Resource

Molecular Cell

Heptad-Specific Phosphorylation of RNA Polymerase II CTD

Highlights

- CTD variants make the entire CTD sequence accessible to mass spectrometry analysis
- Valid CTD phosphosite mapping is based on a synthetic CTD peptide library
- Predominant phosphorylation signatures prevail within neighboring CTD residues
- Inhibition of CDK9 strongly reduces Ser2-P phosphorylation in the CTD

Authors

Roland Schüller, Ignasi Forné, Tobias Straub, ..., Patrick Cramer, Axel Imhof, Dirk Eick

Correspondence

eick@helmholtz-muenchen.de

In Brief

A combined genetic and mass spectrometric approach offers comprehensive insights into CTD phosphorylation patterns and changes in CTD heptad-specific phosphorylation states in mammals and yeast.

Accession Numbers

PXD003159

Schüller et al., 2016, Molecular Cell *61*, 305–314
January 21, 2016
http://dx.doi.org/10.1016/j.molcel.2015.12.003

CellPress

Molecular Cell
Resource

Heptad-Specific Phosphorylation of RNA Polymerase II CTD

Roland Schüller,[1,5] Ignasi Forné,[2,5] Tobias Straub,[2,5] Amelie Schreieck,[3] Yves Texier,[1] Nilay Shah,[1] Tim-Michael Decker,[1] Patrick Cramer,[4] Axel Imhof,[2] and Dirk Eick[1,*]
[1]Department of Molecular Epigenetics, Helmholtz Center Munich and Center for Integrated Protein Science Munich (CIPSM), Marchioninistrasse 25, 81377 Munich, Germany
[2]Biomedical Center Munich, ZFP and Bioinformatic Unit, Großhaderner Strasse 9, 82152 Planegg-Martinsried, Germany
[3]Gene Center, Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 Munich, Germany
[4]Department of Molecular Biology, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany
[5]Co-first author
*Correspondence: eick@helmholtz-muenchen.de
http://dx.doi.org/10.1016/j.molcel.2015.12.003

SUMMARY

The carboxy-terminal domain (CTD) of RNA polymerase II (Pol II) consists of heptad repeats with the consensus motif Y_1-S_2-P_3-T_4-S_5-P_6-S_7. Dynamic phosphorylation of the CTD coordinates Pol II progression through the transcription cycle. Here, we use genetic and mass spectrometric approaches to directly detect and map phosphosites along the entire CTD. We confirm phosphorylation of CTD residues Y_1, S_2, T_4, S_5, and S_7 in mammalian and yeast cells. Although specific phosphorylation signatures dominate, adjacent CTD repeats can be differently phosphorylated, leading to a high variation of coexisting phosphosites in mono- and di-heptad CTD repeats. Inhibition of CDK9 kinase specifically reduces S_2 phosphorylation levels within the CTD.

INTRODUCTION

The carboxy-terminal domain (CTD) of the largest subunit (Rpb1) of RNA polymerase II (Pol II) coordinates the transcription cycle and gene expression (Bentley, 2014; Buratowski, 2009; Corden, 2013; Egloff et al., 2012; Eick and Geyer, 2013). The CTD consists of tandem heptad repeats with the consensus sequence tyrosine-serine-proline-threonine-serine-proline-serine ($Y_1S_2P_3T_4S_5P_6S_7$) and forms a flexible, tail-like extension from the catalytic core of Pol II with regions that are proximal and distal to the Pol II surface (Cramer et al., 2001). The CTD interacts with a wide range of nuclear factors based on its dynamic plasticity and diversity of binding surfaces (Jasnovidova and Stefl, 2013; Meinhart et al., 2005; Phatnani and Greenleaf, 2006) generated by differential phosphorylation at multiple sites, a phenomenon that was referred to as the "CTD code" (Buratowski, 2003, 2009).

Monoclonal antibodies have been used to study in vivo the potentially phosphorylated CTD amino acids (Y_1, S_2, T_4, S_5, and S_7) (Chapman et al., 2007; Descostes et al., 2014; Egloff et al., 2007; Hintermair et al., 2012; Hsin et al., 2011; Lu et al., 1992; Mayer et al., 2012; Zhang and Corden, 1991). However, the epitopes detected by antibodies can be masked by proteins or modifications at neighboring sites. Therefore, the effectiveness of antibodies in western blot or ChIP analysis reflects the number of accessible CTD phosphorylation marks, but not the total number of phosphorylations. Most importantly, CTD phospho-specific antibodies do not provide any heptad (location)-specific information of CTD phosphorylation. Due to these limitations, the principles and patterns of CTD phosphorylation remained elusive. Are all 52 CTD repeats phosphorylated in vivo? What is the frequency of mono- and double-phosphorylated heptad repeats? Do specific double-phospho signatures prevail within neighboring CTD residues? Do mammals and yeast show differences in signatures? What does the phosphorylation profile of all 52 CTD repeats look like? Do CTD kinases introduce phosphosite-specific changes in CTD?

Within this study, we developed a reliable protocol for mapping heptad-specific phosphosites within the entire CTD of Pol II in yeast and mammals, combining genetic manipulation of the CTD with mass spectrometry (MS). This work provides insights into the abundance and spatial patterns found among the five different CTD phosphoresidues in vivo.

RESULTS

Design, Stable Expression, and Purification of CTD Variants for Subsequent MS Analysis

Due to its amino acid sequence, the bottom-up MS analysis of the CTD was so far hampered by the difficulty to digest the CTD into amenable peptides. The proximal part of the mammalian CTD contains only heptads with the canonical sequence, thus lacking the basic residues lysine (K) and arginine (R) that are required for trypsin cleavage. In contrast, K and R residues can be found in the less conserved, distal part of mammalian CTD (Figure S1A), rendering these regions accessible to trypsin cleavage and MS analysis (Figures 1A and 1B).

To generate tryptic peptides covering the complete CTD, we established cell lines, which express CTD variants of Pol II containing additional K and R residues at position 7 of multiple

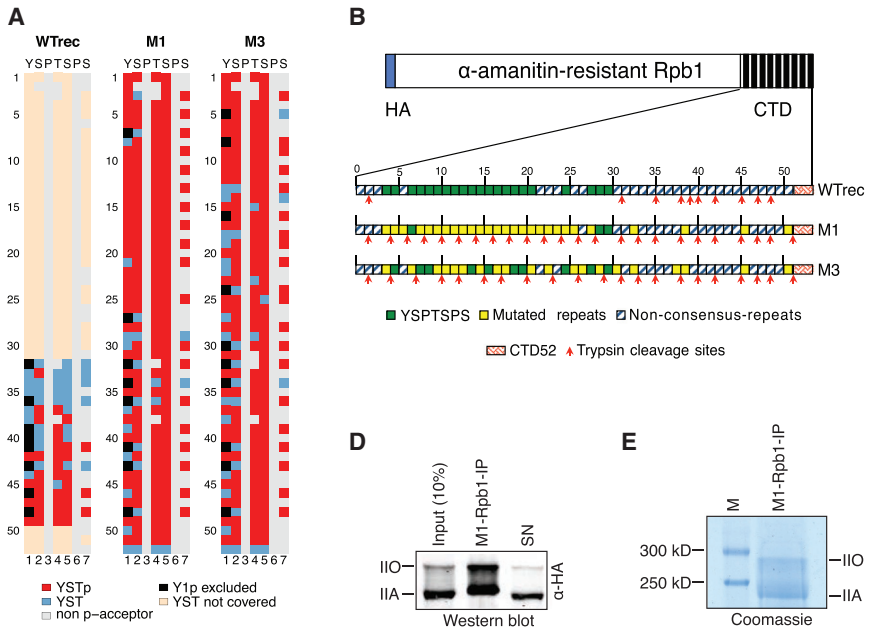


Figure 1. Newly Designed CTD Variants Lead to 100% CTD Sequence Coverage by MS Analysis

(A) Scheme of the mapped phosphosites within the CTD of WTrec, CTD variants M1 and M3. CTD residues are shown by squares. Red: identified phosphoresidue; blue: residue covered in the analysis; gray: non-p-acceptor residue; black: excluded phosphoresidue due to potential Y_1P mis-assignment; and yellow: residue not covered in analysis.

