bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

O-GlcNAc Transferase Activity is Essential for RNA Pol II Pausing in a Human Cell-Free Transcription System

Brian A. Lewis* and David Levens*, Gene Regulation Section, Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, NIH, 9000 Rockville Pike, Bethesda, MD 20892, *Email: brian.lewis@nih.gov; levensd@mail.nih.gov

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Abstract

RNA polymerase II pausing is the major regulatory point in transcription in higher eukaryotes. Despite considerable knowledge of the general transcriptional machinery that are required to recruit RNA pol II to a promoter, much less is known how a paused RNA pol II is established and its release regulated, and the entirety of the machinery is likely not known. In part, this is due to the absence of an appropriate biochemical system that functionally recapitulates RNA pol II pausing and elongation and with which the pausing machinery can be identified. We describe herein a cell-free system (CFS) derived from HeLa cells that recapitulates pausing and elongation events known to occur in vivo. We have used this system to show that O-GlcNAc transferase (OGT) activity is required to establish a paused pol II, without which RNA pol II does not pause and instead enters productive elongation. Coupled with previous observations we show that both O-GlcNAc addition and removal are functionally required for pausing and elongation, respectively. Furthermore, the CFS offers significant inroads into understanding RNA pol II pausing and its regulation.

Introduction

Our understanding of transcriptional regulation by RNA polymerase II has undergone major revisions over the past several years. Next generation sequencing experiments have shed light on the behavior of RNA pol II in vivo and transformed the problem of understanding the regulation of pol II. The initial view of pol II transcriptional regulation assigned the primary regulatory point to the proximal promoter region (consisting of various core promoter elements and activator binding sites) and the congregation of general transcription factors that were then recruited by an activator. This ordered assembly of factors at the promoter is known as the preinitiation complex (PIC). This view in part arose from the identification of the general factors and the eventual reconstitution of pol II-dependent transcription. Consequently, the formation of the PIC was thought to be the major regulatory step at promoters. Once the initiation of transcription occurred, pol II immediately moved into a state of productive elongation.

Only a small minority of promoters were known to be regulated by the restriction of elongation. The best-known examples of such promoters were the heat shock promoters and the MYC promoter, where engaged pol II sit roughly between +30 and +75 downstream of the transcriptional start site[1–3]. However, over the past 10 years, high throughput sequencing

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

technologies and genome-wide studies have shown that a significant majority of promoters, from Drosophila to humans, are regulated not at PIC formation but at the paused pol II step[4]. Thus, it is important to understand the nature and regulation of the paused pol II both as a fundamental problem and because many diseases are now considered "diseases of uncontrolled transcriptional elongation"[5].

A paused pol II is thought to be established by two factors, DSIF and NELF, that together bind to pol II and effectively misalign the active site of pol II from the DNA template preventing nucleotide incorporation and elongation[6–8]. Pol II release occurs via the phosphorylation of DSIF and NELF by P-TEFb, after which NELF is lost, and the polymerase moves into a productive elongation state[9]. More recently, another catalytic activity was found to participate in pol II release, poly-ADP ribose polymerase 1 (PARP-1)[10].

Nevertheless, our understanding of the establishment of a paused polymerase is incomplete. The addition of purified DSIF and NELF does not establish a paused pol II, but only slows its elongation rate[6,7,11,12]. We also do not understand mechanistically how pausing is established or released, now that PARP-1 and OGA are likely regulating these events along with CDK7 and CDK9. These data imply then that additional factors are required to establish a paused pol II. A human cell-free system that recapitulates pol II pausing would enable dissection of the pausing machinery and the sequence of reactions that regulate pause establishment and release[13].

Removal of the O-GlcNAc post-translational modification by the O-GlcNAc aminidase (OGA) is necessary for pol II elongation in a cell-free system[14]. For this to occur, an O-GlcNAcylated protein substrate(s) must have existed prior to OGA action. Additionally, O-GlcNAcylation by the O-GlcNAc transferase (OGT) is required for the recruitment of pol II to the PIC[15]. In vivo, O-GlcNAc peaks are readily discernable at the 5' end of human genes and colocalize with paused pol II[14]. OGT shRNA reduced transcription and pol II promoter occupancy[16]. High densities of OGA, OGT and O-GlcNAc localize to promoters in humans, M. musculus, Drosophila, and C. elegans[14,17,18]. These data all indicate, in some cases directly, that OGT, OGA, and O-GlcNAcylation have functional roles in pol II pausing and elongation.

Described herein is the establishment of a human cell-free transcription system (CFS) that recapitulates known elongation regulatory events and contains an actively engaged paused pol II which can be released through a variety of ways. This system was used to begin to identify other,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

previously unknown components of the pausing machinery: inhibition of the catalytic activity of OGT results in the loss of paused pol II, which instead immediately elongates, effectively bypassing the pausing step. The CFS shows that pol II pausing can be recapitulated in human derived nuclear extracts and that such a system recapitulates known in vivo regulatory and enzymatic behaviors. Additionally, these data show that the decision to pause is an intrinsic part of the initiation process and its establishment occurs at that time via the action of OGT, identifying O-GlcNAcylation as a participant in pausing and release.

Results

The Cell-Free System Recapitulates In Vivo Elongation Events

Prior work showed that pol II elongation in the CFS was blocked with OGA inhibitors[14]. These effects were virtually identical to those of flavopiridol inhibition of P-TEFb[14] and thus indicate that other catalytic activities are required for productive elongation. An interesting from those experiments is that the RNA products below the labeled tRNA band at +75 were not affected by the inhibition of either OGA or P-TEFb. Lastly, the specific localization of these RNAs between +20 and +75 is exactly where one would expect to find a paused pol II population[2]. The following experiments set out to determine the extent to which the cell-free system recapitulated in vivo behaviors and whether these polymerases between +20 and +75 were in fact paused.

CDK7, CDK9, and CDK12 Inhibitors Block Elongation in the CFS

Recently a nuclear extract from HeLa cells was developed to study RNA pol II pausing and elongation[14]. In order to establish the legitimacy of the CFS, it was necessary to determine whether the system recapitulated the effects of known in vivo elongation inhibitors. Several groups have shown that in vitro and in vivo that CDK7 inhibition with THZ1 affects pol II elongation[19,20]. Addition of THZ1 to the CFS showed a clear elongation block of full-length 548 nt RNA product (lane 4, Fig. 1B). Additionally, the RNA population between +20 and +75 decreased while there is an increase in RNAs above the tRNA band at +75. These data suggest that the short RNA population extended and passed beyond +75. A similar effect was also observed by Fisher and colleagues in vivo, who concluded that CDK7 activity was necessary for pausing to occur[21]. The expected elongation block with the CDK9 inhibitor flavopiridol was also observed (lane 3, Fig. 1B)[22].

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

The CFS was assessed for the effects with the CDK12 inhibitor THZ531, also described by Gray and colleagues, who showed that in vivo THZ531 blocked pol II elongation[23]. Titration of increasing concentrations of THZ531 (in the same volume of DMSO) into the CFS showed effects virtually identical to those of THZ1: a reduction of short RNAs between +20 and +75, and increase in RNAs above +75, and a block of full-length RNA synthesis (compare lane 2 to lanes 3-5, Fig. 1C). This experiment suggests that CDK12, like CDK7, functions to establish a paused pol II and that the CFS is recapitulating in vivo behaviors.

PARP-1 Activity Is Required for Elongation in the CFS

Kraus and colleagues have described a positive role for PARP-1 in pol II elongation in vivo, as shown by GRO-seq analysis using the PARP-1 inhibitor PJ34 and PARP-1 shRNA knockdowns[10]. Addition of PJ34 to the cell-free system resulted in a clear block in elongation (lane 3, Fig. 1D), identical to CDK9 (Figure 1A) and OGA inhibition[14], while having no effect on the +20/+75 RNA population; DMSO alone had no effect (lane 2, Fig. 1D). These experiments all show that the cell-free system faithfully recapitulates the in vivo effects of CDK7, CDK9, CDK12, and PARP-1 inhibition.

The Cell-Free System Contains a Paused Pol II Population

These data also show that the RNAs between +20 and +75 display some of the expected properties of a paused pol II population. As indicated above, they are of the expected size range, are not affected by CDK9 inhibition, and aberrantly release after CDK7 inhibition (Figure 1). Lastly, these RNAs represent stable, transcriptionally engaged polymerases, as they can be further elongated on washed (in 1.6M KCl) immobilized templates after the addition of TFIIF and NTPs[14]. If these are paused polymerases, then one would expect that they satisfy several criteria: sensitivity to the detergent Sarkosyl[24,25], evidence of the participation of DSIF (a known pausing factor), and the stimulation of pol II release into productive elongation by P-TEFb, and finally, the paused pol II population should not chase into longer RNAs unless stimulated to do so.

Sarkosyl Addition to the CFS Causes Release of Paused Pol II

The release of paused pol II with the detergent Sarkosyl has been documented previously, in both nuclear run-on assays and in Drosophila nuclear extracts[24,25]. The proposed mechanism of Sarkosyl release is that treating the paused pol II with a mild detergent removes the pausing factors

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

and thereby permits pol II release and elongation. The addition of Sarkosyl to the CFS after a 5' chase (and then extending the chase an additional 5') releases a fraction of +20/+75 RNAs into longer species, in contrast to a 10' chase without Sarkosyl (compare lanes 1 and 2, Fig. 2A). The second important point here is that the +20/+75 RNAs are remarkably stable, with very little chase into longer products even after allowing the chase to occur for 10'. This experiment shows that the +20/+75 RNAs consist of functional, engaged paused polymerases that can elongate after Sarkosyl addition but otherwise do not chase. This is in contrast to other published elongation systems where the entire pol II population will chase and thus do not represent paused pol II populations[11]. We note that, as with the THZ-induced release, the Sarkosyl-induced release pol II does not efficiently elongate as no increase in full-length products was seen. This suggests that this release is an aberrant release and that other steps have been bypassed that would otherwise permit efficient elongation into the 548 nt product.

