Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain

Eric B. Gibbs\textsuperscript{1}, Feiyue Lu\textsuperscript{2,3}, Bede Portz\textsuperscript{3}, Michael J. Fisher\textsuperscript{3}, Brenda P. Medellin\textsuperscript{4,5}, Tatiana N. Laremore\textsuperscript{2}, Yan Jessie Zhang\textsuperscript{4,5}, David S. Gilmour\textsuperscript{3} & Scott A. Showalter\textsuperscript{1,3}

The carboxy-terminal domain (CTD) of the RNA polymerase II (Pol II) large subunit cycles through phosphorylation states that correlate with progression through the transcription cycle and regulate nascent mRNA processing. Structural analyses of yeast and mammalian CTD are hampered by their repetitive sequences. Here we identify a region of the \textit{Drosophila melanogaster} CTD that is essential for Pol II function \textit{in vivo} and capitalize on natural sequence variations within it to facilitate structural analysis. Mass spectrometry and NMR spectroscopy reveal that hyper-Ser5 phosphorylation transforms the local structure of this region via proline isomerization. The sequence context of this switch tunes the activity of the phosphatase Ssu72, leading to the preferential de-phosphorylation of specific heptads. Together, context-dependent conformational switches and biased dephosphorylation suggest a mechanism for the selective recruitment of \textit{cis}-proline-specific regulatory factors and region-specific modulation of the CTD code that may augment gene regulation in developmentally complex organisms.
The carboxy-terminal domain (CTD) of Rpb1, the largest subunit in RNA polymerase II, is an essential regulator of eukaryotic gene expression. This intrinsically disordered protein (IDP), consisting of multiple tandem repeats of the consensus sequence (Y1S2P3T4S5P6S7), acts as a scaffold for the recruitment of factors required for transcription, mRNA biogenesis and modification of the chromatin structure. Tight control over the spatial and temporal recruitment of CTD-associated factors is regulated at the molecular level in part by CTD-specific kinases and phosphatases, which generate dynamic patterns of post-translational modifications (PTMs) collectively referred to as the 'CTD code'. While much is known about how heptads matching the consensus sequence contribute to gene expression, heptads that deviate from the consensus are found in all eukaryotes whose sequences are known. The number and complexity of these non-consensus heptads roughly correlate with developmental complexity. Expression of genes involved in multicellularity were affected by mutating non-consensus heptads of mouse cells in culture, despite data indicating that the non-consensus heptads are not essential for the viability of human cells grown in culture. Likewise, deletion of a small region encompassing several non-consensus heptads caused severe developmental defects and growth retardation in mice, demonstrating that non-consensus heptads may contribute to development and cellular differentiation.

Despite the wealth of information detailing CTD function, little is known about the molecular basis of the CTD code. The established view is that the intrinsically disordered nature of the CTD renders its structure and interactions non-specific, to be dictated predominantly by its PTM status. Prior structural studies, focussing on short CTD peptides composed entirely of consensus heptad repeats, revealed turn and coil structures that have been extrapolated to represent the entirety of the CTD. The repetitive amino-acid sequence comprising the CTD has been a major obstacle in studying its structure because it prevents heptad-specific interpretation of both mass spectra and NMR spectra. Recent mass spectrometry-based investigations have resorted to introducing mutations that facilitate analysis of specific regions of the CTD, but backbone resonance assignment of disordered and repeat-containing polypeptides such as the CTD is often complicated by extreme signal overlap. Analysis of hyper-phosphorylated CTD by mass spectrometry (MS) revealed successful incorporation of up to 12 phosphates per polypeptide (Fig. 2b). Phospho-site identification by tandem MS (MS/MS) led to the conclusion that in vitro phosphorylation protocol predominantly generates high levels of Ser5 phosphorylation (Fig. 2a, c). Preferential phosphorylation in vitro of Ser5 over other amino acids in the CTD by human P-TEFb has been previously observed.

NMR spectroscopy was employed to augment and cross-check our MS-based phospho-site assignment, leading to the pattern depicted in Fig. 2a. NMR is well suited to this task due to its high sensitivity to the local chemical environments of individual residues, but backbone resonance assignment of disordered and repeat-containing polypeptides such as the CTD is often complicated by extreme signal overlap. In this regard, we and others have developed Direct-Detect NMR spectroscopy, which vastly improves resonance dispersion for intrinsically disordered regions, and also provides direct measurement of proline resonances (Fig. 3d). This allowed unambiguous backbone resonance assignment of CTD2, including assignment of the proline residues, which comprise 23% of the polypeptide sequence. Assignments were mapped onto correlation spectra through standard triple resonance experiments. Certain amide proton and side-chain carbon resonances (Figs. 2d, f and 3d) experienced downfield chemical shifts that were consistent with phosphorylation. Correlated with these shifts, several proline residues adjacent to phosphorylation sites showed carbon chemical shift changes characteristic of trans-to-cis isomerization of the peptide plane (Figs. 2d and 3d). In total, 10 bona fide phospho-sites were identified. Real-time NMR (RT-NMR) permitted the kinetic measurement of CTD2 phosphorylation, revealing that for the seven internal heptads containing Ser5-Pro6 pairs phosphorylation proceeded at similar apparent rates and reached comparable levels upon saturation (90%) with human P-TEFb.

