In vitro analysis of RNA polymerase II elongation complex dynamics

Yoo Jin Joo,[1] Scott B. Ficarro,[2,3] Yujin Chun,[1] Jarrod A. Marto,[2,3] and Stephen Buratowski[1]

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA;
[2]Department of Cancer Biology, Dana Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA;
[3]Blais Proteomics Center, Dana Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

RNA polymerase II elongation complexes (ECs) were assembled from nuclear extract on immobilized DNA templates and analyzed by quantitative mass spectrometry. Time-course experiments showed that initiation factor TFIIF can remain bound to early ECs, while levels of core elongation factors Spt4–Spt5, Paf1C, Spt6–Spn1, and Elf1 remain steady. Importantly, the dynamic phosphorylation patterns of the Rpb1 C-terminal domain (CTD) and the factors that recognize them change as a function of postinitiation time rather than distance elongated. Chemical inhibition of Kin28/Cdk7 in vitro blocks both Ser5 and Ser2 phosphorylation, affects initiation site choice, and inhibits elongation efficiency. EC components dependent on CTD phosphorylation include capping enzyme, cap-binding complex, Set2, and the polymerase-associated factor (PAF1) complex. By recapitulating many known features of in vivo elongation, this system reveals new details that clarify how EC-associated factors change at each step of transcription.

The RNA polymerase II (RNApII) transcription cycle has three major stages: initiation, elongation, and termination. During elongation, RNApII interacts directly or indirectly with multiple elongation regulatory factors as well as proteins needed for processing nascent RNA, chromatin modification and remodeling, and, eventually, termination (Buratowski 2009). Elongation regulation is critical for proper gene expression, yet much remains unclear about the dynamics of elongation complexes (ECs).

Of the multiple approaches used to identify factors that interact with RNApII ECs, protein biochemistry and yeast genetics have been particularly fruitful. Early in vitro elongation experiments using purified RNApII led to discovery of TFIIS, a protein that reactivates stalled backtracked ECs (Nakanishi et al. 1981). TFIIS, TFIIF, and the "polymerase-associated factor" complex (Paf1C) were purified by their affinity for RNApII (Nakanishi et al. 1981; Sopta et al. 1985). Mammalian DSIF (the Spt4–Spt5 complex) and FACT (the Spt16–Pob3–Nhp6 complex) were also isolated by their ability to modulate in vitro transcription by purified RNApII factors (Wada et al. 1998; Orphanides and Reinberg 2000). Affinity purifications of these elongation factors uncovered other EC-associated factors such as Spt6, Elf1, and Spn1/Isw1 (Krogan et al. 2002; Prather et al. 2005; Zhang et al. 2008).

These elongation factors were often identified first in genetic screens for transcription regulators. For example, the *SPT* genes were isolated as suppressors of particular transposon insertions (Winston et al. 1984), and many elongation factor mutants are sensitive to drugs that reduce NTP levels, such as 6-azauracil or mycophenolic acid. Chromatin immunoprecipitation (ChIP) experiments show cross-linking of these factors with actively transcribed genes in vivo, and several distinct patterns are seen (Krogan et al. 2002; Kim et al. 2004; Mayer et al. 2010).

In vivo, RNApII ECs must overcome the inhibitory effect of nucleosomes but also restore chromatin integrity after passing through (Orphanides and Reinberg 2000; Li et al. 2007). The elongation factors FACT and Spt6 have histone chaperone activity, and mutations in these genes lead to disrupted chromatin structure, aberrant histone modification, and initiation from cryptic internal promoters (Kaplan et al. 2003). Paf1C is required for H2B ubiquitination and, subsequently, several cotranscriptional histone methylations (Krogan et al. 2003; Wood et al. 2003). The mechanistic details of how these factors function are not yet clear, but recent cryo-EM structures show how several bind to RNApII (Ehara et al. 2017; Xu et al. 2017b; Vos et al. 2018).

Another key component in EC function is the C-terminal domain (CTD) of the RNApII largest subunit, Rpb1. The CTD is comprised of multiple repeats of the heptapeptide sequence Tyr1–Ser2–Pro3–Thr4–Ser5–Pro6–Ser7 (Corden 2013). During transcription, the CTD undergoes a programmed pattern of phosphorylation and dephosphorylation, generating a "CTD code" that creates binding sites for a variety of proteins needed at different stages of transcription (for reviews, see Buratowski 2009; Corden 2013). Factors known to bind phosphorylated Ser5 (Ser5P) during early elongation include mRNA capping enzyme, the non-polyA termination factor Nrd1, and the Set1 histone methyltransferase complex. In contrast, mRNA termination factor Rtt103 and the histone methyltransferase Set2 are coupled to downstream CTD phosphorylation at Ser2 (Ser2P). Mass spectrometry (MS) of factors coimmunoprecipitated with different CTD phosphorylations identified additional candidate EC proteins (Harlen et al. 2016; Ebmeier et al. 2017). It is therefore important to understand how the CTD code is generated and used to regulate cotranscriptional processes.

Although reconstitution with purified factors has been essential for identifying the minimal set of EC proteins, transcription in vivo is coupled to multiple mRNA-processing and chromatin-modifying factors that make full reconstitution difficult. Here we used yeast nuclear extracts to better approximate in vivo conditions. We previously used quantitative proteomics to analyze RNApII preinitiation complexes (PICs) (Sikorski et al. 2012). We now extend this analysis to RNApII ECs formed on DNA templates in vitro. MS identifies a set of core elongation factors (Spt4-Spt5, Spt6-Spn1, Elf1, and Paf1C) as well as EC-associated histone-modifying and mRNA-processing factors. Although elongation is stalled at the end of a short G-less cassette, time-course experiments show that CTD phosphorylations and associated factors continue to exchange as a function of time rather than location along the gene. Chemical inhibition shows that binding of Paf1C, capping enzyme, and Set2 to ECs requires TFIIH kinase (Kin28/Cdk7) activity. As this in vitro system recapitulates many known features of transcription elongation, it can be used to better understand factor dynamics as RNApII transitions from initiation to elongation as well as how transcription is coordinated with nascent RNA processing and chromatin modifications.

Results

MS analysis of RNApII ECs formed on immobilized templates

We sought to characterize RNApII ECs using the immobilized template assay and quantitative MS approach (ITA/MS) used previously for studying PICs (Cho et al. 1997; Sikorski et al. 2012). The DNA template, fixed to magnetic beads via a biotin–streptavidin linkage, has five binding sites for the transcription activator Gal4-VP16 upstream of the CYC1 core promoter (Fig. 1A). Because RNApII can elongate off the end of the linear template, a G-less cas-sette was used to stall ECs. PICs were assembled for 30 min in nuclear extract, and transcription was then initiated by adding three NTPs (ATP, CTP, and UTP) and the chain terminator 3′-O-methyl-GTP (Fig. 1B). Transcription initiates within the G-less cassette just downstream from the promoter and extends until stalling where GTP should be incorporated (Hahn et al. 1985). After 15 min, the beads were washed extensively, and candidate EC components were eluted with the restriction enzyme SspI, which cuts just upstream of the transcription start site (TSS) to leave promoter-bound proteins on the beads (Fig. 1A,B).

Samples were prepared with and without NTPs. To distinguish transcription from other NTP-dependent enzymatic reactions that may affect protein–DNA interactions (e.g., kinases, helicases, and other ATPases), the same experiment was carried out in parallel in the presence of the elongation inhibitor α-amanitin. Quantitative MS using 4-plex iTRAQ labeling identified >1000 proteins (Fig. 1C; Supplemental Fig. 1a; Supplemental Table 1). Signals for the vast majority of proteins were the same under all conditions, thereby defining nonspecific background binding. Using mixed-model analysis to establish a 95% confidence interval (CI) for deviation from this null hypothesis distribution, we identified three groups of proteins with distinct binding profiles (Supplemental Fig. 1b). Binding for the first group (22 total) was strongly increased by NTPs regardless of whether transcription was inhibited (Supplemental Fig. 1d). These include the Swr1 complex (Swr1C), RNApIII subunits, Snf5 and Swp82 of the Swi/Snf complex, the SUMO protein Smt3, Rtt107, multiple DNA repair proteins, and SSA-related heat-shock proteins. While some of these may simply require NTPs for nonspecific DNA binding, it is also possible that the linear DNA template ends are recognized as damage by repair enzymes in the extract. Binding of 55 proteins was reduced in the presence of NTPs both with and without α-amanitin (Supplemental Fig. 1c). These include basal initiation factors for RNApII, negative cofactor 2 complex (NC2), Mot1, nucleotide excision repair factors, replication factors, mRNA-binding proteins, and several sequence-specific transcription factors. The eluted downstream DNA has no consensus TATA, but some AT-rich sequences there probably support low levels of PIC formation. ATP hydrolysis may destabilize these nonproductive PICs via the TFIIH translocase or by the Mot1 ATPase and NC2.

