12, and Cdk13. To analyse, if Cdk10 is also capable to phosphorylate these residues, 50 µM c-Myc (17-167) was incubated with 0.2 µM Cdk10/CycM overnight at room temperature and the phosphosites were determined by mass spectrometry. Cdk10 phosphorylated the Ser/ThrPro sites T58, S62, S71, and T78. Additionally, phosphorylation at S67 was identified. Phosphorylation of S161 cannot been excluded based on these data, since the corresponding peptide was not detected (Figure 5.2.7).

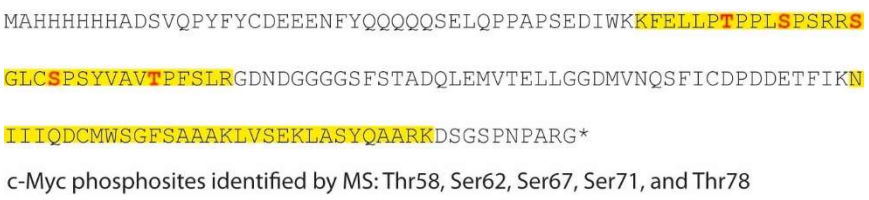


Figure 5.2.7: Identification of c-Myc (17-167) phosphorylation sites by mass spectrometry.
His-c-Myc (17-167) was phosphorylated with 0.2 µM Cdk10/CycM overnight at room temperature. The sample was subjected to SDS-PAGE analysis and the sample subsequently analysed by mass spectrometry. The phosphorylation sites are highlighted in red. The yellow marked sequences represent the amino acids covered by the analysis.

These data demonstrate that Cdk10 is principally able to phosphorylate the same (Ser/Thr)Pro motifs as Cdk9, Cdk12, and Cdk13. For Cdk9, 12, and 13 it has been demonstrated, that among the five phosphorylation sites, Ser62, and Ser71 are the sites of initial phosphorylation while the others are phosphorylated upon longer incubation times (Bösken 2013, Greifenberg 2014). For Cdk10 no such detailed analysis has been performed.

5.2.8 Pharmacologic inhibition of Cdk10

Cyclin-dependent kinases have emerged as promising targets in cancer. Among the transcription associated cyclin-dependent kinases, recent advances comprise the development of inhibitors against Cdk7, Cdk9, Cdk12, and Cdk13.

Cdk10 has been described as both a positive and a negative regulator of cancer (Iorns et al. 2008; Weiswald et al. 2017). Because of the limited published data on Cdk10, there is no evident basis that inhibition of Cdk10 represents a clinically relevant treatment option in cancer. Apart from therapeutic use, identification of specific Cdk10 inhibitors would facilitate future research. In this study, the three ATP competitive small molecular weight compounds Flavopiridol, Dinaciclib, and NVP-2 were analysed for their potential to inhibit Cdk10 (Figure 5.2.8). Flavopiridol is a widely used inhibitor of cyclin-dependent kinases. It has been initially described as a Cdk9 specific inhibitor, but cross-testing between several CDKs revealed inhibition of other CDKs as well. Dinaciclib is a Cdk inhibitor with strong inhibitory potential towards Cdk9, Cdk12, and Cdk13 (Johnson et al. 2016). NVP-2 has been described recently as a potent Cdk9 inhibitor, which also displayed pharmacologic potential towards Cdk10 (Olson et al. 2018).

The inhibitors were tested in a dilution series from 333 µM to 1 nM at an ATP concentration of 1 mM. To determine IC_{50} values of the compounds, results were normalized to DMSO control and data were fitted with GraphPadPrism using a sigmoidal curve fit. Flavopiridol inhibited Cdk10 with an IC_{50} of 1441 nM, while Dinaciclib inhibited Cdk10 at an IC_{50} of 297.3 nM. Among the three inhibitors, NVP-2 displayed the best inhibitory capacity and inhibited Cdk10 activity with an IC_{50} of 131.5 nM.

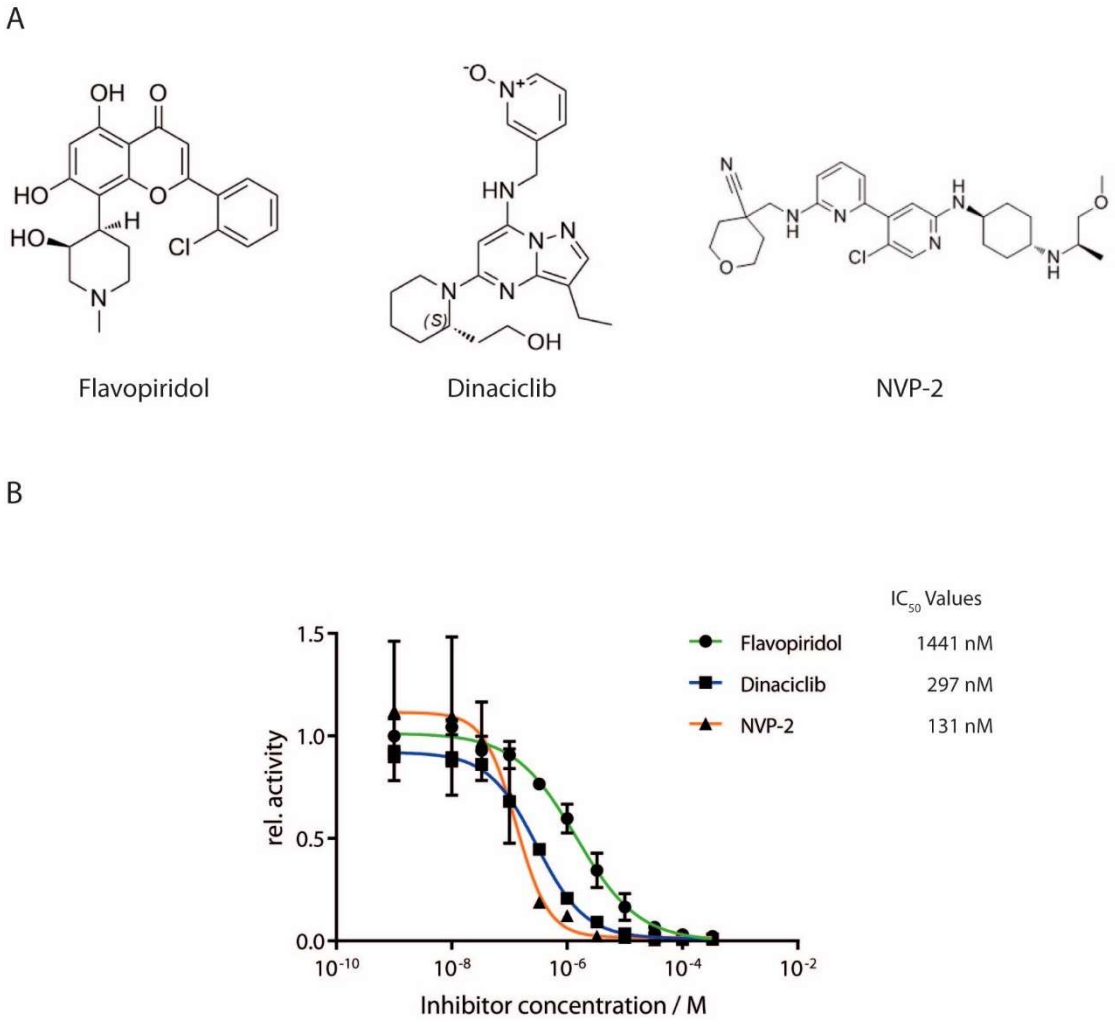


Figure 5.2.8: Pharmacologic inhibition of Cdk10/CycM.
A) Chemical structures of the small molecule inhibitors Flavopiridol, Dinaciclib and NVP-2. **B**) Dose response curve of the Cdk10/CycM activity in presence of potential inhibitors. Cdk10/CycM activity was measured in presence of different concentrations of the respective inhibitor towards 50 µM c-Myc (17-167) with 1 mM ATP for 15 minutes at 30°C. Activity was normalized against a DMSO control. Data are displayed as mean of three individual experiments. Error bars denote standard deviation.

5.2.9 Identification of new Cdk10 substrates by a chemical genetic screen

The cellular function of a kinase is mainly determined by the specificity for its substrates. To date, only two substrates of Cdk10 have been reported, the transcription factor ETS-2 and the kinase PKN2 (Guen et al. 2016; Guen et al. 2013). In this thesis, RNA pol II C-terminal domain and c-Myc were introduced as Cdk10 substrates *in vitro*. To identify cellular Cdk10 substrates we made use of an analogue-sensitive Cdk10 mutant to perform a chemical genetic screen with HeLa full-cell lysate and nuclear extracts. Identification of kinase substrates in cell lysates requires kinase specific, bio-orthogonal labelling and subsequent purification and analysis of labelled substrates. To this end, an analogue-sensitive Cdk10^{M117G}/CycM mutant and the ATP-analogue Phenyethyl-ATP-γ-S were used to label substrates in cell lysates or nuclear extracts. Thio-phosphorylated substrates were affinity purified and analysed by mass spectrometry. Some of the putative substrates were validated *in vitro* using recombinant proteins.

5.2.10 The Cdk10^{M117G} mutant retains activity

The gatekeeper residue of Cdk10 was identified by sequence alignment of different Cdks as methionine 117 (Figure 5.2.9, A). To generate a Cdk10 analogue-sensitive mutant, Met117 was mutated to glycine. Cdk10^{M117G} was co-expressed with Cyclin M and purified with good yield and homogeneity using the protocols established for Cdk10/CycM complexes (Figure 5.2.9, B). Interestingly, expression yields of Cdk10^{M117G}/CycM rather resembled those of kinase-dead Cdk10^{D163N}CycM than wild type Cdk10/CycM.

Next purified Cdk10^{M117G}/CycM activity was compared to Cdk10/CycM to assure, that kinase activity is not abrogated by M117G mutation. In an *in vitro* kinase assay Cdk10^{M117G}/CycM was able to phosphorylate both substrates, GST-CTD$_{[52]}$ and c-Myc but with reduced activity compared to Cdk10 (Figure 5.2.9, C). Using c-Myc as substrate, activity of Cdk10/CycM and Cdk10^{M117G}/CycM at different ATP concentrations was investigated. ATP was serially diluted from 2 mM to 0.125 mM. Kinase activities were measured after 15 minutes, and normalized to 2 mM ATP for analysis. Cdk10/CycM shows no reduction in kinase activity when ATP is reduced to 1 mM. Further dilution of ATP decreases kinase activity to 82% (0.5 mM ATP), 61% (0.25 mM ATP), and 43% (0.125 mM ATP). Cdk10^{M117G}/CycM activity was more susceptible to ATP reduction. Halving ATP concentration resulted in a residual activity of 75% compared to 2 mM ATP. At 0.5 mM ATP, activity was reduced to 54%. The activity further decreased with ATP concentration to 30% (0.25 mM ATP), and 17% (0.125 mM ATP), respectively (Figure 5.2.9, D).

Thus, the Cdk10 M117G mutation results in reduced activity but principally retains enzymatic function. In addition, mutation of the Cdk10 gatekeeper residue to glycine lowers the affinity of the kinase to ATP.

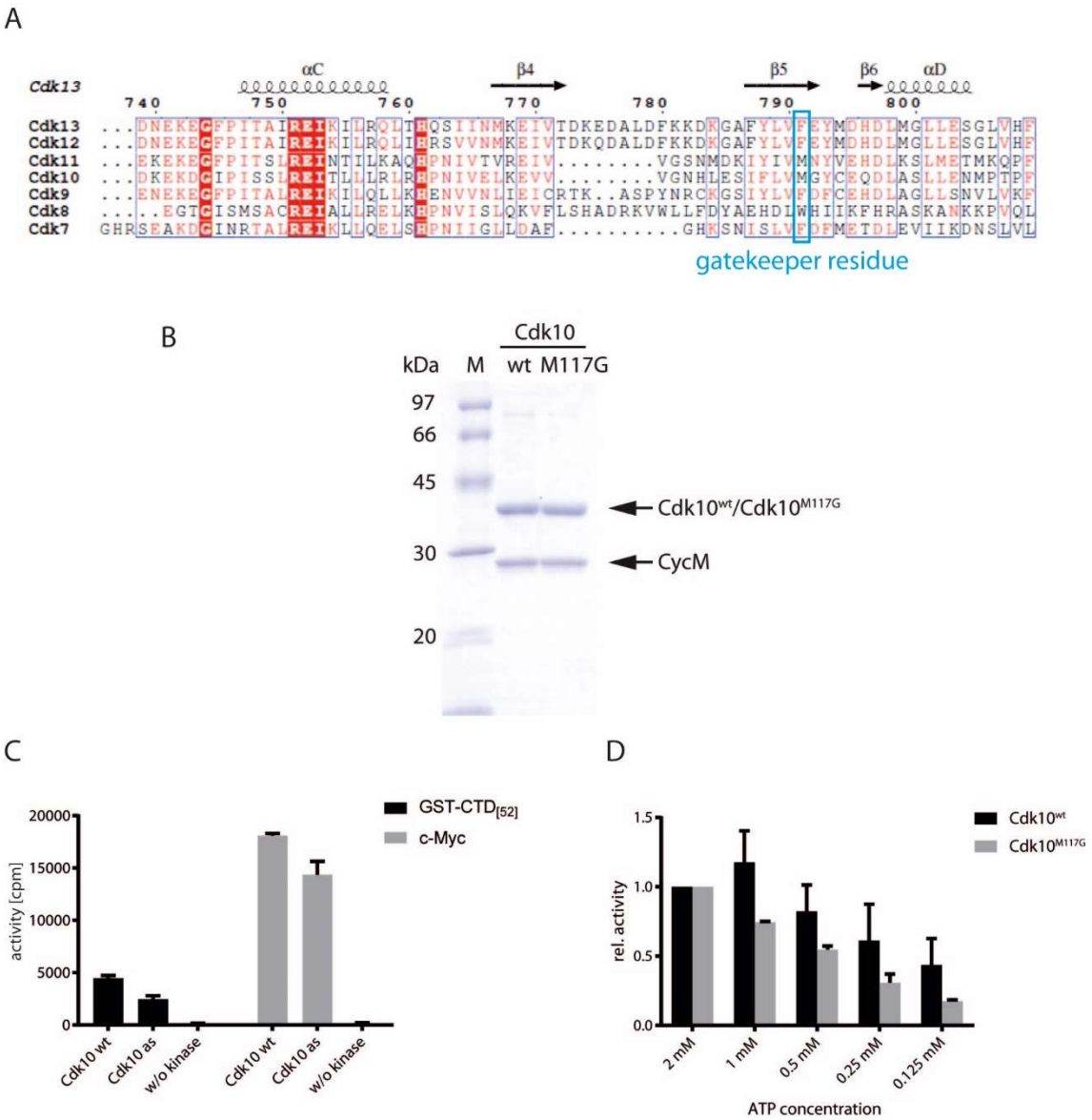


Figure 5.2.9: Expression and analysis of a Cdk10^{M117G}/CycM mutant.
A) Sequence alignment of several CDKs. The secondary structure elements denoted on top are assigned from the Cdk13/CycK crystal structure (PDB: 5EFQ). The sequences have been ordered manually and do not represent the degree of sequence similarity. The gatekeeper residues of the CDKs are highlighted by a rectangle. **B**) Comparison of purified Cdk10wt/CycM and Cdk10^{M117G}/CycM (2 µg each, Coomassie stained SDS-PAGE). **C**) Activity of Cdk10/CycM and Cdk10^{M117G}/CycM towards 10 µM GST-CTD$_{[52]}$ and 50 µM c-Myc (17-167) at 2 mM ATP. F) Cdk10^{M117G}/CycM has a reduced ATP affinity. 0.2 µM of Cdk10wt/CycM and Cdk10^{M117G}/CycM were analysed for their activity towards c-Myc at different ATP concentrations.

5.2.11 Cdk10^{M117G} mutation renders Cdk10 analogue-sensitive

To test if the Cdk10 M117G mutation renders the kinase nucleotide analogue-sensitive and to identify the ATP-analogue which is best suited for further studies, analogues with a modification at the N^6 position of the adenine-ring of ATP-γ-S were tested for their ability to compete with ATP in a radioactive kinase assay. In total, seven N^6-modified analogues were tested which comprise 1-Methyl-Butyl, Phenyl-, Benzyl-, Phenylethyl-, cyclo-Pentyl-, cyclo-Hexyl-, and Furfuryl-ATP-γ-S (Figure 5.2.10, A).

For analysis, [^{32}P]-ATP was mixed with either H_2O, cold ATP, ATP-γ-S or the respective analogue in equimolar concentration of 500 µM each, prior to the kinase assay. The ability to compete out ATP from the ATP binding pocket is then monitored by a reduction of incorporated [^{32}P] into the substrate relative to the H_2O control. In Cdk10/CycM, addition of cold ATP reduced ^{32}P incorporation to 69% and ATP-γ-S to 60% of control levels, respectively. The addition of N^6-modified ATP-γ-S analogues had no considerable effect on Cdk10/CycM. Cdk10^{M117G}/CycM was less affected by ATP and ATP-γ-S with 87% and 89% respectively. In contrast to Cdk10, [^{32}P] incorporation by Cdk10^{M117G}/CycM was reduced by N^6-modified ATP-γ-S analogues. An effect was seen for all analogues tested, but was most prominent for Benzyl- (18%), Phenylethyl- (11%), and cyclo-Hexyl ATP-γ-S (21%) (Figure 5.2.10, B).