**Results**

**Transgenic flies reveal a CTD region needed for development.** To identify a functionally significant region of the CTD for our structural analysis, we tested the ability of ectopically expressed derivatives of Rpb1 to rescue the lethality caused by inhibiting the expression of endogenous Rpb1. When a transgenic fly line expressing RNA-mediated interference (RNAi) against endogenous Rpb1 in response to the GAL4 activator is mated to a fly line that ubiquitously expresses GAL4, no adult progeny are produced. This lethality is overcome by co-expressing a wild-type version of Rpb1 that has been rendered resistant to the RNAi through synonymous substitutions in the Rpb1-coding sequence (Fig. 1a). To identify regions of the CTD essential for the development of an adult fly, several deletions were made in the CTD of the RNAi-resistant form of Rpb1 and tested for their ability to rescue the lethality associated with deleting the endogenous Rpb1 (Fig. 1b, Supplementary Fig. 1b). Western blot analysis with antibody against the ectopically expressed Rpb1 indicated that each derivative was expressed at comparable levels in tissues derived from pupae, indicating that each Rpb1 variant was equally stable (Fig. 1c, Supplementary Fig. 1c). The CTD2.5 mutant was the only one that failed to produce adult flies (Fig. 1d, Supplementary Table 1). Strikingly, this region of the *Drosophila* CTD is the most highly conserved among higher eukaryotes (Fig. 1e, Supplementary Fig. 1b), suggesting functional necessity.

**Hyper-phosphorylation of recombinant CTD2** by P-TEFb. Having identified a region of the *Drosophila* CTD that is essential for normal development, we next sought to incorporate the region removed in the CTD2.5 mutant into a recombinant construct displaying the properties necessary for high-resolution structural biology. We selected a CTD construct, CTD2.5-containing residues 1,657–1,739 of the Rpb1 polypeptide sequence (Fig. 2a). In order to explore how phosphorylation impacts the structure of the CTD, we used *D. melanogaster* positive transcription elongation factor b (Dm-P-TEFb) to hyper-phosphorylate CTD2.5 in vitro. Analysis of hyper-phosphorylated CTD2.5 by mass spectrometry (MS) revealed successful incorporation of up to 12 phosphates per polypeptide (Fig. 2b). Phospho-site identification by tandem MS (MS/MS) led to the conclusion that in vitro phosphorylation protocol predominantly generates high levels of Ser5 phosphorylation (Fig. 2a, c). Preferential phosphorylation in vitro of Ser5 over other amino acids in the CTD by human P-TEFb has been previously observed.

**Hyper-phosphorylation does not alter the scale of CTD2**. We next turned to a detailed investigation of the effects of hyper-phosphorylation on the structure of the CTD2.5 region. In order to test how compact CTD2.5 is in solution, which may
impact its accessibility to CTD-binding factors, we collected small-angle X-ray scattering (SAXS) data on the unphosphorylated and hyper-pSer5 states (Fig. 4, Supplementary Table 3). In the unphosphorylated state, CTD2' displayed an average \( R_g \) of 28.0 ± 0.7 Å, while hyper-pSer5 CTD2' displayed a similar average \( R_g \) of 28.3 ± 0.3 Å using the Guinier approximation (Supplementary Fig. 3). For comparison, the ubiquitous CTD interaction domain (~140 amino acids)22) has an \( R_g \sim 17 \) Å; in contrast, the nucleosome core particle (~800 amino acids and 146 DNA base-pairs) has an \( R_g \sim 41 \) Å, which demonstrates that CTD2' is relatively expanded in solution. Similarly, pair-wise distance distributions revealed no significant increase in the maximum dimension (\( D_{max} \)) upon phosphorylation (Fig. 4b). These results are in good agreement with the dimensions predicted for an excluded volume random coil with the same number of monomers as CTD2' (ref. 23). Independently, \(^{31}P\) NMR spectroscopy revealed that the phosphates in the CTD were in the +2 charge state under our experimental conditions (Supplementary Fig. 4), and yet charge–charge repulsions did not appear to impact the dimensions of the CTD. Incorporation of this data into a model for random-coil structure demonstrates that the median pSer5–pSer5 distance (approximately 18 Å) is likely to be greater than the Debye screening length under our experimental conditions (Supplementary Fig. 5), which accounts for the lack of chain expansion upon hyper-phosphorylation. In summary, our SAXS data demonstrate that the region of the CTD encompassed by CTD2' deletion (CTD2) and recombinant protein CTD2' (black lines, bottom) contain a highly conserved region.

**pSer5 induces sequence-dependent CTD proline isomerization.** On the local scale of several to tens of amino acids, which is the size of the motifs most CTD-interacting domains recognize,
phosphorylation has been shown to strongly perturb backbone dihedral angles\textsuperscript{24–26}, suggesting that phosphorylation could induce local structural perturbations in the CTD. Therefore, we used NMR to probe the backbone conformation of CTD\textsuperscript{2}.

