Methods review: Mass spectrometry analysis of RNAPII complexes

Katlyn Hughes Burriss[a], Amber L. Mosley[a,b,*]

[a] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46402, United States
[b] Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN 46402, United States

A B S T R A C T

RNA Polymerase II (RNAPII) is responsible for transcribing multiple RNA species throughout eukaryotes. A variety of protein-protein interactions occur throughout the transcription cycle for coordinated regulation of transcription initiation, elongation, and/or termination. Taking a proteomics approach to study RNAPII transcription thereby offers a comprehensive view of both RNAPII biology and the variety of proteins that regulate the process itself. This review will focus on how mass spectrometry (MS) methods have expanded understanding of RNAPII and its transcription-regulatory interaction partners. The application of affinity purification mass spectrometry has led to the discovery of a number of novel groups of proteins that regulate an array of RNAPII biology ranging from nuclear import to regulation of phosphorylation state. Additionally, a number of methods have been developed using mass spectrometry to measure protein subunit stoichiometry within and across protein complexes and to perform various types of architectural analysis using structural proteomics approaches. The key methods that we will focus on related to RNAPII mass spectrometry analyses include: affinity purification mass spectrometry, protein post-translational modification analysis, crosslinking mass spectrometry, and native mass spectrometry.

1. Introduction

RNA Polymerase II (RNAPII) is responsible for transcribing multiple RNA species, including messenger RNA (mRNA), noncoding RNAs (ncRNAs), and small nuclear/nucleolar RNAs (sn/snoRNAs) [1,2]. In order to ensure proper transcription of these RNA species regulatory proteins are dynamically recruited to RNAPII. Transcription regulatory proteins may be recruited through interactions with the template DNA, the nascent RNA, or through protein-protein interactions (PPIs) within the transcription complex. The C-terminal domain (CTD) of the largest of the 12 RNAPII subunits, Rpb1, is a critical regulator of these PPIs through its dynamic phosphorylation.

Achieving a comprehensive understanding of transcription on the levels of DNA, RNA and protein requires diverse methodology. High-throughput sequencing method development has revolutionized the analysis of DNA and RNA and has greatly benefited the transcription field. These efforts have been driven by a wide variety of high-throughput nucleic acid sequencing approaches. Proteins are challenging to interrogate using similar high-throughput approaches as a consequence of their diversity in structure, size, and amino acid composition (including post-translational modifications). In recent years, proteomics and complementary structural mass spectrometry

approaches have continued to develop at a rapid rate spurred on by developments in mass spectrometry instrumentation, analysis software, and various chemical tools.

This review will focus on the application of mass spectrometry to the study of RNA Polymerase II (RNAPII) transcription. Mass spectrometry-based approaches to study RNAPII transcription offers a wholistic view of both the core RNAPII machinery and also the proteins that regulate overall transcription including RNA processing and chromatin biology. Therein, this review will discuss how mass spectrometry (MS) methods have expanded understanding of RNAPII and its transcription-regulatory interaction partners.

2. Overview of methods for MS analysis of RNAPII complexes

2.1. Analysis of RNAPII interactors by MS based methods
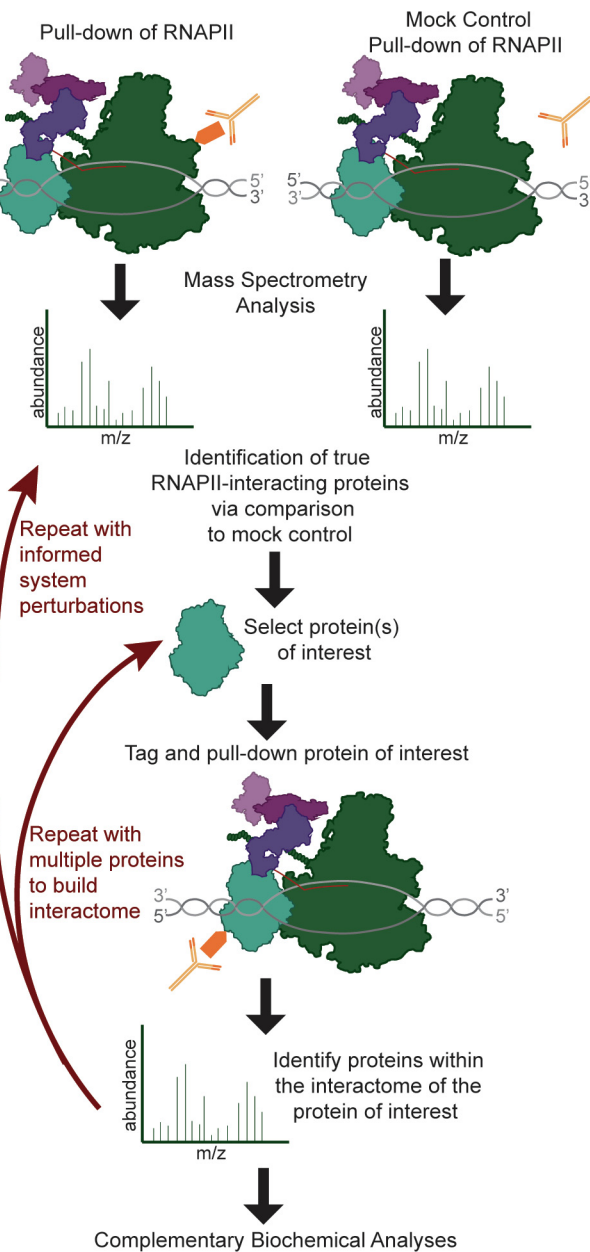
Identification of protein-protein interaction partners of the RNAPII machinery has been a longstanding interest of the molecular biology community. Many of the known RNAPII interacting proteins were isolated based on their affinity for RNAPII, with subsequent identification of the interacting proteins performed using approaches such as cDNA library screening [3]. However, mass spectrometry-based methods

K.H. Burriss and A.L. Mosley

Fig. 1. AP-MS of RNAPII and regulators to build the RNAPII interactome. MS/MS analysis of RNAPII purifications allows identification of RNAPII-interactors. Subsequent purifications of proteins of interest enable interaction mapping.

rapidly gained popularity after FASTA-based protein sequence database searching was enabled by the development of SEQUEST, and other algorithms, allowing researchers to establish a rapid and direct link between protein and gene sequences [4]. Using database sequence-based identification, subunits of large protein complexes, such as the chromatin/transcription related Spt-Ada-Gcn5-acetyltransferase (SAGA) complex, could be rapidly identified after biochemical purification [5]. Many of the proteins within the basic transcription machinery have now been identified as a consequence of wide-spread application of mass spectrometry to RNAPII transcription; however, adoption of new MS techniques and improved instrumentation continues to further the transcription field. MS approaches have not only confirmed and detailed suspected interactions between RNAPII and regulatory proteins, but have also continued to identify novel interactions. Specifically, affinity purification-mass spectrometry (AP-MS) has been a key approach in analyzing and mapping RNAPII protein-protein interactions (PPIs)

(Fig. 1).

The yeast transcription elongation complex has particularly benefitted from the application of AP-MS to studying RNAPII PPIs. Multiple groups used AP-MS pull-downs to identify that Set2, a methyl transferase, is a direct interactor of RNAPII [6–9]. Additional experiments showed that the Set2-RNAPII interaction is dependent upon a serine 2 phosphorylated (Ser2P) CTD. The discovery of this interaction was critical for understanding the coregulation of histone methylation and transcription elongation [6–9]. Similarly, Asr1, a yeast RING finger protein, was found to interact with RNAPII through AP-MS analysis, in which Asr1 was used as the bait protein (Asr1-TAP). Asr1 has some homology to human rA9, which was identified as a candidate RNAPII CTD binding protein through yeast two-hybrid studies prompting the analysis of Asr1 interactors in yeast [10]. Asr1 has been shown to function as a novel E3 ligase for RNAPII and functions through a unique mechanism that is not fully resolved in which two RNAPII subunits, Rpb4 and Rpb7 (hence Rpb4/7), are ejected from the RNAPII complex [10]. Interestingly, AP-MS studies have found that Rpb4/7 levels are reduced when RNAPII complexes are purified using a number of different elongation factors as bait [11]. Another prime example of the impact of AP-MS has had on the discovery of novel transcription regulatory proteins is the CTD phosphatase Rtr1. Mosley et al. [12] identified Rtr1 as a *bona fide* interactor of RNAPII through reciprocal TAP purifications and Multidimensional Protein Identification Technology (MudPIT) MS. Rtr1 was also isolated from human cells through analysis of the RNAPII interactome and named RPAP2 [13]. Additional experiments confirmed Rtr1 as a CTD Ser5 phosphatase and identified changes in RNAPII phosphorylation upon its deletion. Further studies on RPAP2/Rtr1 in yeast and human cells has confirmed its role as a CTD phosphatase that has now also been implicated in the regulation of the unfolded protein response [14–20]. These discoveries are highlighted here as key examples of the utility of AP-MS as a method to identify new interactors and potentially new biology (Fig. 4A).

Furthermore, Krogan et al. [21] demonstrated the usefulness of such an AP-MS approach with an early effort in characterizing the components of RNAPII transcription elongation complexes. Suspected elongation factors (DSIF, FACT, Spt6, TFIIF, Rtf1 and Elongator) were TAP-tagged (tandem affinity purification) and purified from *Saccharomyces cerevisiae* yeast. Different biochemical conditions (e.g. salt concentration) in the purification will determine the degree of interacting proteins that co-purify with the tagged proteins, referred to in AP-MS as baits. The pull-down products, referred to in AP-MS as prey proteins, were then subjected to MALDI-TOF MS for protein identification. These data were able to show the interaction of Spt5 with Spt4 (both members of DSIF) and also with Spt6. Interestingly, Spt6 itself was found to interact with the then uncharacterized protein, Iws1/Spn1, an interaction that is involved in a number of transcription-coupled processes [22–25]. Post-translational modification (PTM) searches of the MS analyses also revealed two phosphorylated casein kinase II (CKII) consensus sites within Spt6 [21]. In fact, PTM mapping depth is enabled through coupling it with AP-MS. Recent papers, utilizing quantitative AP-MS approaches have further detailed the phosphorylation of Spt6 by CKII and how this modification regulates the interaction with Spn1 [22,23]. Additionally, purification of the FACT subunits by Krogan et al. revealed novel interactions of FACT with CKII, both individually and in a complex with the chromatin remodeler Chd1, as well as with all the subunits of the complex now known as the RNAPII associated PAF Complex (PAFC). Bedard et al. further characterized these findings using MS methods to illuminate that FACT interacts with PAFC to facilitate CKII phosphorylation of PAFC [26]. Of note, the two largest subunits of RNAPII were reproducibly identified in purifications of PAFC and FACT; however, the identification of the remaining 12-subunits of RNAPII was stochastic across replicates. These data suggest that the majority of purified PAF-C, for instance, does not associate with RNAPII but rather interacts in a transient fashion during different stages of transcription.