The final group of 43 proteins showed increased binding with NTPs, which was reduced by α-amanitin, indicating that their presence on downstream DNA was transcription-dependent (Fig. 1D; Supplemental Table 1). This cluster contains all of the RNApII subunits detected in this experiment (three small subunits were seen in later experiments described below) as well as known elongation factors Spt5/DSIF, Spt6, Spn1, and Paf1C. Therefore, we conclude that bona fide ECs are enriched in this in vitro system. The Rad26/CSB translocase was also in this group, consistent with its function pushing stalled ECs forward (Xu et al. 2017a). Unexpectedly, the TBP-associated factor (TAF) subunits of TFIID were found on downstream DNA, dependent on transcription. In a
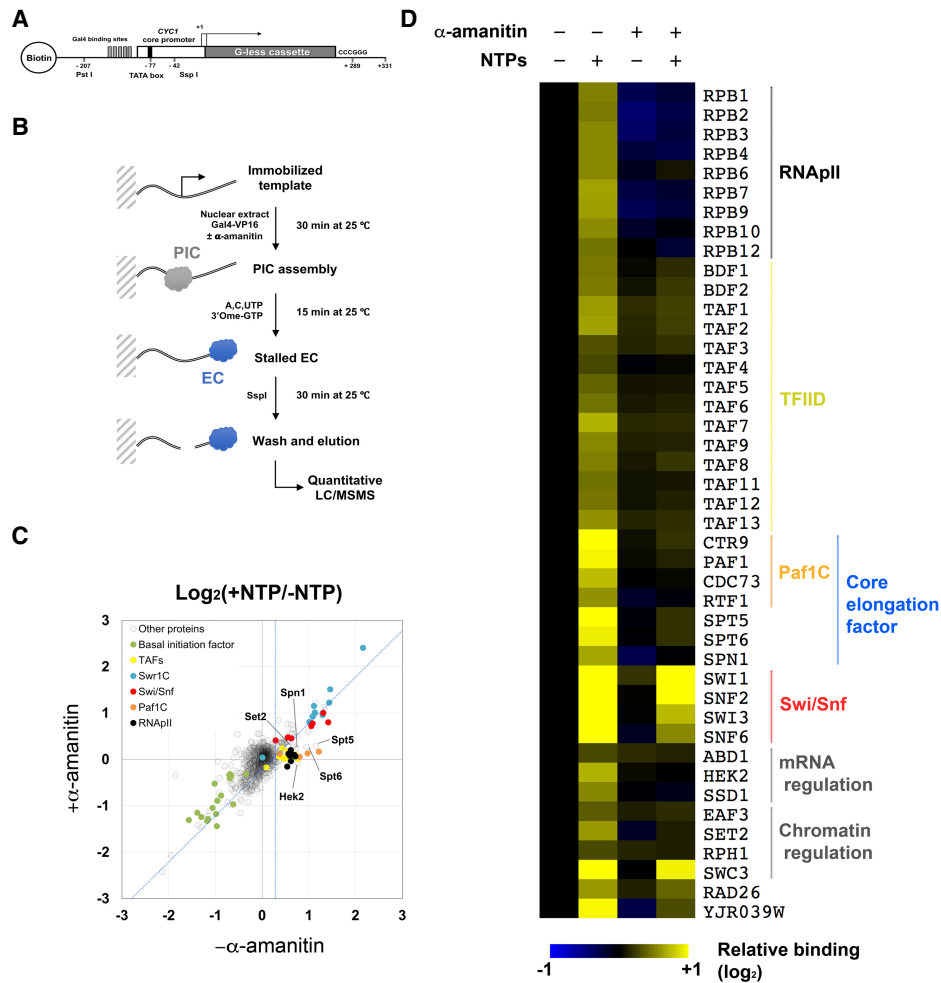
Figure 1. Characterization of RNApII ECs generated in vitro. (*A*) Schematic diagram of DNA template used for RNApII EC isolation and analysis. (*B*) Experimental procedure for purification of RNApII ECs. PICs were preassembled on bead-immobilized DNA templates for 30 min in yeast nuclear extract with Gal4-VP16 in either the absence or presence of 1 μg/mL α-amanitin. Transcription was initiated by addition of NTPs (ATP, CTP, UTP, and 3′-Ome-GTP). After 15 min of transcription, downstream bound proteins were released by SspI digestion and subjected to online 3D chromatography and LC-MS/MS with 4-plex iTRAQ reagents (Ficarro et al. 2011). (*C*) Log$_2$ scaled ±NTP ratios of identified proteins were compared in the absence (*x*-axis) or presence (*y*-axis) of α-amanitin. Each circle represents a single protein, with the value calculated from the sum of all peptide signals assigned to the protein (see details in the Materials and Methods). Color-coding is used to show members of specific complexes or functional subgroups as noted. Proteins specifically enriched by active transcription are defined as those at the *right* of the vertical blue dotted line on the *x*-axis (95% probability for enrichment with NTPs) and *below* the diagonal blue dotted line (95% probability for inhibition by α-amanitin). (*D*) Heat map shows relative binding of transcription-dependent proteins. Signals were normalized by setting the −NTPs without α-amanitin channel to 1.0. See Supplemental Figure 1b and Supplemental Table 1 for additional data.

separate report, we show that these TAFs are not associated with ECs. Instead, TAFs interact posttranscriptionally with sequences just downstream from the TSS to promote reinitiation (Joo et al. 2017).

Although naked DNA templates were used in this experiment, several chromatin-related factors showed transcription-dependent enrichment. Histone chaperone Spt6 and the histone H3 K36 methyltransferase Set2 interact with phosphorylated Rpb1 (Li et al. 2007; Yoh et al. 2007; Sdano et al. 2017). The Eaf3 chromodomain protein, which binds methylated H3 K36 as part of the NuA4 and Rpd3S complexes (Li et al. 2007), and the H3 K36 demethylase Rph1 were also isolated in this group. Finally,

binding of four subunits of Swi/Snf (Swi1, Snf2, Swi3, and Snf6) is reduced by α-amanitin, although not to the same extent as RNApII or other EC components. While normally thought of as being targeted to promoters by activator, there is some evidence for Swi/Snf association with elongation (Schwabish and Struhl 2007). Histones were detected in all four samples. Although these are often present as abundant contaminants in MS-based experiments, we cannot rule out some chromatinization of the DNA templates in the extract.

Several RNA-binding proteins were identified as transcription-associated factors. One is the mRNA cap methyltransferase Abd1. Unlike the capping enzyme

guanyltransferase Ceg1 and triphosphatase Cet1, Abd1 cross-links beyond promoters throughout transcribed regions (Komarnitsky et al. 2000; Schroeder et al. 2000). The hnRNP K protein Hek2/Khd1 binds many mRNAs through a CXXC motif (Hasegawa et al. 2008) that is also present in the G-less cassette. Hek2 coimmunoprecipitates with RNApII (Harlen et al. 2016). Also seen was the RNA-binding protein Ssd1. Ssd1 binds CTD in vitro and shows genetic interactions with many transcription-related factors (Phatnani et al. 2004). Although these RNA-binding proteins might interact with EC proteins, we think it more likely that they bind the nascent transcript. The unusual base composition of the G-less transcript may explain the absence of better-known mRNA-binding factors such as Npl3 and Hrp1. Future experiments with genes containing introns and polyadenylation sequences will be informative.

Compiled results from three biological repeats, each with two technical repeats, show that the immobilized template approach reliably and reproducibly isolates promoter-driven RNApII ECs in nuclear extracts (Supplemental Fig. 1e). Thus, quantitative MS provides sensitive and accurate identification of EC-associated factors even within a complex mixture. The ITA/MS approach complements work with purified factors or in vivo analyses such as ChIP.

RNApII CTD cycle progression is a function of time, not distance

The G-less cassette template used here stalls RNApII 289 bp downstream from the major proximal *CYC1* TSS (Hahn et al. 1985). Based on ChIP experiments (Komarnitsky

et al. 2000; Mayer et al. 2010), we expected ECs isolated at this position to be in the early stages, with high Ser5P and low Ser2P. Contrary to our expectations, identified proteins did not include early binding factors such as capping enzyme but did include the CTD Ser2P-binding protein Set2. We realized that, while ChIP experiments measure EC association as a function of location on the gene, factor exchanges might reflect time elapsed after initiation rather than the distance traveled.

To test whether we had missed early stage complexes, PICs were assembled on bead-immobilized templates as before, and ECs were isolated over a time course of 0, 1, 2, 4, 8, and 16 min after NTP addition (Fig. 2A). CTD phosphorylation was monitored by immunoblotting with specific antibodies, probing either total bead-associated proteins (Fig. 2B [top panels], quantitated in C) or SspI-released downstream DNA (Supplemental Fig. 2a). As PICs form with nonphosphorylated polymerase, neither Ser5P nor Ser2P was detected before NTP addition (0 min). Remarkably, Ser5P increased sharply to its maximum level within the shortest time tested (1 min). After 2 min, it began to diminish until the last point assayed (16 min). Conversely, very low levels of Ser2P were detected 1 min after the start of transcription and increased gradually, reaching maximum level at the final time point. Total Rpb1 binding declined slightly during this time. Therefore, yeast nuclear extracts can reproduce the progression of CTD phosphorylations seen in vivo.