Secondary structure populations for the unphosphorylated state of CTD\textsuperscript{2} were calculated from chemical shifts using the secondary structure calculation program d\textsuperscript{2}D (Supplementary Fig. 6a), revealing small populations of extended \(	ext{b}/\text{PPII}\) character and a strong propensity for random coil-like conformations, consistent with our SAXS results and previous solution studies of CTD peptides\textsuperscript{15}. Further, proline \(C\beta\) and \(C\gamma\) chemical shifts demonstrated a strong preference for the \(\text{trans}\)-proline conformation (~95\% \(\text{trans}\)-proline isomer) (Fig. 5a, Supplementary Fig. 6b). Thus our NMR and SAXS data are strongly consistent with Dm CTD free in solution adopting a spatially heterogeneous ensemble that is highly dynamic on the approximately nanosecond timescale. This observation is consistent with prior reports which concluded that short CTD-derived peptides are predominantly random in solution, lacking long-range order or temporally persistent tertiary structure\textsuperscript{14,15,27,28}. In summary, the unphosphorylated state CTD\textsuperscript{2} is highly disordered, temporally dynamic and contains nearly all \(\text{trans}\)-proline.

Surprisingly, in the hyper-pSer\textsuperscript{5} state, CTD\textsuperscript{2} displayed a twofold enrichment of \(\text{cis}\)-proline content averaged over all peptides (\(\Delta\delta\) AV ~0.5 p.p.m.) (Fig. 2f). This is consistent with the idea that phosphorylation induces an \(\text{cis}\)-to-\(\text{trans}\) isomerization of Pro\textsuperscript{6} (P1718 and P1732; Supplementary Fig. 6b).
19 proline residues (Fig. 5b, Supplementary Fig. 6c). Further, where isomerization occurred, peak splitting into two sets of NMR resonances was observed for Thr4, Ser7, Cys7 and Asn7 residues, accompanied by large chemical shift perturbations (blue bars, Fig. 2f). The presence of two sets of assignable resonances suggests a chemical exchange process on the millisecond timescale or slower. To confirm exchange between cis- and trans-proline isomers, we collected 15N ZZ-exchange NMR spectra, which permits the quantitative observation of conformational exchange on the ms–s timescale, in the presence of the Drosophila prolyl isomerase Dodo 29 (Supplementary Fig. 7). Exchange peaks were observed for all Pro6 residues adjacent to pSer5, but not for Pro3, consistent with Dodo specificity for the pSer5–Pro6 pair30. Interestingly, no exchange could be observed for the pThr5–Pro6 pair in the YTPVpTPS sequence context, suggesting that this heptad is essentially trans-locked on the 100 ms timescale, even in the presence of a prolyl isomerase. Thus, in the hyper-pSer5 state, CTD2 experiences slow exchange between trans- and cis-proline isomers. In this exchange regime, peak intensities correspond to the populations of cis- and trans-proline species, which allowed

![Figure 3](image-url)

**Figure 3** Phospho-sites and proline isomerization in hyper-pSer5 CTD2 probed by NMR spectroscopy. (a) 2D 1H–15N correlation spectra of unphosphorylated (black) and hyper-pSer5 (red) CTD2. (b) 2D 13C0–15N correlation spectra of unphosphorylated (black) and hyper-pSer5 (red) CTD2. (c) Annotation of pSer5/pThr5 resonances in the downfield region of the 2D H–N correlation spectrum. (d) Proline region from the 2D C0–N correlation spectrum of unphosphorylated (black) and hyper-pSer5 (red) CTD2 with some proline resonances annotated.

![Figure 4](image-url)

**Figure 4** Small angle X-ray scattering reveals no significant change in pair-wise distances within CTD2 upon extensive serine 5 phosphorylation. (a) Raw scattering data for unphosphorylated CTD2 (grey circles, bottom) and hyper-pSer5 CTD2 (grey circles, top). Fits for unphosphorylated CTD2 and hyper-pSer5 CTD2 are shown superimposed on the raw data (solid black and red lines, respectively). (b) Representative pair-wise distance distributions for unphosphorylated CTD2 (black) and hyper-pSer5 CTD2 (red) calculated using the autoGNOM function in Primus qt, where the error bars represent the fit error.
us to estimate the magnitude of cis-proline within each heptad (Fig. 6a). Within repeats of YSPTpSPS and the similar cysteine-containing repeat of YSPTpSPC, cis-Pro6 content was enriched threefold (to ~15%) by Ser5 phosphorylation. Further, Pro3 within all heptads containing pSer5 showed a modest enrichment of cis-proline content by ~5%, suggesting some non-local effects. Strikingly, Pro6 within heptads containing Asn7 (YSPTpSPN) showed a sixfold (~35%) enrichment in cis-proline. Thus the proline trans-to-cis switch is sequence-context-dependent and modulated by both phosphorylation and deviations from the consensus heptad sequence.