It is important to note that the impact of AP-MS on the characterization of protein-protein interactions within the transcription complex has not been limited to the yeast basal elongation complex. These contributions include the identification of a number of uncharacterized RNA Polymerase Associated Proteins (RPAPs) in tandem affinity purifications (TAP) of human RNAPII from HEK293 cells [13,27]. Follow up studies on XAB1/Npa3 and the RPAPs have revealed that they play diverse roles in the regulation of RNAPII biology that has been uncovered as consequence of AP-MS experiments. XAB1/Npa3 is a conserved GPN-loop GTPase that has been implicated in RNAPII nuclear import as an RNAPII chaperone [28–31]. RPAP3 has been characterized as a R2TP-like co-chaperone that may regulate RNAPII stability as well as that of other proteins/protein complexes [32–34]. RPAP1 has been characterized to play a role in the interaction between RNAPII and Mediator [35].

However, while AP-MS of RNAPII and/or interacting, proteins has the ability to both identify and aid in the characterization of interactions within the transcription complex. The dynamic nature of these PPIs during transcription can make the determination of specific bait-prey interactions, and any changes in them, hard to distinguish from non-specific or random preys. An early approach to solve this problem was pioneered by Ranish et al. [36] that provided benefits to both the proteomics and transcription fields. Isotope-coded affinity tag (ICAT) reagents were used in both a purification of the pre-initiation complex (PIC) [37] and a control purification. The use of ICAT reagents coupled with MudPIT MS/MS improved the quantitation of relative abundance measurements for prey proteins, allowing the identification of specific interactions over background noise. Thus, the true components of the PIC were able to be identified and comprehensively analyzed [36]. More recent development of database and computational tools such as the CRAPome [38], significant analysis of interactome (SAINT) [39] and the Coon OMSSA Proteomic Analysis Software Suite (COMPASS) [40] have aided in the ability to distinguish real bait-prey interactions from contaminants, and score these interactions, both with probability and fold change values. For example, using SAINT analysis followed by data visualization by ProHits-viz [41] of the dataset from Bedard *et. al*, we performed prey-prey correlation analysis. Using these approaches, the high degree of copurification dynamics of the two largest RNAPII

subunits can be seen across various baits (Fig. 2). As mentioned above, Rpb1 and Rpb2 are readily identified in transcription-related AP-MS studies due to their large size. The reproducible identification of the other subunits of RNAPII is also readily seen when a core RNAPII subunit is used as the bait protein (Fig. 2, [42]). Along these same lines, a high degree of correlation is observed between subunits of individual protein complexes such as PAFC (subunits: Ctr9, Paf1, Cdc73, Leo1, and Rtf1). With the proper controls and statistical tools, AP-MS can provide insightful data on interaction dynamics within and between complexes regulating transcription.

Assuredly, the quality of data AP-MS will continue to improve with the development of new affinity purification mass spectrometry approaches. Emerging methods include, proximity biotin-labeling of interacting proteins through approaches including APEX [43], BioID [44], and antibody recognition [45]. These methods [43–45] have been shown to further improve the sorting true interactors from contaminants, a fundamental problem for AP-MS, as discussed previously. Improvements in MS-based quantitation between samples through isotopic labeling-based multiplexing techniques [46–48] also have the potential to benefit the accuracy and sensitivity of AP-MS. Multiplexing through isotopic labeling allows multiple samples to be analyzed within the same MS run, thereby reducing batch effect and improving quantitation. These approaches have not been broadly applied to the RNAPII interactome at this point in time, and so are not discussed at length in this review. However, Fig. 3 depicts an example RNAPII experiment comparing MS quantitation using a label-free method versus using stable isotope labeling in cell (SILAC)-based multiplexing.

A new MS-based method aimed at achieving deeper mechanistic insights from RNAPII AP-MS, by specifically looking at 5′ and 3′ RNAPII complexes, has recently been developed by Harlen and Churchman [49]. RNA stem-loop sequences were inserted into the untranslated regions (UTRs) of a single gene with 2 PP7 sequences in the 5′ UTR and 2 MS2 sequences in the 3′ UTR. Plasmids containing either GFP-PP7 coat binding protein or RFP-MS2 coat binding protein are expressed in these same cells. After total RNAPII is purified via an epitope tag on Rpb3, antibodies against the RNA stem loop coat proteins are used to isolate RNAPII complexes that were specifically enriched at either the 5′ or 3′ end of the gene. The purifications were then analyzed via MS to

Prey-prey correlation of transcription elongation factor purifications
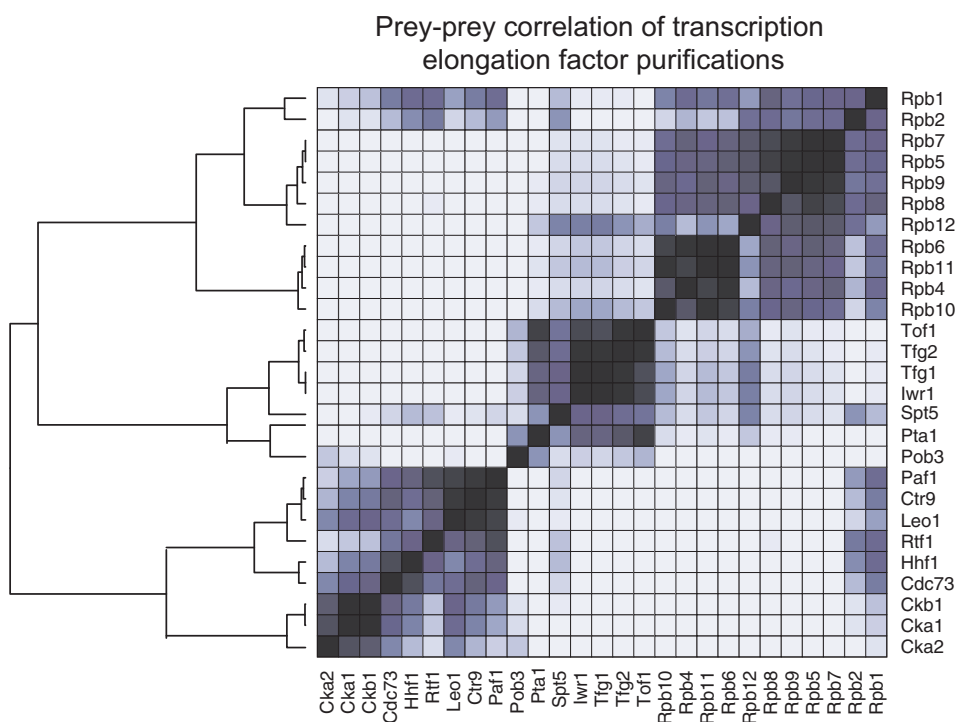
Fig. 2. Prey-prey correlation analysis of RNAPII, FACT, and PAFC purifications. Following significance analysis of interactome (SAINT), the degree of correlation between the quantity of various co-purifying proteins was analyzed using ProHits-viz (Nat Methods. 2017 Jun 29;14(7):645–646.). Quantitative data was re-analyzed from Bedard LG, Dronamraju R, et al. J Biol Chem. 2016 Jun 24;291(26):13410–20.

