Biophysical *Journal*
Biophysical Letter

Biophysical Society

# Substrate Specificity of the Kinase P-TEFb towards the RNA Polymerase II C-Terminal Domain

Eric B. Gibbs,[1] Tatiana N. Laremore,[2] Grace A. Usher,[3] Bede Portz,[3] Erik C. Cook,[1] and Scott A. Showalter[1,3,*]
[1]Department of Chemistry, [2]Huck Institutes of the Life Sciences, and [3]Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania

ABSTRACT The positive transcription elongation factor b (P-TEFb) promotes transcription elongation through phosphorylation of the RNA polymerase II C-terminal domain. This process is not well understood, partly due to difficulties in determining the specificity of P-TEFb toward the various heptad repeat motifs within the C-terminal domain. A simple assay using mass spectrometry was developed to identify the substrate specificity of the *Drosophila melanogaster* P-TEFb (DmP-TEFb) in vitro. This assay demonstrated that DmP-TEFb preferentially phosphorylates Ser5 and, surprisingly, that pre-phosphorylation or conserved amino acid variation at the 7-position in the heptad can alter DmP-TEFb specificity, leading to the creation of distinct double-phosphorylation marks.

Dynamic patterns of post-translational modifications, collectively referred to as the "CTD code", exert tight control over factor association with the RNA polymerase II C-terminal domain (CTD). In this context, the positive transcription elongation factor b (P-TEFb), composed of cyclin-dependent kinase 9 (CDK9) and cyclinT1, promotes transcription elongation through phosphorylation of the CTD. Mechanistic understanding of this process is incomplete, partially because the determinants of P-TEFb specificity toward target motifs found in the CTD are unknown. The primary structure of the CTD features a linear array of heptad motifs bearing the consensus sequence $Y_1S_2P_3T_4S_5P_6S_7$, but substantial variability in the fourth and seventh positions of the heptad sequence is commonly encountered. The canonical substrate recognition motif for CDKs, including CDK9, is a serine or threonine residue followed by a proline (1), with additional amino acids flanking the S/T-P pair providing specificity. Hence, depending on the details of sequence context, P-TEFb should be able to target Ser2, Ser5, or both, within the CTD heptads, and all have been reported (2). Recently, a chemical-genetic screen identified over 100 putative targets of human P-TEFb in vivo, revealing a variety of substrate peptide motifs, many bearing little similarity to the CTD consensus heptad (3), leaving the in vivo specificity of P-TEFb a contested issue.

We have recently reported that hyper-phosphorylation of the *Drosophila melanogaster* CTD by *D. melanogaster* P-TEFb (DmP-TEFb) leads to heptad sequence-specific conformational transitions that impact downstream recruitment of CTD-interacting factors (4). To understand the specificity of DmP-TEFb in vitro, we designed a small peptide library modeled on the consensus heptad sequence that we subjected to phosphorylation by DmP-TEFb. We find a marked preference for phosphorylation of Ser5 within the consensus sequence and show that pre-phosphorylation or conserved sequence variation at the 7-position can alter DmP-TEFb specificity, leading to the creation of distinct double-phosphorylation marks.

In this assay, CTD peptides were treated with DmP-TEFb and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was used to measure the relative populations of each phosphorylated peptide species, accounting for the possibility of multi-phosphorylated peptides (Figs. S1–S9, *A*). MALDI-TOF-TOF tandem mass spectra were then collected for each species. In the resulting spectra, distinct sets of b and y ions were observed, consistent with the presence of multiple phosphoisoforms (Figs. S1–S9, *B*). Using the intensities of the site-determining ions for each set, the percentage of each phosphoisoform within the sample was determined. In this way, all detectable phosphoisoforms produced through DmP-TEFb treatment were directly identified and quantified to give an unbiased sampling of the phosphoisoform distribution. This enabled us to assay the effects of point substitutions and prephosphorylation on DmP-TEFb activity and specificity, in terms of the relative