Heptad-specific proline switches modulate Ssu72 activity. The observation that deviations from the consensus heptad sequence modulate the extent to which Pro6 cis-trans equilibria are affected by pSer5 suggests that, in response to uniform phosphorylation patterns, non-consensus heptads may impart an additional layer of specificity for CTD interacting factors. The CTD phosphatase Ssu72 has been shown to exhibit activity towards heptads containing the pSer5–cysPro6 dipeptide pair. Thus we hypothesized that non-consensus heptads within the hyper-pSer5 CTD modulate the apparent Ssu72 activity through pSer5 induced Pro6 isomerization. To test this possibility, we followed the dephosphorylation of hyper-pSer5 CTD2 by D. melanogaster Ssu72-symplekin using RT-NMR spectroscopy (Fig. 7). Loss in NMR peak intensities relative to the zero time point were observed for all heptads containing pSer5–Pro6 dipeptide pairs (Fig. 7a, Supplementary Fig. 8). Interestingly, in each heptad sequence context, different apparent rates of dephosphorylation were observed (Table 4). For pSer5 residues within the region flanked by the two consensus heptads, similar apparent rates were observed, suggesting that small deviations from the consensus motif (YSPTpSPC or YSPSpSPS) do not dramatically alter Ssu72 activity. However, minimal Ssu72 activity was observed for pS1682 (YSNPpSPS) and no dephosphorylation could be observed for pS1675 (YSPPSSN), consistent with the requirements of Thr4 and Pro6 for Ssu72 activity.

Strikingly, Ssu72 exhibited nearly threefold apparent activity enhancement towards pSer5 residues within the Asn7 heptads, relative to the consensus motifs, strongly suggesting that the higher propensity for cis-Pro6 within these motifs increases the apparent Ssu72 activity.

To understand how residues flanking the pSer5–Pro6 pair affect pSer5 dephosphorylation by Ssu72, we analysed the dephosphorylation of phosphoryl CTD peptides bound to Ssu72. Several structures have been published for Drosophila or human Ssu72 bound to CTD peptides of different phosphorylation states, including Drosophila Ssu72, Drosophila Ssu72-Symplekin and human Ssu72-Symplekin, bound to pSer5 CTD (PDB codes: 3P9Y, 4IMJ and 3O2Q, respectively), and Drosophila Ssu72-symplekin bound to a CTD peptide with Thr4/Ser5 doubly phosphorylated (PDB code: 4IMF). The superimposition of these structures reveals that all known phosphoryl CTD peptides adopt a tight turn facilitated by cis-proline upon binding to Ssu72-symplekin (Fig. 8a). This tight turn is stabilized by three intra-molecular hydrogen bonds formed by the hydroxyl side chain of Thr4 with the main chain carboxylate of Ser7 (2.8 Å) and the amide group of Tyr1 in the following repeat (3.3 Å), as well as the main chain carboxylate of Thr4 and Pro6 (3.2 Å). Due to the high conservation of the tight turn configuration, the identity of the residues flanking Pro6 can be altered for effective Ssu72 recognition as long as the intra-molecular hydrogen bond network is maintained. For example, our NMR results demonstrate that Ssu72 dephosphorylates YSPPpSPS or YSPStpSPC with similar efficacy as it does consensus heptads; molecular modelling suggests that, in all three of these heptad motifs, the intra-molecular hydrogen bond network can be conserved even as the sequence varies (Fig. 8b). On the other hand, little to no Ssu72 activity was observed upon the replacement of Thr4 by Asn. We attribute this loss of activity to the need for Asn4 to adopt an alternative rotameric state to avoid steric clashes, which is incompatible with forming two of the hydrogen bonds that stabilize the Ssu72 recognition conformation (Fig. 8c). We have shown previously that the phosphorylation of Thr4, which also

Figure 5 | Structural characterization of the unphosphorylated and pSer5 CTD2 by NMR spectroscopy. (a) 1HN and 2C chemical shifts from the 3D CCCON spectrum of unphosphorylated CTD2 demonstrate that, when resolved, individual proline side chain resonances show a nearly all-trans state. Blue and red bars represent the range of chemical shifts (mean ± s.d.) for prolines in the trans and cis conformation, respectively. (b) 1HN and 2C chemical shifts from the 3D CCCONH spectrum of hyper-pSer5 CTD2 reveal dramatic trans to cis conformational switches in response to pSer5.
disrupts these two hydrogen bonds, reduces Ssu72 activity fourfold. In the context of the present study, the most significant observation of our NMR analysis of Ssu72-catalysed CTD20 dephosphorylation is that Ssu72 shows its greatest activity toward pSer5-CTD heptads containing Asn7, which we have also shown are the heptads most highly enriched in cis-proline among those observed in this region of the CTD. For this heptad, molecular modelling suggests that the side chain of Asn7 could form an additional intra-molecular hydrogen bond with the carboxylate group of Thr4 (Fig. 8d). With all possible isomeric states of the Asn sidechain, the most favourable configuration is within 3.2 Å away from the backbone carboxylate of Thr4, which further strengthens the tight turn conformation needed for Ssu72 recognition and makes the cis-Pro6 more energetically favourable.

Taken together, this data strongly suggest that, in the presence of uniform phosphorylation patterns, non-consensus CTD heptads encode cryptic structural switches to fine tune the specificity of CTD interacting factors.

**Discussion**

Based on these results, we propose a model in which hyper-pSer5 does little to alter the scaling properties of the CTD20 region of Drosophila CTD; specifically the measured $R_g$ and $D_{max}$ are unaltered. Instead, dramatic structural rearrangements occur on the single heptad scale, driven by sequence context-dependent proline trans-to-cis isomerizations (Fig. 6b). A general conclusion from these observations is that these features allow the CTD to transduce homogenous PTMs into structurally and functionally diverse responses. This discovery predicts multiple potential mechanistic outcomes in the context of the CTD code.