K.H. Burriss and A.L. Mosley
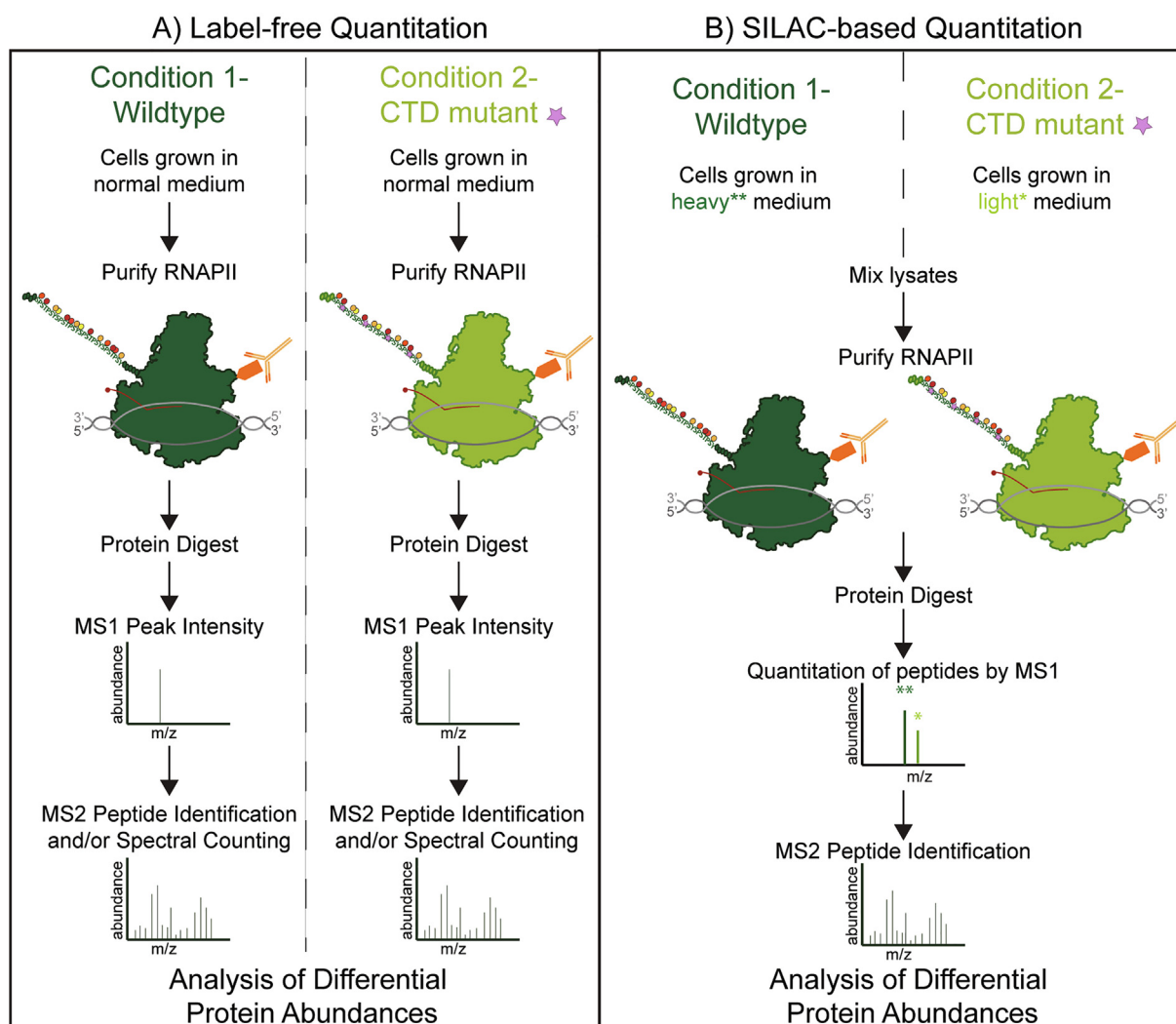
Fig. 3. Strategies for quantitative analysis of RNAPII interactors. (A) Label-free quantitation of RNAPII associated proteins can be done from multiple conditions. Individual mass spectrometry is performed and then quantitation is performed by either MS1 precursor quantitation of ion intensity and/or MS2 fragment ion count based quantitation (referred to as spectral counting or peptide-spectrum match (PSM) counting). (B) SILAC based quantitation can be performed using cells grown in various combinations of light and heavy amino acids typically focused on lysine and arginine. Samples can be mixed together based on equal cell numbers or by equal protein concentration. Quantitation is then performed by analysis of the heavy isotope containing or light isotope containing MS1 ion intensities. MS2 based fragment ion analysis is still used for peptide fingerprinting but is not used for quantitation in SILAC based approaches.

precisely determine the proteins present at early (5′) and late (3′) transcription. The results for the single gene locus used in these experiments implicate a novel, uncharacterized role for the exonuclease complex Rat1 and Rai1—typically associated with termination—during the earlier phases of transcription [49–52]. Additional results from these experiments, in conjunction with other techniques, support a role for the ubiquitin ligase Bre1 later in transcription than previously thought, through regulation of RNAPII pausing. This method provides an avenue for analysis of RNAPII interactors at specific stages of transcription, although it does require multiple genetic changes to the target gene of interest. Defining the stages of transcription that PPIs occur during will help to elucidate the mechanisms and functions of RNAPII-interacting proteins. The specificity of this approach has unique advantages when compared to purifications based on CTD phosphorylation marks, as these marks are not clear-cut between the different stages of transcription and may be heterogeneous across different gene types. In contrast, a disadvantage of analyzing only the 5′ and 3′ ends of a gene, compared to phospho-CTD pulldowns, is the limitation of single gene loci, and the loss of elongation dynamics that may occur in the gene body. For projects interested in looking at the RNAPII interactome at the start and ends of genes of interest, this approach holds the

potential to increase the specificity of RNAPII interactome, increasing the amount of PPI mechanism and function data garnered from AP-MS.

Finally, an important application of AP-MS to the study of RNAPII is the analysis of post-translational modifications that may regulate transcription dynamics. Although the CTD of Rpb1 receives a majority of the focus in the context of post-translational modifications, it is not the only region of RNAPII that is modified. Mohammed et al. [53] digested RNAPII with multiple proteases before multiplexing the digested peptides to be analyzed via MS, using both collision-induced dissociation (CID) and electron-transfer dissociation (ETD) fragmentation techniques. These methods allowed the detection of 19 phosphorylation-sites outside of the CTD, 12 of these being novel, with possible effects on the conformation of the clamp region. The biological pathways that may be regulated by these phosphorylation events could regulate RNAPII PPIs, which could be characterized through a variety of biochemical studies, including AP-MS. Thus, AP-MS has the potential to not only identify interactors of RNAPII, but also provide essential information regarding RNAPII PTM dynamics that may regulate those interactions. For both interactor identification and PTM analysis, AP-MS offers decreased sample population and increased coverage depth compared to global proteomics. The benefits of AP-MS studies may also

be complemented by other methods. For example, structural studies on the RNAPII interacting protein Spt6 have found that previously uncharacterized phosphorylation sites in the linker region of RNAPII (S1493, T1471, and/or Y1473) stabilize the interaction between RNAPII and Spt6, playing an important role in the recruitment of Spt6 throughout the genome [54]. Phosphorylation of RNAPII at T1471 was initially reported by a large scale phosphoproteomics analysis using immobilized metal affinity chromatography (IMAC) in yeast [55] as well as in a focused proteomics study on RNAPII [53]. Further proteomics-based exploration of RNAPII PTMs outside of the CTD will undoubtedly benefit understanding of the complex regulation of the dynamic PPI network that regulates RNAPII transcription.

2.2. Analysis of CTD interactome dynamics

The CTD of the largest subunit of RNAPII is known to be an essential regulator of transcription, as its deletion is not compatible with life [56,57]. Although its composition is simple, consisting of repeats (number varies with organism) of the amino acid consensus sequence $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$, its role in the mechanism of transcription has proved to be dynamic and intricate. The hypothesis of a CTD Code has been proposed, wherein the CTD is differentially phosphorylated at each stage of transcription, and this differential phosphorylation is responsible for recruiting the appropriate transcription factors to RNAPII [58]. Decoding the full range of PTM states found within the native CTD in eukaryotic systems is important to understand how RNAPII PPIs are regulated and has been an intense area of focus in the transcription field. Characterization of both CTD interacting proteins and CTD PTM states has been driven by a variety of MS-based methods (Fig. 4).

One popular approach to identify candidate CTD interactors is to use a purified GST-CTD for binding assays with lysate from cells of interest (Fig. 4C). Carty and Greenleaf [59] used MALDI-MS to identify phospho-CTD associated proteins (PCAPs) after performing an *in vitro* binding assay using a kinase (CTK1) modified CTD as bait and HeLa cell extract as a source for interacting partners (preys). The novel PCAPs identified in this study suggested that the phospho-CTD was responsible for recruiting more than just RNA-processing proteins, and that CTD phosphorylation played a role in the regulation of a variety of nuclear processes. A recent publication from Ebmeier et al. [60] used a similar approach with GST-purified CTD in a binding assay with HeLa nuclear extract. By first phosphorylating GST-CTD in a kinase assay with either Cdk7 (TFIIH) or Cdk9 (P-TEFb) before the binding assay, Orbitrap MS of pCTD-bound proteins was able to identify overlapping, yet somewhat distinct interactomes for the two different pCTD isoforms. Main findings were that 5′ mRNA capping enzymes and SETD1A/B were identified in both interactomes, suggesting they bind to a phosphorylated CTD. However, SETD2 bound specifically to a P-TEFb phosphorylated CTD. This type of approach can help better understand how differential phosphorylation of the CTD recruits specific proteins to the site of transcription. A GST-CTD approach has also aided in the characterization of an understudied CTD PTM, *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) [61]. While it was shown over 25 years ago that the RNAPII CTD is modified by *O*-GlcNAc [62], the function this modification has during transcription has remained elusive. Recent ETD MS analysis of GST-CTD after incubation with *O*-GlcNAc-transferase (OGT) was able to detect *O*-GlcNAc modifications on Ser2 and Ser5 residues [61]. Further experiments demonstrated that *O*-GlcNAc exists on RNAPII at promoters, and inhibition of O-GlcNAcylation prevented RNAPII progression [61]. These data suggest an important role for *O*-GlcNAc in regulating transcription initiation and early elongation, possibly even through a reciprocal relationship with Ser2 and Ser5 phosphorylation. The interplay between the different CTD PTMs and the proteins that regulate them is an exciting and ongoing area of study in the transcription field that has been aided by MS analysis.

Mass spectrometry has helped fine tune the details of how phosphorylation sites of the CTD are regulated by the interplay between protein kinases and phosphatases, using both *in vitro* and *in vivo* approaches. Smith-Kinnaman et al. [17] used an AP-MS approach to focus on investigating the role of phosphorylation on the recruitment of the CTD phosphatase Rtr1, which is enriched in early elongation by ChIP-qPCR. AP-MS was able to identify 20 proteins within the interactome of an epitope tagged version of the Rtr1 phosphatase, including 9 subunits of RNAPII. However, when Rtr1 was affinity purified from *CTK1 deletion* yeast (CTDK-1 Ser2 kinase), the interaction probability of Rtr1 with RNAPII was decreased. These results, along with other experiments, suggest that the interaction of Rtr1 and RNAPII is regulated by CTDK-1 and that hyperphosphorylation of the CTD is required for Rtr1 recruitment. Synthetic peptides have also been a valuable MS application to study CTD kinases and phosphatases considering that it is a highly quantitative approach. Czudnochowski [63] et al. were able to use synthetic CTD peptides that were either hyperphosphorylated, Ser2P, Ser5P or Ser7P in a time-course experiment with P-TEFb, a well described Ser2 kinase *in vivo*. The phosphorylation status of the CTD after incubation with P-TEFb, was analyzed by ESI-MS and it was observed that only the Ser7P CTD peptide was further phosphorylated, and that this phosphorylation occurred on Ser5. These results indicate a Ser7P CTD as the preferred substrate for the P-TEFb kinase to phosphorylate Ser5, challenging the dogma of P-TEFb as a Ser2 kinase. Using the *Drosophila melanogaster* CTD (*Dm*CTD) as a substrate, Gibbs et al. similarly found that *Dm*P-TEFb heavily modifies Ser5P *in vitro* in 12/42 *Dm*CTD repeats [64]. These studies suggest that phosphorylation of the CTD at Ser2P *in vivo* may require additional factors which contribute to the change in P-TEFb specificity from Ser5P *in vitro*. Luo et al. [65] investigated whether Thr4P affects Ser5P phosphatase Ssu72 activity, as the phosphatase seems to require a very specific CTD substrate with Pro6 in a *cis*-confirmation [66]. In their experiments, Ssu72 activity was measured by MS, and it was observed that the presence of Thr4P on the CTD lowered Ssu72 activity by approximately four-fold, but did not completely abolish it. Ssu72 also did not remove Thr4P itself. The authors propose that this activity decrease could be a mechanism of fine-tuning phosphatase activity.