To demonstrate that these changes occur in stalled ECs, gel electrophoresis was used to analyze radioactive transcripts retained on beads or released into the supernatant (Fig. 2B [bottom panels], quantitated in C). The bulk of
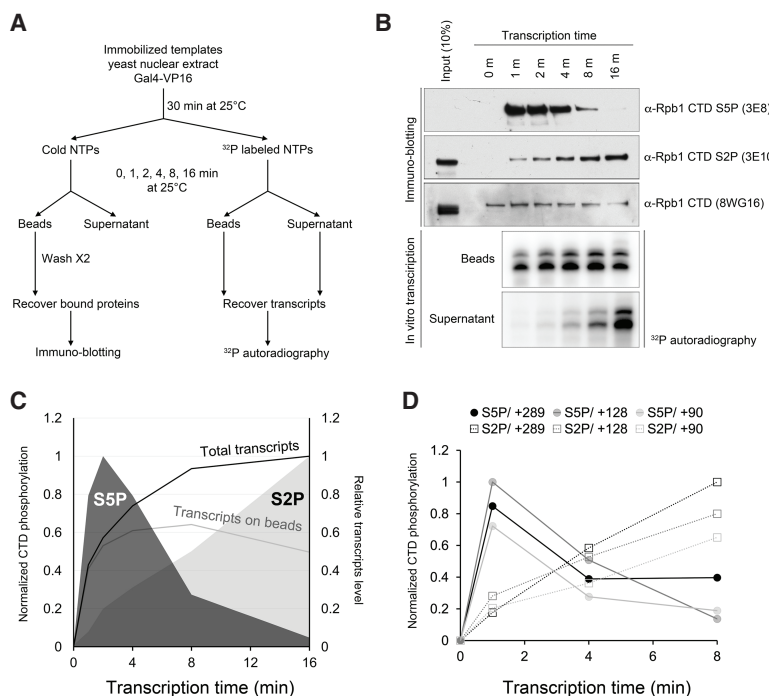

Figure 2. Time-dependent progression of CTD phosphorylation in vitro. (*A*) Experimental workflow of the time-course experiment for in vitro transcription and ITA. After PIC preassembly, samples were split in half. Transcription was started with NTPs (ATP, CTP, UTP, and 3'-Ome-GTP). In one half reaction, ^{32}P-labeled UTP was included for transcript detection. Proteins bound to the immobilized template were recovered magnetically after 0, 1, 2, 4, 8, and 16 min. (*B*, *top* panels) Samples with the nonlabeled half reaction were immunoblotted with antibodies against "total" Rpb1 CTD (8WG16, which recognizes repeats not phosphorylated on Ser2), CTD Ser5P (3E8), or CTD Ser2P (3E10), with 10% of input extract used as a marker. No signal was detected for Ser7P (4E12). (*Bottom* panels) With the ^{32}P-labeled half reactions, nascent (beads) and released (supernatant) transcripts were recovered and detected by autoradiography. (*C*) Levels of the two CTD phosphorylations, normalized to total CTD, were quantitated and graphed as a function of time. Relative levels of total transcripts and nascent transcripts on beads were quantitated by phosphorimager densitometry and are presented in the same graph. (*D*) CTD phosphorylations were measured after 0, 1, 4, and 8 min of in vitro transcription on different length G-less templates (90, 128, and 289 nt) as in *B* and *C*. See Supplemental Figure 2b for additional information.

full-length transcription was observed within the first 4 min, with a slow increase thereafter that may be due to later initiation events. Binding of the EC at the stall site was stable for up to 8 min. Some transcript release was seen at 4 and 8 min and was quite appreciable by 16 min, when Ser2P was highest. While the G-less cassette does not contain any of the known termination consensus sequences, EC-associated Rad26, Nab3–Sen1, and cleavage/polyadenylation factors provide candidate release mechanisms (see below; Fig. 1). Importantly, the kinetics of CTD phosphorylation changes were similar when ECs were stalled at +90, +128, or +289 (Fig. 2D; Supplemental Fig. 2b) but not when PICs were treated with ATP alone (Supplemental Fig. 2c).

These results lead to several conclusions. First, the CTD phosphorylation cycle, previously inferred from ChIP experiments, can be reproduced in vitro. This system will facilitate future studies of the relevant kinases and phosphatases. Second, the progression from Ser5P to Ser2P continues despite the lack of forward movement in stalled ECs. This finding complements ChIP experiments showing that RNApII mutants with accelerated or slowed polymerization cause CTD phosphorylation peaks to shift slightly downstream or upstream, respectively, in vivo (Fong et al. 2017; Soares et al. 2017). Together, these results argue that the CTD cycle is generated as a function of time rather than distance from the start of transcription.

Temporally regulated exchange of EC components

To analyze changes in EC components, ITA/MS was performed over a time course similar to that above (Fig. 3A). Over 1000 proteins were identified on the downstream template after NTP addition, with most showing no significant changes (Supplemental Fig. 3a). As in Figure 1, binding of the basal transcription factors, NC2, and Mot1 decreased (Supplemental Fig. 3c). In contrast, addition of NTPs increased binding of ~10% (105) of detected proteins (Supplemental Fig. 3b). Time-dependent changes in NTP-enriched proteins were quantified by calculating the slope of a line through their post-NTP time-point values (Figs. 3, 4, shown as green/pink scale; Supplemental Table 1). Two biological repeats, each with two technical repeats, showed excellent reproducibility (Supplemental Fig. 4a,b).

RNApII subunits appeared downstream within 1 min and remained steady during the entire reaction (Fig. 3B, C), consistent with levels of nascent transcripts on the beads (Fig. 2B,C). Known elongation factors Paf1C, DSIF (Spt4–Spt5), Spt6–Spn1, and Elf1 were also rapidly and strongly recruited by NTPs, and the ratio of these factors to RNApII remained relatively constant throughout the reaction (Fig. 3B,C). TFIIS (Dst1) also increased after NTP addition but not above the level of statistical significance. Structural studies show that these factors bind directly to RNApII (Ehara et al. 2017; Xu et al. 2017b; Vos et al. 2018). They also cross-link throughout transcribed regions in ChIP studies, and mutations cause phenotypes associated with elongation defects. Based on
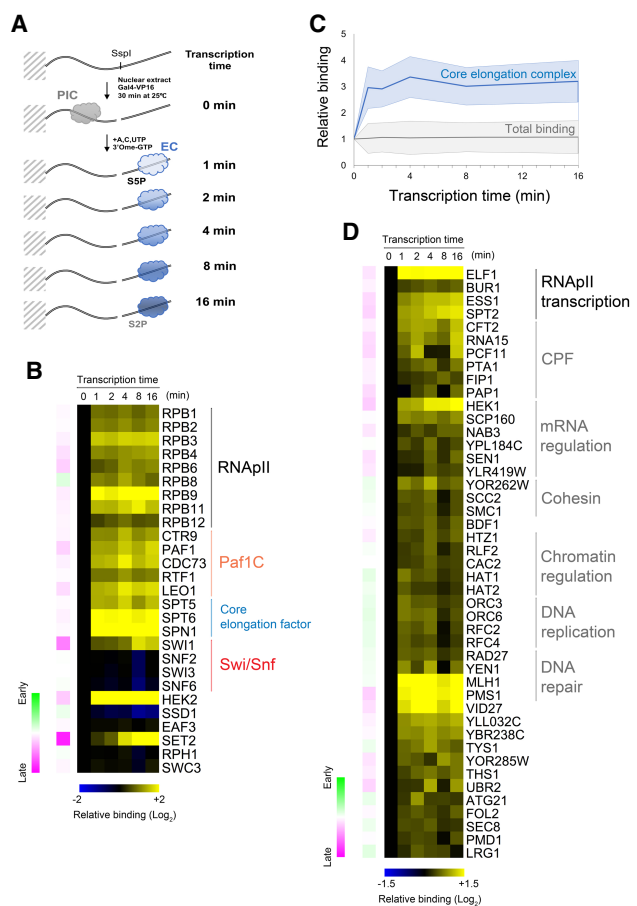

Figure 3. Time-course analysis showing EC factor exchange. (*A*) Schematic diagram of the time-course experiment. After preassembling PICs, transcription was started by NTPs (A, C, UTP, and 3′-Ome-GTP). Downstream bound proteins were eluted by SspI digestion at 0, 1, 2, 4, 8, or 16 min after transcription start. Eluted proteins were analyzed by 6-plex TMT-labeled quantitative 3D LC-MS/MS. (*B*) Heat map shows protein levels relative to the 0 time point (blue/yellow scale) and slope of change for the 1- to 16-min time points (pink/green scale, with pink signifying an increase and green signifying a decrease over time). (*C*) Normalized and averaged signals for RNApII subunits and core elongation factors are graphed (blue line) with standard deviation (blue shading), compared with total identified proteins (gray line and shading). (*D*) Heat map shows relative levels and slope of change for additional proteins significantly and consistently enriched downstream by NTPs. These proteins are defined as having >95% probability for enrichment by NTPs (averaged for 1-, 2-, 4-, 8-, and 16-min time points) but <95% probability for significant slope change compared with all other proteins. See Supplemental Figure 3, b and c, and Supplemental Table 1 for additional data.

these criteria, we designate this group of EC-associated proteins as "core elongation factors."