The first general conclusion supported by our findings is that sequence context-dependent structural switches created through...
enriched cis-proline isomerization have the ability to facilitate or impair the binding of isomer-sensitive CTD interacting factors at specific regions of the CTD. The diversity of CTD sequences across eukaryotes has been recently acknowledged3, though the functional significance of many non-consensus heptads has not been widely investigated. Non-consensus CTD repeats may expand the repertoire of available PTMs, thus increasing the complexity of signalling through the CTD4,35–38. In our minimal system, we observed that non-consensus motifs that contained conservative deviations from the consensus responded similarly to Dm P-TEFb and Ssu72. In contrast, more cryptic variants such as YSPNSPS produced drastically altered outcomes, depending on the modifying enzyme present, and substantial deviation from the consensus heptad sequence rendered some repeats resistant to modification by both enzymes. Even in the context of the heptads that conform the least well to the consensus, we emphasize that our assays did not yield substantial phosphorylation of Ser2 residues, suggesting a strong specificity of Dm P-TEFb for serine residues occupying the 5-position of the heptad in vitro.

Our findings suggest that the relationship between consensus conservation and functional specialization in the CTD may lie on a continuum. In this view, conservative sequence deviations may be tolerated by the majority of regulating enzymes, imparting only modest differences in modification kinetics and patterning. By contrast, more dramatic deviations from the consensus sequence may only support interaction with a subset of regulatory factors. While the unique functions the full set of conserved non-consensus motifs serve during transcription will need to be determined empirically, our first set of observations provide support for the emerging hypothesis that variation in CTD

---

**Figure 7** | Structural switches in hyper-pSer5 CTD2 modulate the apparent Ssu72 activity. (a) Representative kinetic traces of Ssu72 dephosphorylation of pSer5 in CTD2 monitored by RT-NMR. (b) Apparent rate constants for pSer5 dephosphorylation reveal heptad-specific Ssu72 activities. The highest apparent Ssu72 activities are observed for pSer5 residues within heptads containing Asn7. Error bars represent the errors from non-linear least squares fitting. All fitting procedures are described in detail in Methods section.

**Figure 8** | The conserved conformation of CTD peptides recognized by Ssu72. (a) Ssu72 is shown as a ribbon diagram with α-helices in green and β-strands in gold. CTD peptides are shown as coloured sticks with carbon atoms shown in different colours: PDB code 4IMI (yellow), 4IMJ (blue), 3P9Y (salmon), and 3O2Q (magenta). The intra-molecular hydrogen bonds are shown in green dashed lines. The CTD residues are numbered based on consensus sequence and the following repeat residues are labelled with a prime. (b) The intra-molecular hydrogen bond network can be maintained even when Thr4 is replaced by Ser. (c) The replacement of Thr4 by Asn loses two intra-molecular hydrogen bonds. (d) An additional intra-molecular hydrogen bond can be formed (orange dashed line) when Ser7 is replaced by Asn.
emergence of straight-winged adults among the progeny (Actin-GAL4/+; UAS-Rpb1b; UAS-Rpb1b\ ^{2Wtparam}). Western blot analysis for ectopic expression of Rpb1 was done by dissecting late pupae from the pupal case and then homogenizing and boiling the tissue in LDS sample buffer (Invitrogen). Equal numbers of male and female late pupae were selected and pupae of the genotype yw; Actin-GAL4/+; UAS-Rpb1b; UAS-Rpb1b\ ^{2Wtparam} were distinguished from the yw; CyO/+; UAS-Rpb1b; UAS-Rpb1b\ ^{2Wtparam} counterparts by the intensity of the eye (the UAS-Rpb1 and Actin-GAL4 transgenes have white gene markers). For western blotting, tissues equivalent to 0.3 pupae were loaded into each lane on a 3–8% Tris-acetate SDS–polyacrylamide gel electrophoresis (PAGE) gel (21–245 kDa). A broad range (11–245 kDa) anti-Rpb1 antibody (Santa Cruz, sc-13, 1:2000) was used to probe the blots. A rabbit anti-mouse antibody (Santa Cruz, sc-2005, 1:5000) was used to detect the β-actin internal reference. Equal loading was confirmed by staining with Coomassie brilliant blue R-250. Flies were staged according to Drosophila melanogaster developmental times (16). 