Thr4P has also recently been implicated in the recruitment of additional transcription regulators using an AP-MS approach with mutant CTD constructs in yeast. Nemec et al. [67] affinity purified RNAPII with either WT or T4A (Thr4- > Ala) CTDs. Label-free MS-based quantitation was then used to identify significant interacting proteins (Fig. 4A). These studies found that termination factors, including Rtt103, were significantly decreased in the T4A mutant. These data, along with other experiments showing termination defects in the presence of T4A, and the direct binding of Rtt103 to a Thr4P CTD, implicate a role for Thr4P in the regulation of Rtt103 termination at specific genes. Harlen and Churchman [68] have also identified roles for Thr4P in the regulation of global transcription. In an effort to analyze phospho-specific CTD interactomes, sequential immunoprecipitations (IP) with phospho-site specific monoclonal antibodies were used (Fig. 4B). First total RNAPII was purified via epitope tag and then phospho-site CTD antibodies were used to isolate the specific populations containing the phospho-CTD of interest for MS. The IPs were then analyzed via label-free MS to identify prey proteins and the data showed distinct interactomes for each specific phospho-isoform of RNAPII. In agreement with the Nemec, *et al.* study, Rtt103 was found to be enriched in the Thr4P interactome although a different AP-MS approach was used. Additionally, spliceosome proteins were depleted in the RNAPII Thr4P interactome, but were enriched in the Ser5P interactome. Along with other experiments in the study, this data implicates Thr4 in a recruitment and release mechanism for the spliceosome machinery; unphosphorylated Thr4 may allow its recruitment, while Thr4P stimulates its release. The data discussed in this section of the review presents roles for CTD phosphorylation sites other than Ser2 and Ser5 in global regulation of transcription.

Additionally, it is critical to note that in higher eukaryotes, there are heptads within the CTD that vary from the consensus repeat sequence.
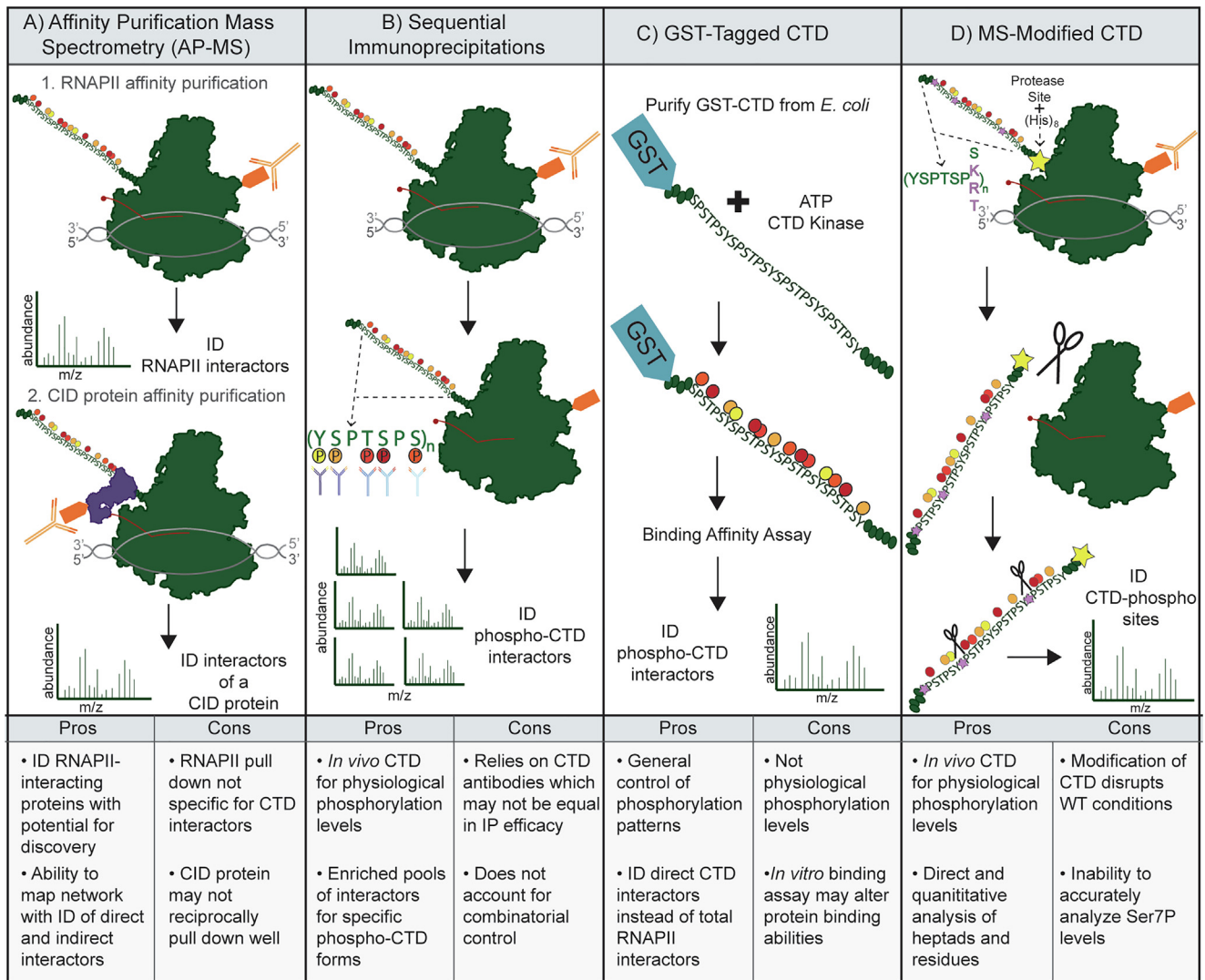
Fig. 4. Strategies for analysis of the CTD and CTD interactors. (A) Affinity purification mass spectrometry (AP-MS) of RNAPII complexes allows identification (ID) of RNAPII interactors. Interactors of interest with CTD Interaction Domains (CIDs) may then be purified to analyze the CTD interactome. (B) Purification of total RNAPII followed by IPs using antibodies specific for CTD phospho-sites. MS analysis to identify protein interactome for CTD phospho-forms. (C) Purification of GST-tagged CTD peptides to be phosphorylated by CTD kinases and used in binding assays. MS analysis of bound proteins to ID CTD-interacting proteins. D) Modification of the CTD in cells to increase suitability for MS analysis. Mutation of select Ser7 residues to lysine to provide trypsin cleavage sites, as well as additional mutations to create unique mass peptides. MS analysis of these msCTDs allows mapping of specific post-translational modifications.

A common variant in the 7th position, normally a serine residue, is lysine. The presence of Lys7 opens up additional possible modifications, such as acetylation and methylation (discussed below). However, it is possible that the lysine residue in the 7th position may also be important to mediate specific metazoan specific PPIs. Recent work has characterized proteins with altered RNAPII interaction following mutation of eight Lys7 residues to arginine using culture SILAC and AP-MS (Fig. 3B). Lys7 to Arg mutations will retain a positive charge but can no longer be modified by enzymes who target primary amines for modification. Using this approach, it was discovered that the RPRD family of proteins (RPRD1A, RPRD1B, RPRD2) are significantly reduced in RNAPII purifications from Lys7 to Arg expressing HEK293T cells [69]. Additionally, proteins that have previously been shown to interact with the RPRDs were also reduced including: RPAP2, RPAP3, MCM7 and RUVB1. The RPRD family of proteins had previously been shown to have an increased affinity for doubly phosphorylated CTD peptides [18]. Interestingly, the combination of Ser2P and Lys7Ac CTD modification also shows an increased affinity for RPRD1B.

The methods described in this section include both *in vitro* and *in vivo* modified CTD approaches. Both approaches have advantages and disadvantages that are important to consider when deciding how to best experimentally answer a question. A major drawback of *in vitro* modified CTD substrates (either GST-CTD or synthetic peptides) is the inability to fully recapitulate the endogenous heterogeneity of the phosphorylated CTD due to contributions *in vivo* by a variety of kinases [70], removal of an unknown portion of phosphorylation sites by phosphatases, and sequential modification/modification removal cycles. For instance, studies on the CTD phosphatase Ssu72 have shown that Ssu72 dephosphorylation may serve as an upstream regulator of another CTD phosphatase, Fcp1 [71,72]. However, a major advantage of *in vitro* approaches, as described earlier in this section by the data gained, is the ability to control how the CTD is modified and analysis of CTD interactors by MS is thereby more likely to be precise and quantitative. In contrast, *in vivo* modified CTDs are more likely to retain their endogenous, and thereby biologically relevant, modification states. However, the difficulty is that endogenous CTD is highly heterogeneously, and thereby it becomes very difficult to analyze specific CTD PTM interactomes quantitatively. Both approaches have also suffered from the lack of a technique to map individual PTM sites on the CTD, limiting the ability to detail exactly where and how much the CTD is

modified at any one time, which will be critical to further characterizing the dynamics of CTD PTMs that regulate PPIs at the site of transcription. Recently developed methods that aim to directly analyze the post-translational modification sites of the CTD are discussed in the next section (see Fig. 4 for a comparison of CTD analysis methods).