GST-SPT5 Stimulates Elongation in the CFS

The DSIF complex is one of two known pausing factors[8]. As such, one might expect that the addition of rSPT5 would stimulate elongation, either by a squelching mechanism where it titrated out other pausing factors, or through its positive elongation activity. To that end, equivalent amounts of either rGST or rGST-SPT5 were titrated into the CFS after PIC formation. After a 10' incubation, the standard pulse-chase assay was done (Fig. 2B). As can be seen, the addition of GST-SPT5 stimulated release of the +20/+75 RNA/paused pol II population, as there is a diminution of RNA around +20 and a concomitant increase of RNA longer than +75 (lanes 6-8), while GST protein alone had little effect (lanes 2-5). This experiment indicates that the CFS contains a pol II population that is responsive to SPT5 and that these pol IIs represent a paused population. Lastly, elongation again was not efficient as no increase in full-length products was seen, indicating that the aberrant release skips steps required for efficient elongation.

Lastly, the strong viral SV40 promoter was used to ask if another promoter could establish a paused pol II population in the CFS. The SV40 promoter was one of the first promoters suggested to have a paused pol II[26]. Titration of the SV40 promoter DNA into the CFS clearly showed the establishment of a paused pol II population between +20 and +75 (lanes 1-5, Figure 2C). This shows that pausing occurs as a default event on two different promoters, but nevertheless, ones that function in human cells and contain a paused pol II in vivo.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

P-TEFb and Super Elongation Complex Stimulate Pause Release in the CFS

P-TEFb activity was originally described by Price and colleagues as necessary for pol II elongation in a DRB-sensitivity assay[27]. The targets of P-TEFb were identified by Handa and colleagues, also using a functional biochemical approach to purify DSIF and NELF[28,29]. This line of investigation showed that DSIF and NELF were necessary for pol II pausing and P-TEFb activity released the paused polymerase into productive elongation. These three factors form the basis for our current understanding of pol II pausing.

Figure 1 shows that elongation in the CFS is sensitive to flavopiridol, a known inhibitor of the CDK9 activity. In order to assay P-TEFb directly, the endogenous activity must be removed. This is easily done by conjugating biotinylated-CMV promoter template to magnetic beads. Using this assay system, one can isolate various steps in the transcription processes wash off soluble factors, and then add back the activity in question to assess its function[14,30,31]. As indicated in Figure 3A, a 30" pulse was used to label RNA and transcription was stopped with EDTA. The beads were separated from the nuclear extract using a magnet, washed several times in a buffer containing 60 mM KCl, and then resuspended in transcription buffer[31]. As shown in lane 1, Figure 3C, there is no full-length product synthesized after the addition of NTPs for the chase step. The addition of a concentrated 0.3M KCl fraction eluted from a P11 column[28] stimulated the production of full-length RNA (lane 2, Fig. 3C). However, the addition of recombinant P-TEFb, purified from Sf9 cells (Fig. 3B) further stimulated elongation (lane 2, Fig. 3C), showing that the paused polymerases isolated on the immobilized templates are active and capable of being released into productive elongation.

The Super Elongation Complex (SEC) was identified by Shilatifard and colleagues and physically connects the ELL elongation factor to the P-TEFb kinase[32,33]. Both can stimulate elongation in vitro by themselves and the SEC can stimulate elongation as a complex[34]. The question remains however, whether the SEC can stimulate pause release. Flag-tagged SEC[33] (F-SEC) was purified from 293 cells and western blot analysis showed that it contained several SEC subunits (E1-E3, Figure 3D). The F-SEC also could phosphorylate rGST-CTD, as did rP-TEFb, as expected (Figure 3E)[33,34]. Next, F-SEC was assayed using immobilized templates (Fig. 3F). Titration of F-SEC relative to the elution buffer (containing FLAG peptide) showed stimulation of isolated elongation complexes into intermediate and full-length RNA products (compare lanes 2-

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

5 with 6-9, Fig. 3G). Thus, these stably isolated, engaged, paused polymerases can enter productive elongation with either P-TEFb or SEC.

The O-GlcNAc Transferase is Necessary for RNA Polymerase II Pausing

O-GlcNAcylation is an important but underappreciated post-translational modification in higher eukaryotes. The modification is found on serine and threonine resides often in a mutually exclusive relationship with phosphorylation. This is problematic, as many serine to alanine substitutions were used to argue that phosphorylation was important in the protein in question. In contrast to the serine/threonine kinases and phosphatases, there is only one transferase, the O-GlcNAc transferase (OGT) and one removal enzyme, O-GlcNAc aminidase (OGA)[35]. Previously it was shown that OGT is required for PIC formation[15,16] and that OGA is required for proper elongation in the CFS[14]. The OGA elongation defect directly implicates an OGT catalytic requirement, which must occur prior to elongation, since OGA activity requires an O-GlcNAcylated substrate.

To establish this requirement for O-GlcNAcylation, the substrate of OGT, UDP-GlcNAc, was titrated into the CFS (Figure 4A). The titration of UDP-GlcNAc (after PIC formation) resulted in a concentration-dependent decrease in full-length RNA (lanes 1-4, Fig 4A; although significant UDP-GlcNAc levels were previously documented in nuclear extracts[16], those levels may not be saturating for the OGT activity or may be depleted by prior steps and/or long-term storage of the extract itself.). A similar result was obtained with immobilized templates and the 0.3M P11 elongation fraction (used previously in Fig. 3C): addition of UDP-GlcNAc decreased synthesis of full-length RNA (compare lanes 2 and 3 in Fig. 4B). That elongation was blocked suggested that pausing had increased via OGT activity and the O-GlcNAcylation of pausing factors, further indicating that O-GlcNAcylation established a paused pol II.

These data also implicated the O-GlcNAcylation of pausing factors. To test this hypothesis, nuclear extract was incubated with wheat germ agglutinin (WGA)-agarose beads to affinity purify O-GlcNAcylated proteins and the bound proteins were assayed by western blot. NELF-A and -E bound the WGA-beads (left-most lane, Fig. 4C) but did not bind beads that had been preincubated with the competitor GlcNAc, indicating that the WGA interaction was specific to O-GlcNAcylation (+GlcNAc, Fig. 4C). As a control for the possible loss of O-GlcNAcylation by endogenous OGA, the OGA inhibitor PUGNAc (+PUG Fig. 4C) was added but did not show any

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

increased accumulation of O-GlcNAcylated NELF subunits. The literature for mass spectroscopy data of O-GlcNAcylated proteins further shows that at least 32 factors that were O-GlcNAcylated (Figure 4D), including the pausing factors DSIF and NELF[36,37].

Previously, addition of the OGT inhibitor ST045849[38] during PIC formation showed a transcription defect by primer extension assay[16]. Examination of the PIC composition suggested that ST045849 (hereafter referred to as STO4) blocked PIC formation as indicated by the absence of pol II[15]. However, one caveat of the primer extension assay used previously is that it does not allow one to observe the complete process of elongation. Therefore, the more informative pulse-chase assay was used to assay ST04. The addition of ST04 along with CMV promoter DNA at the start of PIC formation completely abrogated all RNA synthesis (compare lanes 1 and 2, Fig. 5A). However, the addition of ST04 after PIC formation showed a release of pol II from the paused population into elongation (post-PIC, compare lanes 1 and 4, Figure 5A), while PUGNAc showed the expected elongation defect[14] (lane 5, Fig. 5A). Secondly, titration of ST04 and a second OGT inhibitor ST06 showed similar effects: 0. 1mM ST04 and to a lesser extent 1mM STO6 showed considerable transcription but where the paused pol II population at +20/+75 was released into longer RNAs (compare lane 2 to lanes 3 and 6, Fig. 5B).

More recently, the Walker lab has developed several additional OGT inhibitors[39]. Titration of these inhibitors post-PIC formation showed an increase in full-length RNA relative to the untreated and DMSO-treated nuclear extract with both OSMI-2a and OSMI-4a (compare lanes 1 and 2 to lanes 3 and 4, Fig. 5C). Further refinement of the ST04 titration showed the same effects: a significant increase in the full-length RNA product and the apparent escape of the paused pol II population, especially the population just above +20 (lanes 5-7, Figure 5C). As the ST04 concentration increased, a second effect appeared closer to the full-length RNA at 548 nt (as seen in lane 4, Fig. 5A and lanes 3 and 6, Fig. 5B), suggesting a second OGT requirement for elongation into a full-length RNA at 548nt. These experiments, using four different OGT inhibitors, show that OGT activity is necessary to establish a paused pol II, and in the absence of OGT, the pol II does not pause, but immediately moves into an elongation phase.