Protein expression and purification. Carboxy-terminal domain 2. A synthetic gene for the Drosophila melanogaster RPB1 Carboxy-terminal domain was purchased from GeneArt (Thermo Fisher Scientific). A region corresponding to residues (1657–1739) was amplified by PCR, cut with XhoI (NEB) and XmaI (NEB) and ligated into the pET49b + expression vector (Novagen) using T4 DNA ligase (NEB) to produce a construct containing glutathione S-transferase and His tags. Protein expression was performed in E. coli BL21 DE3 cells. Batch cultures (500 ml) were grown at 37 °C in Luria–Bertani (LB) medium supplemented with 30 μg ml−1 kanamycin. At an optical density at 600 nm (OD600) of 0.8, cells were induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and allowed to incubate at 37 °C for 3 h. Following lysis by sonication on ice, 1 ml (500 mM NaCl, 20 mM imidazole, 2.5 mM β-mercaptoethanol, 10 × EDTA) free protease inhibitor cocktail (Calbiochem) and 10 units of RNase free DNase (NEB), samples were centrifuged at 4 °C for 40 min at 11,500 g. Cleared supernatant was passed over HisPur Ni \(^{2+}\)–NTA resin (Thermo Fisher Scientific) and bound protein was washed with containing 5 column volumes of wash buffer (50 mM Tris/HCl pH 7.5, 500 mM NaCl, 20 mM imidazole, 1.0% Triton-1000, 2.5 mM β-mercaptoethanol). Protein was eluted using elution buffer (50 mM Tris/HCl pH 7.5, 500 mM NaCl, 200 mM imidazole, 2.5 mM β-mercaptoethanol). The glutathione S-transferase and 6 × His tags were removed by adding HindIII, followed by expression of the His-tagged 3C protease (resulting in a N-terminal non-native GPI) and dialysing the mixture overnight against 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 2.5 mM β-mercaptoethanol at 4 °C. The protein was then passed over the Ni \(^{2+}\)–NTA column to remove the protease and non-specifically bound contaminants. A final purification was then performed by size exclusion chromatography in 80 mM Imidazole pH 6.5, 50 mM KCl and 2.5 mM β-mercaptoethanol using P-10 resin (BioRad). 

DM T-TEFb. SD cells were grown in suspension at 27 °C to 1.5 million cells ml−1 and infected with 1/10 culture volume D. mel P-TEFb virus (generous gift from J. T. Lis). Infection was carried out at 27 °C at a shaker speed of 75 r.p.m. for 72 h. Following lysis in 50 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 1% Nonidet P-40, 2.5 mM imidazole and 2.5 mM β-mercaptoethanol and protease inhibitors, by dounce homogenization, lysates were centrifuged at 100,000 g for 30 min at 4 °C. Cleared supernatant was passed through TALON resin (Clontech) and bound protein was washed with 5 column volumes of wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 1% Nonidet P-40, 200 mM imidazole and 2.5 mM β-mercaptoethanol) and flash frozen. Kinase activity towards CTD2′ was confirmed by autoradiography.

Dm Dodo. A synthetic gene corresponding to residues (1–166) of D. mel Dodo was purchased from GeneArt (Thermo Fisher Scientific). The gene was cut with XhoI (NEB) and XmaI (NEB) and ligated into the pET49b + expression vector (Novagen) using T4 DNA ligase (NEB) to produce a His-tagged construct. Protein expression was performed in E. coli BL21 DE3 cells. Batch cultures (500 ml) were grown at 37 °C in LB medium supplemented with 30 μg ml−1 kanamycin. At an OD600 of 0.8, cells were induced with 0.5 mM IPTG and allowed to incubate at 37 °C for 3 h. Following lysis by sonication, samples were centrifuged at 4 °C for 40 min at 11,500 g and purified by affinity chromatography using HisPur Ni \(^{2+}\)–NTA resin (Thermo Fisher Scientific) and eluted using elution buffer. Following removal of the His-tag using His-tagged HRV 3C protease, a final purification was performed by gel filtration in 50 mM Tris/HCl pH 7.5, 150 mM NaCl, 1 mM EDTA and 2.5 mM β-mercaptoethanol using a sephacryl S-100 Hi-prep 16/60 size exclusion column on an Äkta PCLC (GE). 

Methods

Fly procedures. Sequences encoding RNAi-resistant Rpb1\(^{WT}\) or Rpb1 derivatives with a double FLAG-tag at the C-terminus were subcloned into the pUASt-at8 vector, followed by transformation into the attP site on chromosome 3 in the Phc31 attP M66b by line\(^{17}\). UAS-Rpb1 and yw; Actin-GAL4/Cyo were obtained from the Bloomington Stock Center (lines 36,830 and 4,414, respectively). Rpb1\(^{b}\) resistance of the ectopically expressed Rpb1 variants was achieved by changing the part of the coding sequence of Rpb1 that corresponds to the 21 nt RNAi recognition sequence (sense strand: AAGGTGAAACGCTGGAAGA) to AACGGTCAAAGTGGACACAA. The UAS-Rpb1b; UAS-Rpb1\(^{b}\) lines were generated by routine matings and meiotic recombination. The lethality test was done by mating virgin female yw; Actin-GAL4/Cyo to male yw; UAS-Rpb1b; UAS-Rpb1\(^{b}\) at 25 °C. Animals were raised at 21 °C. Rescue was confirmed by the
16°C for an additional 16 h. The cultures were then pelleted, lysed by sonication (lysis buffer containing 50 mM Tris pH 8.0, 500 mM NaCl, 10 mM Imidazole, 0.1% Triton X-100 and 10% glycerol) and centrifuged to separate cell debris. Each protein was purified individually by running the aqueous fraction through a Ni-NTA column (Qiagen) and eluting with imidazole. The tags were removed by thrombin cleavage during dialysis at 4°C overnight. Each protein was further purified by gel filtration (Superdex-75 GE Healthcare) and concentrated. Purity was verified by SDS–PAGE through each step.