To reveal differences in ATP competition between the three most potent analogues, Benz-, PhEt- and cHe-ATP-γ-S were analysed in more detail in a serial dilution. The analogues were serially diluted to 500 µM, 250 µM, 125 µM, 62.5 µM, 31.25 µM 15.625 µM, and 7.8125 µM and then assayed for their ability to compete with 500 µM ATP. ATP competition at equimolar concentrations of analogue and ATP (500 µM) resembles the results obtained before with a residual ^{32}P incorporation of 16% (cHe), 11% (PhEt), and 21% (Benz). PhEt showed the best ATP competition at all concentrations tested. Notably, low concentrations of ATP-analogue stimulated Cdk10^{M117G}/CycM activity above control levels. At an ATP-analogue concentration of 7.8125 µM radioactive ^{32}P incorporation was 1.6-fold of H_2O control for cyclo-Hexyl-, 1.3-fold for Phenylethyl-, and 1.8-fold for Benzyl-ATP-γ-S (Figure 5.2.10, C).

For data interpretation it is important to note, that the data are not corrected for the respective background in absence of kinase. The residual activity seen with PhEt-ATP-γ-S resembles background levels (data not shown). Equimolar concentration of PhEt-ATP-γ-S hence fully competes out non-modified ATP in the Cdk10 M117G mutant without having an effect on Cdk10 wild type. Thus, Cdk10^{M117G} can be considered as an analogue-sensitive Cdk10 mutant (Cdk10^{as}).

To ascertain the utility of PhEt-ATP-γ-S to specifically label Cdk10 substrates *in vitro*, c-Myc was incubated with Cdk10^{wt}/CycM or Cdk10^{as}/CycM and ATP-γ-S or PhEt-ATP-γ-S. Thio-phosphorylation was detected after alkylation with PNBM by western blot using a specific antibody. C-Myc was phosphorylated by Cdk10^{wt} in presence of 500 µM ATP-γ-S but not with 500 µM PhEt-ATP-γ-S. Cdk10^{as}, in contrast, was able to use both ATP-γ-S and modified PhEt-ATP-γ-S for phosphorylation (Figure 5.2.10, E). Activity of Cdk10^{wt} was increased compared to Cdk10^{as}, which is in line with the data from radioactive measurements in Figure 5.2.9, D.

Next, it was assessed if PhEt-ATP-γ-S can be used to specifically thio-phosphorylate proteins in cell lysate. HeLa full-cell lysate was incubated with PhEt-ATP-γ-S or ATP-γ-S without addition of recombinant Cdk10 and resulting thio-phosphorylation detected by western blot (Figure 5.2.10, E).

Addition of PhEt-γ-S resulted in basically no thio-phosphorylation of protein in cell lysate. Only two faint bands at 35 kDa and 40 kDa are visible. In contrast, addition of ATP-γ-S results in strong signals dispersed throughout the lane. This finding confirms that PhEt-ATP-γ-S is not used by endogenous kinases present in the lysate. When lysate was incubated with $Cdk10^{wt}$ in presence of PhEt-γ-S faint bands are visible sparsed throughout the lane, which indicates that also $Cdk10^{wt}$ used the ATP analogue to a small extend. Incubation of HeLa lysate with $Cdk10^{as}$ in presence of analogue leads to a well visible signal. When $Cdk10^{wt}$ and $Cdk10^{as}$ samples were not alkylated with PNBM, no signal was detected, confirming specificity of the antibody. Addition of $Cdk10^{wt}$ led to a mild increase in thio-phosphorylation compared to the endogenous sample, likely due to the high concentration of the kinase. Nonetheless, $Cdk10^{as}$ was considerably more potent in thio-phosphorylating proteins in cell lysate, which reflects the data seen for phosphorylation of recombinant c-Myc. Hence, it can be concluded that proteins in cell lysates are specifically thio-phosphorylated by $Cdk10^{as}$ with a good signal to noise ratio.

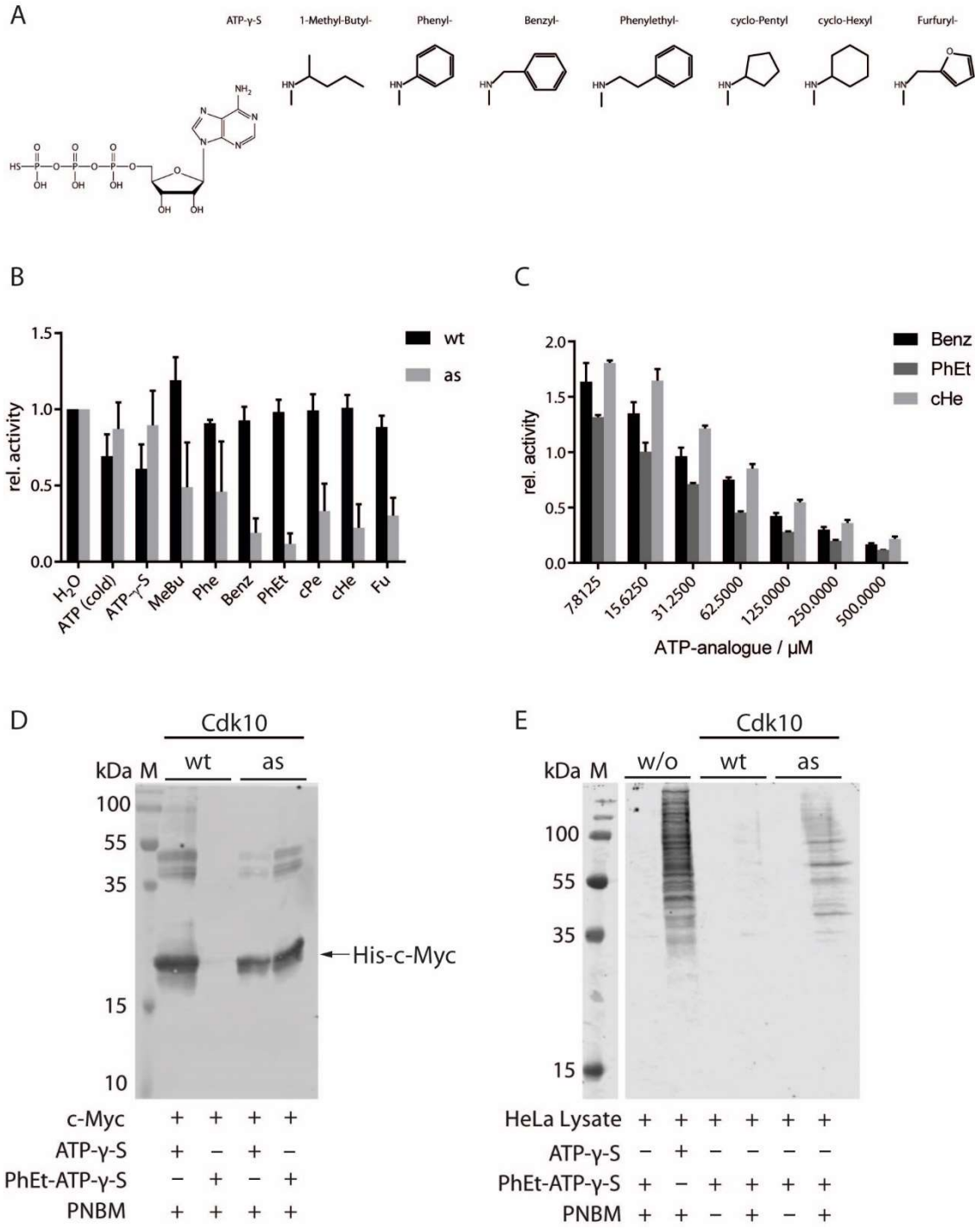


Figure 5.2.10: Cdk10^{M117G} mutation confers analogue-sensitivity to Cdk10.

A) ATP analogues used in this study. **B**) ATP competition assay. 0.2 µM Cdk10wt/CycM or Cdk10as/CycM were analysed for their response to ATP analogues. 500 µM ATP containing [^{32}P]-ATP, was either mixed with H_2O or with equimolar concentrations of ATP, ATP-γ-S, or N^6-modified ATP-γ-S analogues prior to kinase assay. The kinase assay was started by addition of the ATP/analogue mixture to the reaction mixture. Data were normalized to the ATP/H_2O control. **C**) Benyl-ATP-γ-S, Phenylethyl-γ-S, and cyclo-Hexyl-ATP-γ-S were added in different concentrations to 500 µM of 32P-ATP containing ATP and analysed for their potential to compete with ATP. **D**) Recombinant c-Myc (17-167) was phoshphorylated by 0.2 µM Cdk10wt/CycM or Cdk10as/CycM in presence of 0.5 mM ATP-γ-S or 0.5 mM Phenylethyl ATP-γ-S. Reactions were stopped with 50mM EDTA. The samples were alkylated with 2.5 mM PNBM and then analysed with an alkylation specific antibody by western blot. **E**) HeLa nuclear extracts were incubated with PhEt-ATP-γ-S or ATP-γ-S and either without additional kinase, with Cdk10wt/CycM or Cdk10as/CycM. After kinase reaction, samples were alkylated and analysed by western blot.

5.2.12 Identification of Cdk10 substrates by mass spectrometry

Cdk10 is mainly localised in the nucleus. Therefore, we used HEK293 cell nuclear extracts in addition to HeLa full cell lysate to reduce sample complexity for the mass spectrometry (MS) based screen. Nuclear extracts were provided by the group of Prof. Dr. Henning Urlaub, MPI for Biophysical Chemistry, Göttingen.

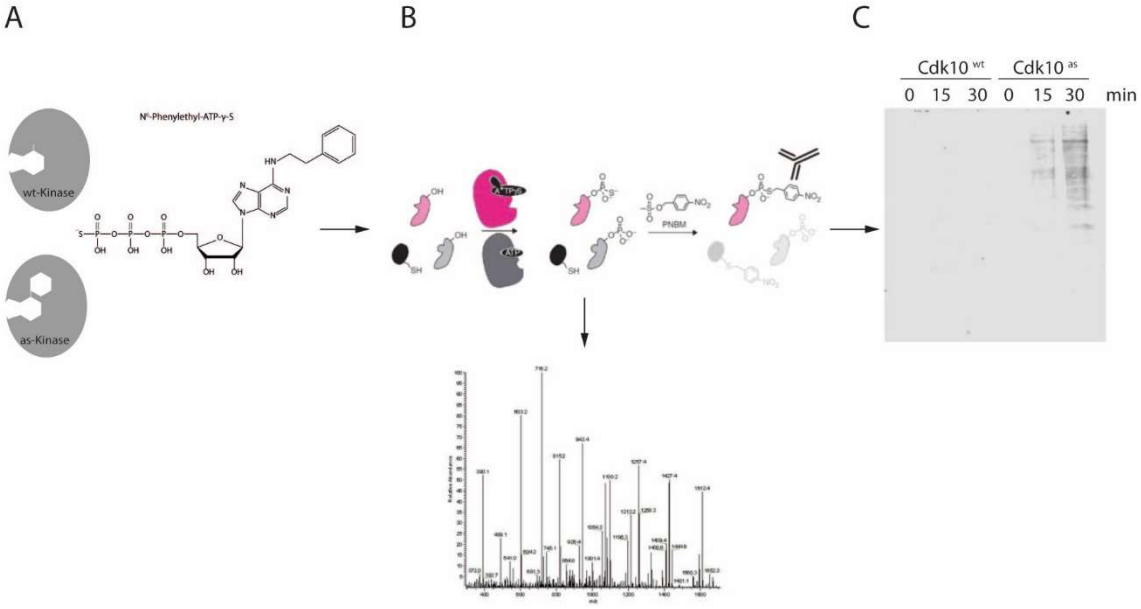


Figure 5.2.11: Sample preparation for substrate identification by mass spectrometry.
A) Cdk10/CycM or the analogue sensitive Cdk10^{M117G}/CycM were incubated with nuclear extracts in presence of N^6-Phenylethyl-ATP-γ-S. **B**) Samples were snap frozen and further processed for substrate identification by mass spectrometry. A small aliquot of the sample was alkylated with PNBM and analysed by western blot to monitor the thio-phosphorylation. **C**) Western blot of the samples subjected to mass spectrometry.

Nuclear extracts were thio-phosphorylated and samples were taken after 15 minutes and 30 minutes (Figure 5.2.11). The rationale behind short incubation times was to prevent non-specific signals by prolonged incubation times. However, the 15 minutes samples did not provide enough signal for valid data generation by mass spectrometry. Therefore, only data for 30 minutes are analysed in this study. For the generation of a second dataset, both time and protein amount of the samples was increased to 2 mg/sample, incubated for 60 minutes, to more robustly identify potential substrates. In total, one dataset was collected from HeLa full cell lysate, two from nuclear extracts after 30 minutes of Cdk10 incubation and another one which was measured in duplicate from nuclear extracts incubated for 60 minutes with Cdk10. In addition to purification via iodoacetyl beads, replicates of this sample were first enriched for phosphopetides by TiO_2 and subsequently thio-phosphorylated peptides were extracted.

For analysis, all datasets were combined. The individual datasets can be found in the appendix (Table S1). The analysis shown in the main body of the thesis includes all peptides which were at least 2-fold enriched over control, or identified in the analogue-sensitive Cdk10 sample but not in the control. By this, 92 peptides were identified in total. NPM1 Ser70, SSB Ser336, and BABAM1 Ser29 were found in two datasets, hence the data comprise 89 different putative Cdk10 phosphorylation sites, originating from 66 different proteins (Figure 5.2.12).