If OGT is required for pausing, then the addition of rOGT and UDP-GlcNAc should reverse the effects of ST04. This experiment is readily done with immobilized templates. PICs were formed for 30', then, where indicated, ST04 was added for 10'. Transcription was initiated with

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

a 30" pulse followed by inhibition with 20 mM EDTA. The beads were then isolated using a magnet, washed with transcription buffer to remove STO4, NTPs, and nuclear extract, and then resuspended in transcription buffer. To the indicated reactions rOGT or rOGT plus UDP-GlcNAc was added for 15', followed by a 5' chase with cold NTPs (Figure 5D). As expected, the untreated templates did not appreciably elongate after washing and the ST04 reaction showed much higher levels of RNAs longer than +75 (lanes 1 and 2, Fig. 5E). The addition of rOGT has a slight effect on the levels of these RNAs (lane 3, Fig. 5E) but the addition of both rOGT and its nucleotide-sugar substrate UDP-GlcNAc dramatically suppressed the levels of RNAs longer than +75 (lane 4, Fig. 5E), indicating that OGT catalytic activity reversed the inhibitory effect of ST04. This experiment shows that OGT catalytic activity is directly necessary for establishment of the paused pol II.

Discussion

Recapitulation of RNA pol II pausing

We present here evidence that a human cell-free system (CFS) recapitulates the expected behaviors of pol II pausing and elongation. It shows the expected sensitivity to CDK7 and CDK9 inhibitors which were previously established in vivo and in other CFSs. The CFS is also dependent on CDK12 and PARP catalytic activities, whose requirements had been established by in vivo experiments[10,23], but which remained to be seen in a direct, biochemical transcription assay. These data also show that the effects of CDK12 and PARP observed in vivo are due to direct effects on pol II elongation and not indirect, genetic effects or system-wide effects due to drug inhibition and shRNA knockdown.

The CFS establishes a bona fide paused pol II, based on ten different criteria (Figures 1-3). 1) The CFS synthesizes a population of paused RNAs found between +20 and +75 downstream of the TSS. 2) These RNAs are stably engaged to the DNA template and remain associated with the pol II. 3) The +20/+75 RNA population does not chase into longer RNAs. 4) Sarkosyl treatment of this population on washed immobilized templates results in the release of the pol II and the synthesis of longer RNAs. 5) The treatment of the CFS with the CDK9 inhibitor flavopiridol inhibits elongation after +75 but does not perturb the +20/+75 RNA population. 6) The

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

conclusions for flavopiridol also apply to the PARP-1 inhibitor PJ34. 7) In contrast, treatment with the CDK7 inhibitor, THZ1, eliminates the +20/+75 RNA population, as observed in other experiments[19,20]. 8) Excess SPT5 results in the synthesis of RNAs longer than +75, as expected if titrating out associated pausing factors. 9) The +20/+75 RNA population is released upon treatment with either rP-TEFb or purified SEC. 10) Price and colleagues recently showed by PRO-seq analysis that in vivo the CMV IE promoter (and all other CMV promoters) are in fact paused[40]. These data directly show that the CFS establishes an engaged, paused RNA pol II population between +20 and +75 and which will enter productive elongation under proper stimuli.

These data indicate that nucleosomes likely do not play a role in establishing or maintaining a paused pol II since there is no chromatin in our system. This is consistent with the CMV genome in vivo which displays pausing at all viral promoters but is not packaged into chromatin[40]. The positioning of a nucleosome downstream of the paused pol II never strongly supported that interpretation either[13].

Additional elongation regulatory steps

In analyzing the elongation behaviors in the CFS, it became apparent that there is a second regulated elongation step further downstream of the paused pol II. The THZ1 and THZ531 inhibitors block elongation at approximately +400 (Figure 1). The Sarkosyl-released RNAs do not extend to full-length nor do the GST-SPT5-released RNAs (Figure 2). ST04 also inhibits elongation past +400 (Figure 5) suggesting not only that there is a regulatory point at +400 but that it is OGT-dependent. These data indicate that pause release is not necessarily equivalent to efficient elongation. Note that not all inhibitors behaved this way. PJ34 and flavopiridol completely block elongation without affecting the paused pol II population, while the THZ1, THZ531, and OGT inhibitors effect an apparently aberrant release but one that does not produce a full-length transcript. Such regulatory "checkpoints" have been noted previously using GRO-seq assays and CDK9 inhibitors[41].

O-GlcNAc cycling is necessary for RNA pol II pausing and elongation

O-GlcNAc ChIP-seq peaks overlap paused pol II positions in human BJAB cells[14]. Additionally, mass spectroscopy data indicated that both DSIF and NELF pausing factors are O-GlcNAcylated[36], as is pol II[16,42], and these might account for those observed ChIP-seq peaks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Previous experiments showed that inhibition of OGA in the CFS blocked elongation[14], as effectively as the CDK9 and PARP-1 inhibitors in this study. Those experiments indicated that OGA catalytic activity was required for elongation and secondly, that OGT must have acted prior to that point in order to create a substrate for the OGA.

The experiments herein confirmed this hypothesis and showed that OGT inhibition allowed pol II to bypass a pausing step and proceed into elongation (Figure 5). The rescue of the OGT inhibition showed that this was a direct consequence of OGT catalytic activity and is supported by the titration of UDP-GlcNAc into the CFS, which blocked synthesis of a full-length RNA product (Figure 4). Likewise, the addition of rOGT and UDP-GlcNAc into an EEC immobilized template assay also blocked elongation (Figure 4). The aforementioned results suggest a model where O-GlcNAcylation addition and removal respectively regulate the establishment of a paused pol II and the subsequent release of the paused pol II into productive elongation. One can imagine two mechanisms by which O-GlcNAcylation regulates these processes. The first is that O-GlcNAcylation regulates phosphorylation by blocking access of kinases to the necessary serine and threonine residues. OGA activity then removes this block and permits phosphorylation to occur, leading to pol II escape into productive elongation. The second mechanism is that the O-GlcNAcylation addition/removal is directly regulating the structures of the pausing and elongation protein machinery and thereby regulating their activities.

A multi-factor model of RNA pol II pausing and release

These data can be further assembled into a multifactor model of pol II pause establishment and release. Pol II pausing is established by DSIF and NELF in conjunction with the catalytic action of CDK7[21], CDK12/13[23], and OGT. Release of the paused pol II into productive elongation is also a multifactor release step requiring P-TEFb[9], OGA[14], and PARP-1[10] catalytic activities (Figure 6). There are likely several protein targets for each of these enzymes, in addition to the ones depicted in the model. Although these catalytic activities are required in concert in this system with the CMV IE promoter, there may be subsets of genes in vivo that do not display requirements for all of these activities[10]. It is important to note that these are not redundant regulatory events. Instead, each is required to stimulate pause release. It is also apparent that pausing consists of many discrete kinetic steps as indicated by the number of RNA products in the CFS assays, where each RNA product between +25 and +75 implies a relatively slow kinetic step

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

in the pausing process. Lastly, these catalytic requirements for pause establishment and release are usually assumed to contribute post-translational modifications, but ATP, UDP-GlcNAc, and NAD+ are all high energy compounds. Hence, we hypothesize that the hydrolysis of these high energy donors defines specific energy requirements for pausing and pause release in addition to the effects of the concomitant post-translational modification.

The functional necessity of RNA pol II pausing

It is unclear why there is pausing in metazoans, but not in evolutionarily more distant organisms. Functionally however, evidence exists that the disruption of pausing leads to an increase in the variance of levels of RNA synthesis[43]. The premise of this noise suppression is that there are numbers of steps that pol II must progress through during the pausing process. A decrease in the number of those steps will increase the variance of expression. Secondly, these steps are, by definition, irreversible[43]. The distribution of RNAs between +25 and+75 in the CFS is evidence of these steps.

We suggest that these steps are a kinetic proofreading mechanism meant to suppress both the temporal variance in regulated expression and to distinguish between specific and nonspecific transcription. These steps effectively filter out nonspecific transcription from progressing throughout the genome, while being permissive to specific transcription from a promoter. Plasticity of the functional moieties in the active site of RNA polymerase provides a means to modulate the efficiency of nucleotide addition[6]. An energy-dependent allosteric transition from a paused to an elongation-efficient polymerase would constitute a promoter checkpoint defined by the kinetic proofreading concept which demands an irreversible energy expenditure to minimize errors.[44,45]. Here however, the kinetic proofreading of pol II does not edit out incorporation errors. Instead, the pausing/elongation machinery uses these kinetic steps to distinguish between a nonspecifically initiated polymerase which would not pass through the kinetic checkpoints, and a specific, paused polymerase that progresses through a series of kinetically defined checkpoints that validate its legitimacy. The nonspecific polymerases do not get these conditional green lights because they were not tagged during initiation as being specific. Hence, a polymerase will encounter steps designed to let specific polymerases proceed to the next step while the nonspecific polymerases will likely be terminated.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

We suggest here that one function of the CDK7, CDK12/13, and/or OGT catalytic activities is to create these irreversible kinetic proofreading steps. The reason that these evolutionarily more distant organisms do not have pausing is that their genomes are small enough that the ratio of specific to nonspecific transcription is such that the nonspecific events are relatively negligible. These catalytic activities all utilize high energy donor compounds from various metabolic pathways. The OGT substrate UDP-GlcNAc is made from fructose-6-phosphate that originates from an early step in glycolysis. The PARP substrate NAD+ is derived from tryptophan and may also reflect the state of the Krebs cycle. Finally, ATP is of course the output from oxidative phosphorylation. Thus, if these enzymes are responsive to the metabolic flux of their substrates, then the preinitiation complex/paused pol II/pol II release at the 5' proximal end of genes may then be a sensor of the nutrient/metabolic state of the cell, modulating transcriptional output in response to changing metabolic flux.