To form the Sus72-symplekin complex, the proteins were incubated at a molar ratio of 1:1.5 (Sus72:symplekin) overnight at 4°C in a dialysis bag. The complex was separated from unbound protein by running through gel filtration (Superdex-200 GE Healthcare) and concentrated. Purity was verified using a Bioanalytical Systems instrument.

**Kinase reactions.** Kinase reactions were generally carried out in 50 mM Tris/HCl pH 7.5, 300 mM MgCl2, 2 mM DTT, 1% (v/v) HPLC-grade acetonitrile containing 0.1% FA. The gradient was delivered by a Dionex Ultimate 3000 nano-LC system (Thermo) with a 90-min 4–60% linear gradient of aqueous 0.1% FA and 60% acetonitrile containing 0.1% FA as a mobile phase. The instrument was calibrated using a matrix-assisted laser desorption ionization–time of flight (MALDI-TOF MS) MALDI TOF mass spectra of intact unphosphorylated CTD2 and phospho CTD2 were acquired on an Ultraflextreme instrument (Bruker, Billerica, MA) in linear positive detection mode using a factory-configured instrument parameters for 5,000–20,000 m/z range. The instrument was calibrated using a protein mixture containing bovine insulin, MW 57335; bovine ubiquitin, MW 85648; bovine RNase A, MW 13682.2; equine heart cytochrome C, MW 12359.9 and equine heart myoglobin, MW 169513.1 (all from Sigma); a 20 mg/ml solution of super-DHB (2,5-dihydroxybenzoic acid and 2-hydroxy-5-methoxybenzoic acid, Sigma) in 50% aqueous acetonitrile (ACN) containing 0.1% 1-H phosphoric acid (EMD Millipore) and 0.1% trifluoroacetic acid (Thermofisher) was used as the matrix for both the calibrants and the CTD samples. The CTD samples for the MALDI TOF MS were prepared by mixing 1 µl of 100 mM protein sample in water and 1 µl of the matrix solution and applying 1 µl of this mixture to a polished stainless steel target. Mass spectra were acquired by summing 2000–3000 shots at a 100 Hz laser repetition rate, average calibration error was 521 p.p.m.

**Chromatopin digest.** A 1 µg −1 µl −1 stock solution of sequencing-grade chromatopin (Thermofisher) prepared in 1 mM hydrochloric acid was diluted 50-fold with 1 mM hydrochloric acid and 10% (v/v) aqueous solution of 2-mercaptoethanol and 0.1% trisfluoroacetic acid (Thermofisher) was added to the solution in a 1:1 vol/vol ratio. The proteolysis was allowed to proceed overnight at 37°C. Samples were acidified with a 1% aqueous solution of formic acid (FA), dried down and re-dissolved in 15 µl of 4% ACN containing 0.1% FA for the nano-LC MS/MS analysis.

**NMR Spectroscopy.** For NMR experiments, expression of CTD2 was performed in M9 minimal media enriched with 15N-NH4Cl and/or 13C-Glucose (Cambridge Isotope Laboratories). Following purification, samples were buffer exchanged in Amicon Ultra-15, 3000 MW centrifugal filters (Merck Millipore Ltd.). Typically, 80 mM imidazole pH 6.5, 50 mM KCl, 10% glycerol, 2 mM DTT and 10% D2O was used for NMR experiments. For 22 time-exchange experiments, 20 mM MES pH 6.5, 50 mM KCl, 10% glycerol, 2 mM DTT and 10% D2O was used. To obtain the desired pH range during 18 experiments, 80 mM Citrate pH 4.0/5.0/5.5, 80 mM imidazole pH 6.2/6.5 and 80 mM Tris/HCl pH 7.2/8.3 containing 50 mM KCl, 10% glycerol, 2 mM DTT and 10% D2O were used. NMR Spectra were collected at the Lloyd-Jackson NMR Facility at the Pennsylvania State University on Bruker Avance III spectrometers operating at proton frequencies of 500, 600 or 850 MHz equipped with TCI single-axis gradient cryoprobes (H2/13C/15N/1H) with enhanced sensitivity for 1H and 13C.

**Phosphoros-detect experiments** were performed on the 80 MHz Bruker Avance-III-HD spectrometer equipped with a broadband (BBO) Prodigy CryoProbe. Chemical shift assignments were made using 1H–Direct Detect methods developed in-house[43,44], as well as standard 1H–Detect triple resonance experiments. 1H and 13C chemical shifts were referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and phosphoric acid standards, respectively. 2D Spectra were processed in Topspin 3.2 (Bruker) and NMRPipe and analysed in Sparky. 1D spectra were processed in Topspin 3.2 (Bruker) and analysed in MzML (Mestrelab Research). Average chemical shift perturbations for CTD2 upon phosphorylation were calculated by:

\[
\Delta \delta_{AV} = \frac{1}{1} \left( \frac{\Delta \delta_{\text{pSer}}}{\sqrt{\Delta \delta_{\text{pSer}}}^2} + \frac{0.102 \Delta \delta_{\text{pThr}}}{\sqrt{0.102 \Delta \delta_{\text{pThr}}}^2} + \frac{0.251 \Delta \delta_{\text{pCys}}}{\sqrt{0.251 \Delta \delta_{\text{pCys}}}^2} \right)^{1/2}
\]

where \(\Delta \delta_{\text{pSer}}, \Delta \delta_{\text{pThr}}\) and \(\Delta \delta_{\text{pCys}}\) are the differences in 1H, 13N and 13C chemical shift between the unphosphorylated and phosphorylated species, respectively.