2.3. Direct analysis of CTD post-translational modification site composition

As discussed in the above section, mass spectrometry has been used to great effect for the identification and characterization of CTD protein interactors, as well as the dynamics of these interactions based on general phosphorylation patterns. However, direct MS analysis of the CTD itself has posed a challenge. Due to the lack of tryptic cleavage sites among consensus heptads and the repetitive nature of the CTD, the mapping of individual modification sites has been problematic. However, recent method development provided an avenue for such analyses. Lys7 residues provide tryptic cleavage sites at the end of variant heptads. While coverage is still not ideal, there is some opportunity for modification mapping. Voss et al. [73] used mass spectrometry to identify acetylation as well as mono-, di-, or tri-methylation of the human CTD. These findings were supported by another group which identified the same modifications using an antibody-based approach [74]. In all, these data found that acetylation or di-/tri-methylation only occurred on phosphorylated CTD peptides and that specific Lys7 residues could be alternatively acetylated and methylated. In studies such as these, mass spectrometry provides an advantage over western blot analysis because the degree of co-occurrence of PTMs on RNAPII CTD peptides can be directly assessed using MS. Voss *et al.* found acetylation in combination with mono-repeat phosphorylation (within two CTD consensus repeats) at Tyr1, Ser2, Thr4, or Ser5 [73]. In contrast, mono-methylation could occur on either mono-repeat phosphorylated or hypo-phosphorylated CTD di-heptad peptides. The combined data shows modification of variant Lys7 residues to be as dynamic as that of consensus residues. As Lys7 is not essential for life, but conserved in vertebrates, the authors speculate that Lys7 modification could regulate the transcription of a subset of vertebrate genes.

Recent method development has taken further advantage of Lys7 residues. These variants do not occur naturally in yeast, and their presence in higher eukaryotes does not occur until well toward the distal end of the CTD, thus limiting tryptic cleavage over the full length of the CTD. To address the compatibility of the CTD for MS analysis, both Suh et al. [75] and Shuller et al. [76] endogenously modified the CTD of Rpb1 to be more suitable for MS analysis, termed the msCTD (Fig. 4D). Within a subset of heptads, the 7th position residue, normally serine, is replaced with either lysine or arginine to facilitate tryptic cleavage. The msCTDs also contain subsets of heptads with additional residue substitutions in order to give cleaved fragments unique precursor masses, thus enabling modification mapping to individual residues within the natively repetitive CTD sequence. Suh et al. [75] used yeast CTD while Shuller et al. [76] focused mainly on mammalian CTD; despite the difference in model systems the MS data show striking similarities. Both studies confirm the presence of phosphorylation on Tyr1, Ser2, Thr4, Ser5, and Ser7, some of which had only previously been identified by antibody-based approaches. One major conclusion from both data sets is that on average, Ser2P and Ser5P levels are fold-change levels higher than Tyr1P, Thr4P, and Ser7P. Authors used this result to caution against direct comparison of different antibodies for relative quantitation of CTD phosphorylation marks. To confirm that MS analysis was able to measure phosphorylation changes at a physiologically relevant level, both studies employed Ser2 kinase inhibition, whether through mutation or drug treatment. Positively, both data sets showed marked decrease in Ser2P levels upon inhibition, while there was very little change in phosphorylation of the other CTD residues.

In both msCTD studies, the total phosphorylation level of the CTD was found to be much lower than expected. Multiply-phosphorylated heptads were found to be rare in both yeast and mammalian cells. With this initial data observing low total phosphorylation dominated by Ser2P and Ser5P, authors speculate that the CTD code might end up being simpler than expected. This idea is somewhat contrary to data presented in the above section, which suggest important roles for Ser7P, Thr4P and Lys7 modifications in metazoans. One caveat for these msCTDs is that Ser7 phosphorylation levels may be measured as low due to the mutations made to make the CTD suitable for mass spectrometry (although this possibility was considered in the published works). A potential drawback to using a modified msCTD is that the CTD is no longer wildtype. Both studies confirmed that cells carrying only msCTD grew in a manner approximate to wildtype. However, it remains to be seen how the mutations and trypsin cleavage may alter the ability to detect wildtype phosphorylation marks. Finally, it is also possible that heavily phosphorylated CTD peptides were refractory to reversed-phase column elution and/or electrospray ionization possibly losing the net positive charge required for guidance of the peptides through the mass spectrometer. Challenges with multiply-phosphorylated peptides in reversed phase chromatography have been reported previously [77] as have ionization issues for heavily phosphorylated sequences [78], although the latter issue has been reported to have a high-degree of peptide sequence dependence. Nevertheless, msCTDs are an exciting new tool for analysis of quantitative PTM mapping of an *in vivo* modified CTD.

Beyond the addition of the phosphate group itself, phosphorylation has the potential to change the structure of the CTD. Recent work has focused on investigating how phosphorylation might alter the CTD and regulate its interactions beyond the addition of the mark itself. Gibbs et al. [64] used a combination of mass spectrometry and NMR spectroscopy to observe that hyper Ser5P alters the local CTD structure via *cis*-proline isomerization in a sequence-dependent manner. This proline isomerization was then shown to modulate the activity of Ssu72, which is known to be *cis*-proline specific CTD interactor [66]. Specific mammalian CTD repeat variants were found to more frequently be in a *cis*-proline state upon Ser5P, suggesting a potential mechanism for increased Ssu72 recruitment to specific regions of the CTD.

2.4. Analysis of RNAPII by structural proteomics

Structural proteomics is a rapidly developing area of mass spectrometry-based method development (reviewed in [79]). Many methods that fall under the umbrella of structural proteomics have been used to analyze RNAPII and its interactions with accessory proteins to provide novel insights into RNAPII biology and to aid in MS method development efforts. Crosslinking-mass spectrometry (XL-MS), for instance, has been used to study a number of macromolecular complexes including TFIIH [80]. The cross-linker Bis (sulphosuccinimidyl) suberate (BS3) has been used to map interactions between RNAPII and TFIIF [81] as well as interactions between TFIIF and TFIIH [82] to assist in modeling protein complex positioning within a cryoEM structure of a 32-protein RNAPII pre-initiation complex (PIC). These structural studies were expanded to a full 52 protein PIC with Mediator using additional crosslinking experiments of Mediator-PIC and Mediator-RNAPII with the 1-hydroxy-7-azabenzotriazole analog of DSS (disuccinimidyl suberate), 1,1′-(suberoyldioxy)bisazabenzotriazole (SBAT) [83]. The data from the BS3 and SBAT XL-MS studies were later utilized in a cryoEM study of the yeast PIC and PIC-core Mediator complexes. This 2017 study additionally obtained novel crosslinks using the 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) crosslinker and were able to retain structures cryoEM with resolutions of 4.7 Å and 5.8 Å, respectively [84].

BS3-based XL-MS has continued to couple well with cryo-electron microscopy based structural biology analysis of large transcription-related complexes and has been used to provide peptide level resolution of amino acid proximity for: RNAPII-PAFC-TFIIS [85], RNAPII–DSIF–NELF [86], and RNAPII-DSIF-PAFC-SPT6 complexes These

approaches provide a high degree of analytical power for coupling with protein complexes used in structural biology studies. An emerging approach for large-scale such as organelle scale XL-MS is using mass-spectrometry cleavable crosslinkers (reviewed in [87]). MS-cleavable crosslinkers have been applied to large-scale protein-protein interaction analysis [88] and could be used for both qualitative and quantitative studies. Although disuccinimidyl sulfoxide cleavable crosslinking has not been used to interrogate the RNAPII interactome directly at this point, recent studies on the histone interactions isolated from intact nuclei could provide novel insights into co-transcriptional histone biology. DSSO crosslinking of intact nuclei and subsequent analysis by a MS3-based method identified a large number of potential histone interacting proteins (n = 778) as well as histone crosslinks with known interactors [89]. While organelle level analysis of protein-protein interactions with DSSO is a powerful approach, there are still challenges to overcome since the depth of coverage is limited by multiple factors including the longer cycle times needed to carry out MS3 analysis.

An additional emerging approach in structural proteomics for analysis of RNAPII complexes involves the application of native spray mass spectrometry. Advances in this area have been driven by changes in sample preparation approaches, as well as technological developments and optimizations that allow for large protein assemblies to both ionize and retain their native quaternary state in gas phase. These approaches can be applied to macromolecular complexes to obtain precise analytical data related to intact protein complex mass, which can reveal protein subunit stoichiometry, post-translational modifications, and protein complex modularity. For RNAPII complexes, native electrospray mass spectrometry has been used to monitor complex engagement of alpha-amanitin [90,91], as well as to investigate the stoichiometry of the yeast capping enzyme alone and in complex with transcribing phosphorylated-RNAPII (with a 20.2 kDa DNA-RNA scaffold) [92]. These native-MS studies revealed that phosphorylated RNAPII can interact with either heterotrimeric (Cet1-Cet1-Ceg1) or heterotetrameric (Ceg1-Cet1-Cet1-Ceg1) forms of the capping enzyme. The capping enzyme-RNAPII complexes were also monitored in the presence or absence of either GTP or the GTP analog, GpCpp. The remarkable utility of the native-MS approach was clearly illustrated by the authors' use of negative ion mode (positive ion mode is typically used for protein analysis in MS) to quantitatively monitor the addition of the RNA modification (5′-triphosphate end) by capping enzyme. While, protein complex modularity of RNAPII has not been monitored by native spray mass spectrometry, this application of the technique is not necessary, due to the wealth of high resolution RNAPII structural data that has already been obtained. However, noncovalent nano-electrospray ionization mass spectrometry-based investigations of the yeast RNAPII-associated cleavage and polyadenylation factor (CPF) complex has revealed a wealth of information regarding the modularity of the complex [93]. The native-MS studies of CPF were able to define polyA-polymerase, nuclease, and phosphatase modules, aidingd in the selection of subunits for cryoEM structural analysis of the polymerase module of CPF [93]. This study, along with the others highlighted in this section demonstrate the utility of MS-based structural proteomics approaches in gaining novel insights into the protein complexes that regulate RNAPII transcription. Native spray MS has also been applied to other transcription related complexes of interest including RNAPI, RNAPIII, and TFIID [94–96].

3. Conclusion

While there is valuable knowledge of transcription dynamics to be gained through high-throughput sequencing methods, mass spectrometry-based analysis of RNAPII and its associated proteins is critical to advance our discovery of new transcription-related biology and related mechanisms. Application of mass spectrometry to the study of RNAPII transcription has shown to be immensely useful in both the qualitative and quantitative analysis of protein–protein interactions that regulate

the process. Advances in MS technologies in sample preparation, instrumentation, and data analysis will increasingly facilitate sensitive and accurate quantitation of protein–protein interactions, post-translational modifications, and enzymatic activity. The emerging area of structural proteomics has already greatly benefitted the transcription field and there is no doubt that technological advancements will continue to change the types of mechanistic questions that can be asked and answered about RNAPII transcription.