Other factors that showed increased binding at all time points include the CTD prolyl isomerase Ess1, the Cdk9 homolog Bur1/Sgv1, and Spt2/Sin1—all thought to function during elongation (Fig. 3D). We did not detect the CTD Ser2 kinase Ctk1/Cdk12 or the CTD phosphatases
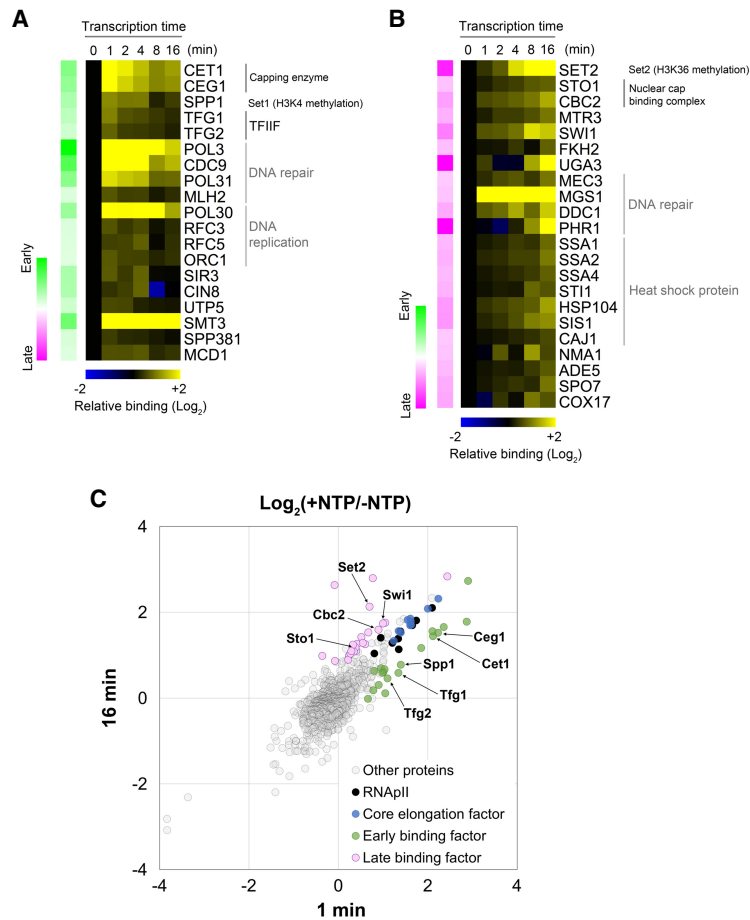
Figure 4. EC-associated factors that change over the time course. Heatmaps show relative protein levels and slope of change for early-enriched (*A*; strong positive slope; green) and late-enriched (*B*; strong negative slope; pink) factors from the quantitative MS experiment shown in Figure 3. Differential binding factors are defined as those having 95% probability for being enriched by NTPs and 95% probability for the slope being outside the distribution of slopes for all proteins. (*C*) Scatter plot showing $\log_2$ scaled ratio for total proteins at the early (1 min on *x*-axis) and late (16 min on *y*-axis) time points relative to the 0 time point. Each circle represents a single protein. The early (green) and late (pink) binding factors shown in *A* and *B* are color-coded along with RNApII (black) and core elongation factors (blue). See also Supplemental Figure 3b and Supplemental Table 1.

Ssu72, Rtr1, or Fcp1. As these are presumably required for progression of CTD phosphorylations, these proteins may be below the level of detection or, like many kinases and phosphatases, may bind their substrates only transiently. Addition of NTPs also increased levels of the SUMO protein Smt3 and many DNA repair factors. However, their binding was not sensitive to α-amanitin and therefore independent of transcription (Fig. 1). Interestingly, although FACT (Spt16–Pob3), Chd1, SAGA, the ISWI complexes, Def1, and Elongator have been implicated in elongation and many peptides from these factors were detected, none of them showed EC enrichment on naked DNA templates.

Multiple RNA-binding factors were again enriched upon NTP addition despite nonspecific competitor RNA being present in the reaction (Fig. 3D). Most strongly recruited were the KH domain proteins Hek2 and Pbp2/Hek1, the yeast homologs of hnRNP K protein. A third KH domain protein, Scp160, was also enriched, as was Mrn1/YPL184C, a protein implicated in mRNA maturation. Other mRNA-processing factors enriched by NTPs include multiple cleavage and polyadenylation factors (Fig. 3D). Lacking a polyadenylation site, there is no apparent cleavage of the G-less RNA transcript (Fig. 2B), and no enrichment of the "torpedo" termination factors Rtt103, Rat1, or Rai1 was detected. The Nab3 and Sen1 proteins, which cooperate with Nrd1 to terminate transcription

on snoRNAs and other small noncoding transcripts (Steinmetz et al. 2001), also scored as EC-associated in this experiment. As in Figure 1, known mRNA-binding proteins such as Hrp1, Npl3, Yra1, Mex67, or THO/TREX were not enriched, suggesting that the ECs generated in vitro on this template lack the appropriate RNA sequences or some other factor needed for their stable binding. In the future, it will be informative to analyze ECs on templates containing introns, polyadenylation sites, or Nrd1/Nab3 terminators.

Some proteins showed strong differential binding over time, as calculated from the time-course slope (Fig. 4C; Supplemental Table 1). For example, mRNA capping enzyme (a heterodimer of the triphosphatase Cet1 and guanylyltransferase Ceg1) showed strongest binding immediately after addition of NTPs (Fig. 4A, 1 min). Previous immobilized template experiments showed that ATP alone triggers CTD phosphorylation and capping enzyme association with PICs (Cho et al. 1997). Paralleling Ser5P (Fig. 2B), capping enzyme association gradually decreased.

Several other proteins related to gene expression showed strong binding to early ECs (Fig. 4A). In marked contrast to other basal initiation factors (Supplemental Fig. 3d), TFIIF (a heterodimer of Tfg1 and Tfg2) was enriched at early time points. While ChIP experiments do not detect TFIIF cross-linking throughout transcribed regions, our reanalysis of ChIP-exo data (Rhee and Pugh

2012) found that TFIIF and TFIIH, but no other basal factors, show a shoulder of cross-linking within the first 100 transcribed base pairs (Supplemental Fig. 3e). While we cannot rule out an interaction of TFIIF with the TAFs that remain bound to downstream promoter sequences (Joo et al. 2017), our results suggest TFIIF can remain bound to RNApII during very early elongation. TFIIF affects promoter clearance, can bind RNApII ECs, and stimulates elongation in vitro (Thomas and Chiang 2006). RNApI and RNApIII have TFIIF-like integral subunits that presumably remain associated throughout transcription (Vannini and Cramer 2012), so dissociation of TFIIF soon after promoter escape may have evolved as an RNApII-specific feature.

Also showing early binding is the Spp1 subunit of the Set1/COMPASS H3K4 histone methyltransferase, which binds the Ser5P-associated CTD. Other COMPASS subunits were enriched in only a subset of replicates, so COMPASS association may be weaker than other interactors. The Sub1/PC4 protein, which associates with PICs and perhaps the transcription bubble (Sikorski et al. 2011), was observed at early time points just below the statistical cutoff for 95% confidence.

Factors preferentially binding late ECs included cap-binding complex (CBC; a heterodimer of Sto1 and Cbc2) and the histone methyltransferase Set2 (Fig. 4B). Set2 directly binds the Ser2P CTD, explaining the increase over time. The CBC is not known to directly bind the CTD, so interaction is most likely with the newly formed cap structure of the nascent mRNA, derived from the 3′-Ome-GTP in the reaction (K. Murakami, pers. comm.). Multiple experiments suggest that the CBC functions in

ECs: CBC ChIPs to transcribed regions and strong genetic interactions of CBC genes with Set2 and other elongation factors have been reported (Hossain et al. 2013). Late binding was also seen for Fkh2, another protein implicated in elongation (Morillon et al. 2003). The exosome subunit Mtr3 also showed a significant increase over time, although other subunits of the complex were not enriched by NTPs. Several heat-shock and DNA repair proteins also increased, but their enrichment was also seen in the absence of transcription (Fig. 1).

Kin28/Cdk7 kinase activity is required for both Ser5 and Ser2 phosphorylation

The time-course experiment showed clear correlations between CTD phosphorylations and differential EC binding of several factors. To directly test the role of CTD phosphorylation, we used the Kin28is double mutant (L83G, V21C), which can be irreversibly inactivated by the chemical inhibitor CMK (Rodríguez-Molina et al. 2016). As reported previously, CMK strongly inhibited growth of Kin28is, but not wild-type, cells (Supplemental Fig. 5a). Upon treatment of Kin28is nuclear extract with CMK, in vitro phosphorylation of Ser5 was effectively blocked (Fig. 5A; Supplemental Fig. 5b).