intensity	gene name	identifier	name	p-site ± 5 amino acids	STY	pos	loc. prob.
NaN	ACAD11	A0A0J9YXS1	Acyl-CoA dehydrogenase family member 11	QRAGKSNPTFY	S	217	0,71
NaN	ACAD11	A0A0J9YXS1	Acyl-CoA dehydrogenase family member 11	GKSNPTFYLQK	T	220	0,71
NaN	ACIN1	S4R3H4	Apoptotic chromatin condensation inducer in the nucleus	QARRLSQPESA	S	652	0,96
1,27	AHNAK	Q09666	Neuroblast differentiation-associated protein AHNAK	YEVTGSDDETG	S	5841	0,99
1,04	AHNAK	Q09666	Neuroblast differentiation-associated protein AHNAK	EAPLPSPKLEG	S	5110	1,00
NaN	ANKRD17	H0YM23	Ankyrin repeat domain-containing protein 17	AAAALTRMRAE	T	105	1,00
NaN	ANKRD32	Q9BQI6	Ankyrin repeat domain-containing protein 32	LDYVVSPQIKE	S	919	0,56
NaN	APC	P25054-2	Adenomatous polyposis coli protein	DNEKHSPRNMG	S	1963	1,00
NaN	ARGLU1	Q9NWB6	Arginine and glutamate-rich protein 1	RERASSPPDRI	S	77	0,96
1,41	ARL14	Q8N4G2	ADP-ribosylation factor-like protein 14	___MGSLGSKN	S	3	0,86
1,13	ATP8A2	F8VRS1	Phospholipid-transporting ATPase	VQELETKSRVL	T	915	0,94
NaN	BABAM1	M0QY17	BRISC and BRCA1-A complex member 1	RPRTRSNPEGA	S	29	0,99
NaN	BABAM1	M0QXG9	BRISC and BRCA1-A complex member 1	RPRTRSNPEGA	S	29	0,99
NaN	CCDC96	Q2M329	Coiled-coil domain-containing protein 96	GRDILTKTKQA	T	479	1,00
NaN	CCDC96	Q2M329	Coiled-coil domain-containing protein 96	DILTKTKQARE	T	481	1,00
NaN	CDK10	Q15131	Cyclin-dependent kinase 10	PVKPMTPKVVT	T	196	1,00
NaN	CENPH	Q9H3R5	Centromere protein H	DRMRLSTALKK	S	109	0,50
NaN	CENPH	Q9H3R5	Centromere protein H	RMRLSTALKKN	T	110	0,50
NaN	CEP63	H0YAE6	Centrosomal protein of 63 kDa	FKPTHSRTTEF	S	236	0,50
NaN	CEP63	H0YAE6	Centrosomal protein of 63 kDa	RIFKPTHSRTT	T	234	0,50
NaN	CKM	P06732	Creatine kinase M-type	ALNSLTGEFKG	T	166	1,00
1,38	DNAH11	U3KQJ8	Dynein heavy chain 11, axonemal	IFAKATPVDRQ	T	4465	1,00
NaN	DUT	P33316-2	Deoxyuridine 5'-triphosphate nucleotidohydrolase	ETPAISPSKRA	S	11	0,75
1,81	EEF1D	E9PK01	Elongation factor 1-delta	NVLEKSSPGHR	S	118	0,61
NaN	EIF3D	O15371-3	Eukaryotic translation initiation factor 3 subunit D	ANGEVSFINIK	S	374	1,00
NaN	EIF3D	O15371-3	Eukaryotic translation initiation factor 3 subunit D	HDGVMTGANGE	T	367	1,00
NaN	EIF4B	E7EX17	Eukaryotic translation initiation factor 4B	HPSWRSEETQE	S	409	0,53
1,03	EIF4B	E7EX17	Eukaryotic translation initiation factor 4B	RERHPSWRSEE	S	406	1,00
1,13	EIF4G1	E7EX73	Eukaryotic translation initiation factor 4 gamma 1	EAEEESDHN__	S	1433	1,00
2,45	GGCT	M0QZK8	Gamma-glutamylcyclotransferase	__MNKSNLNSL	S	4	1,00
NaN	GLIPR1	P48060	Glioma pathogenesis-related protein 1	VKRYYSVVYPG	S	219	0,55
NaN	GPRC5A	Q8NFJ5	Retinoic acid-induced protein 3	ENRAYSQEEIT	S	301	0,99
NaN	HDGF	P51858	Hepatoma-derived growth factor	DLLEDSPKRPK	S	165	1,00
2,28	HIDE1	A8MVS5	Protein HIDE1	VTFNLSGGSSK	S	75	0,91
1,02	HSF1	E9PMQ6	Heat shock factor protein 1	EGRPPSPPPTS	S	363	1,00
1,42	HSPB1	P04792	Heat shock protein beta-1	LSRQLSSGVSE	S	82	0,84
NaN	KCNG1	Q9UIX4	Potassium voltage-gated channel subfamily G member 1	ALRILYVMRLA	Y	352	1,00
NaN	KHSRP	A0A087WTP3	Far upstream element-binding protein 2	LIQDGSQNTNV	S	274	0,97
NaN	LAMA4	A0A0A0MTC7	Laminin subunit alpha-4	NAKKEYMGLAI	Y	887	1,00
NaN	LIMK2	B5MC51	LIM domain kinase 2	LEINGTPVRTL	T	132	0,98
NaN	LMNA	P02545-2	Prelamin-A/C;Lamin-A/C	AQASSTPLSPT	T	19	0,52
NaN	LMNA	P02545	Prelamin-A/C;Lamin-A/C	GSGGGSFGDNL	S	636	1,00
NaN	LRRTM4	Q4KMX1	Leucine-rich repeat transmembrane neuronal protein 4	YNRLRSLSRNA	S	196	1,00
NaN	LRRTM4	Q4KMX1	Leucine-rich repeat transmembrane neuronal protein 4	RLRSLSRNAFA	S	198	1,00
NaN	MAP2K2	M0R1B6	Dual specificity mitogen-activated protein kinase kinase 2	LNQPGTPTRTA	T	98	0,83
NaN	MAP3K5	Q99683	Mitogen-activated protein kinase kinase kinase 5	FEMVNTITEEK	T	652	1,00
NaN	MAP3K5	Q99683	Mitogen-activated protein kinase kinase kinase 5	MVNTITEEKGR	T	654	1,00
NaN	MATR3	A0A0R4J2E8	Matrin-3	HFRRDSFDDRG	S	188	0,99
NaN	MEPCE	Q7L2J0	7SK snRNA methylphosphate capping enzyme	AVGRESPGAAA	S	69	0,98
NaN	NCOA7	Q8NI08-7	Nuclear receptor coactivator 7	IEYYLTKNKEG	T	413	0,81
NaN	NEURL3	A0A087WVI2	E3 ubiquitin-protein ligase NEURL3	REHLPTLHTRR	T	65	1,00
NaN	NEURL3	A0A087WVI2	E3 ubiquitin-protein ligase NEURL3	LPTLHTRRGVC	T	68	1,00
NaN	NOLC1	A0A0A0MRM9	Nucleolar and coiled-body phosphoprotein 1	SIKFDSE_____	S	707	1,00
NaN	NPM1	P06748-2	Nucleophosmin	MNYEGSPIKVT	S	70	0,96
NaN	NPM1	P06748	Nucleophosmin	MNYEGSPIKVT	S	70	0,99
1,34	OPTN	Q96CV9-3	Optineurin	RHGARTSDSDQ	T	468	0,56
NaN	OSBPL8	Q9BZF1-3	Oxysterol-binding protein-related protein 8	LLSTITDPSVI	T	102	0,51
2,44	PALLD	Q8WX93-3	Palladin	TARIASDEEIQ	S	511	1,00
NaN	PHF6	Q8IWS0-5	PHD finger protein 6	SYRDRSPHRSS	S	165	0,91
NaN	PHF6	Q8IWS0-5	PHD finger protein 6	HELEPSSPKSK	S	120	0,60
1,11	PI4K2A	Q9BTU6	Phosphatidylinositol 4-kinase type 2-alpha	AGSGPSPPGSP	S	47	0,94
1,11	PI4K2A	Q9BTU6	Phosphatidylinositol 4-kinase type 2-alpha	PSPPGSPGHDR	S	51	1,00
NaN	PSME1	Q06323	Proteasome activator complex subunit 1	AYAVLYDIILK	Y	227	0,80
3,28	RPL14	P50914	60S ribosomal protein L14	ALLKASPKKAP	S	139	1,00
NaN	RPL5	P46777	60S ribosomal protein L5	MIVRVTNRDII	T	56	1,00
NaN	RRP12	Q5JTH9-2	RRP12-like protein	LRTLITKGCQA	T	570	1,00
NaN	SERBP1	Q8NC51-4	Plasminogen activator inhibitor 1 RNA-binding protein	LHKSKSEEAHA	S	309	0,95
NaN	SRRM2	Q9UQ35	Serine/arginine repetitive matrix protein 2	SRPSMSPTPLD	S	2132	1,00
NaN	SRRM2	Q9UQ35	Serine/arginine repetitive matrix protein 2	VSGRTSPPLLD	S	2398	0,95
NaN	SRRM2	Q9UQ35	Serine/arginine repetitive matrix protein 2	AGRSPSPASGR	S	297	0,77
NaN	SRRM2	Q9UQ35	Serine/arginine repetitive matrix protein 2	AKTHTTALAGR	T	289	0,59
NaN	SRRM2	Q9UQ35	Serine/arginine repetitive matrix protein 2	SERAPSPSSRM	S	2426	0,97
NaN	SRRM2	Q9UQ35	Serine/arginine repetitive matrix protein 2	LKRVPSPTPAP	S	2581	0,93
NaN	SRRT	Q9BXP5-5	Serrate RNA effector molecule homolog	ASEPGTPPLPT	T	507	0,80
NaN	SSB	P05455	Lupus La protein	KTKFASDDEHD	S	366	1,00
1,44	SSB	P05455	Lupus La protein	KTKFASDDEHD	S	366	1,00
NaN	TARS	P26639	Threonine--tRNA ligase, cytoplasmic	EEKASSPSGKM	S	8	0,96
NaN	TCOF1	J3KQ96	Treacle protein	GEASVSPEKTS	S	1340	1,00
NaN	TCOF1	J3KQ96	Treacle protein	APAKESPRKGA	S	381	1,00
NaN	TCOF1	J3KQ96	Treacle protein	LLSGKSPRKSA	S	156	0,96
NaN	TCRBV16S1A1N1	A0A5B0	T cell receptor beta variable 14	VSRLLSLVSLC	S	7	0,50
NaN	TCRBV16S1A1N1	A0A5B0	T cell receptor beta variable 14	LLSLVSLCLLG	S	10	0,50
NaN	THRAP3	Q9Y2W1	Thyroid hormone receptor-associated protein 3	KPFRGSQSPKR	S	406	0,70
NaN	TRA2B;TRA2A	H7BXF3	Transformer-2 protein homolog beta	SYRRRSPSPYY	S	103	0,99
NaN	TRA2B;TRA2A	H7BXF3	Transformer-2 protein homolog beta	RRRSPSPYYSR	S	105	0,86
NaN	TRIM28	Q13263-2	Transcription intermediary factor 1-beta	VKRSRSGEGEV	S	391	0,99
NaN	U2AF2	P26368-2	Splicing factor U2AF 65 kDa subunit	GGLIRSPRHEK	S	79	1,00
3,50	UBAP2L	Q14157-1	Ubiquitin-associated protein 2-like	SAPQMSPGSSD	S	467	0,64
NaN	YBX1	P67809	Nuclease-sensitive element-binding protein 1	EKNEGSESAPE	S	174	0,75
1,03	YBX1	P67809	Nuclease-sensitive element-binding protein 1	QNYQNSESGEK	S	165	0,81
NaN	ZC3HC1	C9J0I9	Nuclear-interacting partner of ALK	PIVSRTRSWDS	T	309	0,59
NaN	ZNF638	*Q14966*	Zinc finger protein 638	__MLQTIYLFS	T	4	0,55

Figure 5.2.12: Identified peptides by mass spectrometry.
Summary of all peptides which were enriched at least 2-fold over wild type Cdk10/CycM control. Proteins are sorted alphabetically by gene name. Intensity: log2-fold intensity, NaN: Not a Number means that the peptide was not detected in the control sample; identifier: uniprot accession code of the protein; p-site: phosphorylation site (highlighted in red); STY: amino acid which carried the phosphorylation; pos: position within the protein; loc. Prob: localisation probability of the phosphosite within the peptide.

Pre-purification of phosphopeptides by TiO_2 and subsequent capture of thio-phosphates results in detection of only two phosphorylation sites, Npm1 Ser70, and NOLC Ser707 (data not shown).

Most of the phosphosites identified in the screen have been described in phosphoproteomic studies to occur *in vivo*.For analysis of the Cdk10/CycM biological function, the entire substrate list was analysed for GO-term enrichment compared to the entire human proteome. Putative Cdk10 substrates were enriched for RNA regulation, transcription, and translation related processes (Figure 5.2.13, A). However, when the data were compared to a reference of 2909 nuclear proteins, no statistically significant enriched GO-terms could be assigned (data not shown).

To determine a Cdk10 consensus sequence, five amino acids N- and C-terminal of the site of phosphorylation were aligned. The identified phosphosites had a slight preference for an (S/T)P motif, but apart from a preference for P in +1 position no apparent consensus sequence can be assigned (Figure 5.2.13, B). When the peptides were manually selected for (S/T)P motifs, K and R in the +3 position become more prominent such that 35% of (S/T)P peptides resemble the consensus SPx(K/R) motif of CDKs and MAP kinases. Moreover, 25% of (S/T)P peptides contained a positive charge in +2 position. In addition, there is a noteworthy preference for Pro in -1 and -2 position, as well as a for Arg at -3 and Glu at -4 position, respectively (Figure 5.2.13, C). Analysing only sequences in which the phosphosite localisation probability within the peptide was >75% did not apparently change the consensus sequence obtained including all peptides (data not shown).

Finally, substrate specificity among different CDKs, the (S/T)P containing substrates from this thesis were compared to published datasets which analysed Cdk9 (Sansó et al. 2016) and Cdk1 (Blethrow et al. 2008) substrates by the same technique. Among the datasets, the Cdk10 dataset contained 24 different putative (S/T)P substrates, the screen for Cdk9 substrates identified 109 different proteins, whereas 65 different proteins were identified for Cdk1. In total, nine proteins were previously found as substrates for Cdk9 and/or Cdk1, leaving a remainder of 15 proteins which can be uniquely assigned to Cdk10. Cdk10 shares four substrates with Cdk1 (AHNAK, DUT, U2AF2, Ubap2L), and one (MePCE) with Cdk9. The remaining 4 proteins (LaminA/C, NPM1, SRRM2, TCOF) were identified in all three datasets. Aside from these targets, Cdk1 and Cdk9 have four substrates (Calnexin, Palladin, Stathmin, SPT5) in common (Figure 5.2.13, D).

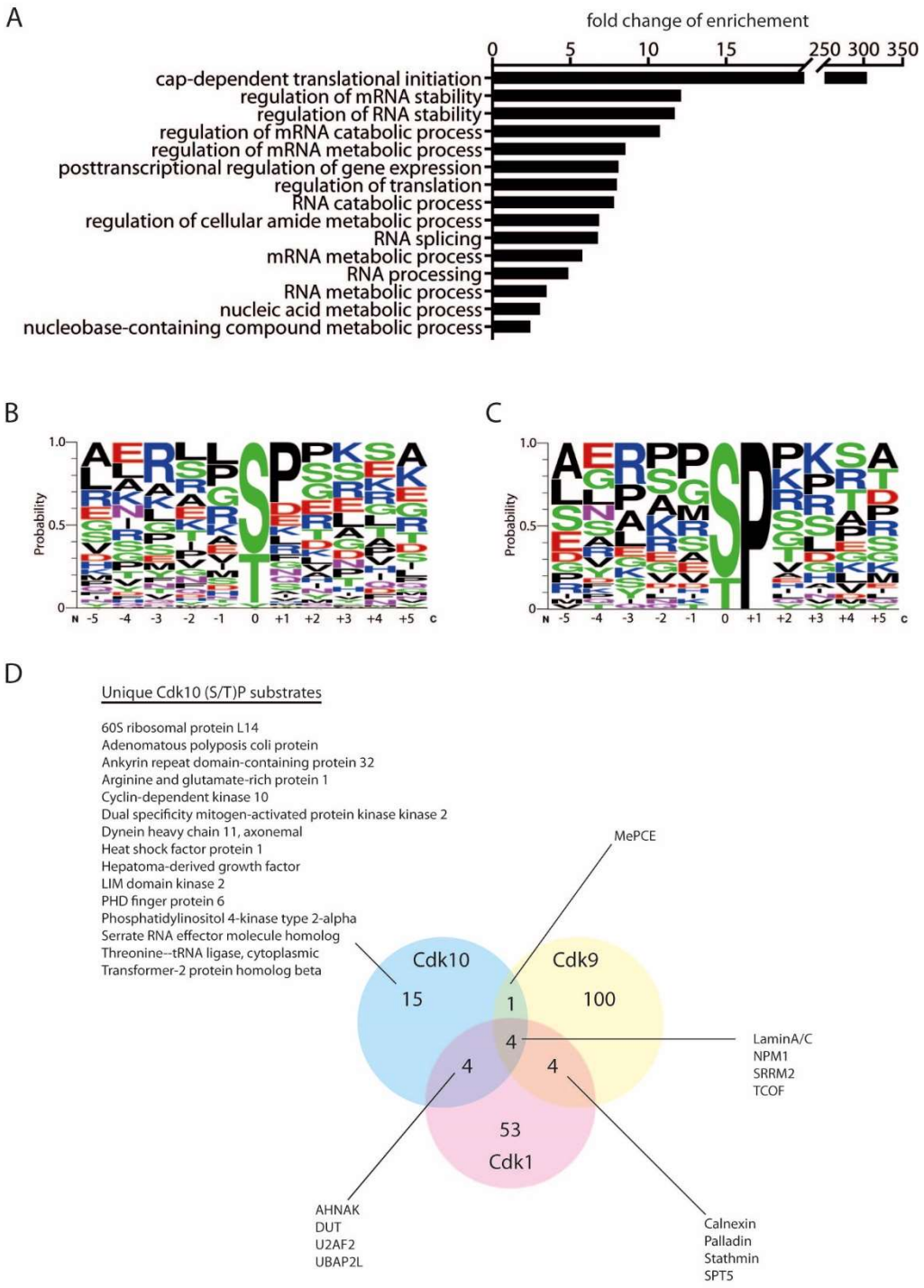


Figure 5.2.13: Analysis of the identified Cdk10 substrates.
A) GO-term enrichment. Proteins were analysed for enriched GO-terms compared to the human proteome using GOrilla. **B**) Analysis of the Cdk10 consensus sequence based on all peptides identified in the screen. **C**) Analysis of the consensus sequence of peptides restricted to (S/T)P motif related phosphosites. **D**) Venn-Diagramm of common and unique Cdk10 (S/T)P protein substrates. (S/T)P substrates identified in this thesis were compared to (S/T)P substrates which were identified by a similar screening for Cdk1/CycB1 (Blethrow et al. 2008) and Cdk9/CycT1 (Sansó et al. 2016). Datasets were only compared for same protein substrates. The substrates might differ in the identified phospho-peptide.

5.2.13 Validation of putative Cdk10/CycM substrates *in vitro*

Six putative substrates of Cdk10/CycM which were found in the chemical genetic screen were validated *in vitro* using recombinant proteins. The proteins selected for validation *in vitro* comprise Ubap2L, Npm1, Hsf1, Limk2, Srrt, and Trim28. For *in vitro* validation, the protein domains containing the putative phosphorylation site were expressed in *E. coli*. The purified proteins used for *in vitro* validation as Cdk10 substrates are shown in Figure 5.2.14, A and comprise Ubap2L (355-652), Npm1 (1-265, isoform 2), MBP-Hsf1 (217-384), GST-Limk2 (149-242), GST-Srrt (495-602), and GST-Trim28 (1-853, full lenght). Phosphorylation of putative substrates was analysed in a radioactive kinase assay by incubation with 0.2 µM Cdk10/CycM and 1 mM ATP for 1 h at 30°C. As control, proteins were incubated in absence of kinase.

All proteins were found to be phosphorylated by Cdk10. Ubap2L, Npm1, and Hsf1 showed a phosphorylation with radioactive counts of nearly 5000 cpm (Hsf1) or higher (Ubap2L, Npm1). Limk2 and Srrt were phosphorylated only weakly with 1135 cpm (Limk2) and 1479 cpm (Srrt), respectively. Trim28, with 2768 cpm, exhibited an intermediate signal (Figure 5.2.14, B). For Ubap2L, Npm1, and Hsf1 it was further analysed, if *in vitro* phosphorylation is specific to Cdk10 or if they represent substrates of other CDKs as well. Therefore, the cyclin-dependent kinases Cdk7, Cdk9, Cdk10, Cdk12, and Cdk13 were tested for their ability to phosphorylate Ubap2L, Npm1, or Hsf1. In addition to Cdk10, all three substrates got phosphorylated by Cdk9, but not by any other Cdk tested. In any case phosphorylation by Cdk10 was better compared to Cdk9. For Ubap2L, phosphorylation by Cdk10 was 2.7-fold stronger than by Cdk9. Npm1 was a 6.6-fold better substrate for Cdk10 than for Cdk9, whereas the difference for Hsf1 was only 1.5-fold. Taken together, despite phosphorylation by Cdk9, the phosphorylation of the substrates Ubap2L and Npm1 appears to be specific for Cdk10. Hsf1 instead serves as an *in vitro* substrate to both Cdk9 and Cdk10 at a comparable level (Figure 5.2.14, C-E).