Acknowledgements

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2019.03.013.

References

[1] S.M. Tan-Wong, J.B. Zaugg, J. Camblong, Z. Xu, D.W. Zhang, H.E. Mischo, A.Z. Ansari, N.M. Luscombe, L.M. Steinmetz, N.J. Proudfoot, Gene loops enhance transcriptional directionality, Science 338 (6107) (2012) 671.

[2] F. Wyers, M. Rougemaille, G. Badis, J.-C. Rousselle, M.-E. Dufour, J. Boulay, B. Régnault, F. Devaux, A. Namane, B. Séraphin, D. Libri, A. Jacquier, Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly (A) polymerase, Cell 121 (5) (2005) 725–737.

[3] M. Sopta, Z.F. Burton, J. Greenblatt, Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II, Nature 341 (6241) (1989) 410–414.

[4] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, J. Am. Soc. Mass Spectrom. 5 (11) (1994) 976–989.

[5] P.A. Grant, D. Schieltz, M.G. Pray-Grant, D.J. Steger, J.C. Reese, J.R. Yates, J.L. Workman, A subset of TAFIIs are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation, Cell 94 (1) (1998) 45–53.

[6] J. Li, D. Moazed, S.P. Gygi, Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation, J. Biol. Chem. 277 (51) (2002) 49383–49388.

[7] B. Li, L. Howe, S. Anderson, J.R. Yates 3rd, J.L. Workman, The Set2 histone methyltransferase functions through the phosphorylated carboxyl-terminal domain of RNA polymerase II, J. Biol. Chem. 278 (11) (2003) 8897–8903.

[8] T. Xiao, H. Hall, K.O. Kizer, Y. Shibata, M.C. Hall, C.H. Borchers, B.D. Strahl, Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast, Genes Dev. 17 (5) (2003) 654–663.

[9] N.J. Krogan, M. Kim, A. Tong, A. Golshani, G. Cagney, V. Canadien, D.P. Richards, B.K. Beattie, A. Emili, C. Boone, A. Shilatifard, S. Buratowski, J. Greenblatt, Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II, Mol. Cell. Biol. 23 (12) (2003) 4207–4218.

[10] A. Daulny, F. Geng, M. Muratani, J.M. Geisinger, S.E. Salghetti, W.P. Tansey, Modulation of RNA polymerase II subunit composition by ubiquitylation, Proc. Natl. Acad. Sci. U. S. A. 105 (50) (2008) 19649–19654.

[11] A.L. Mosley, G.O. Hunter, M.E. Sardiu, M. Smolle, J.L. Workman, L. Florens, M.P. Washburn, Quantitative proteomics demonstrates that the RNA polymerase II subunits Rpb4 and Rpb7 dissociate during transcriptional elongation, Mol. Cell. Proteomics 12 (6) (2013) 1530–1538.

[12] A.L. Mosley, S.G. Pattenden, M. Carey, S. Venkatesh, J.M. Gilmore, L. Florens, J.L. Workman, M.P. Washburn, Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation, Mol. Cell. 34 (2) (2009) 168–178.

[13] C. Jeronimo, D. Forget, A. Bouchard, Q. Li, G. Chua, C. Poitras, C. Therien, D. Bergeron, S. Bourassa, J. Greenblatt, B. Chabot, G.G. Poirier, T.R. Hughes, M. Blanchette, D.H. Price, B. Coulombe, Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme, Mol. Cell. 27 (2) (2007) 262–274.

[14] S. Irani, S.D. Yogesha, J. Mayfield, M. Zhang, Y. Zhang, W.L. Matthews, G. Nie, N.A. Prescott, Y.J. Zhang, Structure of Saccharomyces cerevisiae Rtr1 reveals an active site for an atypical phosphatase, Sci. Signal 9 (417) (2016) p. ra24.

[15] P.L. Hsu, F. Yang, W. Smith-Kinnaman, W. Yang, J.E. Song, A.L. Mosley, G. Varani, Rtr1 is a dual specificity phosphatase that dephosphorylates Tyr1 and Ser5 on the RNA polymerase II CTD, J. Mol. Biol. 426 (16) (2014) 2970–2981.

[16] G.O. Hunter, M.J. Fox, W.R. Smith-Kinnaman, M. Gogol, B. Fleharty, A.L. Mosley, Phosphatase Rtr1 regulates global levels of serine 5 RNA polymerase II C-terminal domain phosphorylation and cotranscriptional histone methylation, Mol. Cell. Biol. 36 (17) (2016) 2236.

[17] W.R. Smith-Kinnaman, M.J. Berna, G.O. Hunter, J.D. True, P. Hsu, G.I. Cabello, M.J. Fox, G. Varani, A.L. Mosley, The interactome of the atypical phosphatase Rtr1 in Saccharomyces cerevisiae, Mol. BioSyst. 10 (7) (2014) 1730–1741.

[18] Z. Ni, C. Xu, X. Guo, G.O. Hunter, O.V. Kuznetsova, W. Tempel, E. Marcon, G. Zhong, H. Guo, W.-H.W. Kuo, J. Li, P. Young, J.B. Olsen, C. Wan, P. Loppnau, M. El Bakkouri, G.A. Senisterra, H. He, H. Huang, S.S. Sidhu, A. Emili, S. Murphy, A.L. Mosley, C.H. Arrowsmith, J. Min, J.F. Greenblatt, RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation, Nat. Struct. Mol. Biol. 21 (8) (2014) 686–695.

[19] S. Egloff, J. Zaborowska, C. Laitem, T. Kiss, S. Murphy, Ser7 phosphorylation of the CTD recruits the RPAP2 Ser5 phosphatase to snRNA genes, Mol. Cell. 45 (1) (2012) 111–122.

[20] T.K. Chang, D.A. Lawrence, M. Lu, J. Tan, J.M. Harnoss, S.A. Marsters, P. Liu, W. Sandoval, S.E. Martin, A. Ashkenazi, Coordination between two branches of the unfolded protein response determines apoptotic cell fate, Mol. Cell. 71 (4) (2018) 629–636.e5.

[21] N.J. Krogan, M. Kim, S.H. Ahn, G. Zhong, M.S. Kobor, G. Cagney, A. Emili, A. Shilatifard, S. Buratowski, J.F. Greenblatt, RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach, Mol. Cell. Biol. 22 (20) (2002) 6979–6992.

[22] R. Dronamraju, J.L. Kerschner, S.A. Peck, A.J. Hepperla, A.T. Adams, K.D. Hughes, S. Aslam, A.R. Yoblinski, I.J. Davis, A.L. Mosley, B.D. Strahl, Casein kinase II phosphorylation of Spt6 enforces transcriptional fidelity by maintaining Spn1-Spt6 interaction, Cell. Rep. 25 (12) (2018) pp. 3476–3489 e5.

[23] E. Gouot, W. Bhat, A. Rufiange, E. Fournier, E. Paquet, A. Nourani, Casein kinase 2 mediated phosphorylation of Spt6 modulates histone dynamics and regulates spurious transcription, Nucleic Acids Res. 46 (15) (2018) 7612–7630.

[24] S. Li, A.R. Almeida, C.A. Radebaugh, L. Zhang, X. Chen, L. Huang, A.K. Thurston, A.A. Kalashnikova, J.C. Hansen, K. Luger, L.A. Stargell, The elongation factor Spn1 is a multi-functional chromatin binding protein, Nucleic Acids Res. 46 (5) (2018) 2321–2334.

[25] A.K. Thurston, C.A. Radebaugh, A.R. Almeida, J.L. Argueso, L.A. Stargell, Genome instability is promoted by the chromatin-binding protein Spn1 in Saccharomyces cerevisiae, Genetics 210 (4) (2018) 1227–1237.

[26] L.G. Bedard, R. Dronamraju, J.L. Kerschner, G.O. Hunter, E.D. Axley, A.K. Boyd, B.D. Strahl, A.L. Mosley, Quantitative analysis of dynamic protein interactions during transcription reveals a role for casein kinase II in polymerase-associated factor (PAF) complex phosphorylation and regulation of histone H2B mono-ubiquitylation, J. Biol. Chem. 291 (26) (2016) 13410–13420.

[27] C. Jeronimo, M.F. Langelier, M. Zeghouf, M. Cojocaru, D. Bergeron, D. Baali, D. Forget, S. Mnaimneh, A.P. Davierwala, J. Pootoolal, M. Chandy, V. Canadien, B.K. Beattie, D.P. Richards, J.L. Workman, T.R. Hughes, J. Greenblatt, B. Coulombe, RPAP1, a novel human RNA polymerase II-associated protein affinity purified with recombinant wild-type and mutated polymerase subunits, Mol. Cell. Biol. 24 (16) (2004) 7043–7058.

[28] S.W. Minaker, M.C. Filiatrault, S. Ben-Aroya, P. Hieter, P.C. Stirling, Biogenesis of RNA polymerases II and III requires the conserved GPN small GTPases in Saccharomyces cerevisiae, Genetics 193 (3) (2013) 853–864.

[29] J. Niesser, F.R. Wagner, D. Kostrewa, W. Mühlbacher, P. Cramer, Structure of GPN-Loop GTPase Npa3 and implications for RNA polymerase II assembly, Mol. Cell. Biol. 36 (5) (2016) 820.

[30] H. Reyes-Pardo, A.A. Barbosa-Camacho, A.E. Perez-Mejia, B. Lara-Chacon, L.A. Salas-Estrada, A.Y. Robledo-Rivera, G.M. Montero-Moran, S. Lara-Gonzalez, M.R. Calera, R. Sanchez-Olea, A nuclear export sequence in GPN-loop GTPase 1, an essential protein for nuclear targeting of RNA polymerase II, is necessary and sufficient for nuclear export, Biochim. Biophys. Acta 1823 (10) (2012) 1756–1766.