(B) Schematic overview of all 52 repeats of WTrec, CTD variants M1 and M3 including consensus and non-consensus repeats, as well as mutated repeats with newly introduced trypsin cleavage sites.

(C) (Left) CTD sequence of variant M1. Red letters: amino acid substitution or addition. Blue letters: non-consensus residues. (Right) List of tryptic CTD peptides of variant M1. Numbers indicate CTD repeats covered by the corresponding peptide.

(D) Western blot analysis of purified Rpb1-HA of CTD variant M1 using CTD phospho-specific antibodies.

(E) Coomassie-stained SDS-PAGE of purified Rpb1-M1-HA. M, marker.

See also Figure S1.

C

CTD variant M1:

1	YSPTSPA	18	YSPTSPK	35	YSPTSPK
2	YEPRSPGG	19	VVYSPTSPS	36	YTPTSPS
3	YTPQSPS	20	YSPTSPK	37	YSPSSPE
4	YSPTSPR	21	LLYSPTSPS	38	YTPASPK
5	AYSPTSPS	22	YSPTSPK	39	AALYSPTSPS
6	YSPTSPK	23	LYSPTSPN	40	YSPTSPK
7	YSPTSPS	24	YTPTSPK	41	YSPTSPT
8	YSPTSPK	25	LVYSPTSPS	42	YSPTTPK
9	VYSPTSPS	26	YSPTSPK	43	YSPTSPT
10	YSPTSPK	27	YTPTSPN	44	YSPTSPV
11	LYSPTSPS	28	YSPTSPK	45	YTPTSPK
12	YSPTSPK	29	YSPTSPS	46	AYSPTSPT
13	AAYSPTSPS	30	YSPTSPS	47	YSPTSPK
14	YSPTSPK	31	YSPSSPR	48	YSPTSPT
15	AVYSPTSPS	32	YTPQSPT	49	YSPTSPKGST
16	YSPTSPK	33	YTPSSPK	50	YSPTSPG
17	ALYSPTSPS	34	YSPSSPS	51	YSPTSPR
				52	YSLTSPAISPDDSDEEN

Trypsin

Predicted peptides:

1-2	YSPTSPAYEPR
2-4	SPGGYTPQSPSYSPTSPR
5-6	AYSPTSPSYSPTSPK
7-8	YSPTSPSYSPTSPK
9-10	VYSPTSPSYSPTSPK
11-12	LYSPTSPSYSPTSPK
13-14	AAYSPTSPSYSPTSPK
15-16	AVYSPTSPSYSPTSPK
17-18	ALYSPTSPSYSPTSPK
19-20	VVYSPTSPSYSPTSPK
21-22	LLYSPTSPSYSPTSPK
23-24	LYSPTSPNYTPTSPK
25-26	LVYSPTSPSYSPTSPK
27-28	YTPTSPNYSPTSPK
29-31	YSPTSPSYSPTSPSYSPSSPR
32-33	YTPQSPTYTPSSPK
34-35	YSPSSPSYSPTSPK
36-38	YTPTSPSYSPSSPEYTPASPK
39-40	AALYSPTSPSYSPTSPK
41-42	YSPTSPTYSPTTPK
43-45	YSPTSPTYSPTSPVYTPTSPK
46-47	AYSPTSPTYSPTSPK
48-49	YSPTSPTYSPTSPK
50-51	GSTYSPTSPGYSPTSPR
52	YSLTSPAISPDDSDEEN

heptad repeats in the proximal part of the mammalian CTD (Figure 1B). To distinguish the detected peptides in the MS, it is essential that all peptides have a unique amino acid composition. Consequently, additional amino acids next to the substituted K and R residues were strategically inserted. In total, five CTD variants were designed that differed in the resulting tryptic peptides (Figures 1C and S1B). CTD variant M1 for example, carries 15 lysine or arginine substitutions and several alanine, leucine, and valine additions (Figure 1C). The tryptic cleavage of CTD variant M1 resulted in 25 peptides with unique amino acid composition (Figure 1C). To reduce the risk of designing non-functional CTDs, the minimal distance between two mutations had to be two heptad repeats (Liu et al., 2008; Schwer et al., 2012) (Figures 1C and S1B).

CTD sequences were cloned into the expression vector LS*mock (Meininghaus et al., 2000), which carries the hemagglutinin (HA)-tagged mouse Rpb1 gene. A point mutation (N793D) in Rpb1 confers resistance to α-amanitin and allows the chemical "knockout" and functional replacement of endogenous Pol II (Bartolomei and Corden, 1987). Stably transfected Raji cells expressed recombinant Rpb1 (Figure S1C) and proliferated in the presence of α-amanitin treatment (Figure S1D). Although the CTD variant cell lines showed slightly reduced cell growth compared to wild-type (WT) Raji cells, their cell viability remained constantly high (Figure S1D). Western blot analysis showed that the recombinant wild-type Pol II CTD (WTrec), Pol II CTD variants M1 and M3, and endogenous Pol II CTD (WT) all showed similar CTD phosphorylation levels of Y_1, S_2, S_5, and S_7 (Figure S1E). In contrast, T_4-P levels were reduced in both CTD variants suggesting that our changes in the CTD sequence either led to a masking of the T_4-P-specific antibody epitope or interfered with the recruitment of T_4-specific kinases or phosphatases. Additionally, the expression of the recombinant Rpb1 protein resulted in a higher abundance of the Pol IIA form, whereas the actively transcribing Pol IIO form was detected in similar levels when comparing the endogenous (WT)—and recombinant (WTrec, M1 and M3)—expression system (Figure S1E).

For MS analysis, Rpb1 was purified by immunoprecipitation (IP) using a combination of two phospho-CTD-specific antibodies (α-S_2P (3E10) and α-S_5P (3E8)) or HA-tag-specific antibodies. The CTD phospho-specific antibodies purified the hyperphosphorylated Pol IIO form to an extent of 80%–90%

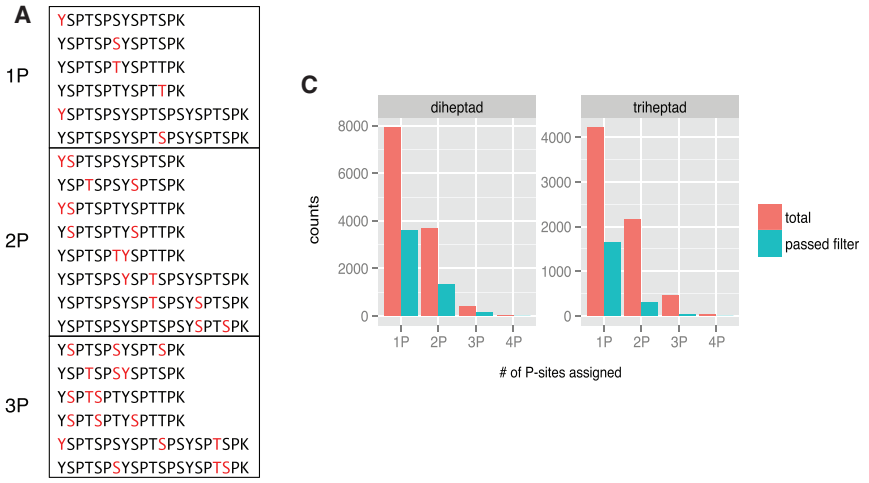


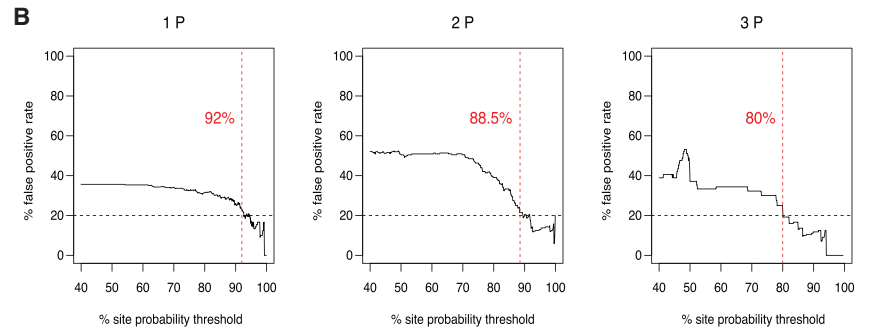
Figure 2. Valid Phosphosite Assignment by Determining Site Probability Thresholds Using Synthetic CTD Peptide Library