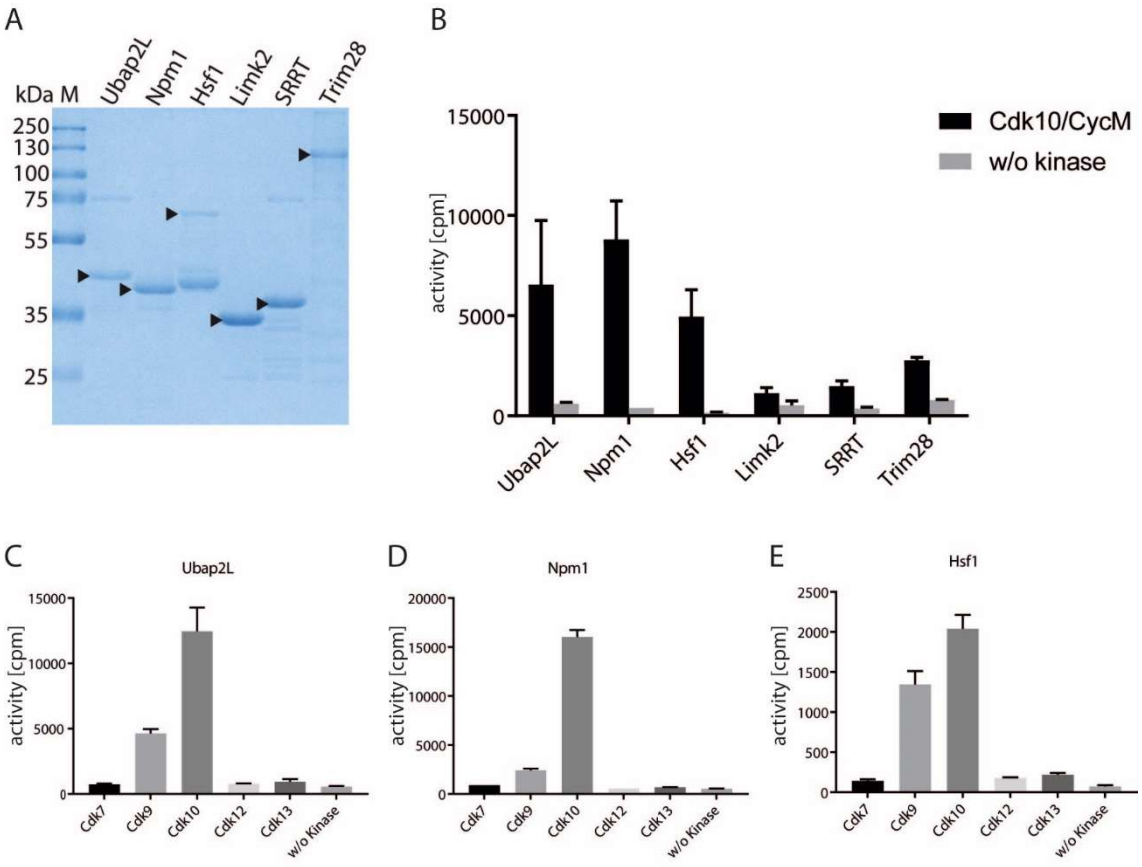


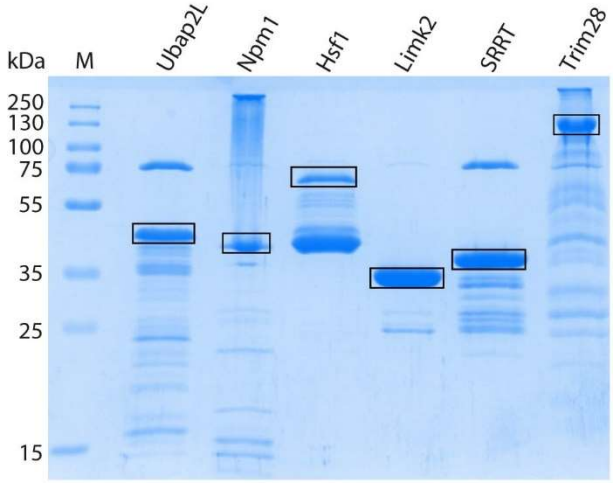
Figure 5.2.14: Cdk10 substrate validation *in vitro*.
A) Coomassie staining of 2 µg of the putative Cdk10 substrates (arrowhead) Ubap2L (355-652), His-Npm1 (1-256, isoform 2), MBP-Hsf1 (217-384), GST-Limk2 (149-242), GST-SRRT (495-602), GST-Trim28 (1-835). **B**) Radioactive kinase assay with 0.2 µM Cdk10/CycM for 60 min at 30°C. Substrate concentration was 40 µM each. **C-D**) Ubap2L, Npm1, and Hsf1 were incubated with 0.2 µM of the respective Cdk using the same assay condition as in B.

5.2.14 Validation of the phosphorylation sites by mass spectrometry

To analyse, if phosphorylation of the recombinant proteins mirrors the phosphosites identified in the chemical genetic screen, proteins were subjected to mass spectrometry after phosphorylation by Cdk10 *in vitro*. Proteins were phosphorylated for 1 h at 30°C with Cdk10 and the reaction stopped by addition of 2xSDS sample buffer. 6 µg of each protein was separated by SDS-PAGE, stained with coomassie blue and the bands excised for proteolytic digest (Figure 5.2.15). Ubap2L was phosphorylated at eleven different serine or threonine residues including (S/T)P and non (S/T)P sites indicating a relaxed substrate recognition which is not restricted to (S/T)P. However, no phosphorylation at Ser467, the phosphosite in the chemical genetic screen was detected. Interestingly, GST-Ubap2L, showed was phosphorylated at different sites compared to Ubap2L without GST. GST-Ubap2L was phosphorylated at 12 sites, but only three of which were overlapping with the data obtained from Ubap2L. Notably, these three sites are (S/T)P motifs. Npm1 was phosphorylated at Ser125 and at Ser237. However, phosphorylation at Ser70 was not detected, although the corresponding, non-phosphorylated peptide was detected more than 100 times. The SDS-PAGE analysis shows, that part of the Npm1 was not resolved in the gel, but got trapped at the beginning of the separation gel instead. It might be, that pSer70 phosphorylated Npm1 can be found in this not resolved fraction.

Mass spectrometric analysis of MBP-Hsf1 (217-384) revealed phosphorylation at Ser303, Ser307, and Ser363. Serine 363 has been identified as phosphorylation site in the Cdk10 chemical genetic screen and could thus been confirmed *in vitro* using recombinant MBP-Hsf1 (217-384) as substrate. Serine 303 phosphorylation has thus far been assigned to GSK-3b and requires priming at Ser307, which is phosphorylated by MAPK3 (Knauf et al. 1996). SRRT got phosphorylated in the GST-tev linker region. In addition, a phosphorylation at Ser540 but not Ser544, the site which was found in the screen, was identified. Limk2 was found to be phosphorylated exclusively at Thr210, the site identified in the screen. Limk2 Thr210 resides within the PDZ-domain of Limk2 and phosphorylation at this site could regulate interaction of Limk2 with other PDZ-domain containing proteins. Trim28 was phosphorylated at 5 positions, Ser19, Thr34, Ser50, Thr574, and Ser601 (Figure 5.2.15, B). With respect to the chemical genetic Cdk10 substrate screen only Hsf1 Ser363, Limk2 Thr210, and Trim28 Ser19 could be confirmed as direct phosphorylation sites *in vitro*. Interestingly, the identified phosphorylation sites of Ubap2L and GST-Ubap2L were almost different. This might reflect a relaxed substrate specificity by Cdk10/CycM in intrinsically disordered protein domains.

In an *in vitro* chemical genetic screen 66 putative Cdk10/CycM substrates were identified. Using recombinant proteins six of the putative substrates were confirmed as Cdk10/CycM in radioactive *in vitro* kinase assays. Mapping of the phosphorylation sites by mass spectrometry revealed differences between the sites identified in the chemical genetic screen using cell extracts and the *in vitro* approach using recombinant proteins. Future studies are needed to address the significance of the identified Cdk10/CycM substrates *in vivo*.



Ubap2L	Ser356, Ser360, Ser380, Ser389, Ser394, Thr396, Ser410, **Ser416**, **Thr425**, Ser427, Ser439, Ser444, Ser445, Thr446, Ser453, Ser454, Ser604, Ser608, **Ser609**
Npm1	Thr237
Hsf1	Ser304, Ser307, Ser363
Limk2	Thr210
SRRT	Ser540
Trim28	Ser19, Thr34, Ser50, Thr574, Ser601

Figure 5.2.15: Phosphosite determination of *in vitro* phosphorylated recombinant substrates.
A) Samples were phosphorylated for 1 h at 30°C with 0.2 µM Cdk10/CycM. The reaction contained 2 mM ATP and 40 µM of each substrate. The reaction was stopped by addition of 2xSDS sample buffer and 6 µg of each protein was resolved by SDS-PAGE. Samples which were analysed by mass spectrometry are indicated with a rectangle. **B**) Phosphosites identified in the samples. For Ubap2L, also data from GST-Ubap2L after overnight incubation with 0.2 µM Cdk10/CycM are included and indicated by underline. Phosphosites which were identified in both samples are highlighted in bold.

6 Discussion

In this thesis the Cyclin-dependent kinases Cdk7 and Cdk10 in complex with their regulatory subunits were analysed by biochemical and biophysical techniques. Purified recombinant proteins expressed in *Sf9* insect cells were used to address the function, substrate specificity and regulation of these Cyclin-dependent kinase complexes. This discussion will place the obtained results within the current scientific debate and highlight technical restrictions of the performed analyses.

6.1 Cdk7

The cellular functions of Cdk7 have been subject to intensive research in the last decades. Nonetheless, many aspects, in particular regarding molecular mechanisms of Cdk7 kinase activity modulation are not entirely understood. In this study a yet overlooked mode of Cdk7 kinase activity regulation, the phosphorylation of Cdk7 at Ser164, was identified. In the following section I will try to highlight technical constraints of the performed analyses and speculate about the significance of Cdk7 Ser164 phosphorylation into a cellular context.

6.1.1 What is the phosphorylation status of recombinant Cdk7?

In the first part of this thesis it was shown that the T-loop phosphorylation status of recombinant Cdk7 expressed in *Sf9* insect cells differs between monomeric Cdk7, dimeric Cdk7/CycH and trimeric Cdk7/CycH/Mat1 complexes. Even if differences in T-loop phosphorylation upon co-expression of Cdk7/CycH/Mat1 have been documented in earlier studies (Larochelle et al. 2001; Lolli et al. 2004) it seems important to me to highlight this fact at the beginning of the Discussion. It is not intuitive that co-expression of dimeric Cdk7/CycH results in a Cdk7 phosphorylation status which is nearly 100% doubly phosphorylated, whereas upon co-expression of all three subunits of the ternary Cdk7/CycH/Mat1 complex the Cdk7 T-loop phosphorylation at both Ser164 and Thr170 is nearly absent. T-loop phosphorylation of monomeric Cdk7 and Cdk7/CycH/Mat1 complexes has been analysed by mass spectrometry in previous studies with different results.

The laboratory of Prof. Dame Dr. Louise Johnson determined the crystal structure of Cdk7 (Lolli et al. 2004). In the study the authors showed by mass spectrometry that monomeric Cdk7, purified similarly as in this thesis, was phosphorylated to 100 % at Thr170 of which 30% were doubly phosphorylated at Ser164 and Thr170, mirroring the observations of this thesis (Figure 5.1.2). However, the phosphorylation status of co-expressed trimeric Cdk7/CycH/Mat1 complexes differed from the observations in this thesis. Whereas co-expressed trimeric Cdk7/CycH/Mat1 in this thesis was virtually devoid of T-loop phosphorylation, the laboratory of Louise Johnson found approximately 60% of the sample being phosphorylated at Thr170. Moreover, 30% was bis-phosphorylated at Ser164 and Thr170. An explanation for the relatively high abundance of T-loop phosphorylation in their study is found in the methodology. For expression, they infected *Sf21* cells with three different viruses containing the individual components at an MOI of 0.3 each. It is therefore likely, that not all cells were co-infected with all three components at the same time and that trimeric complexes have formed from dimeric Cdk7/CycH which was phosphorylated at Ser164 and Thr170 and Mat1 upon lysis of the cells.

Another study analysing the phosphorylation status of Cdk7 in trimeric complexes by mass spectrometry reported, that Cdk7 is phosphorylated at Thr170 in approximately 40% of the sample, while serine 164 phosphorylation was nearly absent in the same sample (Fouillen et al. 2010). The considerable amount of Thr170 phosphorylation upon co-expression might be explained by technical issues again. In their study Cdk7/CycH/Mat1 complexes for mass spectrometrical analysis were affinity purified by a strep-tag at Cdk7. Complexes were not further purified, and thus it is possible that excess non-complexed Cdk7 of high pThr170 status affected the detected pThr170 levels in the study. This idea fits with the WB data shown in this thesis in which Cdk7 T-loop phosphorylation of co-expressed trimeric complexes was detected in GST-purified samples, but not after further purification by SEC (Figure 5.1.3).

It remains an open question why dimeric Cdk7/CycH is doubly T-loop phosphorylated to nearly 100% at Ser164 and Thr170, whereas these phosphorylations are absent in trimeric Cdk7/CycH/Mat1. It suggests that Mat1 either prevents Cdk7 phosphorylation, perhaps by shielding the Cdk7 T-loop from kinases, or that Mat1 induces de-phosphorylation of previously phosphorylated complexes. In either case it is remarkable, that samples do not contain a broad spectrum of non-, singly-, or doubly-phosphorylated Cdk7.

In *Sf9* cells, Mat1 binding and Cdk7 phosphorylation seem to be mutually exclusive events. It remains an open question if this holds true for human cells. To my knowledge, there are no data available about how and when trimeric Cdk7/CycH/Mat1 complexes form in cells. It thus remains an open question if Cdk7 is phosphorylated before or after association with Mat1 or if both ways exist in human. Notably, both Cdk7 Thr170 and Ser164 can be phosphorylated by Cdk1 and Cdk2 *in vitro* (Garrett et al. 2001).

6.1.2 Mat1 increases Cdk7 activity in a phosphorylation dependent manner

Mat1 binding to pSer164/pThr170 Cdk7/CycH increases the kinase activity of Cdk7/CycH by up to 10-fold towards RNA pol II CTD. The increase seems to affect both, phosphorylation at position Ser5 and at Ser7 and thus does not alter phosphorylation preferences within the repeat. Previous studies suggested that Mat1 dependent increase in Cdk7 kinase activity is regulated by Cdk7 Thr170 phosphorylation (Larochelle et al. 2001; Schachter et al. 2013; Larochelle et al. 2012b). However, mutation of Cdk7 Ser164 to alanine or glutamate, respectively, clearly demonstrated that Ser164 phosphorylation is essential for increased activity upon Mat1 binding. Analysis of the Mat1 binding affinity to different Cdk7/CycH complexes by surface plasmon resonance showed, that Ser164 phosphorylation does not affect Mat1 binding affinity.

Whereas the increase of Cdk7/CycH upon Mat1 binding was determined to be approximately 10-fold, Cdk2 phosphorylation by Cdk7 was only increased 3-fold. This might reflect, that Cdk7 activity towards Cdk2 is affected by Thr170 phosphorylation but not by Ser164 phosphorylation. An interesting observation is, that phosphorylation of the synthetic $CTD_{[3]}$ peptides was only increased 3-fold whereas in the same experiment GST-$CTD_{[52]}$ and GST-$CTD_{[9]KKK}$ phosphorylation were increased 6-fold upon Mat1 addition. A reason for this could be the lower number of absolute repeats and hence phosphorylation sites. GST-$CTD_{[52]}$ was used at 10 µM concentration, GST-$CTD_{[9]KKK}$ at 50 µM and cons-$CTD_{[3]}$ at 100 µM resulting in a total of 520, 450, or 300 CTD repeats, respectively.

However, the fold-change is exactly the same for cons-$CTD_{[3]}$ and K7-$CTD_{[3]}$ as well as for Cdk2, even if the absolute phosphorylation strength differs among the samples. This strongly argues against a concentration dependent bias. An explanation could be that efficient substrate recognition requires a prolonged sequence. Another possibility would be, that Mat1 alters the processivity of the Cdk7/CycH complex.

In kinase assays, full activation of Cdk7/CycH by MBP-Mat1 (230-309) required a 4-fold excess of Mat1. Stability measurements with full length Mat1 produced uniform results only at a 10-fold excess. Activity measurements were carried out at 100 nM Cdk7/CycH complex and stability determination at 2.5 µM Cdk7/CycH and 25 µM Mat1. With respect to the K_D for Mat1 binding, which was determined to be at 9.8 nM, mixing in equimolar ratio should have been sufficient to generate Cdk7/CycH/Mat1 complexes at these concentrations. It seems unlikely that Cdk7/CycH/Mat1 does not associate in a 1:1:1 ratio. Rather, it is assumed that a considerable fraction of the MBP-Mat1 (230-309) is not properly folded. Notably, after affinity chromatography approximately one third of the MBP-Mat1 (230-309) was associated with a protein from *Sf9* insect cells which is commonly observed in protein preparations in *Sf9* cells. It can be assumed that this protein is a chaperone. The chaperone-free Mat1 could be separated from chaperone associated fraction by size exclusion chromatography. The even higher requirement of full length Mat1 in thermal stability measurements partially arises from reduced purity compared to MBP-Mat1 (230-309).

Cdk7 serine 164 phosphorylation could not be adequately compensated by a phosphorylation-mimicking mutation to glutamate. However, when complexed to truncated CycH (1-291) instead of full length CycH, $Cdk7^{S164E}$ was activated by Mat1 to the same extend as wild type Cdk7. Consecutive C-terminal truncations of CycH revealed that inclusion of amino acids 292-295, a basic KKRK cluster in the CycH C-terminus prevented the activatory potential of Mat1. This suggests that the basic KKRK cluster is located in a way which is inhibitory in the absence of Ser164 phosphorylation. Release of the inhibitory state requires re-location of this cluster which is coordinated by Cdk7 Ser164 phosphorylation. The fact, that the inhibitory cluster is positively charged offers the possibility that relocation of the cluster is mediated by direct coordination of the basic cluster by Ser164 phosphorylation.

Mutant $Cdk7^{S164A}$ was not able to bind a C-terminally truncated form of CycH (1-291). The absence of CycH (1-291) binding in the $Cdk7^{S164A}$ mutant not necessarily indicates a requirement of Ser164 phosphorylation. Serine 164 could also be important for the formation of hydrogen bonds by its hydroxyl-group which would be abolished by the S164A mutation. Nonetheless, these data revealed an unexpected role of the Cyclin H C-terminus in Cdk7 binding. The fact, that phosphorylated wild type Cdk7 and $Cdk7^{S164E}$ bound CycH (1-291) suggests, that phosphorylated Ser164 interacts with residues outside of the C-terminus of CycH. The fact, that $Cdk7^{S164A}$ bound full length CycH, together with the other data discussed here, leads to the conclusion that the CycH C-terminus can be bound in at least two different ways.