The cell-free transcription system

The lack of a human nuclear extract system that recapitulates pausing has hindered understanding of the regulation of the paused pol II. Pol II pausing is of paramount importance, as most genes now are known to be regulated at that stage and not at the earlier PIC formation/initiation stage. The cell-free system described here will allow one to further explore the depths of pol II pausing, the factors involved in this step, and how pausing and pause release might be regulated. Such a system also allows one to directly probe these events, in contrast to the indirect genetic experiments to which one must resort for in vivo studies. Through the combination of the bottom-up functional biochemistry and in vivo genetic approaches, one should be able to further understand RNA pol II pausing and its regulation.

Author Contributions

B.A.L. conceived of the work, designed and conducted the experiments. B.A.L. and D.L. analyzed the data and wrote the manuscript.

Competing Interests

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Methods

In vitro transcription assays

Assays were all done (NE-based and EEC-based) as in the appropriate figure schematics and as described previously[14], except that a 60 mM KCl salt wash buffer[11] was used to wash the EECs.

Proteins, HeLa nuclear extract, and HeLa P.3 Fraction

rP-TEFb was synthesized in Tni-FNL cells by co-expression of CDK9-his6 and cyclinT1 subunits and then purified via IMAC. rGST and rGST-SPT5 proteins were purified as described previously for other GST-tagged proteins[16]. HeLa nuclear extract was prepared as described previously[14]. F-SEC was purified from 293 Flp-in-TRex cells as indicated[33]. rOGT and GST-CTD were purified as described previously[16]. HeLa P.3 fraction was isolated by fractionation of a HeLa nuclear extract with a P11 column. The column was eluted with 0.1, 0.3, 0.5, and 1M KCl BC buffer (containing 20 mM Tris pH7.9 @ 4ºC, 10% glycerol, 0.2 mM DTT, 0.2 mM EDTA). The 0.3M peak was pooled and precipitated with 50% ammonium sulfate. The resulting precipitate was suspended in 1/10th volume (relative to the starting material volume) BC buffer without KCl and dialyzed against 0.1M KCl BC buffer.

WGA assay

20 µL wheat germ agglutinin (WGA) agarose beads was washed 1x with 500 µL BC100, spun to remove buffer with 26 gauge needle/syringe. 50 µL nuclear extract was mixed with 50 µL H_2O and incubated with the WGA beads for 60 minutes with rotation at room temperature. For the addition of PUGNAc, 10 µL of 20 mM PUGNAc was added to the 100 µL mixture of extract and H_2O. To use GlcNAc as competitor, 50 µL 1M GlcNAc was added to the 50 µL extract and incubated with the WGA beads. After incubating for one hour, beads were washed 3x with BC100, eluted with Laemmli sample buffer, separated by denaturing PAGE, transferred to nitrocellulose, and analyzed for NELF by western blot.

Kinase assay

0.2 µL GST CTD 1-26, 0.5 µL F-AFF1 or 0.5 µL rPTEFb, and 8.5 µL H_2O, 1 uL 10x P-TEFb kinase buffer (50mM Tris pH7.5, 5mM DTT, 5mM $MnCl_2$, 4mM $MgCl_2$[46]) were mixed and incubated for 30 minutes at 30^{0}C. Sample buffer was added to each sample and separated by

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

denaturing PAGE followed by transfer to nitrocellulose and western blotting with anti-phosphoserine 2 antibody.

Antibodies

Anti-pol II CTD phosphoserine S2 (Abcam ab5095)

Anti-ELL2 (Abcam ab194445)

Anti-CDK9 (Bethyl A303-493A)

Anti-Flag (Sigma)

Anti-NELF-A (Santa Cruz sc-32911)

Anti-NELF-E (Santa Cruz sc-32912)

Reagents

ST045849 and ST060266 (TimTec), Flavopiridol (Sigma F3055), PJ34, THZ531 (APExBIO, A8736), PUGNAc (Sigma A7229), ^{32}P-CTP (3000 Ci/mmol, Perkin Elmer), NTPs (Roche), Sarkosyl (Sigma), WGA-agarose, GlcNAc (Sigma), UDP-GlcNAc (Sigma).

Acknowledgments

The authors wish to thank Ali Shilatifard for the F-SEC cell line, Nathaniel Gray for THZ1, Yuki Yamaguchi for the GST-SPT5 expression vector, Jane Jones of the NCI Frederick Protein Synthesis Core Facility for the production and isolation of rP-TEFb, and Suzanne Walker for supplying OSMI-2a and -4a. This work was supported by the Intramural Research Program, Center for Cancer Research, Nation Cancer Institute.

References

1. Bentley, D. L. & Groudine, M. Novel promoter upstream of the human c-myc gene and regulation of c-myc expression in B-cell lymphomas. *Mol Cell Biol* **6**, 3481–3489 (1986).

2. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950–3 (2013).

3. Eick, D. & Bornkamm, G. W. Transcriptional arrest within the first exon is a fast control

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

mechanism in c-myc gene expression. *Nucleic Acids Res.* **14**, 8331–8346 (1986).

4. Adelman, K. & Lis, J. T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* **13**, 720–31 (2012).

5. Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **70**, (2018).

6. Bernecky, C., Plitzko, J. M. & Cramer, P. Structure of a transcribing RNA polymerase II–DSIF complex reveals a multidentate DNA–RNA clamp. *Nat. Struct. Mol. Biol.* **24**, (2017).

7. Vos, S. M. *et al.* Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* (2018). doi:10.1038/s41586-018-0440-4

8. Yamaguchi, Y., Shibata, H. & Handa, H. Transcription elongation factors DSIF and NELF: Promoter-proximal pausing and beyond. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 98–104 (2013).

9. Guo, J. & Price, D. H. RNA polymerase II transcription elongation control. *Chem. Rev.* **113**, 8583–603 (2013).

10. Gibson, B. A. *et al.* Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. *Science* **353**, 45–50 (2016).

11. Cheng, B. & Price, D. H. Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J. Biol. Chem.* **282**, 21901–21912 (2007).

12. Renner, D. B., Yamaguchi, Y., Wada, T., Handa, H. & Price, D. H. A highly purified RNA polymerase II elongation control system. *J. Biol. Chem.* **276**, 42601–9 (2001).

13. Core, L. & Adelman, K. Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* 1–23 (2019). doi:10.1101/gad.325142.119

14. Resto, M. *et al.* O-GlcNAcase is an RNA polymerase II elongation factor coupled to pausing factors SPT5 and TIF 1β. *J. Biol. Chem.* **291**, 22703–22713 (2016).

15. Lewis, B. A., Burlingame, A. L. & Myers, S. A. Human RNA Polymerase II Promoter Recruitment *in Vitro* Is Regulated by *O* -Linked *N* -Acetylglucosaminyltransferase (OGT).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

J. Biol. Chem. **291**, 14056–14061 (2016).

16. Ranuncolo, S. M., Ghosh, S., Hanover, J. a., Hart, G. W. & Lewis, B. a. Evidence of the involvement of O-GlcNAc-modified human RNA polymerase II CTD in transcription in vitro and in vivo. *J. Biol. Chem.* **287**, 23549–23561 (2012).

17. Love, D. C. *et al.* Dynamic O-GlcNAc cycling at promoters of Caenorhabditis elegans genes regulating longevity, stress, and immunity. *Proc. Natl. Acad. Sci.* **107**, 7413–7418 (2010).

18. Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* **32**, 645–655 (2013).

19. Nilson, K. A. A. *et al.* THZ1 Reveals Roles for Cdk7 in Co-transcriptional Capping and Pausing. *Mol. Cell* **59**, 1–13 (2015).

20. Kwiatkowski, N. *et al.* Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature* **511**, 616–620 (2014).

21. Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nat. Struct. Mol. Biol.* **19**, 3–8 (2012).

22. Cheng, B. *et al.* Functional association of Gdown1 with RNA polymerase II poised on human genes. *Mol. Cell* **45**, 38–50 (2012).

23. Zhang, T. *et al.* Covalent targeting of remote cysteine residues to develop CDK12 and CDK13 inhibitors. *Nat. Chem. Biol.* **12**, 876–884 (2016).

24. Rougvie, A. E. & Lis, J. T. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell* **54**, 795–804 (1988).

25. Li, J. *et al.* Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. *Mol. Cell* **50**, 711–722 (2013).

26. Skolnik-David, H. & Aloni, Y. Pausing of RNA polymerase molecules during in vivo transcription of the SV40 leader region. *EMBO J.* **2**, 179–184 (1983).

27. Marshall, N. F. & Price, D. H. Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *Journal of Biological Chemistry* **270**, 12335–12338 (1995).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

28. Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).

29. Yamaguchi, Y. Y. *et al.* NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* **97**, 41–51 (1999).

30. Adamson, T. E., Shore, S. M. & Price, D. H. Analysis of RNA polymerase II elongation in vitro. *Methods Enzymol.* **371**, 264–75 (2003).

31. Cheng, B. & Price, D. H. Isolation and functional analysis of RNA polymerase II elongation complexes. *Methods* **48**, 346–52 (2009).

32. Luo, Z. *et al.* The Super Elongation Complex Family of RNA Polymerase II Elongation Factors: Gene Target Specificity and Transcriptional Output. *Mol. Cell. Biol.* **32**, 2608–2617 (2012).

33. Lin, C. *et al.* AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia. *Mol. Cell* **37**, 429–437 (2010).

34. Biswas, D. *et al.* Function of leukemogenic mixed lineage leukemia 1 (MLL) fusion proteins through distinct partner protein complexes. *Proc. Natl. Acad. Sci.* **108**, 15751–15756 (2011).