(A) List of synthesized CTD peptides (JPT—Innovative Peptide Solutions). 20 mono-, double- or triple-phosphorylated (1P, 2P and 3P) peptides were made. Corresponding phospho-residues are shown in red.

(B) 1P- (left), 2P- (middle) and 3P- (right) synthetic peptides were used to determine the corresponding site probability threshold (92%, 88.5% and 80%, respectively) at a false localization rate (FLR) of 20%. The assigned thresholds set the filter for valid phosphosite-determination through the Proteome Discoverer plugin PhosphoRS.

(C) Total counts of detected 1P-, 2P-, 3P- and 4P-di-heptad-long (left) and -tri-heptad-long (right) peptides with or without a phosphosite filter.

See also Figure S2.



(Figure 1D). The detection of the two main Rpb1 species (IIO- and IIA-form) using Coomassie-stained SDS-PAGE served as a quality control for efficient purification of recombinant Rpb1 (Figure 1E). The two different Pol II forms were separately excised and subjected to in-gel cleavage, followed by TiO_2-phospho-enrichment and liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) (Experimental Procedures).

Valid Phosphosite Assignments Using Site-Specific Probability Thresholds Based on a Synthetic CTD Peptide Library

The phosphosites in CTD peptides were mapped based on MS data using PhosphoRS 2.0 (Thermo Scientific) (Taus et al., 2011), which calculates the localization probabilities for all potential phosphorylation sites. The false localization rate (FLR) (Marx et al., 2013) was determined using 20 synthetic CTD peptides, which were either mono-, di-, or tri-phosphorylated (Figures 2A and 2B). We set an FLR of 20%, which led to site probability thresholds of 92%, 88.5%, and 80% for mono-, double- and triple-phosphorylated peptides, respectively (Figure 2B). CTD peptides had to be two or three heptads in length for appropriate detection by our MS analysis. Accordingly, the highest phosphosite coverage along the CTD was obtained from the three mammalian CTD variants M1, M2, and M3, for which only di- and tri-heptads are generated by tryptic cleavage (Figures 1A and S2). Additionally, due to very similar mass counts between the Y-b_1-ion and the PK-y_2-ion in MS^2, we excluded Y_1-P counts coming from peptides containing both Y in position 1 and terminal PK from the final dataset (942 out of 20,220 phosphopeptides), leading to lower detection levels for Y_1 phosphorylation (Figures 1A and S2). Similarly, due to the many mutations made in the S_7 position within all five CTD variants, our MS data of S_7 phosphorylation has to be considered as incomplete.

We detected up to four phosphosites in mammalian CTD peptides, although peptides were predominantly mono- and di-phosphorylated (Figure 2C). CTD variant M1 showed the highest number of mapped phosphosites within the proximal part of the CTD because it is only cleaved into di-heptad-long peptides (Figure 1C), which pass the phosphosite filter more often than the tri-heptad-long peptides (Figure 2C). In contrast, the high number of identical CTD peptides in the distal part of CTD variants M1, M2, and M3 led to similar phosphosite coverage in this region (Figures 1A and S2). Our MS analysis confirmed that all five residues, which were previously described as being phosphorylated based on antibody reactivity, are indeed phosphorylated. It also revealed that the five resulting phosphoresidues, Y_1-P, S_2-P, T_4-P, S_5-P, and S_7-P, are generally found along the entire CTD (Figures 1A and S2).

Frequency of Mono- and Double-Phosphorylated CTD Heptad Repeats

To determine the relative abundance among the mono (1P)- and double (2P)-phosphorylated heptad repeats, consensus repeats along the entire CTD from all five CTD variants were taken into account. S_2-P and S_5-P predominated in mono-phosphorylated repeats contributing ~75% to the total phospho-counts (Figure 3A), reflecting the prevailing importance of these two phosphoresidues during the transcription of all coding and non-coding RNAs. In contrast, the frequency of both Y_1-P and S_7-P

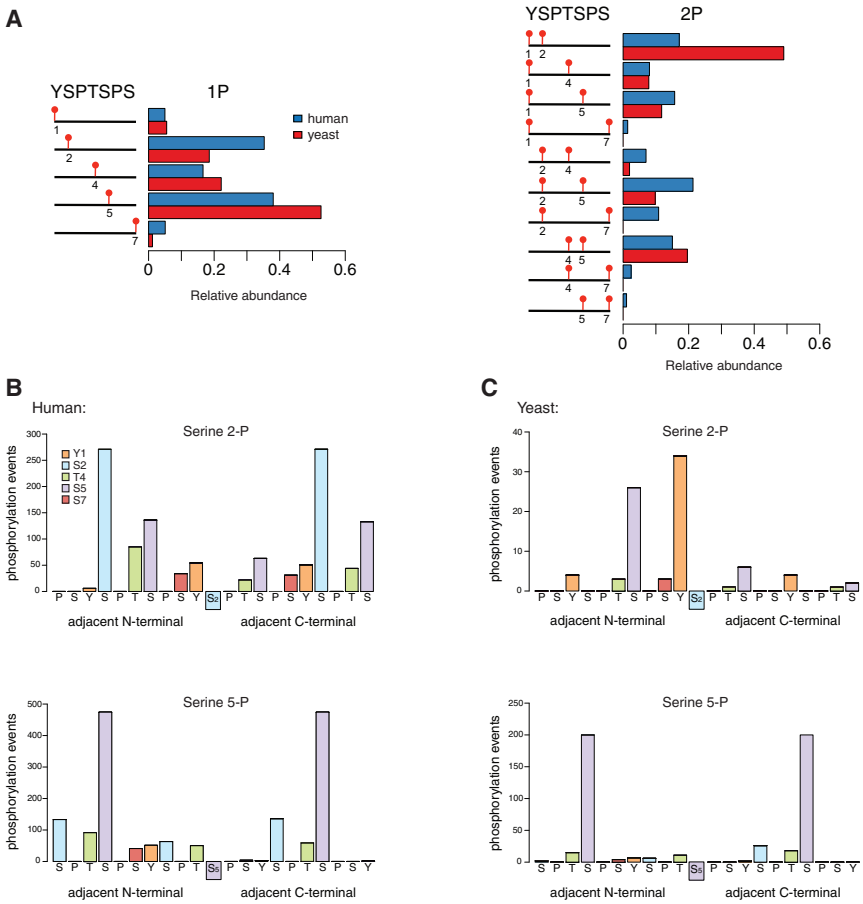


Figure 3. Phosphorylation Patterns within Neighboring CTD Residues

(A) (Left) Relative abundance of the five different CTD phosphosites (1P) within mono-heptads in human (blue) and yeast (red). (Right) Relative abundance of all ten possible double-phosphosite combinations (2P) within mono-heptads in human (blue) and yeast (red). All repeats with the sequence YSPTSPX (X = S, T, K, R, N, E, or G) were included in the final dataset.