These observations might explain why Ser164 phosphorylation cannot be mimicked by glutamate. In contrast to glutamate which introduces a single negative charge, serine phosphorylation introduces three negatively charged oxygens and is more spacious compared to glutamate. The lack of another charge in the S164E mutant might prevents coordination of the CycH C-terminus and hence

relocation of the inhibitory KKRK cluster. Based on these data and assumptions, the following hypotheses of the Cyclin H C-terminus can be postulated which should be tested in future biochemical studies:

- the Cyclin H C-terminus is able to bind Cdk7 in a Cdk7 Ser164 independent manner
- the Cyclin H C-terminus is reallocated upon Cdk7 Ser164 phosphorylation
- the Cyclin H C-terminus contains an inhibitory region, presumably the KKRK cluster.

It will be of interest to characterise the minimal CycH C-terminus that is still able to bind $Cdk7^{S164A}$. This would give insights into the C-terminal part required for Cdk7 binding in absence of phosphorylation. Additionally, mutation of the KKRK cluster to AAAA in the context of full length Cyclin H could give valuable insights into the inhibitory nature of this cluster.

Notably, regulation of Cdk7 activity by phosphorylation of the CycH C-terminus has been described and suggested as a regulatory mechanism (Akoulitchev et al. 2000; Lolli et al. 2004; Schneider et al. 2002). Whereas the sites of Cdk7/CycH autophopshorylation have not been determined, phosphorylation within the Cyclin H C-terminus has been described to occur upon Cdk8/CycC or Casein Kinase 2 incubation. Cyclin H phosphorylation at Ser5 and Thr304 by Cdk8/CycC inhibited transcriptional activity of TFIIH and its CTD kinase activity (Akoulitchev et al. 2000). In contrast, phosphorylation of CycH at Ser315 by casein kinase increased Cdk7/CycH/Mat1 activity (Schneider et al. 2002). Phosphorylation of CycH by Cdk7/CycH autophosphorylation, Cdk8/CycC, and casein kinase 2 was confirmed in this thesis, but no experiments regarding the resulting consequences on activity were performed (data not shown).

It is important to highlight that differences in activity of both, Cdk7 serine 164 mutants and Cyclin H truncations, only manifest upon Mat1 binding. None of the modifications had a detectable effect on dimeric Cdk7/CycH activity. It remains an open question if the suggested conformational changes are already formed in absence of Mat1 or if they are induced upon Mat1 binding.

6.1.3 Stability of Cdk7/CycH/Mat1 complexes

In this thesis, the thermal stability of several Cdk7 complexes was determined. Analysis showed, that in dimeric Cdk7/CycH complexes phosphorylation of Ser164 increases thermal stability by 1°C. Comparison of co-expressed and hence non-phosphorylated Cdk7/CycH/Mat1 (230-309) with its phosphorylated counterpart, which was generated by co-purification of Cdk7/CycH and Mat1 from separate expressions, revealed, that the phosphorylated complex was more stable with a difference of 9.5°C. It is important to highlight, that both complexes have exactly the same subunit composition and that increased stability is a consequence of post-translational modification, most likely the documented differences in T-loop phosphorylation. The respective $Cdk7^{S164A}$/CycH + Mat1 complex was less stable by 2°C but principally retained high stability. This suggests, that Thr170 phosphorylation is the main determinant for increased stability. In the same analysis truncation of Cyclin H produced a bi-phasic denaturation in thermal stability experiments. This could be a consequence of sub-stoichiometric amounts of Mat1 or significantly reduced Thr170 phosphorylation. That Mat1 is bound at a substoichiometric ratio seems rather unlikely, because all complexes were affinity purified via the MBP-tag fused to Mat1. Truncation of CycH did not affect

Mat1 binding affinity in SPR measurements. Therefore, it seems unlikely that a considerable loss of Mat1 during SEC could account for the effect. Decreased levels of pThr170 cannot be excluded but it is striking, that also in the non-related *in vitro* reconstitution experiments with full length Mat1 only Cdk7/CycH (1-291) but not any of the Cdk7/CycH (1-323) complexes produced a non-homogenous melting point. This suggests that the C-terminus is required to stabilise trimeric complexes.

6.1.4 Is Cdk7 Ser164 phosphorylation a regulatory switch *in vivo*?

In this thesis it could be unequivocally demonstrated, that Cdk7 Ser164 phosphorylation is required for full activation of the Cdk7 kinase activity in the trimeric Cdk7/CycH/Mat1 complex. This study does not provide any cellular data that allow to claim biological significance of this finding. Even if there are only few studies analysing the role of Cdk7 Ser164 phosphorylation in cells it is tempting to speculate that Cdk7 indeed is regulated by Cdk7 Ser164 phosphorylation. The studies are described in the introduction of this thesis and will not reconciled here. Instead, this part of the discussion will speculate about possible pathways involving Cdk7 Ser164 phosphorylation.

6.1.5 Is Cdk7 activity regulation at Ser164 specific for RNA pol II CTD?

Increased activity towards RNA pol II CTD upon association with TFIIH or Mat1 has been described early on (Rossignol et al. 1997; Yankulov und Bentley 1997). In contrast Cdk2 phosphorylation by Cdk7/CycH is not or only mildly affected by Mat1 association. This observation has led to the proposal that Mat1 association regulates the dichotomous activities of Cdk7 as CAK and in transcription. However, the phosphorylation efficiency of several Cdk7 substrates has been described to be Mat1 association dependent and comprise Cdk4, p53, TFIIE, RNA pol II CTD, and SPT5 (Larochelle et al. 2006; Schachter et al. 2013; Chi et al. 2011; Ko et al. 1997; Larochelle et al. 2012b; Rochette-Egly et al. 1997; Lu et al. 1997). In the light of this thesis it has to be established if these substrates depend on Cdk7 Ser164 phosphorylation as well. This highlights the possibility that Cdk7 Ser164 phosphorylation serves as a general regulatory switch for phosphorylation of several substrates.

6.1.6 An inhibitory function of Ser164 phosphorylation?

In 1998, two groups independently discovered that transcription associated Cdk7 activity is reduced in mitosis (Long et al. 1998; Akoulitchev und Reinberg 1998). The laboratory of Danny Reinberg assigned this reduction in activity to phosphorylation of Cdk7 at Ser164 contrasting the activity data in this thesis as well as published activity and stability data (Martinez et al. 1997; Larochelle et al. 2001). Importantly, the reduced Cdk7 activity was only found in TFIIH associated Cdk7 but not in free Cdk7/CycH/Mat1 complexes. Moreover, the study did not determine Cdk7 pThr170 levels of the TFIIH associated Cdk7 protein but detected a general decrease in Cdk7 pThr170 and an increase of pSer164 in mitotic extracts.

This raises the possibility, that Cdk7 Ser164 phosphorylation might be inhibitory to Cdk7 activity in the absence of Thr170 phosphorylation. In this study, Cdk7 Ser164 was mainly analysed in combination with Cdk7 Thr170 phosphorylation. Interestingly, a Cdk7^{S164E}/CycH (1-291)/Mat1

(230-309) complex expressed during the course of this study exhibited lower activity than all other complexes tested (data not shown). As it was co-expressed it can be assumed that Cdk7 was not phosphorylated at Thr170 in this complex. However, the complex was only purified once and the role of Ser164 phosphorylation in absence of Thr170 was not systematically analysed.

6.1.7 Does Cdk7 T-loop phosphorylation affect binding to interaction partners?

Another possibility how Ser164 phosphorylation could affect Cdk7 function would be by affecting association of Cdk7/CycH/Mat1 with its binding partners. Association of Mat1 with Thr170 phosphorylated Cdk7/CycH results in a strong increase in thermal stability. It is intriguing to speculate that the increased stability is a consequence of structural changes in the complex. This raises the question if phosphorylation of Cdk7 Serine 164 and/or Thr170 affects the association with Cdk7/CycH/Mat1 binding partners. The best studied Cdk7/CycH/Mat1 binding partner is the TFIIH core complex. Association is mediated by the TFIIH XPD and XPB subunits. Binding to XPD depends on the interaction of the central helical Mat1 domain with the XPD ARCH domain. A recent study demonstrated that the release of Cdk7/CycH/Mat1 from TFIIH, a pre-requisite for nucleotide excision repair (NER), is induced by binding of XPA and a concomitant structural change of the XPD ARCH domain (Kokic et al. 2019). In agreement with a reversible mechanism, addition of excess Cdk7/CycH/Mat1 complex was able to inhibit NER function. In their study they could show, that this effect is not dependent on Cdk7 kinase activity. Altered binding affinity could thus directly affect TFIIH function in NER and transcription. XPD also forms a quaternary complex with Cdk7/CycH/Mat1 which seems to titrate Cdk7/CycH/Mat1 away from cell cycle targets in mitosis. Interestingly, Cdk7 Ser164 phosphorylation is high in mitosis arrested nocodazole treated cells. Recently, TFIIE was found to recruit Cdk7/CycH/Mat1 complexes to the pre-initiation complex in a TFIIH independent manner (Compe et al. 2019). Notably, the inhibitory KKRK cluster in CycH was shown to be a functional nuclear localisation sequence and interacts with importin α (Krempler et al. 2005). Re-organization of the KKRK cluster in Cdk7/CycH complexes after phosphorylation of Cdk7 Ser164 could affect importin association or vice versa. Future studies should address the Cdk7 T-loop phosphorylation status when analysing interaction partners.

6.1.8 Cdk7 Ser164 phosphorylation – a positive feedback loop for cell cycle commitment?

In 2013, the laboratory of Robert Fisher showed that Cdk4 activation by Cdk7 regulates cell cycle progression at the restriction point. Interestingly, Cdk4 activation was Mat1 dependent *in vitro*. Increase of activity upon Mat1 addition has so far been only attributed to Cdk7 Thr170 phosphorylation and it was therefore speculated, that Cdk7 activity towards Cdk4 is regulated by Cdk7 T-loop phosphorylation at Thr170. Indeed, Cdk7 Thr170 phosphorylation was absent in serum starved cells and increased upon cell cycle re-entry. In the light of the data obtained in this thesis, which show that Ser164 phosphorylation is required for activity regulation by Mat1, it is intriguing to speculate that also Cdk4 activation is dependent on Cdk7 Ser164 phosphorylation. Active Cdk4/CycD phosphorylates Rb proteins which inhibit E2F transcription factors. Upon

phosphorylation by Cdk4, Rb-E2F interaction is released and E2F binds to its target genes. Among the E2F target genes is Cyclin E which upon expression associates with Cdk2 leading to hyperphosphorylation of the Rb protein. Together with the finding, that Cdk2 phosphorylates Cdk7 Ser164 and Thr170 *in vitro* (Garrett et al. 2001) this raises the possibility of a positive feedback loop at the restriction point. The Cdk2/CycE complex would not only phosphorylate Rb protein but also Cdk7 Ser164 resulting in increased Cdk4 activation.

6.1.9 Does Ser164 phosphorylation account for CAK differences in yeast?

Finally, a highly speculative but interesting aspect of Cdk7 Ser164 phosphorylation comes from observations in yeast. As outlined in the Introduction section of this thesis, *S. cerevisiae* and *S. pombe* differ with respect to Cdk7 CAK function. In *S. cerevisiae* the Cdk7 homologue Kin28 does not possess any CDK-activating kinase activity and CAK function is carried out by a separate kinase, Cak1. In *S. pombe* the Cdk7 homologue Mcs6 contains CAK function but is assisted by another kinase termed Csk1.

Comparison of human Cdk7 with its counterparts in *S. pombe* and *S. cereviseae* reveals, that Mcs6 but not Kin28 contains an additional phosphorylatable Thr/Pro site in its activation loop. Accordingly, the *S. pombe* Cyclin H homologue Mcs2, but not the *S. cerevisiae* Cyclin H homologue Ccl1 contains a basic KRK cluster in a similar position in its C-terminus. Serine 164 phosphorylation increases Cdk7 activity towards RNA pol II CTD but not or only to a lesser extent to cyclin-dependent kinases. Even if counterintuitive, the requirement of a separate cdk-activating kinase in humans and *S. pombe* is maybe circumvented by an additional layer of regulation mediated by phosphorylation of the Cdk7 T-loop and an inhibitory basic cluster following the cyclin-boxes in Cyclin H.

6.2 Cdk10

In this study an expression and purification strategy was developed which allows expression and purification of recombinant human Cdk10/CycM complexes. Cdk10/CycM expression levels in *Sf9* cells could be improved by two means, fusion of Cdk10 to MBP and generation of a kinase-dead mutant. MBP-Cdk10/His-CycM was expressed much better than its GST-Cdk10/GST-CycM counterpart suggesting improved stability due to MBP-association. In accordance with this idea, Cdk10/CycM was only stable in buffers containing 300 mM NaCl, whereas a salt concentration of 150 mM NaCl results in nearly complete precipitation upon MBP-tag removal. Moreover, Cdk10 has been described to strongly interact with HSP90 in human cells, indicating that it is non-stable (Taipale et al. 2012). Expression of Cdk10/CycM in *Sf9* cells generated active kinase complexes without the need of special means to activate the kinase. *In vitro* kinase assays identified RNA pol II CTD and c-Myc as substrates of this kinase. The *in vitro* activity towards GST-CTD was low compared to known CTD kinases Cdk7 and Cdk9. The much higher activity towards c-Myc compared to CTD indicates, that low activity towards CTD is substrate specific and not due to inefficient T-loop phosphorylation. Additionally, co-expression of Cdk10/CycM with *S. cerevisiae* Cak1 did not increase the activity (data not shown).

In addition to wild type and kinase dead Cdk10/CycM complexes, an analogue-sensitive Cdk10 variant was generated by mutation of the gatekeeper residue methionine 117 to glycine. Cdk10^{M117G}/CycM was less active than wild type and showed a lower affinity for ATP (Figure 5.2.9). The reduced activity and ATP affinity might explain why Cdk10^{M117G}/CycM expression levels were higher compared to wild type Cdk10 and resembled the expression levels of the kinase dead Cdk10^{D163N} mutant. An ATP competition assay testing seven different ATP analogues identified benzyl-, phenylethyl-, and cyclo-hexyl-ATP-γ-S as most potent ATP competitors using Cdk10^{M117G}/CycM. None of the analogues tested had an effect on wild type Cdk10/CycM. Titration of these analogues revealed that phenylethyl-ATP-γ-S was the most potent ATP competitor. Interestingly, all three analogues stimulated kinase activity at low concentration. An explanation for this might be that that the analogues had a stabilising effect on the kinase. This idea could be further addressed in nanoDSF thermal stability studies testing different analogues as well as the inhibitor 1-NM-PP1, which specifically inhibits analogue sensitive kinases. If the analogues or the inhibitor indeed lead to a considerable stabilization of the complex this would serve as another option for crystallography of the Cdk10/CycM complex.

6.2.1 Cdk10 substrate specificity

Cdk10/CycM phosphorylated all three serine positions within the RNA pol II CTD hepta repeats. Substitution of Ser7 by Lys increased CTD phosphorylation by Cdk10/CycM. This is different for the known transcriptional CTD kinases P-TEFb in which K7 modification did not affect activity (Bösken, Thesis), and DYRK1A in which activity is halved by Ser7 to Lys substitution (unpublished data, data not shown). Interestingly, serine to lysine substitutions are found in distal human CTD repeats and are absent in yeast. This finding is in line with the absence of a yeast Cdk10 homologue. Attempts to analyse if K7 modification directs Cdk10/CycM to Ser2 or Ser5 position with phosphorylation specific antibodies by western blot were not conclusive (data not shown). It is possible, that the K7 modification affects recognition of the phosphorylation by the antibodies. The

presence of the lysine at position 7 of every repeat in the GST-CTD$_{[9]}$K7 construct allows for analysis of the phosphorylation pattern by mass spectrometry in future studies.

To identify novel Cdk10 substrates a chemical genetic screen was applied. The identification of putative substrates allows to suggest a consensus sequence for Cdk10/CycM substrate recognition. Analysis of the residues adjacent to the site of phosphorylation revealed a preference for (S/T)P motifs. However, only 35% of the putative substrates contained this minimal Cdk consensus sequence. A similar result was found in a screen for Cdk9 substrates (Sansó et al. 2016) and contrasts data for Cdk1/CycB and Cdk2 in complex with a viral cyclin in which substrates nearly exclusively contained (S/T)P motifs (Blethrow et al. 2008; Umaña et al. 2018). This suggests, that Cdk10 and Cdk9 do not strictly rely on proline at position +1 for efficient substrate recognition whereas it is a requirement for Cdk1 and Cdk2. In agreement with this, both Cdk9 and Cdk10 are able to phosphorylate serine7 within a CTD heptad, which is no SP motif. Moreover, identification of the phosphorylation sites in recombinant Ubap2L after incubation with Cdk10/CycM revealed several non-(S/T)P sites. Interestingly, analysis of GST-tagged Ubap2L revealed different sites of phosphorylation as Ubap2L without GST-tag. Of the 19 phosphosites which were found, only three overlapped between GST-Ubap2L and Ubap2L. Notably, the three overlapping sites were (S/T)P motifs. Ubap2L is a mainly unstructured protein, and the part which was expressed in this thesis is also supposed to be unstructured. The differences in GST-Ubap and Ubap2L phosphorylation might argue that non-(S/T)P phosphorylation of Cdk10 is determined by general structural features and not by local sequence. In contrast, substrate recognition of (S/T)P sites is directly dependent on the sequence. This could explain, why the phosphosites differ except for those which are (S/T)P motifs. The hypothesis that non-canonical phosphorylation by CDKs is dependent on distant structural constraints but not local sequence has been raised for CTD phosphorylation at position Ser7 by Cdk9. A further example for the phosphorylation of non-(S/T)P motifs by a Cdk which is dependent on distal structure features instead of the local sequence is the activation of Cdk1, Cdk2, and Cdk6 by Cdk7. Future studies should take this into consideration when evaluating possible substrates.