Surprisingly, Ser2P was also completely blocked by Kin28 inhibition. This was unexpected, as in vivo inhibition of Kin28/Cdk7 in budding or fission yeast cells has been reported to have little effect on Ser2P (Tietjen et al. 2010; Bataille et al. 2012; Mbogning et al. 2015; Rodríguez-Molina et al. 2016; Booth et al. 2018). By comparison, both Ser5P and Ser2P were partly reduced upon Cdk7
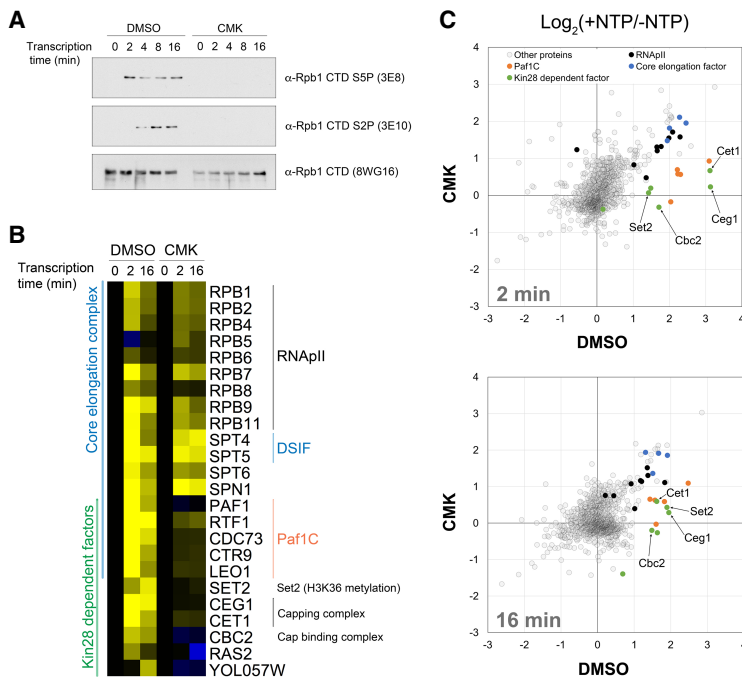

Figure 5. Kin28/Cdk7 activity is required for CTD phosphorylation and association of a subset of EC factors. (*A*) PICs were assembled in Kin28is nuclear extract in the presence of 250 nM CMK inhibitor or DMSO solvent. Total bound proteins were recovered 0, 2, 4, 8, and 16 min after transcription initiation with NTPs (A, C, UTP, and 3′-Ome-GTP). CTD phosphorylation was monitored with specific antibodies against CTD Ser5P (3E8), CTD Ser2P (3E10), and "total" CTD (8WG16). (*B*) Reactions were performed as in *A*, except proteins bound to template were eluted by PstI digestion at 0, 2, or 16 min after transcription start. Proteins were subjected to 6-plex TMT-labeled quantitative 3D LC-MS/MS analysis. The heat map shows relative levels for RNApII subunits, core elongation factors, and Kin28-dependent factors. Color scale is as in Figure 4. Kin28-dependent factors were defined as those with at least 95% confidence level for enrichment by NTPs (log$_2$ of the 2- or 16-min +DMSO signal divided by the 0 time point signal) and also a 95% confidence level for significant reduction upon CMK treatment [log$_2$ (2 or 16 min/0 min) with DMSO −log$_2$ (2 or 16 min/0 min) with CMK inhibition]. Note that Ras2 and YOL057W did not score as transcription-enriched in other experiments and are likely false positives. (*C*) Scatter plots showing log$_2$ scaled ±NTP ratios of 1240 identified proteins in the presence of DMSO (*x*-axis) versus 250 nM CMK (*y*-axis) during either early elongation (2 min; *top* panel) or late elongation (16 min; *bottom* panel). Colors designate RNApII subunits (black), core elongation factors (blue), Paf1C subunits (orange), and other Kin28-dependent factors (green). See also Supplemental Figure 5 and Supplemental Table 1.

inhibition in mammalian cells carrying a mutant kinase sensitized to a noncovalent inhibitor (Larochelle et al. 2012; Ebmeier et al. 2017), perhaps due to incomplete kinase inactivation or other kinases. While the covalent Cdk7 inhibitor THZ1 strongly blocked both Ser5 and Ser2 phosphorylation in vivo (Kwiatkowski et al. 2014) and in vitro (Nilson et al. 2015), this chemical also inhibits the Ser2 kinase Cdk12.

Loss of Ser2P upon Kin28is inactivation is not due to absence of transcription (Fig. 6A). It is unlikely that Kin28 directly phosphorylates Ser2, given its preference for Ser5 (Komarnitsky et al. 2000; Eick and Geyer 2013). Presumably, Kin28 activity on the CTD or some other substrate promotes Ser2 phosphorylation by either recruiting or activating the Ser2 kinases (Larochelle et al. 2012; Eick and Geyer 2013). Although the mechanism remains unclear, the requirement of Kin28 activity for subsequent Ser2 phosphorylation argues for an obligatory sequential sequence of CTD modifications.
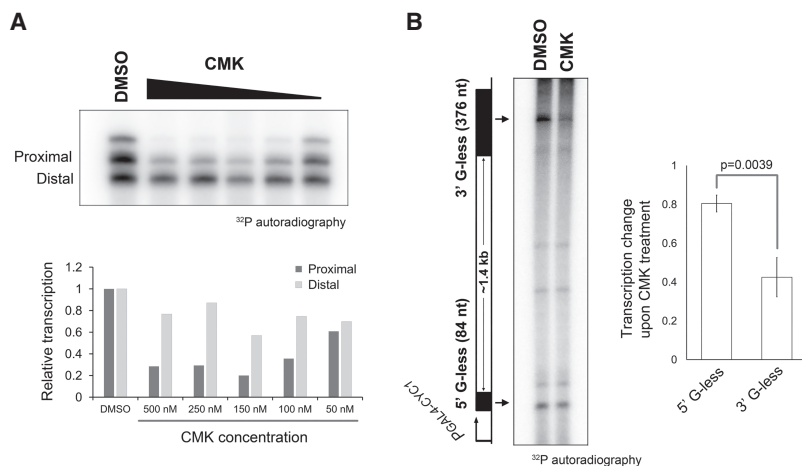
Kin28/Cdk7 kinase activity is required for recruiting a subset of EC components

To identify which EC components are dependent on CTD phosphorylation, Kin28is nuclear extract was treated with CMK or solvent (DMSO) for 30 min during PIC formation on immobilized templates. ECs were recovered by PstI digestion at 0, 2, and 16 min after addition of NTPs, followed by multiplexed quantitative MS (Fig. 5B,C; Supplemental Table 1). NTP-dependent stimulation of RNApII and core elongation factors Spt4–Spt5, Spt6, and Spn1 remained high. Only a limited set of EC factors was strongly reduced upon Kin28is inhibition (Fig. 5B,C). The two mRNA capping enzyme subunits and Set2 topped the list, as expected from their interactions with Ser5P and Ser2P, respectively. CBC was also reduced, presumably due to the lack of capping. Interestingly, Spt6 was unaffected, consistent with recent data suggesting that it binds

to the Rpb1 linker rather than the CTD (Yoh et al. 2007; Sdano et al. 2017; Vos et al. 2018). Finally, all five subunits of Paf1C were strongly reduced by Kin28 inhibition.

The drop in Paf1C was unexpected because Paf1C ChIP is independent of Ser2P phosphorylation by Ctk1, yet its cross-linking pattern does not match Ser5P (Ahn et al. 2004). Furthermore, recent structures show Paf1C interacting with the body of RNApII, with no apparent CTD contacts (Ehara et al. 2017; Xu et al. 2017b; Vos et al. 2018). Paf1C can bind the phosphorylated CTD in vitro (Phatnani et al. 2004; Qiu et al. 2012), but Paf1C recruitment to ECs is thought to primarily involve binding to an Spt5 domain that is phosphorylated by Bur1/Cdk9 early in elongation (Liu et al. 2009; Zhou et al. 2009). Qiu et al. (2009, 2012) showed that this Spt5 phosphorylation is blocked by Kin28 inhibition and suggested that this effect was due to Bur1 directly binding CTD Ser5P. However, Bur1–Bur2 association with ECs was not blocked by CMK in our experiment (Supplemental Table 1). Therefore, how Kin28 activity promotes Paf1C binding remains an open question.

Loss of Paf1C from the EC is predicted to affect elongation. Indeed, Kin28 inhibition dramatically reduced the ratio of distal to proximal RNA on a long in vitro transcription template (Fig. 6B). Paf1C loss may also explain the elongation defect seen upon in vivo inhibition of Kin28 (Kim et al. 2010; Rodríguez-Molina et al. 2016; Booth et al. 2018). Interestingly, the small drop in transcription upon CMK treatment, seen with Kin28is but not wild-type nuclear extract, primarily reduced the proximal *CYC1* TSS (Fig. 6A; Supplemental Fig. 6a,b). Similar TSS effects are seen with certain TFIIB mutants and RNApII mutants that slow elongation rate (Kaplan et al. 2012). We note that Murakami et al. (2015) also reported a role for Kin28 in determining TATA box–TSS spacing, but that function is clearly distinct, as it shifts TSSs downstream and is kinase-independent. Future experiments will explore the roles of Kin28 kinase activity in TSS selection.