(B) Next-neighbor phosphorylation study of S_2-P and S_5-P in human.

(C) Next-neighbor phosphorylation study of S_2-P and S_5-P in yeast. Only doubly phosphorylated CTD peptides were taken into account, and the defined sequence area comprised −10 to +10 CTD residues.

See also Figure S3.

heptad repeats didn't exceed 5%, whereas T_4-P heptad repeats showed significantly higher counts with a relative abundance of ~15% (Figure 3A). When analyzing double-phosphorylated heptad repeats, all possible double-phospho (2P) combinations (10 out of 10) were detected within single heptads (Figure 3A) revealing the high diversity of different 2P-combinations found within a single seven-amino-acid interval. However, the total counts of 2P-heptad repeats were ~30 times lower compared to 1P-heptad repeats, suggesting that the parallel existence of two phosphosites within a single heptad repeat seems to be a rather temporary rare event in vivo (Figure S3A). Additionally no triple-phosphorylated heptad repeats were found, indicating that multiple phosphosites found in close proximity to each other might lead to unfavorable structural changes due to strong charge repulsions.

Specific Double-Phospho Signatures Prevail within Neighboring CTD Residues

We next analyzed the co-occurrence of Y_1-P, S_2-P, T_4-P, S_5-P, and S_7-P with a second phosphorylated site in the same heptad repeat or in the adjacent N- or C-terminal heptad repeats (Figures 3B and S3B). Only double-phosphorylated CTD peptides were taken into account, and the defined sequence area comprised −10 to +10 CTD residues. S_2-P and S_5-P predominantly coexisted with a second phosphorylation at the same position in the adjacent repeat (± 7 position). Additionally, the combination S_2-P/S_5-P frequently occurred within neighboring repeats and with lower frequency also within the same repeat (Figure 3B). The predominance of 2P-combinations of S_2-P and S_5-P marks coincides with several, previously described, known CTD binding motifs of mammalian CTD interacting factors such as Mce1, SCP1, SCAF8, or Pin 1 (Jasnovidova and Stefl, 2013). T_4-P frequently co-occurred with S_2-P, S_5-P or a second T_4-P. Y_1-P and S_7-P showed only a very low number of phosphorylation events in our study (Figure S3B), although we note that we underestimate Y_1-P and S_7-P occurrence due to an ambiguity in the fragment masses derived from Y_1-P containing peptides (Experimental Procedures) and several mutations made in the S_7 position, respectively. All in all, 2P-patterns that were found along adjacent heptad repeats occurred more frequently compared to 2P-signatures coexisting within the same repeat, suggesting that the majority of CTD recognition motifs of interacting protein factors might comprise more than one heptad repeat in length.

Differences in CTD Phosphorylation Signatures between Yeast and Mammals

To investigate the possible evolutionary conservation of CTD phosphorylations, we extended our analysis to the yeast system. Two yeast CTD variants, called Yst1 and Yst2, were generated (Figures S3C–S3E). The CTD variant Yst2 led to a higher phosphosite coverage along the entire CTD (Figure S3F). As for the mammalian CTD, all five phosphoresidues were detected (Figure S3F). In mono-phosphorylated heptads, S_5-P was by far the most frequently detected phosphosite, whereas S_2-P was less abundant, in contrast to observations for the mammalian CTD (Figure 3A). The higher S_5-P/S_2-P ratio for the yeast CTD may result from the dramatically different gene lengths in yeast and human. Due to the lack of long introns, genes are much

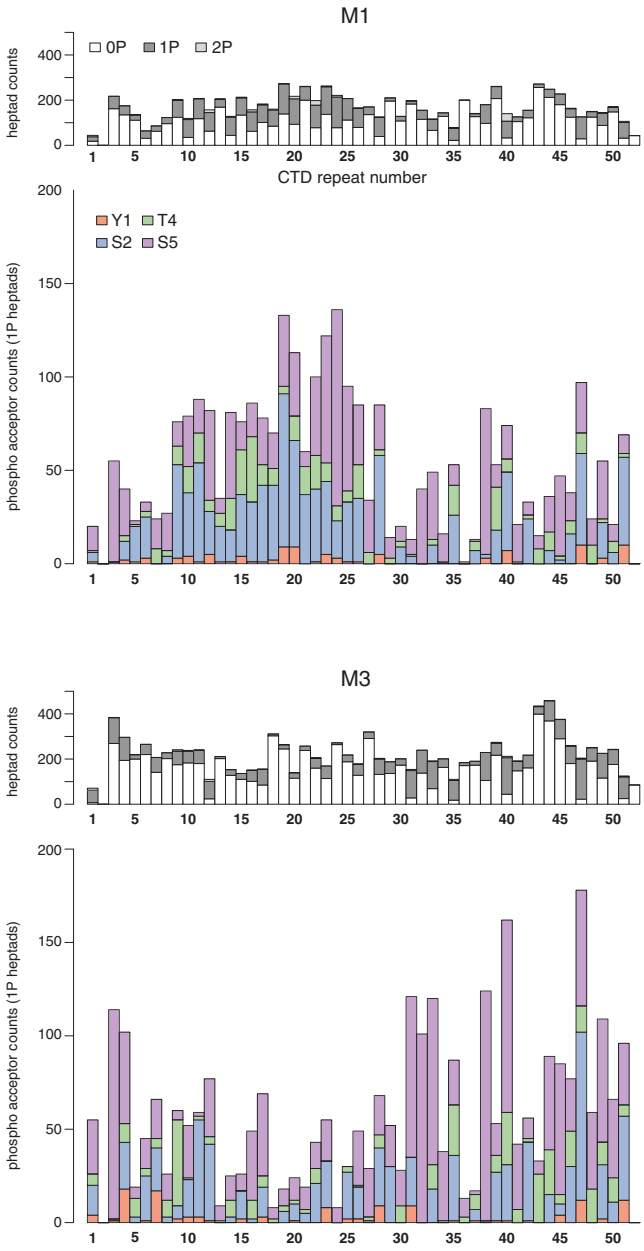


Figure 4. Phosphorylation Profile of All 52 CTD Repeats in Mammals

Comparison of the phosphorylation profile of all 52 CTD repeats in CTD variant M1 (top) and M3 (bottom). Total counts of un-phosphorylated (white), mono-phosphorylated (dark gray), and double-phosphorylated (light gray) counts of each CTD repeat are shown. Total counts of phosphosites Y_1-P (red), S_2-P (blue), T_4-P (green), and S_5-P (purple) within each repeat are shown. Y_1-P counts are under-represented in distinct repeats due to data exclusion from Y_1-P-false-positive biased peptides. S_7-P counts were excluded due to many amino acid substitutions at position 7. No data are shown for repeat 2, which is split in half after trypsin digest due to an R residue at position 4.

shorter in yeast and, thus, on average there are more transcribing, S_2-phosphorylated polymerases on genes in human, whereas S_5-P levels might be similar in yeast and human because this phosphorylation is restricted to the 5′-end of genes (Bataille et al., 2012). Y_1-P and S_7-P heptad repeats displayed comparable low counts in both species. Surprisingly, T_4-P repeats showed a similar frequency to S_2-P repeats (~20% each) (Figure 3A), although no biological function of T_4-P in yeast has been described so far.