Determination of the Cdk10 consensus sequence restricted to SP motifs revealed a preference for P, K or R in the +2 position. This could provide an explanation for increased phosphorylation of the CTD by Cdk10 when Ser7 was substituted by Lys (Figure 5.2.6). Moreover, it suggests that K7 CTD is favourably phosphorylated at Ser5 but not Ser2. The preference for a negatively charged glutamate at position -4 offers the possibility to further increase substrate recognition of K7 repeats by phosphorylation of Tyr1. Another interesting aspect based on the derived consensus sequence would be that pre-phosphorylation of Ser5 in the adjacent N-terminal repeat, but not in the same repeat could facilitate Ser2 phosphorylation (Figure 6.2.1).

-5	-4	-3	-2	-1	p-site	+1	+2	+3	+4	+5	
A	E	R	P	P	S	P	P/K/R	K	S	A	consensus sequence
T	S	P	S	Y	S	P	T	S	P	S	Ser2
S	Y	S	P	T	S	P	S	Y	S	P	Ser5
S	P	T	S	P	S	Y	S	P	T	S	Ser7
T	S	P	K	Y	S	P	T	S	P	K	Ser2 in K7
K	Y	S	P	T	S	P	K	Y	S	P	Ser5 in K7

Figure 6.2.1: Alignment of CTD phosphorylation sites with the Cdk10 consensus sequence.
The possible phosphorylation sites within a CTD hepta repeat and the adjacent amino acids were aligned to the Cdk10 consensus sequence derived from data analysis of the chemical genetic screen. P-site: site of phosphorylation.

6.2.2 Identification of Cdk10 substrates by mass spectrometry

Cdk10 is supposed to act mainly nuclear. In this study nuclear extracts were used to enrich the sample for relevant substrates and to reduce sample complexity. However, despite dominant nuclear distribution of several cyclin-dependent kinases, non-nuclear functions have been ascribed to Cdk10. Cdk10 was found to localise to the basal body of primary cilia at the cell membrane (Guen et al. 2016). The use of nuclear extracts maybe prohibited the identification of important non-nuclear Cdk10 substrates. In future studies, fractionated cell extract should be used as an addition after initial screening of full cell lysates.

In this study, thio-phosphorylated peptides were enriched by a covalent capture and release protocol (Blethrow et al. 2008). The detection of kinase substrates by purification of thio-phosphorylated peptides faces two method inherent confounding issues. The first is, that thio-phosphorylated cysteine-rich peptides are not covered by the screen as they tend to bind the iodoacetyl matrix by their cysteine residues and are subsequently removed during the washing steps. This issue has been raised already by the initial description of the protocol (Blethrow et al. 2008) and could be observed in this study as attempts to recover *in vitro* thio-phosphorylated, cysteine rich c-Myc (17-167) from full cell lysate were not successful. Only the use of large quantities of thio-phosphorylated c-Myc allowed its identification. This observation may explains, why c-Myc was not identified in this screen and also not in a similar screen for Cdk9 substrates (Sansó et al. 2016) even if it serves as a potent substrate of these kinases in recombinant *in vitro* assays. The second confounding issue is, that some substrates are not detected because they are phosphorylated within a sequence which is not applicable to mass spectrometry analysis. A good example for this is the RNA pol II CTD which, due to the repetitive nature and low abundance of the amino acids arginine or lysine in the sequence, does not allow for tryptic digest. Even if RNA pol II CTD is probably one of the best studied substrates of Cdk9 it was not detected by this kind of screen (Sansó et al. 2016).

A possibility to address both issues would be to immunoprecipitate the thio-phosphorylated proteins from cell lysate after alkylation with the alkylation specific antibody (Allen et al. 2007). Putative substrates could be identified by enrichment of a protein in the AS sample over the control sample. This technique would allow to identify proteins which are not covered by the covalent capture and release method but would lose phosphorylation site specific information. Future studies should combine both ways of substrate enrichment to achieve a more complete picture of the kinase substrates.

In the Cdk10 substrate screen, a total of 89 phosphosites originating from 66 different proteins were identified. Analysis of GO-term enrichment against the human proteome using GOrilla (Eden et al. 2009) revealed that the proteins are enriched for RNA associated processes. However, when the sample was analysed for GO-term enrichment against nuclear proteins no significant enrichment was found. The observed enrichment in RNA regulatory proteins could therefore be a consequence of the use of nuclear extracts. The analysis assigned a 300-fold enrichment of proteins involved in cap-dependent translational initiation. This strong enrichment arises from the identification of the eukaryotic translation initiation factor 3, subunit d (EIF3D), and the eukaryotic translation initiation factor 4 gamma 1 (EIF4G1) in the screen. Because EIF3D and EIF4G1 are the only two proteins in humans which are included in this GO-term, the strong enrichment in the analysis should be considered with care. Concomitantly, it was not annotated as significant when the dataset was analysed with PANTHER (Mi et al. 2019).

Interestingly, Cdk10 was found to phosphorylate itself at its activatory threonine in the T-loop. Autophosphorylation of Cdk10 during kinase assays was observed in this study but has not been further characterized (data not shown). No Cdk10 activating kinase has been identified yet. It has to be established if Cdk10 auto-activation serves as a regulatory mechanism *in vivo*. In this study autophosphorylation of Cdk10 at Thr196 might was caused by the high abundance of recombinant $Cdk10^{M117G}$/CycM in the sample.

In higher organisms, interference with Cdk10 activity by mutation, genetic ablation, or morpholino based downregulation affects development of the central nervous system. In contrast, Cdk10 is not essential for maintenance of the cell cycle or cell viability for cultured cells (Windpassinger et al. 2017; Yeh et al. 2013; Guen et al. 2018). In cell culture, phenotypic effects of Cdk10 are often observed after inhibition or stimulation with growth factors or differentiating agents (Iorns et al. 2008; Windpassinger et al. 2017; Guen et al. 2016). Additionally, Cdk10 was found to promote transcription in the butterfly *Helicoverpa armigera* after stimulation with the growth hormone 20-hydroxyecdysone (Liu et al. 2014). It thus seems, as if Cdk10 exerts important function in development and in response to growth factors. Due to the limited number of known Cdk10 substrates, data interpretation in earlier studies was restricted to effects based on the Cdk10-ETS2 axis and subsequent alteration of the Ras-Raf-MAPK pathway. The identification of Cdk10 substrates in this thesis provides a first step to identify other pathways by which Cdk10 could affect the cellular response to external stimuli.

7 References

Akhtar, Md Sohail; Heidemann, Martin; Tietjen, Joshua R.; Zhang, David W.; Chapman, Rob D.; Eick, Dirk; Ansari, Aseem Z. (2009): TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. In: *Molecular cell* 34 (3), S. 387–393. DOI: 10.1016/j.molcel.2009.04.016.

Akoulitchev, S.; Chuikov, S.; Reinberg, D. (2000): TFIIH is negatively regulated by cdk8-containing mediator complexes. In: *Nature* 407 (6800), S. 102–106. DOI: 10.1038/35024111.

Akoulitchev, S.; Reinberg, D. (1998): The molecular mechanism of mitotic inhibition of TFIIH is mediated by phosphorylation of CDK7. In: *Genes & development* 12 (22), S. 3541–3550. DOI: 10.1101/gad.12.22.3541.

Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Morgan, David; Raff, Martin; Roberts, Keith; Walter, Peter (2015): Molecular biology of the cell. Unter Mitarbeit von John Wilson und Tim Hunt. Sixth edition. New York, NY: Garland Science Taylor and Francis Group.

Allen, Jasmina J.; Li, Manqing; Brinkworth, Craig S.; Paulson, Jennifer L.; Wang, Dan; Hübner, Anette et al. (2007): A semisynthetic epitope for kinase substrates. In: *Nature methods* 4 (6), S. 511–516. DOI: 10.1038/NMETH1048.

Andersen, G.; Busso, D.; Poterszman, A.; Hwang, J. R.; Wurtz, J. M.; Ripp, R. et al. (1997): The structure of cyclin H: common mode of kinase activation and specific features. In: *The EMBO journal* 16 (5), S. 958–967. DOI: 10.1093/emboj/16.5.958.

Avery, O. T.; Macleod, C. M.; McCarty, M. (1944): STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. In: *The Journal of experimental medicine* 79 (2), S. 137–158. DOI: 10.1084/jem.79.2.137.

Bao, Zhao Qin; Jacobsen, Douglas M.; Young, Matthew A. (2011): Briefly bound to activate: transient binding of a second catalytic magnesium activates the structure and dynamics of CDK2 kinase for catalysis. In: *Structure (London, England : 1993)* 19 (5), S. 675–690. DOI: 10.1016/j.str.2011.02.016.

Baumli, Sonja; Lolli, Graziano; Lowe, Edward D.; Troiani, Sonia; Rusconi, Luisa; Bullock, Alex N. et al. (2008): The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. In: *The EMBO journal* 27 (13), S. 1907–1918. DOI: 10.1038/emboj.2008.121.

Bertoli, Cosetta; Skotheim, Jan M.; Bruin, Robertus A. M. de (2013): Control of cell cycle transcription during G1 and S phases. In: *Nature reviews. Molecular cell biology* 14 (8), S. 518–528. DOI: 10.1038/nrm3629.

Bieniossek, Christoph; Richmond, Timothy J.; Berger, Imre (2008): MultiBac: multigene baculovirus-based eukaryotic protein complex production. In: *Current protocols in protein science* Chapter 5, Unit 5.20. DOI: 10.1002/0471140864.ps0520s51.

Bishop, A. C.; Ubersax, J. A.; Petsch, D. T.; Matheos, D. P.; Gray, N. S.; Blethrow, J. et al. (2000): A chemical switch for inhibitor-sensitive alleles of any protein kinase. In: *Nature* 407 (6802), S. 395–401. DOI: 10.1038/35030148.

Blethrow, Justin D.; Glavy, Joseph S.; Morgan, David O.; Shokat, Kevan M. (2008): Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. In: *Proceedings of the National Academy of Sciences of the United States of America* 105 (5), S. 1442–1447. DOI: 10.1073/pnas.0708966105.

Bösken, Christian A.; Farnung, Lucas; Hintermair, Corinna; Merzel Schachter, Miriam; Vogel-Bachmayr, Karin; Blazek, Dalibor et al. (2014): The structure and substrate specificity of human Cdk12/Cyclin K. In: *Nature communications* 5, S. 3505. DOI: 10.1038/ncomms4505.

Bösken, Christian Andreas (2013): Strukturelle und funktionelle Charakterisierung der transkriptionsregulierenden Kinasen Cdk9 und Cdk12.

Bradner, James E.; Hnisz, Denes; Young, Richard A. (2017): Transcriptional Addiction in Cancer. In: *Cell* 168 (4), S. 629–643. DOI: 10.1016/j.cell.2016.12.013.

Brambilla, R.; Draetta, G. (1994): Molecular cloning of PISSLRE, a novel putative member of the cdk family of protein serine/threonine kinases. In: *Oncogene* 9 (10), S. 3037–3041.

Brotherton, D. H.; Dhanaraj, V.; Wick, S.; Brizuela, L.; Domaille, P. J.; Volyanik, E. et al. (1998): Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19INK4d. In: *Nature* 395 (6699), S. 244–250. DOI: 10.1038/26164.

Bullrich, F.; MacLachlan, T. K.; Sang, N.; Druck, T.; Veronese, M. L.; Allen, S. L. et al. (1995): Chromosomal mapping of members of the cdc2 family of protein kinases, cdk3, cdk6, PISSLRE, and PITALRE, and a cdk inhibitor, p27Kip1, to regions involved in human cancer. In: *Cancer research* 55 (6), S. 1199–1205.

Busso, D.; Keriel, A.; Sandrock, B.; Poterszman, A.; Gileadi, O.; Egly, J. M. (2000): Distinct regions of MAT1 regulate cdk7 kinase and TFIIH transcription activities. In: *J. Biol. Chem.* 275 (30), S. 22815–22823. DOI: 10.1074/jbc.M002578200.

C Quaresma, Alexandre J.; Bugai, Andrii; Barboric, Matjaz (2016): Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. In: *Nucleic acids research* 44 (16), S. 7527–7539. DOI: 10.1093/nar/gkw585.

Chapman, Rob D.; Heidemann, Martin; Hintermair, Corinna; Eick, Dirk (2008): Molecular evolution of the RNA polymerase II CTD. In: *Trends in genetics : TIG* 24 (6), S. 289–296. DOI: 10.1016/j.tig.2008.03.010.

Chen, Jian; Larochelle, Stéphane; Li, Xiaoming; Suter, Beat (2003): Xpd/Ercc2 regulates CAK activity and mitotic progression. In: *Nature* 424 (6945), S. 228–232. DOI: 10.1038/nature01746.

Chi, Yayun; Zhang, Chunyi; Zong, Hongliang; Hong, Yi; Kong, Xiangfei; Liu, Haiou et al. (2011): Thr-370 is responsible for CDK11(p58) autophosphorylation, dimerization, and kinase activity. In: *The Journal of biological chemistry* 286 (3), S. 1748–1757. DOI: 10.1074/jbc.M110.107367.

Chipumuro, Edmond; Marco, Eugenio; Christensen, Camilla L.; Kwiatkowski, Nicholas; Zhang, Tinghu; Hatheway, Clark M. et al. (2014): CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. In: *Cell* 159 (5), S. 1126–1139. DOI: 10.1016/j.cell.2014.10.024.

Coin, Frédéric; Oksenych, Valentyn; Mocquet, Vincent; Groh, Stefanie; Blattner, Christine; Egly, Jean Marc (2008): Nucleotide excision repair driven by the dissociation of CAK from TFIIH. In: *Molecular cell* 31 (1), S. 9–20. DOI: 10.1016/j.molcel.2008.04.024.

Colleoni, B.; Paternot, S.; Pita, J. M.; Bisteau, X.; Coulonval, K.; Davis, R. J. et al. (2017): JNKs function as CDK4-activating kinases by phosphorylating CDK4 and p21. In: *Oncogene* 36 (30), S. 4349–4361. DOI: 10.1038/onc.2017.7.

Compe, Emmanuel; Genes, Carlos M.; Braun, Cathy; Coin, Frederic; Egly, Jean-Marc (2019): TFIIE orchestrates the recruitment of the TFIIH kinase module at promoter before release during transcription. In: *Nature communications* 10 (1), S. 2084. DOI: 10.1038/s41467-019-10131-1.

Core, Leighton; Adelman, Karen (2019): Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. In: *Genes & development* 33 (15-16), S. 960–982. DOI: 10.1101/gad.325142.119.

Czudnochowski, Nadine; Bösken, Christian A.; Geyer, Matthias (2012): Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. In: *Nature communications* 3, S. 842. DOI: 10.1038/ncomms1846.

Devault, A.; Martinez, A. M.; Fesquet, D.; Labbé, J. C.; Morin, N.; Tassan, J. P. et al. (1995): MAT1 ('menage à trois') a new RING finger protein subunit stabilizing cyclin H-cdk7 complexes in starfish and Xenopus CAK. In: *The EMBO journal* 14 (20), S. 5027–5036.

Drapkin, R.; Le Roy, G.; Cho, H.; Akoulitchev, S.; Reinberg, D. (1996): Human cyclin-dependent kinase-activating kinase exists in three distinct complexes. In: *Proceedings of the National Academy of Sciences of the United States of America* 93 (13), S. 6488–6493. DOI: 10.1073/pnas.93.13.6488.

Eden, Eran; Navon, Roy; Steinfeld, Israel; Lipson, Doron; Yakhini, Zohar (2009): GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. In: *BMC bioinformatics* 10, S. 48. DOI: 10.1186/1471-2105-10-48.

El-Aouar Filho, Rachid A.; Nicolas, Aurélie; Paula Castro, Thiago L. de; Deplanche, Martine; Carvalho Azevedo, Vasco A. de; Goossens, Pierre L. et al. (2017): Heterogeneous Family of Cyclomodulins: Smart Weapons That Allow Bacteria to Hijack the Eukaryotic Cell Cycle and Promote Infections. In: *Frontiers in cellular and infection microbiology* 7, S. 208. DOI: 10.3389/fcimb.2017.00208.

Endicott, Jane A.; Noble, Martin E. M.; Johnson, Louise N. (2012): The structural basis for control of eukaryotic protein kinases. In: *Annual review of biochemistry* 81, S. 587–613. DOI: 10.1146/annurev-biochem-052410-090317.

Evans, T.; Rosenthal, E. T.; Youngblom, J.; Distel, D.; Hunt, T. (1983): Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. In: *Cell* 33 (2), S. 389–396. DOI: 10.1016/0092-8674(83)90420-8.

Fesquet, D.; Labbé, J. C.; Derancourt, J.; Capony, J. P.; Galas, S.; Girard, F. et al. (1993): The MO15 gene encodes the catalytic subunit of a protein kinase that activates cdc2 and other cyclin-dependent kinases (CDKs) through phosphorylation of Thr161 and its homologues. In: *The EMBO journal* 12 (8), S. 3111–3121.

Fisher, R. P.; Jin, P.; Chamberlin, H. M.; Morgan, D. O. (1995): Alternative mechanisms of CAK assembly require an assembly factor or an activating kinase. In: *Cell* 83 (1), S. 47–57. DOI: 10.1016/0092-8674(95)90233-3.

Fisher, R. P.; Morgan, D. O. (1994): A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. In: *Cell* 78 (4), S. 713–724. DOI: 10.1016/0092-8674(94)90535-5.

Fisher, Robert P. (2005): Secrets of a double agent: CDK7 in cell-cycle control and transcription. In: *Journal of cell science* 118 (Pt 22), S. 5171–5180. DOI: 10.1242/jcs.02718.

Fisher, Robert P. (2016): Getting to S: CDK functions and targets on the path to cell-cycle commitment. In: *F1000Research* 5, S. 2374. DOI: 10.12688/f1000research.9463.1.