**RT-NMR kinetics and data processing.** Dm P-TEFb was used to perform RT-NMR experiments in 50 mM HEPES pH 6.8, 50 mM KCl, 20 mM MgCl2, 12 mM ATP, 2 mM DTT and 10% D2O with 250 µM CTD2 and ~100 µg ml −1 Dm P-TEFb. Standard H, 13N–HSQC spectra were collected at 850 MHz with 1,256 (m) × 256 (n) points, 4096 transients were accumulated with recycle delay of 15 s after 16 scans were collected. To measure slow sites, 16 scans were acquired for each experiment over the entire time course. As recalibration of the instrument was required following enzyme addition, the first data point was acquired in an effective dead time of ~20 min. Phosphatase reactions were performed in 80 mM imidazole pH 6.5, 50 mM KCl, 2 mM DTT and 10% D2O with 1.0 mM hyper-pSer5 CTD2 and ~10 µg of D. mel Pes72-symplekin complex. For the phosphatase reactions, spectra were collected as described using four traversals through the time course. Spectra were processed in MzML (Mestrelab Research). Extracted peak intensities for pSer/pThr the resonances and effect resonances of the neighbouring species were plotted as a function of time and fit in MATLAB. Single exponential decays and build-up curves were fit as irreversible first-order reactions and intermediate species were fit as two consecutive irreversible first-order reactions by the method of non-linear least squares analysis using:

\[
y = y_0 + S_0 e^{-k_{-pThr}} + S_0 (1 - e^{-k_{-pThr}}) + S_1 (1 - e^{-k_{-pSer}}) + S_1 k_{-pSer} e^{-k_{-pSer} t} + S_{0,1} (1 - e^{-k_{-pThr}})
\]

Where, \(k_{-pThr}\) was calculated as the average between a given pSer resonance and the resonances of neighbouring residues. Reported errors represent the 95% confidence intervals, or the propagation of error where \(k_{-pThr}\) represents an average.

**Small-angle X-ray scattering.** SAXS experiments for CTD2 were performed in 80 mM Trips/HCl pH 7.5, 50 mM KCl, 10% glycerol and 5 mM DTT. SAXS data was collected at the Cornell High Energy Synchrotron Source (CHESS) on the G1 beamline. Incident radiation was produced at 9.963 keV with a flux of 8 × 1010 photons s −1 at 51 mA, providing a q-space range of 0.007–0.7 Å −1. Scattering from a silver bhenate standard was used for q-axis mapping. Data collection was performed using dual Pilatus 100K S detectors. Reduction of the 2D images to 1D scattering profiles was performed using BioXtasy Raw. Scattering profiles and uncertainty estimates were computed for each sample, and data from all 3 exposures, with each exposure comprising 20 1 s frames. Solvent blanks were collected immediately before and after each protein sample exposure by measuring the scattering from the spin column flow through from each sample, and solvent subtraction was performed using equivalent nondoped samples. Protein concentrations were 2.5–11 mg ml −1 for both unphosphorylated and pSer5 CTD2. No signs of aggregation, inter-particle effects or radiation damage were observed. Average radius of gyration (\(R_g\)) values were determined for each sample using the Guinier approximation with \(qR_g\) of 0.8, as suggested for disordered systems, such as CTD2 (ref. 43). Guinier fitting and pair-wise distance distribution calculations were performed using the method of non-linear least squares in MATLAB and the auto-GRAM function in Primus qt, respectively.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon request.
References


Acknowledgements

This work was funded by a US National Science Foundation grant to S.A.S. (MCB-1515974), NIH grant GM047477 to D.S.G., NIH grant GM104896 and Welch Foundation grant F-1778 to Y.I.Z., NIH training grant T32-AI074551 to B.P. and a MHPD scholarship from the Alfred P. Sloan Foundation to E.B.G. We thank the TRIP at Harvard Medical School (NIH GM084947) for providing the transgenic RNAi fly stock. The Cornell High Energy Synchrotron Source (CHESS) is supported by the National Science Foundation under awards DMR-1332208 and DMR-0936884; the Macromolecular Diffraction at CHESS (MacCHESS) facility is supported by the National Institutes of Health under award GM-103485.

Author contributions

E.B.G. and T.N.L. designed experiments, performed experiments, analysed data and wrote the manuscript. F.L. designed experiments, performed experiments and analysed data. B.P., M.I.F. and B.P.M. designed experiments and performed experiments. Y.I.Z. analysed data and wrote the manuscript. D.S.G. and S.A.S. designed experiments, analysed data and wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Gibbs, E. B. et al. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. Nat. Commun. 8, 15233 doi: 10.1038/ncomms15233 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.