Figure 6. Kin28 kinase inactivation affects TSS selection and elongation processivity. (*A*) Kin28is nuclear extract was preincubated with the indicated concentrations of CMK for 10 min followed by in vitro transcription on template pUC18-G5CYC1 G-(SB649). (*Top* panel) Labeled transcripts were separated on a 6% urea–acrylamide and detected using autoradiography. The *bottom* panel shows normalized quantitation of transcripts from proximal and distal start sites. Note that the *top* band results from readthrough transcription through the full G-less cassette. (*B*) In vitro transcription was performed with Kin28is nuclear extract preincubated with 500 nM CMK or DMSO solvent. The template was pGCYC1-402 (F647) (Rondon et al. 2003), which produces a 1.4-kb transcript containing two different length G-less cassettes—one near the 5′ end and one near the 3′ end. (*Left* panel) After RNase T1 digestion, resistant fragments were separated on a 6% urea–acrylamide gel. Relative change upon CMK treatment for each fragment is shown in the *right* panel. Error bars designate standard deviation in three independent reactions, calculated using the Student's *t*-test. See also Supplemental Figure 6.

Discussion

Using quantitative proteomics to analyze RNApII complexes on immobilized template DNA, we show that yeast nuclear extract reproduces many key aspects of elongation, including CTD phosphorylation changes and dynamic exchange of associated factors. We identified both "core elongation factors" that track with RNApII levels at all time points (Figs. 1, 3) and other EC proteins that change over time (Fig. 4). Remarkably, ECs continue progression through the CTD cycle even when elongation is stalled (Fig. 2), demonstrating that these transitions are a function of time, rather than distance traveled, after initiation. Unless blocking mechanisms exist, ECs stalled at DNA damage or pause sites may progress to late phosphorylation patterns, perhaps promoting termination or displacement of the EC.

The ITA/MS approach will be very useful for analyzing the dynamics of RNApII elongation. The ability to isolate and characterize in vitro ECs at various postinitiation times provides a powerful system for studying the interplay between CTD modifications, elongation factors, and the DNA template.

Materials and methods

Yeast strains, plasmids, and oligonucleotides

Yeast strains, oligonucleotides, and plasmids used this study are listed in Supplemental Tables 2–4.

Yeast nuclear extract preparation and ITA

Nuclear extracts were prepared as described previously (Sikorski et al. 2012) from wild-type (BY4741 or YF702) or Kin28is (YSB3566) yeast strains. ITAs for ECs were performed as described for PICs (Sikorski et al. 2012) with the modifications described below. Biotinylated DNA templates (Fig. 1A) were amplified by PCR from pUC18-G5CYC1 G-(SB649) with primers O#3151 and O#1477. Naked templates (200 ng per 1× reaction) were immobilized onto 50 µg/5 µL Dynabeads Streptavidin T1 slurry (Invitrogen) in 1× TE (10 mM Tris-HCl at pH 8, 1 mM EDTA) supplemented with 1 M NaCl for 1 h at 25°C with rotation. To reduce background binding, template-coupled beads were blocked for 30 min at 25°C with rotation in 50 µL of transcription buffer (20 mM HEPES-KOH at pH 7.6, 100 mM potassium acetate, 1 mM EDTA, 5 mM magnesium acetate) complemented with 60 mg/mL casein, 5 mg/mL polyvinylpyrrolodone, and 2.5 mM DTT. After washing three times in 400 µL of transcription buffer, the immobilized templates were resuspended in 10 µL of transcription buffer. Recombinant transcription activator Gal4-VP16 (400 ng) was preincubated with immobilized templates for 5 min at 25°C prior to the addition of yeast nuclear extract. Extract (~1 mg) and remaining components (20 U of Rnasin, 2 U of creatine phosphokinase, 10 mM phospho-creatine, 0.03% NP-40, 16.67 µg/mL tRNA, 160 µM S-adenosylmethionine, 20 µM acetyl-CoA, 16.67 µg/mL HaeIII-digested *Escherichia coli* genomic DNA) were added to a total reaction volume of 60 µL. After 30 min of incubation at 25°C to form PICs, transcription was initiated by adding 2.4 µL of NTP mix (final concentration of 400 µM each ATP, CTP, and UTP and 40 µM 3′-Ome-GTP) to the mixture followed by an incubation at 25°C with occasional gentle mixing every 5 min to keep beads suspended. After the time indicated, immobilized templates were recovered with a magnetic stand, and

the reaction was stopped by quickly washing twice in 200 µL of wash buffer (transcription buffer complemented with 0.05% NP40, 2.5 mM DTT). For analysis of total bound proteins, samples were recovered by boiling with 40 µL of transcription buffer and 40 µL of 2× SDS loading dye (100 mM Tris-HCl at pH 6.8, 4% SDS, 0.2% Bromophenol Blue, 20% glycerol, 200 mM DTT). To analyze proteins bound to specific regions on the DNA template, complexes were eluted for 30 min at 25° with rotation in 40 µL of the transcription buffer with 60 U of either SspI-HF or PstI-HF (New England Biolabs) as indicated (see Fig. 1).

Sample preparation for quantitative MS

Quantitative MS was performed as described previously with minor modifications (Sikorski et al. 2012). Reactions were the same as described above for ITA but scaled up fivefold (300-µL reaction volume and 200 µL of final elute). EC formation was validated by immunoblotting 40 µL of the final restriction enzyme eluate with specific antibodies against Rpb1-CTD (8WG16), Rpb1-CTD Ser5P (3E8), and Rpb1-CTD Ser2P (3E10). The remaining eluate (160 µL) was processed for MS. Ammonium bicarbonate was added to a final concentration of 50 mM to each sample. Protein cysteines were reduced with 10 mM DTT for 1 h at 56°C and then alkylated with 55 mM iodoacetamide for 1 h at 25°C in the dark. Trypsin was added at 1:20 (enzyme:substrate) ratio. After incubation overnight at 37°C, samples were acidified with trifluoroacetic acid (TFA) to pH ~2 (measured by pH strip) and microfuged at 13,000 rpm to remove insoluble particles. Tryptic peptides were desalted and purified on a C18 solid phase extraction cartridge, eluted with 80% acetonitrile/0.1% TFA, and dried. For 4-plex iTRAQ reagent, each dried peptide pellet was resuspended in 30 µL of 0.5 M triethylammonium bicarbonate, and then 70 µL of 100% ethanol was added. Each sample was mixed with a specific iTRAQ label and incubated for 1 h at room temperature. For 6-plex TMT labeling, each dried pellet was resuspended in 100 µL of 0.5 M triethylammonium bicarbonate. Each sample was then mixed with a specific TMT reagent reconstituted in 40 µL of acetonitrile followed by incubation for 1 h at room temperature. After labeling, samples were combined and desalted with a C18 solid phase extraction cartridge and dried. The final peptide pellet was reconstituted in 80 µL of 10 mM ammonium formate. Labeling efficiency of >95% was confirmed by 1D reverse-phase LC-MS screening before online 3D chromatography and LC-MS/MS. For each technical repeat, half of each sample was injected. To estimate reproducibility across biological replicates, ±NTP samples (15 min, naked DNA template) from three independent experiments (representing three biological replicates, each with two technical replicates) were analyzed. Over 700 proteins were identified in all six replicates (Supplemental Table 5). For each protein, the *P*-value for the six replicates (three biological times two technical repeats) falling inside the normal distribution for the entire data set was plotted against the +NTP enrichment score, showing very strong statistical significance for enrichment of elongation factors and the large majority of RNApII, Paf1C, and TFIID subunits (Supplemental Fig. 1e).

Online 3D chromatography and LC-MS/MS data acquisition

The 3D chromatography system used for peptide separation was essentially as described (Ficarro et al. 2011; Zhou et al. 2011). Our RP-SAX-RP (reverse phase-strong anion exchange-reverse phase) platform was based on a Waters NanoACQUITY UHPLC system equipped with a third, six-port, two-position valve (VICI-Valco). True nanoflow rates in the third dimension were achieved through use of a passive split located prior to the third-dimension PC. The valve then switched to engage the passive split such that

a majority of LC effluent was diverted to waste prior to the third-dimension PC. Thirteen RP-SAX-RP fractions were analyzed. Either an LTQ Velos (Fig. 1 only) or Q-Exactive HF (Thermo Scientific) machine was used to acquire MS/MS spectra in the data-dependent mode as described (Zhang et al. 2009) using the following parameters: MS spectra were acquired with a target value of 2×10^6, maximum ion injection time of 500 msec, and a mass resolution of 60,000 (FWHM at $m/z = 400$). A minimum threshold of 1×10^5 was used for MS/MS, with a charge state range of 2+ to 4+. Up to eight precursors could be selected for each cycle, with a 20-sec exclusion time for each precursor selected for MS/MS. One linear trap CAD and one HCD MS/MS spectrum were acquired in series for each precursor. An optimized collision energy was used for HCD scans.