Fouillen, Laetitia; Abdulrahman, Wassim; Moras, Dino; van Dorsselaer, Alain; Poterszman, Arnaud; Sanglier-Cianférani, Sarah (2010): Analysis of recombinant phosphoprotein complexes with complementary mass spectrometry approaches. In: *Analytical biochemistry* 407 (1), S. 34–43. DOI: 10.1016/j.ab.2010.07.006.

Fuda, Nicholas J.; Ardehali, M. Behfar; Lis, John T. (2009): Defining mechanisms that regulate RNA polymerase II transcription in vivo. In: *Nature* 461 (7261), S. 186–192. DOI: 10.1038/nature08449.

Garrett, S.; Barton, W. A.; Knights, R.; Jin, P.; Morgan, D. O.; Fisher, R. P. (2001): Reciprocal activation by cyclin-dependent kinases 2 and 7 is directed by substrate specificity determinants outside the T loop. In: *Molecular and cellular biology* 21 (1), S. 88–99. DOI: 10.1128/MCB.21.1.88-99.2001.

Gervais, V.; Busso, D.; Wasielewski, E.; Poterszman, A.; Egly, J. M.; Thierry, J. C.; Kieffer, B. (2001): Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family. In: *J. Biol. Chem.* 276 (10), S. 7457–7464. DOI: 10.1074/jbc.M007963200.

Graña, X.; Claudio, P. P.; Luca, A. de; Sang, N.; Giordano, A. (1994): PISSLRE, a human novel CDC2-related protein kinase. In: *Oncogene* 9 (7), S. 2097–2103.

Greber, Basil J.; Nguyen, Thi Hoang Duong; Fang, Jie; Afonine, Pavel V.; Adams, Paul D.; Nogales, Eva (2017): The cryo-electron microscopy structure of human transcription factor IIH. In: *Nature* 549 (7672), S. 414–417. DOI: 10.1038/nature23903.

Greifenberg, Ann K. (2014): Biochemische Untersuchung der Phosphorylierungsmuster im Transkriptionszyklus der RNA Polymerase II. Dortmund.

Greifenberg, Ann Katrin; Hönig, Dana; Pilarova, Kveta; Düster, Robert; Bartholomeeusen, Koen; Bösken, Christian A. et al. (2016): Structural and Functional Analysis of the Cdk13/Cyclin K Complex. In: *Cell reports* 14 (2), S. 320–331. DOI: 10.1016/j.celrep.2015.12.025.

Guen, V. J.; Gamble, C.; Flajolet, M.; Unger, S.; Thollet, A.; Ferandin, Y. et al. (2013): CDK10/cyclin M is a protein kinase that controls ETS2 degradation and is deficient in STAR syndrome. In: *Proceedings of the National Academy of Sciences* 110 (48), S. 19525–19530. DOI: 10.1073/pnas.1306814110.

Guen, Vincent J.; Edvardson, Simon; Fraenkel, Nitay D.; Fattal-Valevski, Aviva; Jalas, Chaim; Anteby, Irene et al. (2018): A homozygous deleterious CDK10 mutation in a patient with agenesis of corpus callosum, retinopathy, and deafness. In: *American journal of medical genetics. Part A* 176 (1), S. 92–98. DOI: 10.1002/ajmg.a.38506.

Guen, Vincent J.; Gamble, Carly; Perez, Dahlia E.; Bourassa, Sylvie; Zappel, Hildegard; Gärtner, Jutta et al. (2016): STAR syndrome-associated CDK10/Cyclin M regulates actin network architecture and ciliogenesis. In: *Cell cycle (Georgetown, Tex.)* 15 (5), S. 678–688. DOI: 10.1080/15384101.2016.1147632.

Hanahan, Douglas; Weinberg, Robert A. (2011): Hallmarks of cancer: the next generation. In: *Cell* 144 (5), S. 646–674. DOI: 10.1016/j.cell.2011.02.013.

Harlen, Kevin M.; Churchman, L. Stirling (2017): The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. In: *Nature reviews. Molecular cell biology* 18 (4), S. 263–273. DOI: 10.1038/nrm.2017.10.

Hastie, C. James; McLauchlan, Hilary J.; Cohen, Philip (2006): Assay of protein kinases using radiolabeled ATP: a protocol. In: *Nature protocols* 1 (2), S. 968–971. DOI: 10.1038/nprot.2006.149.

Holstege, F. C.; Jennings, E. G.; Wyrick, J. J.; Lee, T. I.; Hengartner, C. J.; Green, M. R. et al. (1998): Dissecting the regulatory circuitry of a eukaryotic genome. In: *Cell* 95 (5), S. 717–728. DOI: 10.1016/s0092-8674(00)81641-4.

Hu, Dongli; Mayeda, Akila; Trembley, Janeen H.; Lahti, Jill M.; Kidd, Vincent J. (2003): CDK11 complexes promote pre-mRNA splicing. In: *The Journal of biological chemistry* 278 (10), S. 8623–8629. DOI: 10.1074/jbc.M210057200.

Hu, Shanhu; Marineau, Jason J.; Rajagopal, Nisha; Hamman, Kristin B.; Choi, Yoon Jong; Schmidt, Darby R. et al. (2019): Discovery and Characterization of SY-1365, a Selective, Covalent Inhibitor of CDK7. In: *Cancer research* 79 (13), S. 3479–3491. DOI: 10.1158/0008-5472.CAN-19-0119.

Iorns, Elizabeth; Turner, Nicholas C.; Elliott, Richard; Syed, Nelofer; Garrone, Ornella; Gasco, Milena et al. (2008): Identification of CDK10 as an important determinant of resistance to endocrine therapy for breast cancer. In: *Cancer cell* 13 (2), S. 91–104. DOI: 10.1016/j.ccr.2008.01.001.

Johnson, Shawn F.; Cruz, Cristina; Greifenberg, Ann Katrin; Dust, Sofia; Stover, Daniel G.; Chi, David et al. (2016): CDK12 Inhibition Reverses De Novo and Acquired PARP Inhibitor Resistance in BRCA Wild-Type and Mutated Models of Triple-Negative Breast Cancer. In: *Cell reports* 17 (9), S. 2367–2381. DOI: 10.1016/j.celrep.2016.10.077.

Kim, K. K.; Chamberlin, H. M.; Morgan, D. O.; Kim, S. H. (1996): Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase. In: *Nature structural biology* 3 (10), S. 849–855.

Knauf, U.; Newton, E. M.; Kyriakis, J.; Kingston, R. E. (1996): Repression of human heat shock factor 1 activity at control temperature by phosphorylation. In: *Genes & development* 10 (21), S. 2782–2793. DOI: 10.1101/gad.10.21.2782.

Ko, L. J.; Shieh, S. Y.; Chen, X.; Jayaraman, L.; Tamai, K.; Taya, Y. et al. (1997): p53 is phosphorylated by CDK7-cyclin H in a p36MAT1-dependent manner. In: *Molecular and cellular biology* 17 (12), S. 7220–7229.

Kokic, Goran; Chernev, Aleksandar; Tegunov, Dimitry; Dienemann, Christian; Urlaub, Henning; Cramer, Patrick (2019): Structural basis of TFIIH activation for nucleotide excision repair. In: *Nature communications* 10 (1), S. 2885. DOI: 10.1038/s41467-019-10745-5.

Kornberg, R. D.; Thomas, J. O. (1974): Chromatin structure; oligomers of the histones. In: *Science (New York, N.Y.)* 184 (4139), S. 865–868. DOI: 10.1126/science.184.4139.865.

Krajewska, Malgorzata; Dries, Ruben; Grassetti, Andrew V.; Dust, Sofia; Gao, Yang; Huang, Hao et al. (2019): CDK12 loss in cancer cells affects DNA damage response genes through premature cleavage and polyadenylation. In: *Nature communications* 10 (1), S. 1757. DOI: 10.1038/s41467-019-09703-y.

Krempler, A.; Kartarius, S.; Günther, J.; Montenarh, M. (2005): Cyclin H is targeted to the nucleus by C-terminal nuclear localization sequences. In: *Cellular and molecular life sciences : CMLS* 62 (12), S. 1379–1387. DOI: 10.1007/s00018-005-5023-5.

Kwiatkowski, Nicholas; Zhang, Tinghu; Rahl, Peter B.; Abraham, Brian J.; Reddy, Jessica; Ficarro, Scott B. et al. (2014): Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. In: *Nature* 511 (7511), S. 616–620. DOI: 10.1038/nature13393.

Larochelle, S.; Chen, J.; Knights, R.; Pandur, J.; Morcillo, P.; Erdjument-Bromage, H. et al. (2001): T-loop phosphorylation stabilizes the CDK7-cyclin H-MAT1 complex in vivo and regulates its CTD kinase activity. In: *The EMBO journal* 20 (14), S. 3749–3759. DOI: 10.1093/emboj/20.14.3749.

Larochelle, S.; Pandur, J.; Fisher, R. P.; Salz, H. K.; Suter, B. (1998): Cdk7 is essential for mitosis and for in vivo Cdk-activating kinase activity. In: *Genes & development* 12 (3), S. 370–381. DOI: 10.1101/gad.12.3.370.

Larochelle, Stéphane; Amat, Ramon; Glover-Cutter, Kira; Sansó, Miriam; Zhang, Chao; Allen, Jasmina J. et al. (2012a): Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. In: *Nature structural & molecular biology* 19 (11), S. 1108. DOI: 10.1038/nsmb.2399.

Larochelle, Stéphane; Amat, Ramon; Glover-Cutter, Kira; Sansó, Miriam; Zhang, Chao; Allen, Jasmina J. et al. (2012b): Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. In: *Nature structural & molecular biology* 19 (11), S. 1108–1115. DOI: 10.1038/nsmb.2399.

Larochelle, Stéphane; Batliner, Jasmin; Gamble, Matthew J.; Barboza, Nora M.; Kraybill, Brian C.; Blethrow, Justin D. et al. (2006): Dichotomous but stringent substrate selection by the dual-function Cdk7 complex revealed by chemical genetics. In: *Nature structural & molecular biology* 13 (1), S. 55–62. DOI: 10.1038/nsmb1028.

Larochelle, Stéphane; Merrick, Karl A.; Terret, Marie-Emilie; Wohlbold, Lara; Barboza, Nora M.; Zhang, Chao et al. (2007): Requirements for Cdk7 in the assembly of Cdk1/cyclin B and activation of Cdk2 revealed by chemical genetics in human cells. In: *Molecular cell* 25 (6), S. 839–850. DOI: 10.1016/j.molcel.2007.02.003.

Lee, Karen M.; Miklos, Ida; Du, Hongyan; Watt, Stephen; Szilagyi, Zsolt; Saiz, Julia E. et al. (2005): Impairment of the TFIIH-associated CDK-activating kinase selectively affects cell cycle-regulated gene expression in fission yeast. In: *Molecular biology of the cell* 16 (6), S. 2734–2745. DOI: 10.1091/mbc.e04-11-0982.

Li, S.; MacLachlan, T. K.; Luca, A. de; Claudio, P. P.; Condorelli, G.; Giordano, A. (1995): The cdc-2-related kinase, PISSLRE, is essential for cell growth and acts in G2 phase of the cell cycle. In: *Cancer research* 55 (18), S. 3992–3995.

Li, Xiaoming; Urwyler, Olivier; Suter, Beat (2010): Drosophila Xpd regulates Cdk7 localization, mitotic kinase activity, spindle dynamics, and chromosome segregation. In: *PLoS genetics* 6 (3), e1000876. DOI: 10.1371/journal.pgen.1000876.

Liu, Wen; Cai, Mei-Juan; Wang, Jin-Xing; Zhao, Xiao-Fan (2014): In a nongenomic action, steroid hormone 20-hydroxyecdysone induces phosphorylation of cyclin-dependent kinase 10 to promote gene transcription. In: *Endocrinology* 155 (5), S. 1738–1750. DOI: 10.1210/en.2013-2020.

Liu, Ying; Xiao, Zhibo; Yang, Daping; Ren, Lihong; Liu, Guofeng; Yang, Lin (2012): Effects of the cyclin-dependent kinase 10 (CDK10) on the tamoxifen sensitivity of keloid samples. In: *Molecules (Basel, Switzerland)* 17 (2), S. 1307–1318. DOI: 10.3390/molecules17021307.

Lodish, Harvey F. (2008): Molecular cell biology. 6th ed. New York: W.H. Freeman.

Lolli, Graziano (2009): Binding to DNA of the RNA-polymerase II C-terminal domain allows discrimination between Cdk7 and Cdk9 phosphorylation. In: *Nucleic acids research* 37 (4), S. 1260–1268. DOI: 10.1093/nar/gkn1061.

Lolli, Graziano; Johnson, Louise N. (2005): CAK-Cyclin-dependent Activating Kinase: a key kinase in cell cycle control and a target for drugs? In: *Cell cycle (Georgetown, Tex.)* 4 (4), S. 572–577.

Lolli, Graziano; Johnson, Louise N. (2007): Recognition of Cdk2 by Cdk7. In: *Proteins* 67 (4), S. 1048–1059. DOI: 10.1002/prot.21370.

Lolli, Graziano; Lowe, Edward D.; Brown, Nick R.; Johnson, Louise N. (2004): The crystal structure of human CDK7 and its protein recognition properties. In: *Structure (London, England : 1993)* 12 (11), S. 2067–2079. DOI: 10.1016/j.str.2004.08.013.

Long, J. J.; Leresche, A.; Kriwacki, R. W.; Gottesfeld, J. M. (1998): Repression of TFIIH transcriptional activity and TFIIH-associated cdk7 kinase activity at mitosis. In: *Molecular and cellular biology* 18 (3), S. 1467–1476. DOI: 10.1128/mcb.18.3.1467.

Lu, H.; Fisher, R. P.; Bailey, P.; Levine, A. J. (1997): The CDK7-cycH-p36 complex of transcription factor IIH phosphorylates p53, enhancing its sequence-specific DNA binding activity in vitro. In: *Molecular and cellular biology* 17 (10), S. 5923–5934. DOI: 10.1128/mcb.17.10.5923.

Malumbres, Marcos (2014): Cyclin-dependent kinases. In: *Genome Biol* 15 (6), S. 122. DOI: 10.1186/gb4184.

Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. (2002): The protein kinase complement of the human genome. In: *Science (New York, N.Y.)* 298 (5600), S. 1912–1934. DOI: 10.1126/science.1075762.

Martinez, A. M.; Afshar, M.; Martin, F.; Cavadore, J. C.; Labbé, J. C.; Dorée, M. (1997): Dual phosphorylation of the T-loop in cdk7: its role in controlling cyclin H binding and CAK activity. In: *The EMBO journal* 16 (2), S. 343–354. DOI: 10.1093/emboj/16.2.343.

Merrick, Karl A.; Larochelle, Stéphane; Zhang, Chao; Allen, Jasmina J.; Shokat, Kevan M.; Fisher, Robert P. (2008): Distinct activation pathways confer cyclin-binding specificity on Cdk1 and Cdk2 in human cells. In: *Molecular cell* 32 (5), S. 662–672. DOI: 10.1016/j.molcel.2008.10.022.

Mi, Huaiyu; Muruganujan, Anushya; Huang, Xiaosong; Ebert, Dustin; Mills, Caitlin; Guo, Xinyu; Thomas, Paul D. (2019): Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). In: *Proceedings of the National Academy of Sciences* 14 (3), S. 703–721. DOI: 10.1038/s41596-019-0128-8.

Murray, Brion W.; Guo, Chuangxing; Piraino, Joseph; Westwick, John K.; Zhang, Cathy; Lamerdin, Jane et al. (2010): Small-molecule p21-activated kinase inhibitor PF-3758309 is a potent inhibitor of oncogenic signaling and tumor growth. In: *Proceedings of the National Academy of Sciences of the United States of America* 107 (20), S. 9446–9451. DOI: 10.1073/pnas.0911863107.

Nilson, Kyle A.; Guo, Jiannan; Turek, Michael E.; Brogie, John E.; Delaney, Elizabeth; Luse, Donal S.; Price, David H. (2015): THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. In: *Molecular cell* 59 (4), S. 576–587. DOI: 10.1016/j.molcel.2015.06.032.

Olson, Calla M.; Jiang, Baishan; Erb, Michael A.; Liang, Yanke; Doctor, Zainab M.; Zhang, Zinan et al. (2018): Pharmacological perturbation of CDK9 using selective CDK9 inhibition or degradation. In: *Nature chemical biology* 14 (2), S. 163–170. DOI: 10.1038/nchembio.2538.

Olson, Calla M.; Liang, Yanke; Leggett, Alan; Park, Woojun D.; Li, Lianbo; Mills, Caitlin E. et al. (2019): Development of a Selective CDK7 Covalent Inhibitor Reveals Predominant Cell-Cycle Phenotype. In: *Cell chemical biology* 26 (6), 792-803.e10. DOI: 10.1016/j.chembiol.2019.02.012.

Poon, R. Y.; Yamashita, K.; Adamczewski, J. P.; Hunt, T.; Shuttleworth, J. (1993): The cdc2-related protein p40MO15 is the catalytic subunit of a protein kinase that can activate p33cdk2 and p34cdc2. In: *The EMBO journal* 12 (8), S. 3123–3132.

Rochette-Egly, C.; Adam, S.; Rossignol, M.; Egly, J. M.; Chambon, P. (1997): Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7. In: *Cell* 90 (1), S. 97–107. DOI: 10.1016/s0092-8674(00)80317-7.

Rodgers, Griffin; Austin, Christopher; Anderson, James; Pawlyk, Aaron; Colvis, Christine; Margolis, Ronald; Baker, Jenna (2018): Glimmers in illuminating the druggable genome. In: *Nature reviews. Drug discovery* 17 (5), S. 301–302. DOI: 10.1038/nrd.2017.252.