*Submitted May 11, 2017, and accepted for publication September 11, 2017.*
*Correspondence: sas76@psu.edu
Editor: Enrique De La Cruz.
https://doi.org/10.1016/j.bpj.2017.09.011

Please cite this article in press as: Gibbs et al., Substrate Specificity of the Kinase P-TEFb towards the RNA Polymerase II C-Terminal Domain, Biophysical Journal (2017), https://doi.org/10.1016/j.bpj.2017.09.011



amount of each product generated. All reported results are based on 16 h exposures to DmP-TEFb, which we confirmed, through time-course experiments on representative peptides, was a sufficient incubation time to reach a plateau in the extent of phosphorylation (Figs. S10 and S11).

Our study began with the analysis of a consensus peptide (SPSYSPTSPSYSPT) designed to present one complete heptad, insulated from end effects by neighboring half-heptad motifs. Analysis of the DmP-TEFb-treated peptides revealed a mixture of mono-phosphorylated products with detectable populations of pSer2- (22%), pSer5- (74%), and pSer7-labeled species (2%) (Fig. 1; Fig. S1). A di-phosphorylated product (pSer2, pSer5) was also observed; however, this species was only present to ~1% of the total population (Fig. S1, *A* and *C*). These data show that within the context of a consensus CTD peptide, DmP-TEFb preferentially places only one phosphate per heptad and primarily targets the Ser5 residue. This preferential targeting of Ser5 by DmP-TEFb is consistent with human P-TEFb (5) and our results using the *D. melanogaster* CTD (4,6).

As a control to test our phosphorylation site assignment, we next tested peptides bearing single alanine substitutions at either the 2- or 5-position of the central heptad. As expected, the S2A mutation resulted in a mixture of mono-phosphorylated products, with a slight enrichment in pSer5 (80%) and a twofold increase in pSer7 (5%) (Fig. 1; Fig. S2). Further, treatment of the S5A peptide generated a mixture of pSer2- (13%) and pSer7-labeled species (5%) (Fig. 1; Fig. S3). Thus, in the absence of its preferred Ser5 substrate, DmP-TEFb can maintain basal activity toward Ser2 and Ser7.

Having established Ser5 as the preferred target for phosphorylation by DmP-TEFb, we sought to understand what sequence features are responsible for this specificity. In a study of the human P-TEFb, Czudnochowski et al. (5) hypothesized that the Tyr residue within the $S_5PSY_1$ motif plays an important role in defining substrate specificity. In our studies of the *D. melanogaster* CTD, which only maintains two consensus heptads, we observed that the preferential targeting of Ser5 by DmP-TEFb is maintained within the context of heptads that deviate substantially from the consensus sequence (4). In contrast, *Drosophila* maintains strong conservation of the Tyr1 position. Therefore, to test the hypothesis that Tyr1 is necessary for the preferential targeting of Ser5 by DmP-TEFb, we introduced tyrosine-to-alanine substitutions to either side of the central heptad motif in our peptide. When presented with peptides bearing an $A_1SPTS_5PS$ motif, DmP-TEFb produced dramatically less pSer5-labeled species (28%), with slightly greater amounts of pSer2 (32%) and pSer7 (7%) (Fig. 1; Fig. S4). Peptides bearing a $YSPTS_5PSA_{1'}$ motif were phosphorylated to a lesser extent at Ser5 (21%), whereas they unexpectedly produced ~9-fold more pSer7-labeled species (18%) (Fig. 1; Fig. S5). Finally, mutation of both tyrosine residues to generate an $A_1SPTS_5PA_{1'}$ motif resulted in the most significant reduction in the amount of pSer5- (2%) and pSer2-labeled species (9%) generated, yet a fourfold increase in pSer7-labeled species (10%) was observed (Fig. 1; Fig. S6). These data suggest that removal of Tyr1 reduces DmP-TEFb activity toward Ser5, leading to lower efficiency overall and a lack of specificity.