35. Bond, M. R. & Hanover, J. A. *O*- GlcNAc Cycling: A Link Between Metabolism and Chronic Disease. *Annu. Rev. Nutr.* **33**, 205–229 (2013).

36. Hahne, H., Gholami, A. M. & Kuster, B. Discovery of O -GlcNAc-modified Proteins in Published Large-scale Proteome Data. *Mol. Cell. Proteomics* **11**, 843–850 (2012).

37. Hahne, H. *et al.* Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry. *J. Proteome Res.* **12**, 927–36 (2013).

38. Gross, B. J., Kraybill, B. C. & Walker, S. Discovery of O-GlcNAc Transferase Inhibitors. *J. Am. Chem. Soc.* **127**, 14588–14589 (2005).

39. Ortiz-Meoz, R. F. *et al.* A Small Molecule That Inhibits OGT Activity in Cells. *ACS Chem.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Biol. **10**, 150309142052000 (2015).

40. Parida, M. *et al.* Nucleotide Resolution Comparison of Transcription of Human Cytomegalovirus and Host Genomes Reveals Universal Use of RNA Polymerase II Elongation Control Driven by Dissimilar Core Promoter Elements. *MBio* **10**, 1–20 (2019).

41. Laitem, C. *et al.* CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II–transcribed genes. *Nat. Struct. Mol. Biol.* **22**, 396–403 (2015).

42. Kelly, W. G., Dahmus, M. E. & Hart, G. W. RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J. Biol. Chem.* **268**, 10416–24 (1993).

43. Weber, A., Liu, J., Collins, I. & Levens, D. TFIIH operates through an expanded proximal promoter to fine-tune c-myc expression. *Mol Cell Biol* **25**, 147–161 (2005).

44. Hopfield, J. J. Kinetic Proofreading. *Proc Natl Acad Sci U S A* **71**, 4135–4139 (1974).

45. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. (Chapman and Hall/CRC, 2019).

46. Zhou, M. *et al.* Tat Modifies the Activity of CDK9 To Phosphorylate Serine 5 of the RNA Polymerase II Carboxyl-Terminal Domain during Human Immunodeficiency Virus Type 1 Transcription. *Mol. Cell. Biol.* **20**, 5077–5086 (2000).

47. Fujinaga, K. *et al.* Dynamics of human immunodeficiency virus transcription: P-TEFb phosphorylates RD and dissociates negative effectors from the transactivation response element. *Mol. Cell. Biol.* **24**, 787–95 (2004).

48. Ping, Y. H. & Rana, T. M. DSIF and NELF Interact with RNA Polymerase II Elongation Complex and HIV-1 Tat Stimulates P-TEFb-mediated Phosphorylation of RNA Polymerase II and DSIF during Transcription Elongation. *J. Biol. Chem.* **276**, 12951–12958 (2001).

49. Yamada, T. *et al.* P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* **21**, 227–37 (2006).

50. Peterlin, B. M. & Price, D. H. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**, 297–305 (2006).

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figures

Figure 1. Cell-free system is dependent on CDK7, CDK9, CDK12/13, and PARP-1.

A. Schematic of steps in the experiments in panels B-D.
B. Shown is a pulse-chase assay using the DMSO solvent alone, 1 μM CDK9 inhibitor flavopiridol (FP), or 4 μM CDK7 inhibitor THZ1 in DMSO.
C. Shown is a pulse-chase assay using the DMSO solvent alone and a titration of the CDK12 inhibitor THZ531 (0.4 μM, 4 μM, 40 μM final concentrations). Each THZ531 addition was in the same volume of DMSO as in lane 2.
D. Shown is a pulse-chase assay using the DMSO solvent alone or the PARP-1 inhibitor PJ34 (80 μM final concentration).

Figure 2. Sarkosyl and excess SPT5 elicit release of paused pol II.

A. Pausing system is sensitive to Sarkosyl and paused pol II does not chase. On the left is a schematic of steps in the experiment in this panel. Shown are pulse-chase assays allowed to chase for either 10' (lane 1) or for 5', followed by the addition of 0.6% final concentration of Sarkosyl, followed by an additional 5' chase.
B. Pol II is released by excess SPT5. Shown on the left is a schematic of steps in the pulse-chase assay on the right. Equivalent amounts of rGST or rGST-SPT5 were titrated into PICs that had been previously formed for 30'. After a 15' incubation, the pulse-chase assay was performed on each titration.
C. SV40 promoter establishes a paused pol II. SV40 promoter DNA (10, 25, 50, 100, 200 ng) was incubated with nuclear extracts for 30' followed by a standard pulse-chase assay.

Figure 3. Super Elongation Complex (SEC) and P-TEFb stimulate release of paused RNA pol II

A. Shown is a schematic of steps in the immobilized template assays in panels C and F.
B. Purification of rPTEFb. Baculovirus expression vectors containing human P-TEFb and cyclin T1 were expressed in Sf9 cells and purified by Ni-Sepharose affinity chromatography.

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

C. Pol II elongation is stimulated by rP-TEFb. Pulsed immobilized templates were isolated after they had been isolated and washed to remove ^{32}P-CTP/GUA and nuclear extract. The indicated factors (P.3c is a concentrated P11 0.3M elution that contains some elongation activity-see lane 2) were added to immobilized templates and incubated for 15', followed by a chase with unlabeled rNTPs. rP-TEFb plus a P11 elongation fraction (P.3c) was added to immobilized templates and compared to either no additional factors (lane 1) or the P.3c fraction alone (lane 2) as indicated in panel A.

D. Purification of F-SEC. Flag-tagged SEC subunit AFF1 was purified from 293 cells using M2-agarose and eluted by excess Flag peptide. Input extract contained undetectable levels of SEC subunits, but which were concentrated by affinity purification (E1-E3), as indicated by western blot analysis for SEC subunits AFF1, ELL2, and CDK9.

E. Kinase activity of F-SEC was assessed by incubating purified F-SEC (lane 2) or rP-TEFb (positive control, lane 3) with the known substrate GST-CTD, containing the first 26 repeats of human pol II CTD. Phosphorylation was detected by western blot using an anti-phosphoserine 2 CTD antibody. Lane 1 contains only the GST-CTD and the kinase reaction buffer.

F. Schematic shows the steps in the immobilized template assay in panel G.

G. F-SEC stimulates release of paused pol II. Either purified F-SEC or the control elution buffer were titrated into pulsed immobilized templates after they had been isolated and washed to remove ^{32}P-CTP /GUA and nuclear extract. After a 15' incubation, unlabeled NTPs were added for 5' to chase any elongation competent pol II into making full-length RNAs.

Figure 4. O-GlcNAcylation activity blocks RNA pol II elongation.

A. UDP-GlcNAc causes a decrease in elongation in nuclear extracts. Standard pulse-chase assays were performed after incubating preformed PICs with increasing amounts of the OGT substrate UDP-GlcNAc (0.1, 0.2, 0.5, and 1 mM final concentrations).

B. UDP-GlcNAc decreases elongation of EEC templates. Immobilized templates and pulsed EECs were isolated and washed in low salt transcription buffer. To the indicated reactions,

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

either P.3c or P.3c plus 0.25 μg rOGT and 0.4 mM UDP-GlcNAc were added for 15' prior to chasing labeled RNAs for 5' with unlabeled NTPs.

C. Pausing factor NELF is O-GlcNAcylated. Nuclear extracts (incubated as indicated with either 3 mM PUGNAc or 0.5M GlcNAc) were incubated with wheat germ agglutinin-agarose for 1 hour and washed several times. The recovered agarose beads were eluted in sample buffer, separated by PAGE, and transferred to nitrocellulose. NELF subunits were detected by western blot.

D. Shown is a list of O-GlcNAcylated pausing and elongation factors that have been identified by mass spectroscopy[36,37].

Figure 5. OGT is required to establish a paused pol II.

A. OGT inhibitor addition after PIC formation causes release of paused pol II. 40 μM STO4 or 3mM PUGNAc were added simultaneously with CMV promoter DNA or 30' afterwards (to allow PICs to form). These reactions were then assayed by standard pulse-chase assay.

B. OGT inhibition causes release of paused pol II. The indicated final concentrations of either STO4 or STO6 were added to preformed PICs which were then assayed by pulse-chase.

C. Three OGT inhibitors affect pause release. Either DMSO, 80 μM OSMI-2a or 160 μM OSMI-4a were added after PIC formation followed by a standard pulse-chase assay (lanes 1-4). Lanes 5-7 show a titration of STO4 (40, 80, 200, 400 μM final concentrations).

D. Schematic shows the steps in the immobilized template assay in panel E.

E. rOGT rescues the OGT inhibitor STO4 aberrant elongation effect. Immobilized templates were created by allowing PICs for form for 30', followed by a 15' incubation with 0.1 mM STO4 and then a 30'' pulse with ^{32}P-CTP /GUA followed by addition of EDTA to stop transcription. Beads were washed as indicated to remove all rNTPs, STO4, and nuclear extract. Individual reactions were suspended in transcription buffer and either rOGT or rOGT and 0.4 mM UDP-GlcNAc were added as indicated, for 15', followed by a 5' chase with unlabeled NTPs.