Rossignol, M.; Kolb-Cheynel, I.; Egly, J. M. (1997): Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH. In: *The EMBO journal* 16 (7), S. 1628–1637. DOI: 10.1093/emboj/16.7.1628.

Russo, A. A.; Jeffrey, P. D.; Patten, A. K.; Massagué, J.; Pavletich, N. P. (1996a): Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. In: *Nature* 382 (6589), S. 325–331. DOI: 10.1038/382325a0.

Russo, A. A.; Jeffrey, P. D.; Pavletich, N. P. (1996b): Structural basis of cyclin-dependent kinase activation by phosphorylation. In: *Nature structural biology* 3 (8), S. 696–700. DOI: 10.1038/nsb0896-696.

Rustici, Gabriella; Mata, Juan; Kivinen, Katja; Lió, Pietro; Penkett, Christopher J.; Burns, Gavin et al. (2004): Periodic gene expression program of the fission yeast cell cycle. In: *Nature genetics* 36 (8), S. 809–817. DOI: 10.1038/ng1377.

Sainsbury, Sarah; Bernecky, Carrie; Cramer, Patrick (2015): Structural basis of transcription initiation by RNA polymerase II. In: *Nature reviews. Molecular cell biology* 16 (3), S. 129–143. DOI: 10.1038/nrm3952.

Sansó, Miriam; Levin, Rebecca S.; Lipp, Jesse J.; Wang, Vivien Ya-Fan; Greifenberg, Ann Katrin; Quezada, Elizabeth M. et al. (2016): P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. In: *Genes & development* 30 (1), S. 117–131. DOI: 10.1101/gad.269589.115.

Schachter, Miriam Merzel; Fisher, Robert P. (2013): The CDK-activating kinase Cdk7: taking yes for an answer. In: *Cell cycle (Georgetown, Tex.)* 12 (20), S. 3239–3240. DOI: 10.4161/cc.26355.

Schachter, Miriam Merzel; Merrick, Karl A.; Larochelle, Stéphane; Hirschi, Alexander; Zhang, Chao; Shokat, Kevan M. et al. (2013): A Cdk7-Cdk4 T-loop phosphorylation cascade promotes G1 progression. In: *Molecular cell* 50 (2), S. 250–260. DOI: 10.1016/j.molcel.2013.04.003.

Schilbach, S.; Hantsche, M.; Tegunov, D.; Dienemann, C.; Wigge, C.; Urlaub, H.; Cramer, P. (2017): Structures of transcription pre-initiation complex with TFIIH and Mediator. In: *Nature* 551 (7679), S. 204–209. DOI: 10.1038/nature24282.

Schneider, Eberhard; Kartarius, Sabine; Schuster, Norbert; Montenarh, Mathias (2002): The cyclin H/cdk7/Mat1 kinase activity is regulated by CK2 phosphorylation of cyclin H. In: *Oncogene* 21 (33), S. 5031–5037. DOI: 10.1038/sj.onc.1205690.

Schwer, Beate; Shuman, Stewart (2011): Deciphering the RNA polymerase II CTD code in fission yeast. In: *Molecular cell* 43 (2), S. 311–318. DOI: 10.1016/j.molcel.2011.05.024.

Serizawa, H.; Mäkelä, T. P.; Conaway, J. W.; Conaway, R. C.; Weinberg, R. A.; Young, R. A. (1995): Association of Cdk-activating kinase subunits with transcription factor TFIIH. In: *Nature* 374 (6519), S. 280–282. DOI: 10.1038/374280a0.

Shiekhattar, R.; Mermelstein, F.; Fisher, R. P.; Drapkin, R.; Dynlacht, B.; Wessling, H. C. et al. (1995): Cdk-activating kinase complex is a component of human transcription factor TFIIH. In: *Nature* 374 (6519), S. 283–287. DOI: 10.1038/374283a0.

Shuttleworth, J.; Godfrey, R.; Colman, A. (1990): p40MO15, a cdc2-related protein kinase involved in negative regulation of meiotic maturation of Xenopus oocytes. In: *The EMBO journal* 9 (10), S. 3233–3240.

Solomon, M. J.; Harper, J. W.; Shuttleworth, J. (1993): CAK, the p34cdc2 activating kinase, contains a protein identical or closely related to p40MO15. In: *The EMBO journal* 12 (8), S. 3133–3142.

Stettler, Karin; Li, Xiaoming; Sandrock, Björn; Braga-Lagache, Sophie; Heller, Manfred; Dümbgen, Lutz; Suter, Beat (2015): A Drosophila XPD model links cell cycle coordination with neuro-development and suggests links to cancer. In: *Disease models & mechanisms* 8 (1), S. 81–91. DOI: 10.1242/dmm.016907.

Taipale, Mikko; Krykbaeva, Irina; Koeva, Martina; Kayatekin, Can; Westover, Kenneth D.; Karras, Georgios I.; Lindquist, Susan (2012): Quantitative Analysis of Hsp90-Client Interactions Reveals Principles of Substrate Recognition. In: *Cell* 150 (5), S. 987–1001. DOI: 10.1016/j.cell.2012.06.047.

Tassan, J. P.; Jaquenoud, M.; Fry, A. M.; Frutiger, S.; Hughes, G. J.; Nigg, E. A. (1995): In vitro assembly of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein. In: *The EMBO journal* 14 (22), S. 5608–5617.

Taylor, W. R.; Stark, G. R. (2001): Regulation of the G2/M transition by p53. In: *Oncogene* 20 (15), S. 1803–1815. DOI: 10.1038/sj.onc.1204252.

Trembley, Janeen H.; Hu, Dongli; Hsu, Li-Chung; Yeung, Cho-Yau; Slaughter, Clive; Lahti, Jill M.; Kidd, Vincent J. (2002): PITSLRE p110 protein kinases associate with transcription complexes and affect their activity. In: *The Journal of biological chemistry* 277 (4), S. 2589–2596. DOI: 10.1074/jbc.M109755200.

Umaña, Angie C.; Iwahori, Satoko; Kalejta, Robert F. (2018): Direct Substrate Identification with an Analog Sensitive (AS) Viral Cyclin-Dependent Kinase (v-Cdk). In: *ACS chemical biology* 13 (1), S. 189–199. DOI: 10.1021/acschembio.7b00972.

Unger, Sheila; Böhm, Detlef; Kaiser, Frank J.; Kaulfuss, Silke; Borozdin, Wiktor; Buiting, Karin et al. (2008): Mutations in the cyclin family member FAM58A cause an X-linked dominant disorder characterized by syndactyly, telecanthus and anogenital and renal malformations. In: *Nature genetics* 40 (3), S. 287–289. DOI: 10.1038/ng.86.

Vaughn, J. L.; Goodwin, R. H.; Tompkins, G. J.; McCawley, P. (1977): The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae). In: *In vitro* 13 (4), S. 213–217. DOI: 10.1007/bf02615077.

Vos, Seychelle M.; Farnung, Lucas; Boehning, Marc; Wigge, Christoph; Linden, Andreas; Urlaub, Henning; Cramer, Patrick (2018): Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. In: *Nature* 560 (7720), S. 607–612. DOI: 10.1038/s41586-018-0440-4.

WATSON, J. D.; CRICK, F. H. (1953): Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. In: *Nature* 171 (4356), S. 737–738. DOI: 10.1038/171737a0.

Weiswald, Louis-Bastien; Hasan, Mohammad R.; Wong, John C. T.; Pasiliao, Clarissa C.; Rahman, Mahbuba; Ren, Jianhua et al. (2017): Inactivation of the Kinase Domain of CDK10 Prevents Tumor Growth in a Preclinical Model of Colorectal Cancer, and Is Accompanied by Downregulation of Bcl-2. In: *Molecular cancer therapeutics* 16 (10), S. 2292–2303. DOI: 10.1158/1535-7163.MCT-16-0666.

West, M. L.; Corden, J. L. (1995): Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. In: *Genetics* 140 (4), S. 1223–1233.

Windpassinger, Christian; Piard, Juliette; Bonnard, Carine; Alfadhel, Majid; Lim, Shuhui; Bisteau, Xavier et al. (2017): CDK10 Mutations in Humans and Mice Cause Severe Growth Retardation, Spine Malformations, and Developmental Delays. In: *American journal of human genetics* 101 (3), S. 391–403. DOI: 10.1016/j.ajhg.2017.08.003.

Wood, Daniel J.; Endicott, Jane A. (2018): Structural insights into the functional diversity of the CDK-cyclin family. In: *Open biology* 8 (9). DOI: 10.1098/rsob.180112.

Yankulov, K. Y.; Bentley, D. L. (1997): Regulation of CDK7 substrate specificity by MAT1 and TFIIH. In: *The EMBO journal* 16 (7), S. 1638–1646. DOI: 10.1093/emboj/16.7.1638.

Yee, A.; Nichols, M. A.; Wu, L.; Hall, F. L.; Kobayashi, R.; Xiong, Y. (1995): Molecular cloning of CDK7-associated human MAT1, a cyclin-dependent kinase-activating kinase (CAK) assembly factor. In: *Cancer research* 55 (24), S. 6058–6062.

Yeh, Chi-Wei; Kao, Shoa-Hsuan; Cheng, Yi-Chuan; Hsu, Li-Sung (2013): Knockdown of cyclin-dependent kinase 10 (cdk10) gene impairs neural progenitor survival via modulation of raf1a gene expression. In: *The Journal of biological chemistry* 288 (39), S. 27927–27939. DOI: 10.1074/jbc.M112.420265.

Zeng, Mei; Kwiatkowski, Nicholas P.; Zhang, Tinghu; Nabet, Behnam; Xu, Mousheng; Liang, Yanke et al. (2018): Targeting MYC dependency in ovarian cancer through inhibition of CDK7 and CDK12/13. In: *eLife* 7. DOI: 10.7554/eLife.39030.

Zhang, David W.; Rodríguez-Molina, Juan B.; Tietjen, Joshua R.; Nemec, Corey M.; Ansari, Aseem Z. (2012): Emerging Views on the CTD Code. In: *Genetics research international* 2012, S. 347214. DOI: 10.1155/2012/347214.

Zhang, Tinghu; Kwiatkowski, Nicholas; Olson, Calla M.; Dixon-Clarke, Sarah E.; Abraham, Brian J.; Greifenberg, Ann K. et al. (2016): Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. In: *Nature chemical biology* 12 (10), S. 876–884. DOI: 10.1038/nchembio.2166.

8 Appendix

Proteins identified in the chemical genetic screen which were enriched at least two-fold over control or exclusively identified in the AS sample.

Table S1 Putative Cdk10/CycM substrates identified in the chemical genetic screen

full cell lysate - 60 min

log intensity ratio	gene name	uniprot identifier	aa	pos	Loc. Prob.
NaN	BABAM1	M0QY17	S	29	0,99
NaN	TARS	P26639	S	8	0,96
NaN	DUT	P33316-2	S	11	0,75
NaN	YBX1	P67809	S	174	0,75
NaN	GPRC5A	Q8NFJ5	S	301	0,99
NaN	LMNA	P02545-2	T	19	0,52
NaN	RPL5	P46777	T	56	1,00
NaN	RRP12	Q5JTH9-2	T	570	1,00
NaN	SRRT	Q9BXP5-5	T	507	0,80
NaN	OSBPL8	Q9BZF1-3	T	102	0,51
NaN		Q06323	Y	227	0,80
3,50	UBAP2L	Q14157-1	S	467	0,64
3,28	RPL14	P50914	S	139	1,00
2,45		M0QZK8	S	4	1,00
2,44	PALLD	Q8WX93-3	S	511	1,00
2,28	HIDE1	A8MVS5	S	75	0,91
1,81	EEF1D	E9PK01	S	118	0,61
1,44	SSB	P05455	S	366	1,00
1,41	ARL14	Q8N4G2	S	3	0,86
1,38	DNAH11	U3KQJ8	T	4465	1,00
1,34	OPTN	Q96CV9-3	T	468	0,56
1,13	ATP8A2	F8VRS1	T	915	0,94
1,11	PI4K2A	Q9BTU6	S	47	0,94
1,11	PI4K2A	Q9BTU6	S	51	1,00
1,04	AHNAK	Q09666	S	5110	1,00
1,03	EIF4B	E7EX17	S	406	1,00
1,03	YBX1	P67809	S	165	0,81
1,02	HSF1	Q00613	S	363	1,00
0,92	Trim28	Q13263	S	19	1,00

Nuclear extracts 1st dataset - 30 min

log intensity ratio	gene name	uniprot identifier	aa	pos	Loc. Prob.
NaN	CEP63	H0YAE6	S	236	0,50
NaN	NPM1	P06748-2	S	70	0,96
NaN	APC	P25054-2	S	1963	1,00
NaN	MEPCE	Q7L2J0	S	69	0,98
NaN	CENPH	Q9H3R5	S	109	0,50
NaN	SRRM2	Q9UQ35	S	2132	1,00
NaN	LIMK2	B5MC51	T	132	0,98
NaN	CEP63	H0YAE6	T	234	0,50
NaN	CENPH	Q9H3R5	T	110	0,50
NaN	ACAD11	A0A0J9YXS1	S	217	0,71
NaN	ACAD11	A0A0J9YXS1	T	220	0,71
NaN	MAP3K5	Q99683	T	652	1,00
NaN	MAP3K5	Q99683	T	654	1,00

Nuclear extracts 2nd dataset - 30 min

log intensity ratio	gene name	uniprot identifier	aa	pos	Loc. Prob.
NaN	ACIN1	S4R3H4	S	652	0,96
NaN	BABAM1	M0QXG9	S	29	0,99
NaN	GLIPR1	P48060	S	219	0,55
NaN	HDGF	P51858	S	165	1,00
NaN	MAP2K2	M0R1B6	T	98	0,83
NaN	NOLC1	A0A0A0MRM9	S	707	1,00
NaN	NPM1	P06748	S	70	0,99
NaN	SERBP1	Q8NC51-4	S	309	0,95
NaN	SRRM2	Q9UQ35	S	2398	0,95
NaN	SRRM2	Q9UQ35	S	297	0,77
NaN	SRRM2	Q9UQ35	T	289	0,59
NaN	SSB	P05455	S	366	1,00
NaN	TCOF1	J3KQ96	S	1340	1,00
NaN	TCRBV16S1A1N1	A0A5B0	S	7	0,50
NaN	TCRBV16S1A1N1	A0A5B0	S	10	0,50
NaN	U2AF2	P26368-2	S	79	1,00

Nuclear extracts 3rd dataset - 60 min

log intensity ratio	gene name	uniprot identifier	aa	pos	Loc. Prob.
NaN	KHSRP	A0A087WTP3	S	274	0,97
NaN	PHF6	Q8IWS0-5	S	165	0,91
NaN	MATR3	A0A0R4J2E8	S	188	0,99
NaN	TCOF1	J3KQ96	S	381	1,00
NaN	TCOF1	J3KQ96	S	156	0,96
NaN	EIF4B	E7EX17	S	409	0,53
NaN	LMNA	P02545	S	636	1,00
NaN	TRIM28	Q13263-2	S	391	0,99
NaN	PHF6	Q8IWS0-5	S	120	0,60
NaN	ANKRD32	Q9BQI6	S	919	0,56
NaN	ARGLU1	Q9NWB6	S	77	0,96
NaN	SRRM2	Q9UQ35	S	2426	0,97
NaN	SRRM2	Q9UQ35	S	2581	0,93
NaN	THRAP3	Q9Y2W1	S	406	0,70
NaN	ZNF638	A0A096LPB4	T	4	0,55
NaN	ZC3HC1	C9J0I9	T	309	0,59
NaN	CKM	P06732	T	166	1,00
NaN	CDK10	Q15131	T	196	1,00
NaN	NCOA7	Q8NI08-7	T	413	0,81
NaN	LAMA4	A0A0A0MTC7	Y	887	1,00
NaN	KCNG1	Q9UIX4	Y	352	1,00
NaN	LRRTM4	Q4KMX1	S	196	1,00
NaN	LRRTM4	Q4KMX1	S	198	1,00
NaN	TRA2B	H7BXF3	S	103	0,99
NaN	TRA2B	H7BXF3	S	105	0,86
NaN	EIF3D	O15371-3	S	374	1,00
NaN	NEURL3	A0A087WVI2	T	65	1,00
NaN	NEURL3	A0A087WVI2	T	68	1,00
NaN	ANKRD17	H0YM23	T	105	1,00
NaN	EIF3D	O15371-3	T	367	1,00
NaN	CCDC96	Q2M329	T	479	1,00
NaN	CCDC96	Q2M329	T	481	1,00
1,42	HSPB1	P04792	S	82	0,84
1,27	AHNAK	Q09666	S	5841	0,99
1,13	EIF4G1	E7EX73	S	1433	1,00
0.94	U2AF2	P26368-2	S	79	1,00

9 Acknowledgement

Bedanken möchte ich mich bei Prof. Dr. Matthias Geyer für das Thema und die Betreuung der Arbeit. Insbesondere jedoch für die Möglichkeit, meine Arbeitszeiten flexibel zu gestalten. Ohne diese Flexibilität wäre diese Arbeit nicht in dieser Form zustande gekommen. Bedanken möchte ich mich ebenfalls bei den weiteren Mitgliedern meiner Prüfungskommission, Prof. Dr. Oliver Gruss, Prof. Dr. Michael Hölzel, sowie Prof. Dr. Jan Hasenauer.

Zudem gilt mein Dank Prof. Dr. Henning Urlaub für die Kollaboration bezüglich der Massenspektrometrie und die freundliche Aufnahme in seine Arbeitsgruppe für eine Woche der Probenvorbereitung

Weiterhin möchte ich mich bei allen aktuellen und früheren Mitgliedern der AG Geyer für die Diskussion der Ergebnisse, sowie eine gute Arbeitsatmosphäre bedanken. Besonderer Dank gilt Melanie Specht für das Klonieren u. a. zahlreicher Cyclin H C-termini, sowie Dr. Kanchan Anand für ihre unermüdlichen Versuche die Kinasen zu kristallisieren.