In yeast, all four possible 2P combinations with S_7-P being one of the two phosphosites were non-detectable (Figure 3A), which may be explained by the low numbers of S_7 residues found in the two yeast CTD variants. Similar to mammals, yeast 2P repeats were found much less frequently compared to 1P repeats (~15 times lower) (Figure S3A). However, doubly phosphorylated repeats apparently reflect polymerase intermediates that are functional at certain stages of transcription. For example, the double-phosphorylation Y_1-P/S_2-P was the most frequently detected 2P signature within single yeast CTD repeats (Figures 3A and 3C), although the total count numbers were rather low (34 counts). Within adjacent repeats, we predominantly detected the co-existence of S_5-P marks (~200 counts) (Figure 3C) that correspond to the known binding pattern of the capping enzyme Cgt1 (Jasnovidova and Stefl, 2013). The second most abundant 2P signature along neighboring repeats in yeast was S_2-P/S_5-P (~25 counts), in which S_5-P was located at the N-terminal to S_2-P (Figure 3C), whereas in mammals, S_2-P combined similarly frequently with an N- or a C-terminal S_5-P. Surprisingly, compared to the mammalian CTD, adjacent yeast CTD repeats showed no S_2-P/S_2-P 2P-patterns (Figure 3C). This finding might result from the design of our yeast CTD variants, in which several S_2 residues were replaced by other amino acid residues (Figure S3C).

Phosphorylation Profile of All 52 CTD Repeats

Next, we analyzed the distribution of phosphorylations over all 52 mammalian CTD repeats. We first compared the heptad-specific phosphorylation profiles between CTD variants M1 and M3. The distal CTD repeats 32–52 showed a highly similar distribution in the relative abundance of phosphosites between CTD variants M1 and M3 (Figure 4), likely due to the high peptide sequence overlap in this region (Figure 1C; Figure S1B). In contrast, the fragmentation of the proximal part of the CTD from variant M1, resulting in only di-heptad-long peptides and our ability to clearly distinguish and detect Y_1-P (except for repeat 7), led to a more uniform and reliable phosphosite distribution and to higher phospho-counts in this region compared to CTD variant M3 (Figure 4).

By analyzing the heptad-specific phosphorylations in CTD variant M1 in more detail, the proximal consensus CTD part (repeats 4–31) displayed a very similar distribution in the relative abundance of the phosphorylations (Figure 4). S_2-P and S_5-P showed constant high counts, whereas Y_1-P and T_4-P were detected less frequently. S_5-P was the predominant phosphosite in repeats 3, 32, and 38, which all contain a threonine instead of serine at position 2 and glutamine (Q) or alanine (A) at position 4 (Figure 4). The distal part of the CTD showed higher variations in its phosphorylation profile as well as lower total phosphorylation counts, mainly due to the absence of S_2-P in repeats 32, 34, 36, 41, 43, and 48 (Figure 4). This can at least partly be explained by our exclusion of Y_1-P in these repeats, due to a fragment ion mass ambiguity. Indeed, distal CTD repeats 39, 40, 42, 46, 47, 49, and 50, which were unaffected by any false Y_1-P

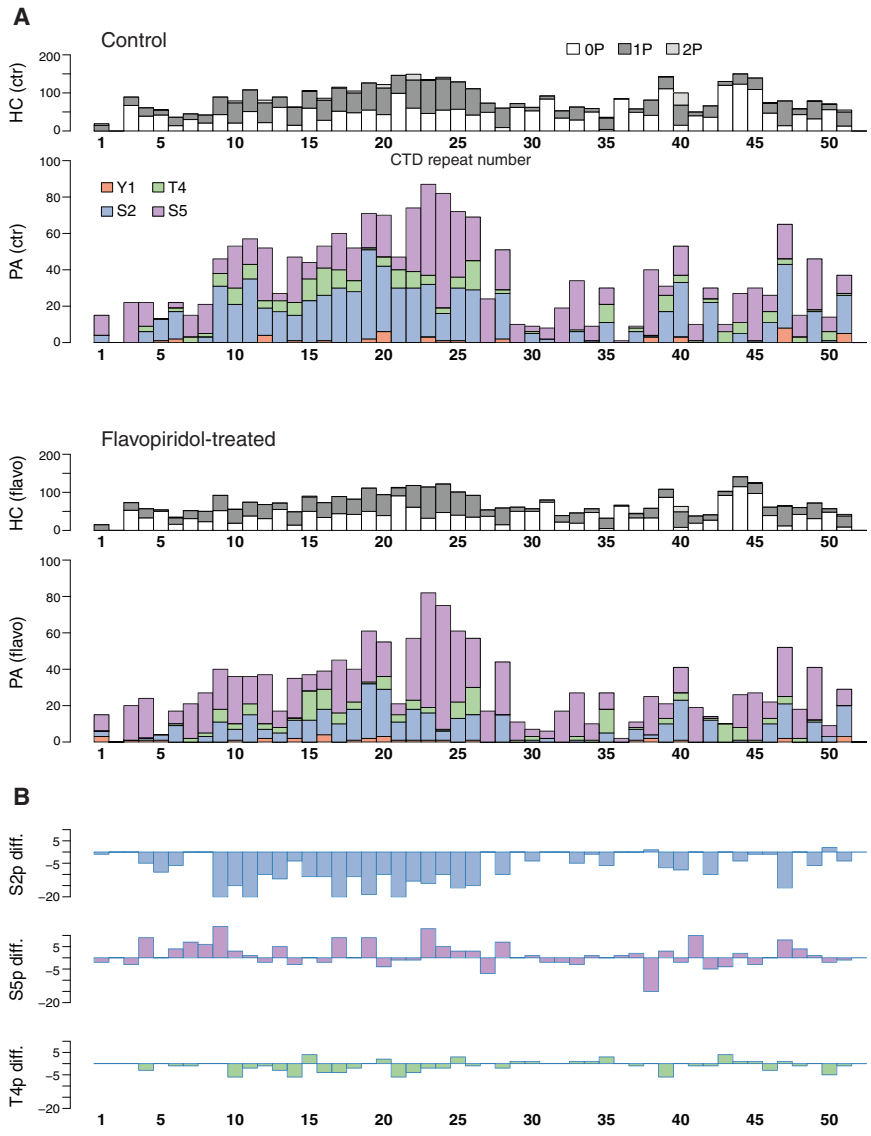


Figure 5. S_2-P Is Strongly Reduced in the CTD upon CDK9 Inhibition

(A) Comparative study of phosphorylation profiles of all CTD repeats between control (ctr) and flavopiridol-treated samples of CTD variant M1 (n = 4) (HC: heptad counts; PA: phospho-acceptor counts).

(B) Increase/decrease (+10/−20 counts) of S_2-P, S_5-P, and T_4-P comparing total counts between control samples and flavopiridol samples for each repeat. Y_1-P counts are under-represented in distinct repeats due to data exclusion from Y_1-P-false-positive biased peptides. S_7-P counts were excluded due to many amino acid substitutions at position 7.

See also Figure S4.

assignments, showed comparable phosphorylation profiles as found in the proximal CTD consensus repeats (Figure 4). Consequently, the mammalian CTD seems to be phosphorylated in a rather uniform manner along its entire sequence with S_2-P and S_5-P being the two predominant phosphosites in almost all CTD repeats. Nevertheless, we cannot exclude that more diverse heptad specific phosphorylation profiles might exist within distinct subpopulations of RNA Pol II.