[31] L. Staresincic, J. Walker, A.B. Dirac-Svejstrup, R. Mitter, J.Q. Svejstrup, GTP-dependent binding and nuclear transport of RNA polymerase II by Npa3 protein, J. Biol. Chem. 286 (41) (2011) 35553–35561.

[32] C. Maurizy, M. Quinternet, Y. Abel, C. Verheggen, P.E. Santo, M. Bourguet, A.C.F. Paiva, B. Bragantini, M.-E. Chagot, M.-C. Robert, C. Abeza, P. Fabre, P. Fort, F. Vandermoere, P.M.F. Sousa, J.-C. Rain, B. Charpentier, S. Cianférani, T.M. Bandeiras, B. Pradet-Balade, X. Manival, E. Bertrand, The RPAP3-Cterminal domain identifies R2TP-like quaternary chaperones, Nat. Commun. 9 (1) (2018) 2093.

[33] J. Henri, M.E. Chagot, M. Bourguet, Y. Abel, G. Terral, C. Maurizy, C. Aigueperse, F. Georgescauld, F. Vandermoere, R. Saint-Fort, I. Behm-Ansmant, B. Charpentier, B. Pradet-Balade, C. Verheggen, E. Bertrand, P. Meyer, S. Cianferani, X. Manival, M. Quinternet, Deep structural analysis of RPAP3 and PIH1D1, two components of the HSP90 Co-chaperone R2TP complex, Structure 26 (9) (2018) 1196–1209.e8.

[34] F. Martino, M. Pal, H. Muñoz-Hernández, C.F. Rodríguez, R. Núñez-Ramírez, D. Gil-Carton, G. Degliesposti, J.M. Skehel, S.M. Roe, C. Prodromou, L.H. Pearl, O. Llorca, RPAP3 provides a flexible scaffold for coupling HSP90 to the human R2TP co-chaperone complex, Nat. Commun. 9 (1) (2018) 1501.

[35] C.J. Lynch, R. Bernad, I. Calvo, S. Nobrega-Pereira, S. Ruiz, N. Ibarz, A. Martinez-Val, O. Grana-Castro, E. Gomez-Lopez, E. Andres-Leon, V. Espinosa Angarica, A. Del, Sol, S. Ortega, O. Fernandez-Capetillo, E. Rojo, J. Munoz, M. Serrano, The RNA polymerase II factor RPAP1 Is critical for mediator-driven transcription and cell identity, Cell. Rep. 22 (2) (2018) 396–410.

[36] J.A. Ranish, E.C. Yi, D.M. Leslie, S.O. Purvine, D.R. Goodlett, J. Eng, R. Aebersold, The study of macromolecular complexes by quantitative proteomics, Nat. Genet. 33 (3) (2003) 349–355.

[37] J.A. Ranish, N. Yudkovsky, S. Hahn, Intermediates in formation and activity of the RNA polymerase II preinitiation complex: holoenzyme recruitment and a post-recruitment role for the TATA box and TFIIB, Genes Dev. 13 (1) (1999) 49–63.

[38] D. Mellacheruvu, Z. Wright, A.L. Couzens, J.-P. Lambert, N.A. St-Denis, T. Li, Y.V. Miteva, S. Hauri, M.E. Sardiu, T.Y. Low, V.A. Halim, R.D. Bagshaw, N.C. Hubner, A. al-Hakim, A. Bouchard, D. Faubert, D. Fermin, W.H. Dunham, M. Goudreault, Z.-Y. Lin, B.G. Badillo, T. Pawson, D. Durocher, B. Coulombe, R. Aebersold, G. Superti-Furga, J. Colinge, A.J.R. Heck, H. Choi, M. Gstaiger, S. Mohammed, I.M. Cristea, K.L. Bennett, M.P. Washburn, R. Raught, R.M. Ewing, A.-C. Gingras, A.I. Nesvizhskii, The CRAPome: a contaminant repository for affinity purification–mass spectrometry data, Nat. Methods 10 (2013) 730.

[39] H. Choi, B. Larsen, Z.-Y. Lin, A. Breitkreutz, D. Mellacheruvu, D. Fermin, Z.S. Qin, M. Tyers, A.-C. Gingras, A.I. Nesvizhskii, SAINT: probabilistic scoring of affinity purification-mass spectrometry data, Nat. Methods 8 (1) (2011) 70–73.

[40] C.D. Wenger, D.H. Phanstiel, M.V. Lee, D.J. Bailey, J.J. Coon, COMPASS: a suite of pre- and post-search proteomics software tools for OMSSA, Proteomics 11 (6) (2011) 1064–1074.

[41] J.D.R. Knight, H. Choi, G.D. Gupta, L. Pelletier, B. Raught, A.I. Nesvizhskii, A.C. Gingras, ProHits-viz: a suite of web tools for visualizing interaction proteomics data, Nat. Methods 14 (7) (2017) 645–646.

[42] A.L. Mosley, M.E. Sardiu, S.G. Pattenden, J.L. Workman, L. Florens, M.P. Washburn, Highly reproducible label free quantitative proteomic analysis of RNA polymerase complexes, Mol. Cell. Proteomics (2010) p. mcp.M110.000687.

[43] J.C. Braisted, S. Kuntumalla, C. Vogel, E.M. Marcotte, A.R. Rodrigues, R. Wang, S.T. Huang, E.S. Ferlanti, A.I. Saeed, R.D. Fleischmann, S.N. Peterson, R. Pieper, The APEX quantitative proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results, BMC Bioinf. 9 (2008) 529.

[44] K.J. Roux, D.I. Kim, B. Burke, BioID: a screen for protein-protein interactions, Curr. Protoc. Protein Sci. 74 (2013) p. Unit 19.23.

[45] D.Z. Bar, K. Atkatsh, U. Tavarez, M.R. Erdos, Y. Gruenbaum, F.S. Collins, Biotinylation by antibody recognition-a method for proximity labeling, Nat. Methods 15 (2) (2018) 127–133.

[46] A. Thompson, J. Schäfer, K. Kuhn, S. Kienle, J. Schwarz, G. Schmidt, T. Neumann, C. Hamon, Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS, Anal. Chem. 75 (8) (2003) 1895–1904.

[47] S. Wiese, K.A. Reidegeld, H.E. Meyer, B. Warscheid, Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research, Proteomics 7 (3) (2007) 340–350.

[48] G.C. McAlister, D.P. Nusinow, M.P. Jedrychowski, M. Wuhr, E.L. Huttlin, B.K. Erickson, R. Rad, W. Haas, S.P. Gygi, MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes, Anal. Chem. 86 (14) (2014) 7150–7158.

[49] K.M. Harlen, L.S. Churchman, Subgenic Pol II interactomes identify region-specific transcription elongation regulators, Mol. Syst. Biol. 13 (1) (2017) 900.

[50] M. Kim, N.J. Krogan, L. Vasiljeva, O.J. Rando, E. Nedea, J.F. Greenblatt, S. Buratowski, The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II, Nature 432 (7016) (2004) 517–522.

[51] S. Jimeno-Gonzalez, L.L. Haaning, F. Malagon, T.H. Jensen, The yeast 5'-3' exonuclease Rat1p functions during transcription elongation by RNA polymerase II, Mol. Cell. 37 (4) (2010) 580–587.

[52] K. Brannan, H. Kim, B. Erickson, K. Glover-Cutter, S. Kim, N. Fong, L. Kiemele, K. Hansen, R. Davis, J. Lykke-Andersen, D.L. Bentley, mRNA decapping factors and the exonuclease Xrn2 function in widespread premature termination of RNA polymerase II transcription, Mol. Cell. 46 (3) (2012) 311–324.

[53] S. Mohammed, K. Lorenzen, R. Kerkhoven, B. van Breukelen, A. Vannini, P. Cramer, A.J. Heck, Multiplexed proteomics mapping of yeast RNA polymerase II and III allows near-complete sequence coverage and reveals several novel phosphorylation sites, Anal. Chem. 80 (10) (2008) 3584–3592.

[54] M.A. Sdano, J.M. Fulcher, S. Palani, M.B. Chandrasekharan, T.J. Parnell, F.G. Whitby, T. Formosa, C.P. Hill, A novel SH2 recognition mechanism recruits Spt6 to the doubly phosphorylated RNA polymerase II linker at sites of transcription, Elife (2017) 6.

[55] C.P. Albuquerque, M.B. Smolka, S.H. Payne, V. Bafna, J. Eng, H. Zhou, A multi-dimensional chromatography technology for in-depth phosphoproteome analysis, Mol. Cell. Proteomics 7 (7) (2008) 1389–1396.

[56] M. Nonet, D. Sweetser, R.A. Young, Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II, Cell 50 (6) (1987) 909–915.

[57] W.A. Zehring, J.M. Lee, J.R. Weeks, R.S. Jokerst, A.L. Greenleaf, The C-terminal repeat domain of RNA polymerase II largest subunit is essential in vivo but is not required for accurate transcription initiation in vitro, Proc. Natl. Acad. Sci. 85 (11) (1988) 3698.

[58] S. Buratowski, The CTD code, Nat. Struct. Biol. 10 (9) (2003) 679–680.

[59] S.M. Carty, A.L. Greenleaf, Hyperphosphorylated C-terminal repeat domain-associating proteins in the nuclear proteome link transcription to DNA/chromatin modification and RNA processing, Mol. Cell. Proteomics 1 (8) (2002) 598–610.

[60] C.C. Ebmeier, B. Erickson, B.L. Allen, M.A. Allen, H. Kim, N. Fong, J.R. Jacobsen, K. Liang, A. Shilatifard, R.D. Dowell, W.M. Old, D.L. Bentley, D.J. Taatjes, Human TFIIH kinase CDK7 regulates transcription-associated chromatin modifications, Cell. Rep. 20 (5) (2017) 1173–1186.

[61] B.A. Lewis, A.L. Burlingame, S.A. Myers, Human RNA polymerase II promoter recruitment in vitro is regulated by O-linked N-acetylglucosaminyltransferase (OGT), J. Biol. Chem. 291 (27) (2016) 14056–14061.

[62] W.G. Kelly, M.E. Dahmus, G.W. Hart, RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc, J. Biol. Chem. 268 (14) (1993) 10416–10424.