Having gained insight into the sequence determinants of DmP-TEFb specificity, we next wanted to assay the effect of priming phosphorylations on this specificity. When presented with a consensus CTD peptide pre-phosphorylated at Ser2, DmP-TEFb was unable to generate a doubly modified product (~1%) (Fig. 2). Similarly, DmP-TEFb could not phosphorylate a consensus peptide containing pSer5 (Fig. 2). In contrast, but consistent with other recent work (7), when presented with a substrate peptide 100% pre-phosphorylated at Ser7, DmP-TEFb converted a significant fraction to doubly phosphorylated peptides (59%), including a large population of $YSPTpS_5PpS_7$ (46%) and a smaller amount of $YpS_2PTSPpS_7$ labeled species (14%) (Fig. 2; Fig. S9, *A* and *B*).

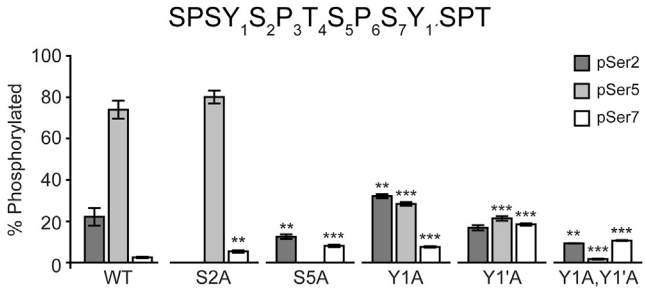


**FIGURE 1 When presented with CTD peptides, DmP-TEFb produces mixtures of phosphorylated species in vitro. Shown are the percentages of each species, where the error bars represent the standard deviation from three replicates, for the consensus motif YSPTSPS and the mutants S2A, S5A, Y1A, $S_5PSA_1$, as well as the Y1A, Y1′A double mutant. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.**

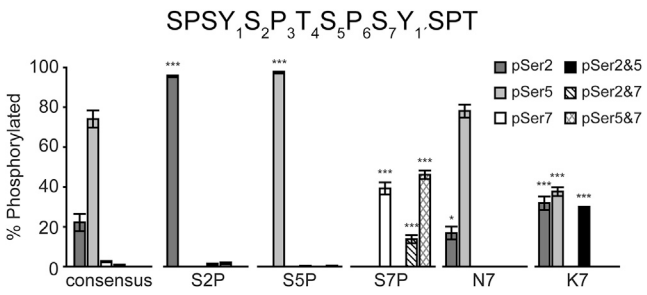


**FIGURE 2 When presented with modified CTD peptides, DmP-TEFb produces mixtures of phosphorylated species in vitro. Shown are the percentages of each species, where the error bars represent the standard deviation from three replicates for the consensus motif, the pSer2 consensus, the pSer5 consensus, the pSer7 consensus, and the naturally occurring S7N and S7K sequence variants. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.**

Please cite this article in press as: Gibbs et al., Substrate Specificity of the Kinase P-TEFb towards the RNA Polymerase II C-Terminal Domain, Biophysical Journal (2017), https://doi.org/10.1016/j.bpj.2017.09.011



Previously, Czudnochowski et al. (5) have shown increased activity of human P-TEFb toward CTD peptides containing pSer7 over other pre-phosphorylated substrates; taken together with our data, this hints at the importance of the 7-position for modulation of CTD-modifying enzyme activity. Indeed, the 7-position is one of the most variable positions in the CTD heptad (8). In our analysis of the *D. melanogaster* CTD, we observed that pSer5-labeled heptads containing Asn in the 7-position had an ~6-fold higher population of *cis*-Pro6, resulting in faster rates of dephosphorylation by the *cis*-Pro6-specific phosphatase Ssu72 (4). Hence, we wanted to test whether other naturally occurring amino acid substitutions in the 7-position of the CTD heptad could alter the specificity of DmP-TEFb. When presented with the Asn7 variant, DmP-TEFb produced a mixture of mono-phosphorylated products, including populations of pSer2- (16%) and pSer5-labeled species (78%) in a distribution similar to the consensus (Fig. 2; Fig. S7). Thus, Asn7 heptads are recognized by DmP-TEFb similarly to the consensus heptad.