S_2-P Is Strongly Reduced within the CTD upon CDK9 Inhibition

Finally, we tested whether our MS approach was able to detect physiological changes in CTD phosphorylation. To do this, we treated the cells with the CDK9 inhibitor flavopiridol, which was reported to decrease S_2-P levels (Chao et al., 2000; Chao and Price, 2001). Using western blot analysis, we confirmed a strong decrease in the signal intensity of S_2-P after 2.5 hr of flavopiridol treatment and a strong reduction of the hyper-phosphorylated IIO-form of Rbp1, whereas the phosphorylation signals for S_5 and S_7 remained constant over 6 hr (Figures S4A and S4B). Consistent with this, our MS analysis also revealed a decrease in the total counts of S_2-P for the majority of CTD repeats of M1 in the flavopiridol-treated samples (Figures 5A and 5B). A strong reduction of S_2-P was observed in almost all S_2-P-containing repeats showing that CDK9 inhibition led to phosphosite-specific changes along the CTD. In contrast, S_5-P and T_4-P counts were only very slightly reduced or even showed increased signals upon CDK9 inhibition (Figures 5A and 5B). The majority of repeats retained comparably high phospho-counts (Figure 5A). Surprisingly, we did not detect in vivo changes in S_5-P upon flavopiridol treatment, although it was reported that CDK9 phosphorylates S_5 in vitro (Bösken et al., 2014; Czudnochowski et al., 2012), emphasizing a need for assessing the specificity of CTD kinases and phosphatases in vivo.

DISCUSSION

Studies of phosphorylation states of the CTD of Pol II have been exclusively performed using CTD-phospho-specific antibodies. Our combined genetic and mass spectrometric approach overcomes the limitations of antibody detection and offers a comprehensive map of CTD phosphorylations that allowed the detection of the prevailing phosphorylation patterns in CTD heptad repeats. Due to the highly repetitive sequence of the CTD, we were faced with several technical challenges. First, we introduced a series of lysines or arginines to allow the tryptic cleavage of the CTD into very short peptides, which do not extend the length of four heptad repeats (Figures 1C and S1B). Longer peptides are multiply charged

and are therefore not suitable for MS analysis (Steen and Mann, 2004). Second, to precisely map phosphosites by LC-MS/MS, we replaced or introduced additional amino acids, which generated peptides with a unique mass for MS^2 analysis (Figures 1C and S1B). Third, we introduced a phosphopeptide enrichment step by TiO_2, taking advantage of the strong affinity of negatively charged phosphate groups toward metal ions (Larsen et al., 2005). Fourth, we used a training set of synthetic phosphopeptides (Figure 2A) to determine the false discovery rate of the assigned phosphosites. The site probability threshold varied depending on the amount of introduced phosphorylation sites per peptide (Figure 2B). The fact that peptides with two or three heptad repeats contain up to 9–14 potential phospho-acceptor sites impedes the correct phosphosite assignment. The probability further decreases with longer peptides. Fifth, the design of five CTD variants allowed us to obtain a large spectrum of different CTD peptides based on sequence and length.

This comprehensive map of CTD phosphorylations in yeast and mammals revealed that the five CTD phosphoresidues are generally distributed along the entire CTD in both species, albeit with different frequencies. S_2-P and S_5-P prevailed, suggesting that these two phosphosites play a predominant role during transcription. Y_1-P and T_4-P were present at lower levels in our MS approach indicating that these CTD residues might play an important role that is restricted to more specific functions. We should emphasize that Y_1-P counts are underestimated in our study due to technical reasons. Consequently, our Y_1-P data cannot be used to support or disprove a function of Y_1-P in initiation or elongation control (Descostes et al., 2014; Mayer et al., 2012). Moreover, we cannot rule out that the found ratios between the different phosphosites might change when analyzing specific subpopulations of RNA Pol II. Another common finding in both species was that the CTD is not heavily phosphorylated. In more detail, double-phosphorylated CTD repeats were rarely found suggesting that multiply phosphorylated repeats more likely represent polymerase intermediates that are functional at certain stages of transcription. Nevertheless, predominant 2P signatures were found within neighboring CTD residues that might serve as recognition motifs of CTD interacting factors and thereby help to orchestrate the transcription cycle as well as transcription-coupled RNA processing events. The approach also enables a more detailed assessment of alterations in CTD phosphorylations during cellular physiology as shown by our CDK9 inhibition experiment in the mammalian CTD. In more detail, the strong reduction of S_2-P within the CTD through the inhibition of the CTD S_2 kinase CDK9 demonstrates that our combined genetic-LC-MS/MS strategy can detect heptad-specific changes in phosphorylation states within the CTD.

A critical step for the successful application of this protocol was the design of CTD variants, which enables cell growth and proliferation. The minimal length and sequence requirements of CTD that are needed for normal cell viability have not been determined yet for mammalian cells. However, our successful design of five stable CTD variants revealed the high plasticity of the mammalian CTD and confirmed that substantial manipulation of CTD is possible (Bartolomei et al., 1988). Consequently, our established CTD variants are a powerful tool to gain deeper insights into CTD heptad-specific phosphorylation after the knock out of cellular kinases or phosphatases. For example, by establishing specific CTD kinase analog sensitive mammalian cell lines (Bartkowiak et al., 2015; Laitem et al., 2015) using the genomic engineering tool CRISPR/Cas (Cong et al., 2013; Fu et al., 2013; Mali et al., 2013), CTD-specific kinase knockdowns will possibly link different CTD subregions to specific kinase activities. In the future, it may also allow us to monitor changes in CTD phosphorylation after activation of specific signaling pathways. Finally, analysis of the heptad specific phosphorylation patterns of RNA Pol II subforms that are located at different positions along genes during the transcription cycle might lead to a better understanding of their functional properties.

EXPERIMENTAL PROCEDURES

Plasmids and Mammalian Cell Culture

CTD sequences of WTrec and CTD variants M1-M5 were synthesized (GeneArt) using human codon optimization and cloned into the LS*mock vector (Meininghaus et al., 2000). All final constructs were sequenced before usage. Full-length Rpb1 expression vectors (WTrec, M1–M5) were transfected into the Epstein-Barr-virus-positive Burkitt lymphoma cell line Raji by electroporation (10 μg plasmid DNA/10^7 cells; 960 μF; 250 V). Selection of a positive polyclonal cell line was achieved by the addition of G418 (1 mg/ml) for 14–20 days to recover a viability of ~95%. Established cell lines were maintained in RPMI 1640 medium supplemented with 10% fetal calf serum, 1% penicillin streptomycin (Gibco, Invitrogen), 2 mM L-glutamine (Gibco, Invitrogen), and 2 μg/ml α-amanitin at 37°C with 5% CO_2.

Yeast Culture

The newly synthesized CTD variants (GeneArt) Yst1 and Yst2 were cloned into the pRS315-Rpb1 plasmid (Sun et al., 2012). After ligation, the product was transformed into XL1 blue cells (Stratagene). Positively sequenced Rpb1 mutant plasmids were then transformed into the BY Δrpb1 pRS316-Rpb1 strain and plated onto -leu/-ura plates (0.69% [w/v] yeast nitrogen base, 0.067% [w/v] drop-out -leu/-ura, 2% [w/v] glucose, 2% [w/v] agar) to select for pRS315 and pRS316 plasmids. Single clones were then transferred to -leu plates to preserve only the pRS315 plasmid. Spot dilutions of single clones were spotted onto YPD and 5-FOA plates (SC [-ura], 0.01% [w/v] uracil. 0.2% [w/v] 5-FOA, 2% [w/v] agar) in parallel. Strains that grew on 5-FOA plates only contained the correct pRS315 shuffle plasmid. Plates were generally incubated at 30°C, and 5-FOA plates were incubated at 25°C. Single clones from 5-FOA plates were sequenced again to ensure that the strain did not change its CTD sequence by self-recombination. The TAP tag on Rpb3 was introduced by homologous recombination of a PCR product amplified from the plasmid pYM-TAP-His3 (Euroscarf) using the Rpb3-TAP primers (fw:5′TCTCAAATGGGTAATACTGGATCAGGAGGGTATGATAAT GCTTGGCG TACGCTGCAGGTCGAC3′, rev:5′TCGTTCACTTGTTTTTTTTCC TCTATTACG CCCACTTGAGAACTA ATCGATGAATTCGAGCTCG 3′).