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

[64] E.B. Gibbs, F. Lu, B. Portz, M.J. Fisher, B.P. Medellin, T.N. Laremore, Y.J. Zhang,

D.S. Gilmour, S.A. Showalter, Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain, Nat. Commun. 8 (2017) 15233.

[65] Y. Luo, S.D. Yogesha, J.R. Cannon, W. Yan, A.D. Ellington, J.S. Brodbelt, Y. Zhang, novel modifications on C-terminal domain of RNA polymerase II can fine-tune the phosphatase activity of Ssu72, ACS Chem. Biol. 8 (9) (2013) 2042–2052.

[66] J.W. Werner-Allen, C.J. Lee, P. Liu, N.I. Nicely, S. Wang, A.L. Greenleaf, P. Zhou, cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72, J. Biol. Chem. 286 (7) (2011) 5717–5726.

[67] C.M. Nemec, F. Yang, J.M. Gilmore, C. Hintermair, Y.H. Ho, S.C. Tseng, M. Heidemann, Y. Zhang, L. Florens, A.P. Gasch, D. Eick, M.P. Washburn, G. Varani, A.Z. Ansari, Different phosphoisoforms of RNA polymerase II engage the Rtt103 termination factor in a structurally analogous manner, Proc. Natl. Acad. Sci. U. S. A. 114 (20) (2017) E3944–E3953.

[68] K.M. Harlen, K.L. Trotta, E.E. Smith, M.M. Mosaheb, S.M. Fuchs, L.S. Churchman, Comprehensive RNA polymerase II interactomes reveal distinct and varied roles for each phospho-CTD residue, Cell. Rep. 15 (10) (2016) 2147–2158.

[69] I. Ali, D. Garrido-Ruiz, Z. Ni, J.R. Johnson, H. Zhang, P.-C. Li, R.J. Conrad, X. Guo, J. Min, J. Greenblatt, M. Jacobson, N.J. Krogan, M. Ott, Crosstalk between RNA Pol II C-terminal domain acetylation and phosphorylation via RPRD proteins. bioRxiv, 2018: p. 442491.

[70] C.M. Nemec, A.K. Singh, A. Ali, S.C. Tseng, K. Syal, K.J. Ringelberg, Y.H. Ho, C. Hintermair, M.F. Ahmad, R.K. Kar, A.P. Gasch, M.S. Akhtar, D. Eick, A.Z. Ansari, Noncanonical CTD kinases regulate RNA polymerase II in a gene-class-specific manner, Nat. Chem. Biol. 15 (2) (2019) 123–131.

[71] D.W. Zhang, A.L. Mosley, S.R. Ramisetty, J.B. Rodriguez-Molina, M.P. Washburn, A.Z. Ansari, Ssu72 phosphatase-dependent erasure of phospho-Ser7 marks on the RNA polymerase II C-terminal domain is essential for viability and transcription termination, J. Biol. Chem. 287 (11) (2012) 8541–8551.

[72] A.R. Bataille, C. Jeronimo, P.E. Jacques, L. Laramee, M.E. Fortin, A. Forest, M. Bergeron, S.D. Hanes, F. Robert, A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes, Mol. Cell. 45 (2) (2012) 158–170.

[73] K. Voss, I. Forne, N. Descostes, C. Hintermair, R. Schuller, M.A. Maqbool, M. Heidemann, A. Flatley, A. Imhof, M. Gut, I. Gut, E. Kremmer, J.C. Andrau, D. Eick, Site-specific methylation and acetylation of lysine residues in the C-terminal domain (CTD) of RNA polymerase II, Transcription 6 (5) (2015) 91–101.

[74] J.D. Dias, T. Rito, E. Torlai Triglia, A. Kukalev, C. Ferrai, M. Chotalia, E. Brookes, H. Kimura, A. Pombo, Methylation of RNA polymerase II non-consensus Lysine residues marks early transcription in mammalian cells, Elife (2015) 4.

[75] H. Suh, S.B. Ficarro, U.B. Kang, Y. Chun, J.A. Marto, S. Buratowski, Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II, Mol. Cell. 61 (2) (2016) 297–304.

[76] R. Schuller, I. Forne, T. Straub, A. Schreieck, Y. Texier, N. Shah, T.M. Decker, P. Cramer, A. Imhof, D. Eick, Heptad-specific phosphorylation of RNA polymerase II CTD, Mol. Cell. 61 (2) (2016) 305–314.

[77] A. Fleitz, E. Nieves, C. Madrid-Aliste, S.J. Fentress, L.D. Sibley, L.M. Weiss, R.H. Angeletti, F.-Y. Che, Enhanced detection of multiply phosphorylated peptides and identification of their sites of modification, Anal. Chem. 85 (18) (2013) 8566–8576.

[78] H. Steen, J.A. Jebanathirajah, J. Rush, N. Morrice, M.W. Kirschner, Phosphorylation analysis by mass spectrometry: myths, facts, and the consequences for qualitative and quantitative measurements, Mol. Cell. Proteomics 5 (1) (2006) 172–181.

[79] U. Kaur, D.T. Johnson, E.E. Chea, D. Deredge, J.A. Espino, L.M. Jones, Evolution of structural biology through the lens of mass spectrometry, Anal. Chem. (2018).

[80] J. Luo, P. Cimermancic, S. Viswanath, C.C. Ebmeier, B. Kim, M. Dehecq, V. Raman, C.H. Greenberg, R. Pellarin, A. Sali, D.J. Taatjes, S. Hahn, J. Ranish, Architecture of

the human and yeast general transcription and DNA repair factor TFIIH, Mol. Cell. 59 (5) (2015) 794–806.

[81] Z.A. Chen, A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.C. Bukowski-Wills, M. Nilges, P. Cramer, J. Rappsilber, Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry, EMBO J. 29 (4) (2010) 717–726.

[82] K. Murakami, H. Elmlund, N. Kalisman, D.A. Bushnell, C.M. Adams, M. Azubel, D. Elmlund, Y. Levi-Kalisman, X. Liu, B.J. Gibbons, M. Levitt, R.D. Kornberg, Architecture of an RNA polymerase II transcription pre-initiation complex, Science 342 (6159) (2013) 1238724.

[83] P.J. Robinson, M.J. Trnka, D.A. Bushnell, R.E. Davis, P.J. Mattei, A.L. Burlingame, R.D. Kornberg, Structure of a complete mediator-RNA polymerase II pre-initiation complex, Cell 166 (6) (2016) 1411–1422.e16.

[84] S. Schilbach, M. Hantsche, D. Tegunov, C. Dienemann, C. Wigge, H. Urlaub, P. Cramer, Structures of transcription pre-initiation complex with TFIIH and Mediator, Nature 551 (2017) 204.

[85] Y. Xu, C. Bernecky, C.T. Lee, K.C. Maier, B. Schwalb, D. Tegunov, J.M. Plitzko, H. Urlaub, P. Cramer, Architecture of the RNA polymerase II-Paf1C-TFIIS transcription elongation complex, Nat. Commun. 8 (2017) 15741.

[86] S.M. Vos, L. Farnung, H. Urlaub, P. Cramer, Structure of paused transcription complex Pol II-DSIF-NELF, Nature 560 (7720) (2018) 601–606.

[87] C. Yu, L. Huang, Cross-linking mass spectrometry: an emerging technology for interactomics and structural biology, Anal. Chem. 90 (1) (2018) 144–165.

[88] C. Yu, A. Huszagh, R. Viner, E.J. Novitsky, S.D. Rychnovsky, L. Huang, Developing a multiplexed quantitative cross-linking mass spectrometry platform for comparative structural analysis of protein complexes, Anal. Chem. 88 (20) (2016) 10301–10308.

[89] D. Fasci, H. van Ingen, R.A. Scheltema, A.J.R. Heck, Histone interaction landscapes visualized by crosslinking mass spectrometry in intact cell nuclei, Mol. Cell. Proteomics (2018) p. mcp.RA118.000924.

[90] J. Lu, M.J. Trnka, S.-H. Roh, P.J.J. Robinson, C. Shiau, D.G. Fujimori, W. Chiu, A.L. Burlingame, S. Guan, Improved peak detection and deconvolution of native electrospray mass spectra from large protein complexes, J. Am. Soc. Mass Spectrom. 26 (12) (2015) 2141–2151.

[91] S. Guan, M.J. Trnka, D.A. Bushnell, P.J.J. Robinson, J.E. Gestwicki, A.L. Burlingame, Deconvolution method for specific and nonspecific binding of ligand to multiprotein complex by native mass spectrometry, Anal. Chem. 87 (16) (2015) 8541–8546.

[92] Fuensanta W. Martinez-Rucobo, R. Kohler, M. van de Waterbeemd, Albert J.R. Heck, M. Hemann, F. Herzog, H. Stark, P. Cramer, Molecular basis of transcription-coupled pre-mRNA capping, Mol. Cell 58 (6) (2015) 1079–1089.

[93] A. Casañal, A. Kumar, C.H. Hill, A.D. Easter, P. Emsley, G. Degliesposti, Y. Gordiyenko, B. Santhanam, J. Wolf, K. Wiederhold, G.L. Dornan, M. Skehel, C.V. Robinson, L.A. Passmore, Architecture of eukaryotic mRNA 3'-end processing machinery 358(6366) (2017) 1056–1059.

[94] K. Lorenzen, A. Vannini, P. Cramer, A.J. Heck, Structural biology of RNA polymerase III: mass spectrometry elucidates subcomplex architecture, Structure 15 (10) (2007) 1237–1245.

[95] L.A. Lane, C. Fernandez-Tornero, M. Zhou, N. Morgner, D. Ptchelkine, U. Steuerwald, A. Politis, D. Lindner, J. Gvozdenovic, A.C. Gavin, C.W. Muller, C.V. Robinson, Mass spectrometry reveals stable modules in holo and apo RNA polymerases I and III, Structure 19 (1) (2011) 90–100.

[96] S. Trowitzsch, C. Viola, E. Scheer, S. Conic, V. Chavant, M. Fournier, G. Papai, I.O. Ebong, C. Schaffitzel, J. Zou, M. Haffke, J. Rappsilber, C.V. Robinson, P. Schultz, L. Tora, I. Berger, Cytoplasmic TAF2-TAF8-TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules, Nat. Commun. 6 (2015) 6011.