MS/MS data files were directly accessed and converted to .mgf files with ytraq4plex.mz to extract CAD and HCD scans into separate files for subsequent MASCOT searches (Zhang et al. 2009). In addition, iTRAQ and TMT reporter ion peak intensity values were extracted from HCD scans and then inserted into the header for corresponding CAD and HCD scans of the same precursor; this providing for convenient retrieval of both peptide sequence identification and reporter ion quantification data from the MASCOT search results. The two files were searched using Mascot version 2.3.1 against a yeast database from the *Saccharomyces* Genome Database (downloaded April 2010), with an appended cRAP (common repository of adventitious proteins) database. The mass tolerance values were set to 10 ppm for each precursor, 0.6 Da for LTQ-CAD scans, and 25 mmu for HCD scans. After searching, an Excel spreadsheet containing the Mascot search results was generated using Multiplierz software (Askenazi et al. 2009; Parikh et al. 2009). Scans with FDR of <1% were extracted for further analysis, and HCD and CAD results were compressed into a single worksheet. For relative quantitation, iTRAQ or TMT reporter ion peak intensity values were extracted from HCD scans, corrected for known isotopic impurities, and included in the Multiplierz spreadsheet. The iTRAQ or TMT reporter signal intensity values of all peptide scans for a given protein were summed prior to calculation of ratios. Only peptides that were unique to the matched protein and detected in both replicates were used for quantitation. Detailed data and prenormalized iTRAQ/TMT values for all individual peptides in each replicate are in Supplemental Table 6. The total number of peptide sequence matches was used to normalize each channel, with restriction enzyme or bovine serum albumin peptides from the restriction buffer acting as a "spike-in" control. Normalized data were statistically analyzed using the mixed model analysis within the R software package. Using a null hypothesis of no differences between conditions, ratios for all proteins were fit to a normal distribution. The mixed model calculated the confidence level for individual protein ratios being outside the null hypothesis distribution.

In vitro transcription

For in vitro transcription on immobilized templates shown in Figure 2B, 400 ng of Gal4-VP16 and 200 ng of bead-linked DNA were mixed with ~1 mg of yeast nuclear extract using the same conditions as for the ITA. For transcript detection, ^{32}P-labeled NTPs (500 µM ATP, 500 µM CTP, 20 µM UTP, 50 µM of 3′-Ome-GTP, and 3 µCi of ^{32}P-UTP) were added. After the time indicated, beads were separated from the supernatant on a magnetic stand. RNA was recovered from both fractions and detected on 8 M urea–6% polyacrylamide gel. In vitro transcription on plasmids was performed as described previously (Sikorski et al. 2012). After a 15-min preincubation of CMK with 100 ng of template plasmid, 200 ng of Gal4-VP16, and ~500 µg of yeast nuclear extract, tran-

scription was initiated by adding ^{32}P-labeled NTPs (500 µM ATP, 500 µM CTP, 20 µM UTP, 50 µM 3′-Ome-GTP, and 3 µCi of ^{32}P-UTP) for 30 min at 25°C. For the in vitro elongation assay shown in Figure 6B, pGCYC1-402 (F647) plasmid was used (Rondon et al. 2003). For the CMK titration assay in Figure 6A, pUC18-G5CYC1 G-(SB649) plasmid was used.

References

Ahn SH, Kim M, Buratowski S. 2004. Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3′ end processing. *Mol Cell* **13:** 67–76. doi:10.1016/S1097-2765(03)00492-1

Askenazi M, Parikh JR, Marto JA. 2009. mzAPI: a new strategy for efficiently sharing mass spectrometry data. *Nat Methods* **6:** 240–241. doi:10.1038/nmeth0409-240

Bataille AR, Jeronimo C, Jacques P-É, Laramée L, Fortin M-È, Forest A, Bergeron M, Hanes SD, 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. doi:10.1016/j.molcel.2011.11.024

Booth GT, Parua PK, Sansó M, Fisher RP, Lis JT. 2018. Cdk9 regulates a promoter-proximal checkpoint to modulate RNA polymerase II elongation rate in fission yeast. *Nat Commun* **9:** 679. doi:10.1038/s41467-018-03006-4

Buratowski S. 2009. Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36:** 541–546. doi:10.1016/j.molcel.2009.10.019

Cho EJ, Takagi T, Moore CR, Buratowski S. 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11:** 3319–3326. doi:10.1101/gad.11.24.3319

Corden JL. 2013. RNA polymerase II C-terminal domain: tethering transcription to transcript and template. *Chem Rev* **113:** 8423–8455. doi:10.1021/cr400158h

Ebmeier CC, Erickson B, Allen BL, Allen MA, Kim H, Fong N, Jacobsen JR, Liang K, Shilatifard A, Dowell RD, et al. 2017. Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications. *Cell Rep* **20:** 1173–1186. doi:10.1016/j.celrep.2017.07.021

Ehara H, Yokoyama T, Shigematsu H, Yokoyama S, Shirouzu M, Sekine SI. 2017. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357:** 921–924. doi:10.1126/science.aan8552

Eick D, Geyer M. 2013. The RNA polymerase II carboxy-terminal domain CTD code. *Chem Rev* **113:** 8456–8490. doi:10.1021/cr400071f

Ficarro SB, Zhang Y, Carrasco-Alfonso MJ, Garg B, Adelmant G, Webber JT, Luckey CJ, Marto JA. 2011. Online nanoflow multidimensional fractionation for high efficiency phosphopeptide analysis. *Mol Cell Proteomics* **10:** O111.011064. doi:10.1074/mcp.O111.011064

Fong N, Saldi T, Sheridan RM, Cortazar MA, Bentley DL. 2017. RNA Pol II dynamics modulate co-transcriptional chromatin modification, CTD phosphorylation, and transcriptional direction. *Mol Cell* **66:** 546–557 e543. doi:10.1016/j.molcel.2017.04.016

Hahn S, Hoar ET, Guarente L. 1985. Each of three 'TATA elements' specifies a subset of the transcription initiation sites at the CYC-1 promoter of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci* **82:** 8562–8566. doi:10.1073/pnas.82.24.8562

Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. 2016. Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue. *Cell Rep* **15:** 2147–2158. doi:10.1016/j.celrep.2016.05.010

Hasegawa Y, Irie K, Gerber AP. 2008. Distinct roles for Khd1p in the localization and expression of bud-localized mRNAs in yeast. *RNA* **14:** 2333–2347. doi:10.1261/rna.1016508

Hossain MA, Chung C, Pradhan SK, Johnson TL. 2013. The yeast cap binding complex modulates transcription factor recruitment and establishes proper histone H3K36 trimethylation during active transcription. *Mol Cell Biol* **33:** 785–799. doi:10.1128/MCB.00947-12

Joo YJ, Ficarro SB, Soares LM, Chun Y, Marto JA, Buratowski S. 2017. Downstream promoter interactions of TFIID TAFs facilitate transcription reinitiation. *Genes Dev* **31:** 2162–2174. doi:10.1101/gad.306324.117

Kaplan CD, Laprade L, Winston F. 2003. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* **301:** 1096–1099. doi:10.1126/science.1087374

Kaplan CD, Jin H, Zhang IL, Belyanin A. 2012. Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. *PLoS Genet* **8:** e1002627. doi:10.1371/journal.pgen.1002627

Kim M, Ahn SH, Krogan NJ, Greenblatt JF, Buratowski S. 2004. Transitions in RNA polymerase II elongation complexes at the 3′ ends of genes. *EMBO J* **23:** 354–364. doi:10.1038/sj.emboj.7600053

Kim H, Erickson B, Luo W, Seward D, Graber JH, Pollock DD, Megee PC, Bentley DL. 2010. Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nat Struct Mol Biol* **17:** 1279–1286. doi:10.1038/nsmb.1913

Komarnitsky P, Cho EJ, Buratowski S. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14:** 2452–2460. doi:10.1101/gad.824700

Krogan NJ, Kim M, Ahn SH, Zhong G, Kobor MS, Cagney G, Emili A, Shilatifard A, Buratowski S, Greenblatt JF. 2002. RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol Cell Biol* **22:** 6979–6992. doi:10.1128/MCB.22.20.6979-6992.2002

Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, et al. 2003. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11:** 721–729. doi:10.1016/S1097-2765(03)00091-1

Kwiatkowski N, Zhang T, Rahl PB, Abraham BJ, Reddy J, Ficarro SB, Dastur A, Amzallag A, Ramaswamy S, Tesar B, et al. 2014. Targeting transcription regulation in cancer with a co-valent CDK7 inhibitor. *Nature* **511:** 616–620. doi:10.1038/nature13393

Larochelle S, Amat R, Glover-Cutter K, Sansó M, Zhang C, Allen JJ, Shokat KM, Bentley DL, Fisher RP. 2012. Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat Struct Mol Biol* **19:** 1108–1115. doi:10.1038/nsmb.2399

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* **128:** 707–719. doi:10.1016/j.cell.2007.01.015

Liu Y, Warfield L, Zhang C, Luo J, Allen J, Lang WH, Ranish J, Shokat KM, Hahn S. 2009. Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol Cell Biol* **29:** 4852–4863. doi:10.1128/MCB.00609-09