Western Blot Analysis

A 2x laemmli buffer was added to protein samples, which were then separated by SDS-PAGE (6.5% gel) and transferred to a nitrocellulose membrane (GE Healthcare) afterward. Unspecific binding of antibodies was blocked by incubation with 5% milk in TBS-T before being incubated with primary antibodies diluted in blocking solution. The primary antibodies we used are as follows: α-Rpb1 (POL 3/3); α-HA (3F10, Roche); CTD phospho-specific antibodies α-Y_1-P, S_2-P, T_4-P, S_5-P, and S_7-P (5G9, 3D12, 3E10, 6D7, 3E8, and 4E12) (Chapman et al., 2007); and α-tubulin (T6557) (Sigma). Afterward, the membranes were either incubated with IRDye-labeled, secondary antibodies against rat (680 nm; Alexa, Invitrogen) and mouse (800 nm; Rockford, Biomol) and analyzed using an Odyssey Imaging System (Li-Cor) or they were stained

with HRP-conjugated secondary antibodies against rat (Sigma) or mouse (Promega) and revealed by enhanced chemiluminescence.

Purification of Recombinant Rpb1 Protein from Mammalian Cells

All of the following steps were performed at 4°C. 3 × 10^8 Raji cells were lysed in 10 ml IP-lysis buffer (50 mM Tris-HCl [pH 8.0], 150mM NaCl, 1% NP-40 [Roche], 1x PhosSTOP [Roche], 1x protease inhibitor cocktail [Roche]) and incubated for 30 min on an end-over-end shaker. Samples were sonified (Sonifier 250 [Branson], 3 × 20 pulses, output 7, duty cycle 70%), incubated for 1 hr on an end-over-end shaker, and centrifuged (10,500 × g, 15 min). Cell lysates were incubated with antibodies, coupled protein G-sepharose beads (20–40 μg of either αHA [3F10, Roche] or αSer2-P [3E10]/ αSer5-P [3E8] in a 1:1 ratio) on an end-over-end shaker overnight. Beads were washed three times with IP-lysis buffer and subsequently boiled in 2x laemmli buffer at 95°C for 8 min. The released proteins were loaded onto an SDS-PAGE gel. After separation by electric current, gels were stained with EZBlue Coomassie Brilliant Blue G-250 (Sigma).

Purification of Recombinant Rpb1 Protein from Yeast

2 l YPD media were inoculated with 500 μl of a pre-culture of BY Δrpb1 pRS315-Rpb1-CTD-variant Yst1 or -variant Yst2 with a TAP-tag on Rpb3 and grown to OD_{600} 3.0–3.5 (30°C/160 rpm) overnight. Cells were harvested, and they were washed once with 500 ml water and once with 25 ml TAP lysis buffer (LB; 50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 0.15 mM $MgCl_2$, 0.15% NP40). Pellets were flash frozen in liquid nitrogen and stored at −80°C. The following steps were all performed at 4°C: After thawing, an equal volume of LB containing protease inhibitors (0.28 μg/ml leupeptin, 1.37 μg/ml pepstatin A, 0.17 μg/ml PMSF, 3.3 μg/ml benzamidine) and 1x PhosSTOP (Roche) was added to the pellet. The same volume of glass beads was added to the cell suspension. Cell lysis was performed in 15 ml Falcon centrifuge tubes using the FastPrep24 (MP) and the 12x15 adapter for 3 × 30 s at 5.5 m/s with 3 min breaks on ice. Lysates were centrifuged at 4,000 rpm for 10 min. The supernatant was transferred to pre-cooled ultracentrifugation tubes and centrifuged at 27,000 rpm for 1 hr. The lysate was transferred from underneath the fatty top phase to new tubes. 200 μl Protein G Dynabeads Magnetic Beads (Invitrogen), pre-coated with 110 μl rabbit IgG (1.5 mg/ml, Sigma/Aldrich), were added to the cell extract and incubated for 1 hr on a turning wheel (8 ml cell extract per 200 μl beads). TAP-tagged proteins were eluted by addition of 70 μl LB and 2 μl TEV protease (2 mg/ml). The beads were re-suspended by vortexing, and the suspension was incubated on a rotating wheel for 90 min. Subsequently, the supernatant was collected followed by the addition of 30 μl LB to the beads (additional elution fraction). 25 μl SDS-PAGE loading buffer (5x) were added to complete the elution, and the suspension was boiled for 2 min at 95°C. Protein samples were separated by SDS-PAGE according to their size, using pre-cast NuPAGE 4%–12% Bis-Tris gels (Life Technologies) in a Novex Mini-Cell chamber (Invitrogen) at 160–200V. Gels were subsequently stained using Instant Blue (Expedeon).

In-Gel Cleavage and Phosphopeptide Enrichment

A standardized protocol was used for in-gel digestion, including four main steps: (1) de-staining, (2) cysteine reduction and alkylation, (3) tryptic cleavage, and (4) extraction of the generated peptides. Subsequent to in-gel cleavage, the extracted peptides were subjected to phosphopeptide enrichment. The sample volume was reduced to 5 μl by speed-vac and afterward dissolved in 100 μl loading buffer (80% acetonitrile, 5% TFA, and 1M glycolic acid). 10 μl of TiO_2 bead solution (0.03 mg/μl TiO_2 beads in 100% ACN) were added, and the samples were incubated on a shaker at room temperature for 10 min. After a short centrifugation, two washing steps were performed (washing buffer 1: 80% acetonitrile, 1% TFA; washing buffer 2: 10% acetonitrile, 0.2% TFA). Afterward, the supernatant was removed completely and the beads were dried for at least 10 min. For elution of the bound phosphopeptides, 50 μl elution buffer (40 μl ammonia solution [28%] in 960 μl H_2O) was added and the samples were incubated on the shaker at room temperature for 10 min. The supernatant containing the phosphopeptides was transferred to a new tube, and the sample volume was reduced to 5 μl using a speed-vac. 30 μl 0.1% formic acid (FA) was added, and the samples were stored at −20°C until measured.

Synthetic Peptide Library

20 Maxi SpikeTides L peptides (amount: 0.5 mg) were synthesized by JPT-Innovative Peptide Solutions. All peptides were heavy isotope-labeled on their C-termini (K) and were dissolved in DMSO (stock solution: 1 nmol/μl). Peptides were further diluted adding 100% ACN before being measured (working solution: 1.5 pmol/μl).

Liquid Chromatography Coupled to Tandem Mass Spectrometry

For LC-MS/MS, TiO2-enriched phosphopeptides were automatically injected into an Ultimate 3000 HPLC system (Dionex), desalted online (300 μm i.d. × 5 mm cm, PepMap100 C18 5 μm, 100 Å, Dionex), and separated on an analytical column (75 μm i.d. × 10 cm, packed in house with ReproSil-Pur C18 AQ 2.4 μm, Dr. Maisch) using a 60 min gradient from 5% to 30% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into a LTQ Orbitrap XL mass spectrometer (Thermo), operated in data-dependent mode to automatically switch between full-scan MS and MS/MS acquisition. Survey full-scan MS spectra (from m/z 300 to 2000) were acquired in the Orbitrap with a resolution of R = 60,000 at m/z 400 (after accumulation to a "target value" of 500,000 in the linear ion trap). The three most intense peptide ions with a charge state higher than 1 and lower than 4 were sequentially isolated to a target value of 10,000 and fragmented in the linear ion trap by collision-induced dissociation (CID). Multistage activation was selected as the fragmentation method to automatically select and fragment the ions originated from the loss of one or two phosphate groups from the parent ion. For all measurements with the Orbitrap mass analyzer, three lock-mass ions from ambient air (m/z = 371.10123, 445.12002, and 519.13882) were used for internal calibration (Olsen et al., 2005). Typical mass spectrometric conditions were the following: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 200°C; and normalized collision energy, 35% for CID in LTQ. The ion selection threshold was 10,000 counts for MS^2. An activation q = 0.25 and activation time of 30 ms were used. Mass spectrometry data were analyzed using Proteome Discoverer 1.4/PhosphoRS 2.0 (Taus et al., 2011) (MS tol., 10 ppm; MS/MS tol., 0.8 Da, Variable modifications, oxidation [M] and phosphorylation [S,T,Y]; fixed modifications, carbamidomethyl [C]; FDR peptide, 0.01).