Recently, extensive mutagenesis yielding full CTDs compatible with MS analysis enabled phosphosite mapping across the CTD in yeast and human cells (9,10). Despite structural data identifying pSer2/pSer5 doubly phosphorylated repeats as binding substrates for multiple CTD-interacting factors, these MS studies indicated that pSer2/pSer5 marks were only present at very low abundance in cells. In contrast, when presented with the Lys7 variant in our peptide study, DmP-TEFb produced a mixture of mono- and di-phosphorylated species in a 72:28 ratio, with populations of pSer2- (33%), pSer5- (39%), and pSer2/pSer5-labeled species (28%) (Fig. 2; Fig. S8). Taken together, pSer2/pSer5 doubly phosphorylated heptads may be important contributers to the CTD code, but our data suggest that they are formed transiently or in low abundance by P-TEFb during transcription. Significantly, these results suggest that variation in the identity of the amino acid occupying the seventh position of the heptad can be used to generate spatially heterogeneous phosphorylation patterns throughout the CTD.

Our data show that full DmP-TEFb activity is dependent on the presence of Tyr1 residues within the heptad. Further, specificity is tuned by the amino acid in the seventh heptad position, which is also the most variable in metazoan CTD sequences. The assay reported here provides confirmation that CTD code "writers" can create cryptic, possibly transient, phosphovariants that go undetected with current in vivo assays. The question of whether the CTD code is more simple than previously thought has recently been raised (9,10); our data provide a rationale to continue searching for technologically elusive marks.

## SUPPORTING MATERIAL

Supporting Materials and Methods and eleven figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(17)31022-6.

## AUTHOR CONTRIBUTIONS

E.B.G., T.N.L., G.A.U., E.C.C., and S.A.S. designed research and analyzed data. E.B.G., B.P., G.A.U., E.C.C., and T.N.L. performed research. and all authors wrote the manuscript.

## ACKNOWLEDGMENTS

Supported by National Science Foundation grant no. MCB-1515974 (to S.A.S.) and an MPHD scholarship from the Alfred P. Sloan Foundation to E.B.G.

## REFERENCES

1. Ubersax, J. A., and J. E. Ferrell, Jr. 2007. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* 8:530–541.
2. Paparidis, N. F., M. C. Durvale, and F. Canduri. 2017. The emerging picture of CDK9/P-TEFb: more than 20 years of advances since PITALRE. *Mol. Biosyst.* 13:246–276.
3. Sansó, M., R. S. Levin, …, R. P. Fisher. 2016. P-TEFb regulation of transcription termination factor Xrn2 revealed by a chemical genetic screen for Cdk9 substrates. *Genes Dev.* 30:117–131.
4. Gibbs, E. B., F. Lu, …, S. A. Showalter. 2017. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat. Commun.* 8:15233.
5. Czudnochowski, N., C. A. Bösken, and M. Geyer. 2012. Serine-7 but not serine-5 phosphorylation primes RNA polymerase II CTD for P-TEFb recognition. *Nat. Commun.* 3:842.
6. Portz, B., F. Lu, …, D. S. Gilmour. 2017. Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat. Commun.* 8:15231.
7. Bösken, C. A., L. Farnung, …, M. Geyer. 2014. The structure and substrate specificity of human Cdk12/Cyclin K. *Nat. Commun.* 5:3505.
8. Corden, J. L. 2013. RNA polymerase II C-terminal domain: tethering transcription to transcript and template. *Chem. Rev.* 113:8423–8455.
9. Suh, H., S. B. Ficarro, …, S. Buratowski. 2016. Direct analysis of phosphorylation sites on the Rpb1 C-terminal domain of RNA polymerase II. *Mol. Cell.* 61:297–304.
10. Schüller, R., I. Forné, …, D. Eick. 2016. Heptad-specific phosphorylation of RNA polymerase II CTD. *Mol. Cell.* 61:305–314.