Mayer A, Lidschreiber M, Siebert M, Leike K, Söding J, Cramer P. 2010. Uniform transitions of the general RNA polymerase II transcription complex. *Nat Struct Mol Biol* **17:** 1272–1278. doi:10.1038/nsmb.1903

Mbogning J, Pagé V, Burston J, Schwenger E, Fisher RP, Schwer B, Shuman S, Tanny JC. 2015. Functional interaction of Rpb1 and Spt5 C-terminal domains in co-transcriptional histone modification. *Nucleic Acids Res* **43:** 9766–9775. doi:10.1093/nar/gkv837

Morillon A, O'Sullivan J, Azad A, Proudfoot N, Mellor J. 2003. Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science* **300:** 492–495. doi:10.1126/science.1081379

Murakami K, Mattei PJ, Davis RE, Jin H, Kaplan CD, Kornberg RD. 2015. Uncoupling promoter opening from start-site scanning. *Mol Cell* **59:** 133–138. doi:10.1016/j.molcel.2015.05.021

Nakanishi Y, Mitsuhashi Y, Sekimizu K, Yokoi H, Tanaka Y, Horikoshi M, Natori S. 1981. Characterization of three proteins stimulating RNA polymerase II. *FEBS Lett* **130:** 69–72. doi:10.1016/0014-5793(81)80667-9

Nilson KA, Guo J, Turek ME, Brogie JE, Delaney E, Luse DS, Price DH. 2015. THZ1 reveals roles for Cdk7 in co-transcriptional capping and pausing. *Mol Cell* **59:** 576–587. doi:10.1016/j.molcel.2015.06.032

Orphanides G, Reinberg D. 2000. RNA polymerase II elongation through chromatin. *Nature* **407:** 471–475. doi:10.1038/35035000

Parikh JR, Askenazi M, Ficarro SB, Cashorali T, Webber JT, Blank NC, Zhang Y, Marto JA. 2009. Multiplierz: an extensible API based desktop environment for proteomics data analysis. *BMC Bioinformatics* **10**. doi:10.1186/1471-2105-10-364

Phatnani HP, Jones JC, Greenleaf AL. 2004. Expanding the functional repertoire of CTD kinase I and RNA polymerase II: novel phosphoCTD-associating proteins in the yeast proteome. *Biochemistry* **43:** 15702–15719. doi:10.1021/bi048364h

Prather D, Krogan NJ, Emili A, Greenblatt JF, Winston F. 2005. Identification and characterization of Elf1, a conserved transcription elongation factor in *Saccharomyces cerevisiae*. *Mol Cell Biol* **25:** 10122–10135. doi:10.1128/MCB.25.22.10122-10135.2005

Qiu H, Hu C, Hinnebusch AG. 2009. Phosphorylation of the Pol II CTD by KIN28 enhances BUR1/BUR2 recruitment and Ser2 CTD phosphorylation near promoters. *Mol Cell* **33:** 752–762. doi:10.1016/j.molcel.2009.02.018

Qiu H, Hu C, Gaur NA, Hinnebusch AG. 2012. Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *EMBO J* **31:** 3494–3505. doi:10.1038/emboj.2012.188

Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483:** 295–301. doi:10.1038/nature10799

Rodríguez-Molina JB, Tseng SC, Simonett SP, Taunton J, Ansari AZ. 2016. Engineered covalent inactivation of TFIIH-kinase reveals an elongation checkpoint and results in widespread mRNA stabilization. *Mol Cell* **63:** 433–444. doi:10.1016/j.molcel.2016.06.036

Rondon AG, Garcia-Rubio M, Gonzalez-Barrera S, Aguilera A. 2003. Molecular evidence for a positive role of Spt4 in transcription elongation. *EMBO J* **22:** 612–620. doi:10.1093/emboj/cdg047

Schroeder SC, Schwer B, Shuman S, Bentley D. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev* **14:** 2435–2440. doi:10.1101/gad.836300

Schwabish MA, Struhl K. 2007. The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. *Mol Cell Biol* **27:** 6987–6995. doi:10.1128/MCB.00717-07

Sdano MA, Fulcher JM, Palani S, Chandrasekharan MB, Parnell TJ, Whitby FG, Formosa T, Hill CP. 2017. A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription. *Elife* **6:** e28723. doi:10.7554/eLife.28723

Sikorski TW, Ficarro SB, Holik J, Kim T, Rando OJ, Marto JA, Buratowski S. 2011. Sub1 and RPA associate with RNA polymerase II at different stages of transcription. *Mol Cell* **44:** 397–409. doi:10.1016/j.molcel.2011.09.013

Sikorski TW, Joo YJ, Ficarro SB, Askenazi M, Buratowski S, Marto JA. 2012. Proteomic analysis demonstrates activator- and chromatin-specific recruitment to promoters. *J Biol Chem* **287:** 35397–35408. doi:10.1074/jbc.M112.391581

Soares LM, He PC, Chun Y, Suh H, Kim T, Buratowski S. 2017. Determinants of histone H3K4 methylation patterns. *Mol Cell* **68:** 773–785. doi:10.1016/j.molcel.2017.10.013

Sopta M, Carthew RW, Greenblatt J. 1985. Isolation of three proteins that bind to mammalian RNA polymerase II. *J Biol Chem* **260:** 10353–10360.

Steinmetz EJ, Conrad NK, Brow DA, Corden JL. 2001. RNA-binding protein Nrd1 directs polyA-independent 3′-end formation of RNA polymerase II transcripts. *Nature* **413:** 327–331. doi:10.1038/35095090

Thomas MC, Chiang CM. 2006. The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol* **41:** 105–178. doi:10.1080/10409230600648736

Tietjen JR, Zhang DW, Rodríguez-Molina JB, White BE, Akhtar MS, Heidemann M, Li X, Chapman RD, Shokat K, Keles S, et al. 2010. Chemical-genomic dissection of the CTD code. *Nat Struct Mol Biol* **17:** 1154–1161. doi:10.1038/nsmb.1900

Vannini A, Cramer P. 2012. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol Cell* **45:** 439–446. doi:10.1016/j.molcel.2012.01.023

Vos SM, Farnung L, Boehning M, Wigge C, Linden A, Urlaub H, Cramer P. 2018. Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* **560:** 607–612. doi:10.1038/s41586-018-0440-4

Wada T, Takagi T, Yamaguchi Y, Ferdous A, Imai T, Hirose S, Sugimoto S, Yano K, Hartzog GA, Winston F, et al. 1998. DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev* **12:** 343–356. doi:10.1101/gad.12.3.343

Winston F, Chaleff DT, Valent B, Fink GR. 1984. Mutations affecting Ty-mediated expression of the HIS4 gene of *Saccharomyces cerevisiae*. *Genetics* **107:** 179–197.

Wood A, Schneider J, Dover J, Johnston M, Shilatifard A. 2003. The Paf1 complex is essential for histone monoubiquitination by the Rad6–Bre1 complex, which signals for histone methylation by COMPASS and Dot1p. *J Biol Chem* **278:** 34739–34742. doi:10.1074/jbc.C300269200

Xu J, Lahiri I, Wang W, Wier A, Cianfrocco MA, Chong J, Hare AA, Dervan PB, DiMaio F, Leschziner AE, et al. 2017a. Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature* **551:** 653–657. doi:10.1038/nature24658

Xu Y, Bernecky C, Lee CT, Maier KC, Schwalb B, Tegunov D, Plitzko JM, Urlaub H, Cramer P. 2017b. Architecture of the RNA polymerase II–Paf1C–TFIIS transcription elongation complex. *Nat Commun* **8:** 15741. doi:10.1038/ncomms15741

Yoh SM, Cho H, Pickle L, Evans RM, Jones KA. 2007. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev* **21:** 160–174. doi:10.1101/gad.1503107

Zhang L, Fletcher AG, Cheung V, Winston F, Stargell LA. 2008. Spn1 regulates the recruitment of Spt6 and the Swi/Snf complex during transcriptional activation by RNA polymerase II. *Mol Cell Biol* **28:** 1393–1403. doi:10.1128/MCB.01733-07

Zhang Y, Ficarro SB, Li S, Marto JA. 2009. Optimized Orbitrap HCD for quantitative analysis of phosphopeptides. *J Am Soc Mass Spectrom* **20:** 1425–1434. doi:10.1016/j.jasms.2009.03.019

Zhou K, Kuo WHW, Fillingham J, Greenblatt JF. 2009. Control of transcriptional elongation and cotranscriptional histone modification by the yeast BUR kinase substrate Spt5. *Proc Natl Acad Sci* **106:** 6956–6961. doi:10.1073/pnas.0806302106

Zhou F, Sikorski TW, Ficarro SB, Webber JT, Marto JA. 2011. Online nanoflow reversed phase-strong anion exchange-reversed phase liquid chromatography-tandem mass spectrometry platform for efficient and in-depth proteome sequence analysis of complex organisms. *Anal Chem* **83:** 6996–7005. doi:10.1021/ac200639v