Figure 6. A multi-factor release model of RNA polymerase II pausing and elongation. The model proposes that there is a paused pol II complex in an O-GlcNAcylated state, partially defined by O-GlcNAcylated pausing factors DSIF and NELF and established by the concomitant catalytic activities of CDK7[21], CDK12/13, and OGT. Release into productive elongation occurs by the

bioRxiv preprint doi: https://doi.org/10.1101/2020.01.23.917237; this version posted January 24, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

action of P-TEFb[9], PARP-1[10], and OGA[14]. This removes GlcNAc from the targeted proteins, ribosylates NELF[10] (and possibly other factors), possibly inducing its release and/or inactivation, and phosphorylates DSIF[47–50], by which it is converted into a positive elongation factor associated with an elongating pol II. The orange "bursting star" surrounding the various metabolites indicates that these hydrolysis products are associated with the release of free energy.

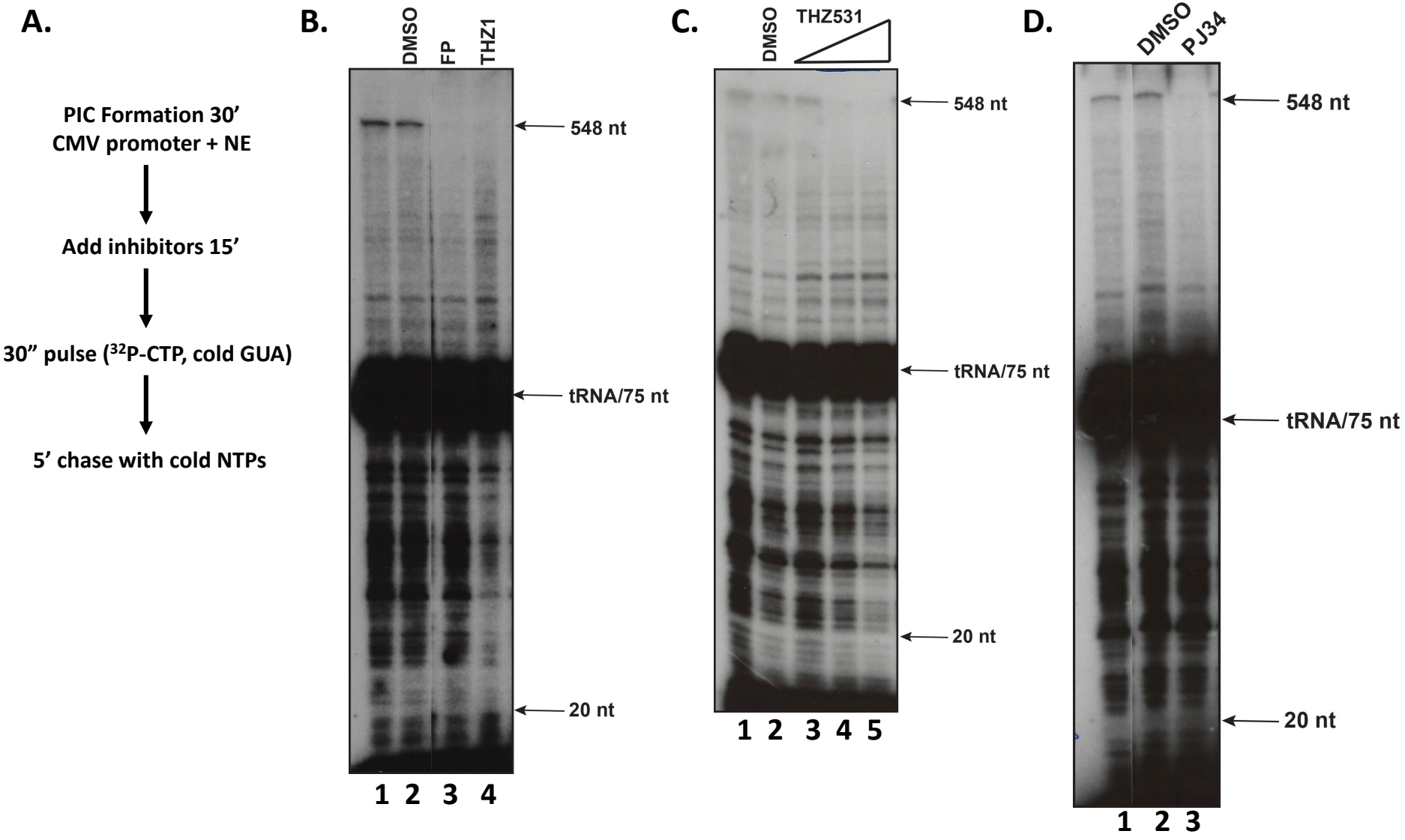
A.
PIC Formation 30'
CMV promoter + NE
Add inhibitors 15'
30" pulse (^{32}P-CTP, cold GUA)
5' chase with cold NTPs
B.
DMSO
FP
THZ1
548 nt
tRNA/75 nt
20 nt
1 2 3 4
C.
DMSO
THZ531
548 nt
tRNA/75 nt
20 nt
1 2 3 4 5
D.
DMSO
PJ34
548 nt
tRNA/75 nt
20 nt
1 2 3

Figure 1

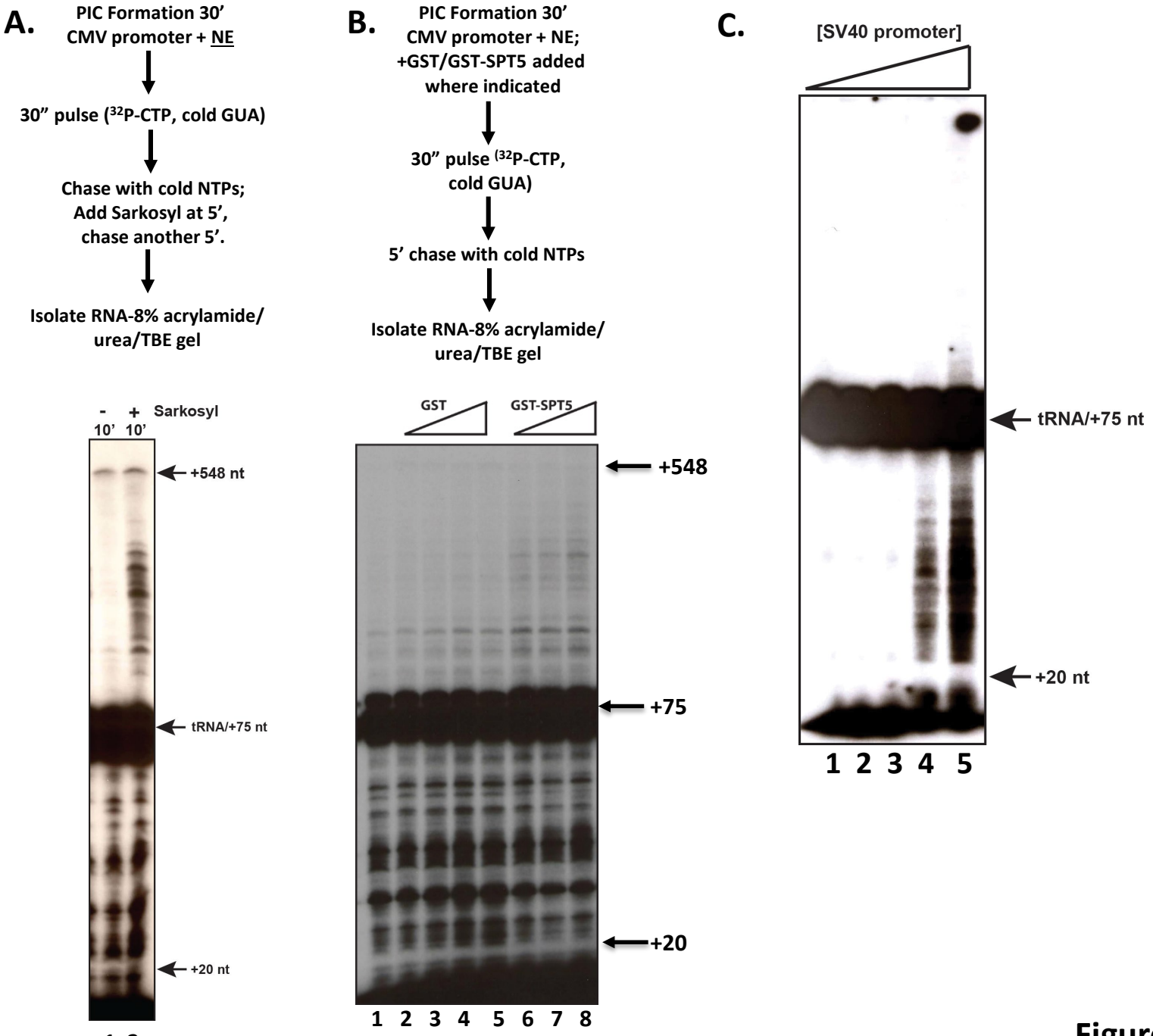
A.
PIC Formation 30'
CMV promoter + NE
30" pulse (^{32}P-CTP, cold GUA)
Chase with cold NTPs;
Add Sarkosyl at 5',
chase another 5'.
Isolate RNA-8% acrylamide/
urea/TBE gel
- 10'
+ 10'
Sarkosyl
+548 nt
tRNA/+75 nt
+20 nt
1 2
B.
PIC Formation 30'
CMV promoter + NE;
+GST/GST-SPT5 added
where indicated
30" pulse (^{32}P-CTP,
cold GUA)
5' chase with cold NTPs
Isolate RNA-8% acrylamide/
urea/TBE gel
GST
GST-SPT5
+548
+75
+20
1 2 3 4 5 6 7 8
C.
[SV40 promoter]
tRNA/+75 nt
+20 nt
1 2 3 4 5