Data Analysis

Proteome Discoverer output was processed for downstream analyses and visualizations using R (http://www.r-project.org). All data obtained from the same variant and same experimental condition were pooled. Only peptides with a search engine rank of 1 were considered for downstream analysis. Phosphosite assignments were made by comparing site-specific phosphorylation probabilities, as calculated by PhosphoRS with the probability thresholds defined by analyzing synthetic peptides as follows: MS output of synthetic peptide runs were processed in the same way as for the biological samples. We determined the fraction of false-positive phosphorylation assignments by calculating the ratio of unexpected, false-positive versus total phosphorylation assignment as a function of a range of threshold levels on the PhosphoRS site probabilities. For this procedure, we separated the data based on the total number of phosphorylation events per peptide (P_{total}) because the site probabilities were inversely correlated with P_{total}. Accordingly, we defined specific thresholds for each level of P_{total} to yield a false-positive rate of 20%. In the biological samples, PhosphoRS sites not passing this filter were marked as ambiguous. In addition, peptides with both Y_1P and C-terminal PK were removed. Context-independent reporting of site-specific phosphorylation included only unambiguous assignments. Phosphorylation pattern analysis and context-dependent reporting excluded instances in which ambiguous sites were present in their respective context. For example, analysis of phosphorylation patterns in mono-heptads included only heptads that did not have ambiguous phosphorylation calls.

ACCESSION NUMBERS

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaíno et al., 2014) via the PRIDE partner repository with the dataset identifier PXD: PXD003159. We acknowledge

the PRIDE team for support of data deposition to the ProteomeXchange Consortium.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2015.12.003.

AUTHOR CONTRIBUTIONS

R.S., A.I., P.C., and D.E. designed experiments and analyzed results; R.S, I.F., A.S., Y.T., N.S., and T.-M.D. carried out experiments; R.S. and I.F. performed and analyzed mass spectrometry experiments; R.S. and T.S. analyzed proteomic data; and R.S., P.C., and D.E. wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft, SFB 1064 collaborative research center (CRC)–Chromatin Dynamics. We thank R.D. Chapman for technical support.

Received: August 31, 2015
Revised: October 30, 2015
Accepted: November 11, 2015
Published: January 21, 2016

REFERENCES

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

Bartolomei, M.S., and Corden, J.L. (1987). Localization of an alpha-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II. Mol. Cell. Biol. *7*, 586–594.

Bartolomei, M.S., Halden, N.F., Cullen, C.R., and Corden, J.L. (1988). Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. Mol. Cell. Biol. *8*, 330–339.

Bataille, A.R., Jeronimo, C., Jacques, P.E., Laramée, L., Fortin, M.E., Forest, A., Bergeron, M., Hanes, S.D., and Robert, F. (2012). A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes. Mol. Cell *45*, 158–170.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Bösken, C.A., Farnung, L., Hintermair, C., Merzel Schachter, M., Vogel-Bachmayr, K., Blazek, D., Anand, K., Fisher, R.P., Eick, D., and Geyer, M. (2014). The structure and substrate specificity of human Cdk12/Cyclin K. Nat. Commun. *5*, 3505.

Buratowski, S. (2003). The CTD code. Nat. Struct. Biol. *10*, 679–680.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Chao, S.H., and Price, D.H. (2001). Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. J. Biol. Chem. *276*, 31793–31799.

Chao, S.H., Fujinaga, K., Marion, J.E., Taube, R., Sausville, E.A., Senderowicz, A.M., Peterlin, B.M., and Price, D.H. (2000). Flavopiridol inhibits P-TEFb and blocks HIV-1 replication. J. Biol. Chem. *275*, 28345–28348.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. Science *318*, 1780–1782.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Corden, J.L. (2013). RNA polymerase II C-terminal domain: tethering transcription to transcript and template. Chem. Rev. *113*, 8423–8455.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863–1876.

Czudnochowski, N., Bösken, C.A., and Geyer, M. (2012). Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. Nat. Commun. *3*, 842.

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife *3*, e02105.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. Science *318*, 1777–1779.

Egloff, S., Dienstbier, M., and Murphy, S. (2012). Updating the RNA polymerase CTD code: adding gene-specific layers. Trends Genet. *28*, 333–341.

Eick, D., and Geyer, M. (2013). The RNA polymerase II carboxy-terminal domain (CTD) code. Chem. Rev. *113*, 8456–8490.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D. (2013). High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. *31*, 822–826.

Hintermair, C., Heidemann, M., Koch, F., Descostes, N., Gut, M., Gut, I., Fenouil, R., Ferrier, P., Flatley, A., Kremmer, E., et al. (2012). Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation. EMBO J. *31*, 2784–2797.

Hsin, J.P., Sheth, A., and Manley, J.L. (2011). RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3′ end processing. Science *334*, 683–686.

Jasnovidova, O., and Stefl, R. (2013). The CTD code of RNA polymerase II: a structural view. Wiley Interdiscip. Rev. RNA *4*, 1–16.

Laitem, C., Zaborowska, J., Isa, N.F., Kufs, J., Dienstbier, M., and Murphy, S. (2015). CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. Nat. Struct. Mol. Biol. *22*, 396–403.

Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P., and Jørgensen, T.J. (2005). Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. Mol. Cell. Proteomics *4*, 873–886.

Liu, P., Greenleaf, A.L., and Stiller, J.W. (2008). The essential sequence elements required for RNAP II carboxyl-terminal domain function in yeast and their evolutionary conservation. Mol. Biol. Evol. *25*, 719–727.

Lu, H., Zawel, L., Fisher, L., Egly, J.M., and Reinberg, D. (1992). Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. Nature *358*, 641–645.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Marx, H., Lemeer, S., Schliep, J.E., Matheron, L., Mohammed, S., Cox, J., Mann, M., Heck, A.J., and Kuster, B. (2013). A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. Nat. Biotechnol. *31*, 557–564.

Mayer, A., Heidemann, M., Lidschreiber, M., Schreieck, A., Sun, M., Hintermair, C., Kremmer, E., Eick, D., and Cramer, P. (2012). CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. Science *336*, 1723–1725.

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

Meininghaus, M., Chapman, R.D., Horndasch, M., and Eick, D. (2000). Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. J. Biol. Chem. *275*, 24375–24382.

Olsen, J.V., de Godoy, L.M., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005). Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol. Cell. Proteomics *4*, 2010–2021.

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

Schwer, B., Sanchez, A.M., and Shuman, S. (2012). Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. Proc. Natl. Acad. Sci. USA *109*, 18024–18029.

Steen, H., and Mann, M. (2004). The ABC's (and XYZ's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. *5*, 699–711.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. *22*, 1350–1359.

Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011). Universal and confident phosphorylation site localization using phosphoRS. J. Proteome Res. *10*, 5354–5362.

Vizcaíno, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Ríos, D., Dianes, J.A., Sun, Z., Farrah, T., Bandeira, N., et al. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. Nat. Biotechnol. *32*, 223–226.

Zhang, J., and Corden, J.L. (1991). Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. J. Biol. Chem. *266*, 2290–2296.