Figure 2

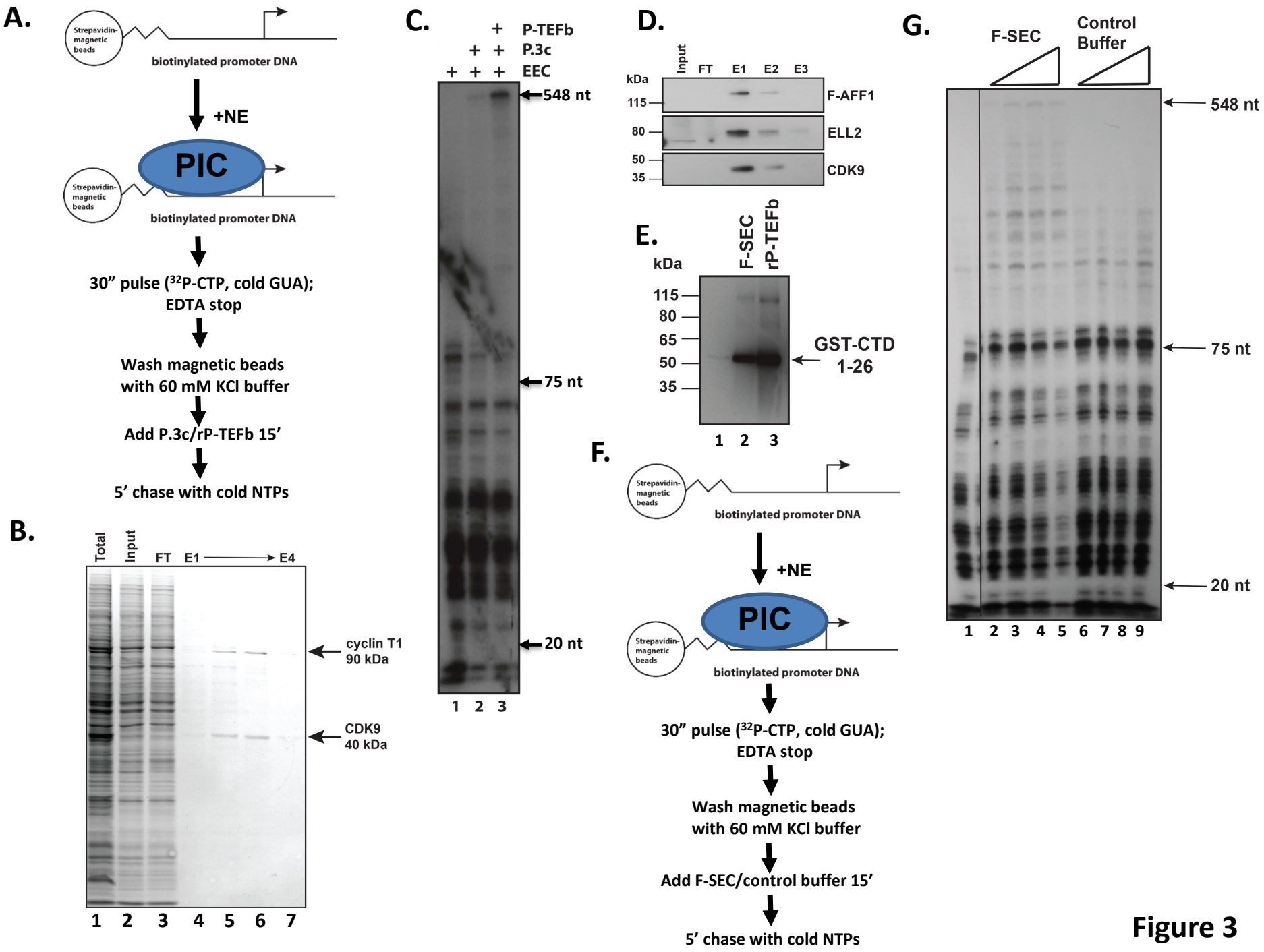
A.
Strepavidin-magnetic beads
biotinylated promoter DNA
+NE
PIC
Strepavidin-magnetic beads
biotinylated promoter DNA
30" pulse ([32]P-CTP, cold GUA); EDTA stop
Wash magnetic beads with 60 mM KCl buffer
Add P.3c/rP-TEFb 15'
5' chase with cold NTPs
B.
Total
Input
FT
E1
E4
cyclin T1 90 kDa
CDK9 40 kDa
1 2 3 4 5 6 7
C.
P-TEFb
P.3c
EEC
548 nt
75 nt
20 nt
1 2 3
D.
Input
FT
E1
E2
E3
kDa
115
80
50
35
F-AFF1
ELL2
CDK9
E.
kDa
F-SEC
rP-TEFb
115
80
65
50
35
GST-CTD 1-26
1 2 3
F.
Strepavidin-magnetic beads
biotinylated promoter DNA
+NE
PIC
Strepavidin-magnetic beads
biotinylated promoter DNA
30" pulse ([32]P-CTP, cold GUA); EDTA stop
Wash magnetic beads with 60 mM KCl buffer
Add F-SEC/control buffer 15'
5' chase with cold NTPs
G.
F-SEC
Control Buffer
548 nt
75 nt
20 nt
1 2 3 4 5 6 7 8 9

Figure 3

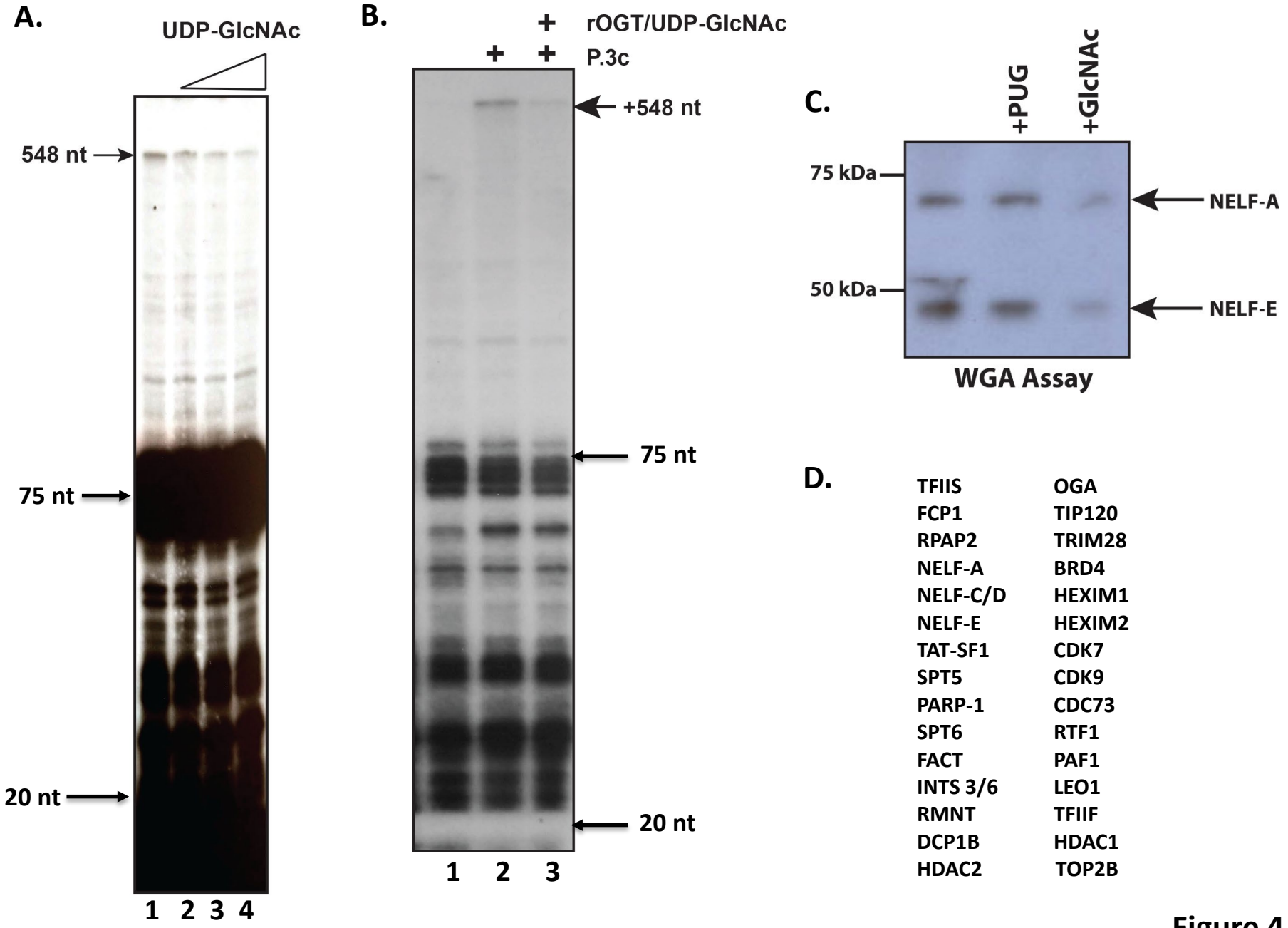
A.
UDP-GlcNAc
548 nt
75 nt
20 nt
1 2 3 4
B.
+ rOGT/UDP-GlcNAc
+ + P.3c
+548 nt
75 nt
20 nt
1 2 3
C.
+PUG
+GlcNAc
75 kDa
50 kDa
NELF-A
NELF-E
WGA Assay
D.
TFIIS
FCP1
RPAP2
NELF-A
NELF-C/D
NELF-E
TAT-SF1
SPT5
PARP-1
SPT6
FACT
INTS 3/6
RMNT
DCP1B
HDAC2
OGA
TIP120
TRIM28
BRD4
HEXIM1
HEXIM2
CDK7
CDK9
CDC73
RTF1
PAF1
LEO1
TFIIF
HDAC1
TOP2B

Figure 4

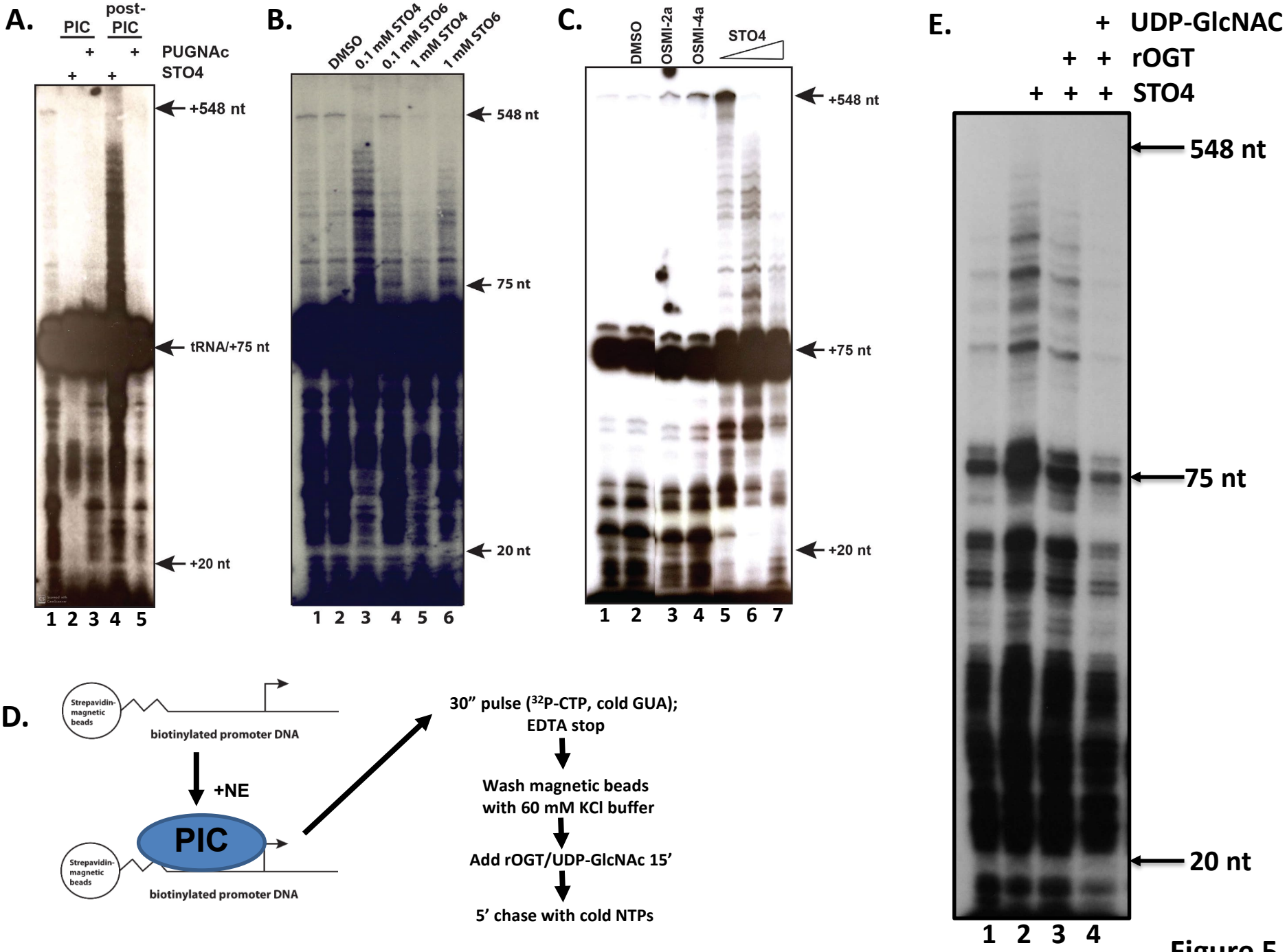
A.
PIC
post-PIC
PUGNAc
STO4
+548 nt
tRNA/+75 nt
+20 nt
1 2 3 4 5
B.
DMSO
0.1 mM STO4
0.1 mM STO6
1 mM STO4
1 mM STO6
548 nt
75 nt
20 nt
1 2 3 4 5 6
C.
DMSO
OSMI-2a
OSMI-4a
STO4
+548 nt
+75 nt
+20 nt
1 2 3 4 5 6 7
D.
Strepavidin-magnetic beads
biotinylated promoter DNA
+NE
PIC
Strepavidin-magnetic beads
biotinylated promoter DNA
30” pulse ([32]P-CTP, cold GUA); EDTA stop
Wash magnetic beads with 60 mM KCl buffer
Add rOGT/UDP-GlcNAc 15’
5’ chase with cold NTPs
E.
+ UDP-GlcNAC
+ + rOGT
+ + + STO4
548 nt
75 nt
20 nt
1 2 3 4

Figure 5

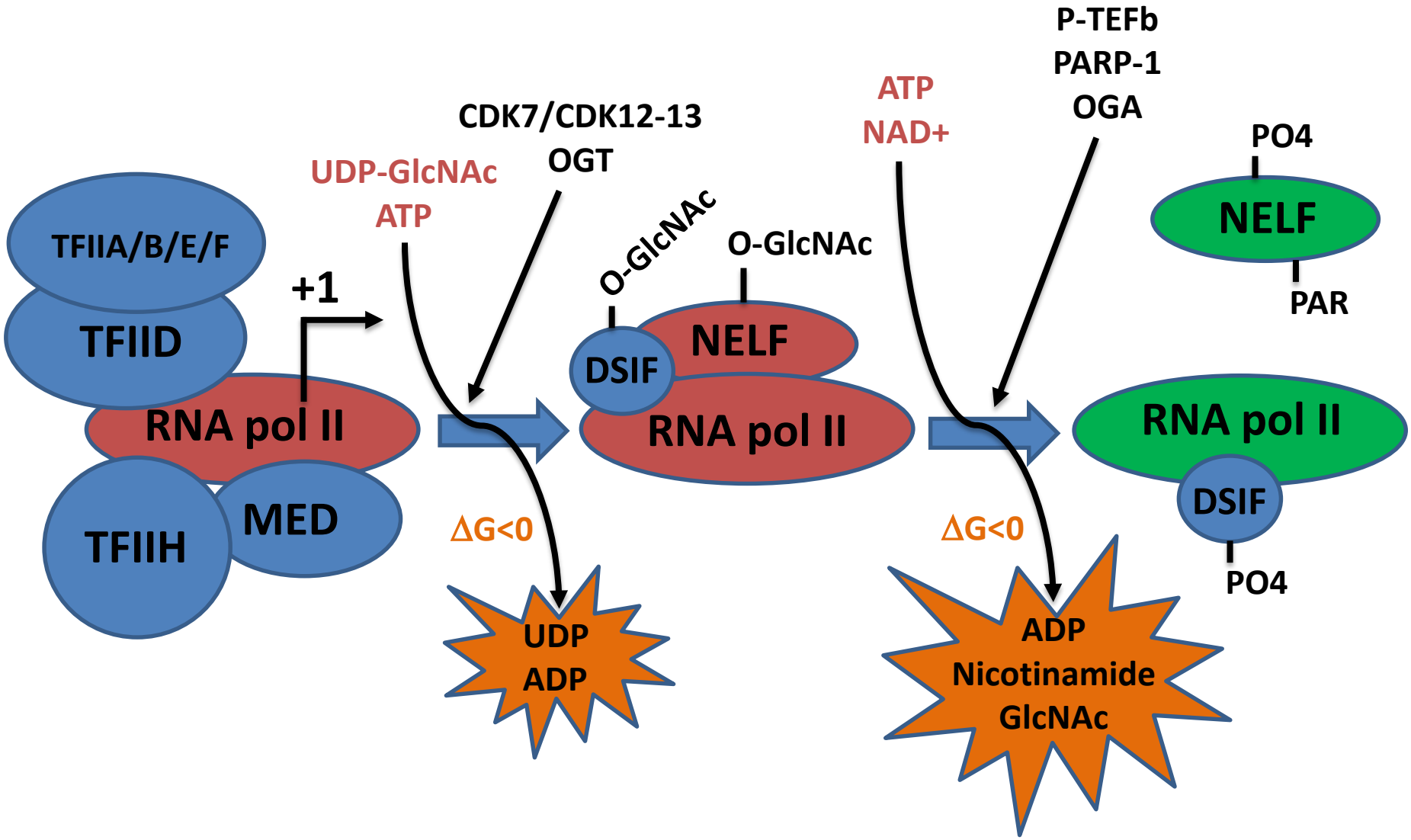
TFIIA/B/E/F
TFIID
RNA pol II
TFIIH
MED
+1
UDP-GlcNAc
ATP
CDK7/CDK12-13
OGT
ΔG<0
UDP
ADP
O-GlcNAc
O-GlcNAc
DSIF
NELF
RNA pol II
ATP
NAD+
P-TEFb
PARP-1
OGA
ΔG<0
ADP
Nicotinamide
GlcNAc
PO4
NELF
PAR
RNA pol II
DSIF
PO4
PIC@Promoter/TSS
Paused Pol II, O-GlcNAcylated State
Elongating Pol II State

Figure 6