Termination of non-coding transcription in yeast relies on both a CTD-interaction domain and a CTD-mimic in Sen1

Han, Z[1,2]; Jasnovidova, O[3,4]; Tudek, A[1,5]; Kubicek, K[3]; Libri, D[1]; Stefl, R[3] and O. Porrua[1*]

[1] Institut Jacques Monod, CNRS, UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, F-75205 Paris, France.

[2] Université Paris-Saclay, 91190 Gif sur Yvette, France.

[3] CEITEC-Central European Institute of Technology, Masaryk University, 62500 Brno, Czechia.

[4] Current address: Max Planck Institute for Molecular Genetics, Ihnestraße 63-73, 14195 Berlin, Germany.

[5] Current address: Department of Biophysics, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland.

*Correspondence should be addressed to: odil.porrua@ijm.fr

ABSTRACT

Pervasive transcription is a widespread phenomenon leading to the production of a plethora of non-coding RNAs (ncRNAs) without apparent function. Pervasive transcription poses a risk that needs to be controlled to prevent the perturbation of gene expression. In yeast, the highly conserved helicase Sen1 restricts pervasive transcription by inducing termination of non-coding transcription. However, the mechanisms underlying the timely recruitment of Sen1 to ncRNAs are poorly understood. Here we employ combined biochemical, structural and genomic approaches to shed light on the protein interactions responsible for the specific function of Sen1 in non-coding transcription. We show that Sen1-dependent termination strictly depends on the recognition of a particular phosphorylated form of the carboxy terminal domain (CTD) of RNA polymerase II by the N-terminal domain of Sen1. In addition, we identify a short motif in an intrinsically disordered region of Sen1 that mimics the phosphorylated CTD and we show that this motif is bound by the CTD-interacting domain of Nrd1, an RNA-binding protein that recognizes specific sequences enriched in ncRNAs. This unique protein configuration in which both a “reader” and a “mimic” of the phosphorylated CTD coexist provides a novel paradigm for the study of protein-protein interactions controlling the action of transcription-related factors.

INTRODUCTION

The concept of pervasive transcription emerged over a decade ago upon the discovery that a large fraction of both the prokaryotic and the eukaryotic transcriptomes is composed of non-coding RNAs (ncRNAs) without any obvious function. Pervasive transcription is potentially harmful for cell homeostasis since it can interfere with normal transcription of canonical genes and provoke the accumulation of toxic RNAs. Therefore, all organisms studied to date have evolved different mechanisms to circumvent the negative consequences of pervasive transcription. These mechanisms often rely on transcription termination and RNA degradation (for review, see Jensen et al, 2013[1]).

In *S. cerevisiae* there are two major pathways for termination of RNA polymerase II (RNAPII) transcription. A pathway that depends on a macromolecular complex including the cleavage and polyadenylation factor (CPF) is essentially responsible for transcription termination at protein-coding genes, whereas the Nrd1-Nab3-Sen1 (NNS) complex targets a large fraction of the non-coding RNAs (ncRNAs) produced in the cell. Specifically, the NNS complex terminates transcription of most sn- and snoRNAs and a class of ncRNAs dubbed CUTs, for cryptic unstable transcripts, that constitutes the major product of pervasive transcription (for review, see Porrua and Libri, 2015[2]). While sn- and snoRNAs are important for splicing and for the correct modification of rRNA, respectively, CUTs are generally considered as non-functional molecules [3–6].

Each pathway is associated with distinct nuclease and polyA-polymerase activities that determine the stability and functionality of the RNAs they take care of. Precursors of mRNAs are cleaved at their 3' ends at the so-called polyA site and polyadenylated by Pap1, which stimulates subsequent export to the cytoplasm and translation (for review, see Porrua and Libri, 2015[2]). In contrast, ncRNAs terminated by the NNS-dependent pathway are polyadenylated by Trf4, a component of the TRAMP complex. These RNAs are then targeted by the nuclear form of the exosome bearing the Rrp6 exonuclease, which catalyses either 3' end maturation, in the case of sn- and snoRNAs, or complete degradation, in the case of CUTs [3,7,8]. Given the very divergent fate of the RNAs terminated by each pathway, several mechanisms have evolved to ensure the specific action of the different protein complexes on the right targets. These mechanisms involve both protein-protein and nucleic acid-protein interactions. Some of the protein interactions that are crucial for coordinated and efficient transcription-related processes as termination and 3' end processing are mediated by the C-terminal domain (CTD) of the largest subunit of RNAPII. The CTD is a large and flexible domain composed of 26 repeats of the heptapeptide YSPTSPS that undergoes dynamic phosphorylation throughout the transcription cycle (for review, see Harlen and Churchman 2017[9]).

The CPF complex interacts preferentially with the S2P-form of the CTD, which is more prominent during middle-late elongation, via the CTD-Interaction domain (CID) of its associated factor Pcf11[10]. In addition, the recognition of several AU-rich sequences, among which the polyA site, by other subunits mediates the specific recruitment of the CPF complex to mRNAs[11]. Similarly, the CID of Nrd1 within the NNS complex interacts with the CTD phosphorylated at S5, which is a mark of early elongation [12,13], and both Nrd1 and Nab3 recognize specific sequence motifs that are enriched at the target ncRNAs [6,14–17]. Both the interaction with the CTD and with the nascent RNA contribute to the early recruitment of the Nrd1-Nab3 heterodimer and the efficiency of transcription termination [13,18,19]. The current model posits that Nrd1-Nab3 would recruit the helicase Sen1 that in turn promotes the final step in transcription termination (i.e. the release of RNAPII and the nascent RNA from the DNA[20]). Subsequently, a complex network of possibly redundant protein-protein interactions involving Nrd1, Nab3 and different components of the TRAMP and exosome complexes promotes efficient degradation of the released ncRNA [19,21,22].

As mentioned above, Sen1 is responsible for dissociation of the elongation complex. Sen1 is a highly conserved RNA and DNA helicase belonging to the superfamily 1 of helicases [23,24]. Transcription termination by Sen1 involves its translocation along the nascent RNA towards RNAPII and possibly subsequent contacts between specific regions of Sen1 helicase domain and the polymerase [20,24,25]. Neither *in vitro* nor *in vivo* Sen1 exhibits any sequence-specific RNA-binding capability [20,26]; implying that Sen1 activity should be regulated in order to ensure its specific action on ncRNAs. Among the mechanisms that might contribute to keep Sen1 under control we can cite: i) the relatively low levels of Sen1 protein (63 to 498 molecules/cell depending on the study[27–30]); ii) its low-processivity as a translocase, which makes termination highly dependent on Sen1 efficient recruitment to the elongation complex and RNAPII pausing[24] and iii) the interaction of Sen1 with the ncRNA targeting proteins Nrd1-Nab3, as evoked above.

Although it has been proposed that Nrd1 and Nab3 function as adaptors that provide the necessary specificity to Sen1, the precise regions involved in the interaction between these factors and whether these interactions are actually sufficient for timely Sen1 recruitment remains unclear. In this study, we identify and characterize the key interactions involved in Sen1 function. Sen1 is composed of a central helicase domain (aa 1095-1876) that is sufficient for transcription termination *in vitro* [24,25], together with a large N-terminal domain (aa 1-975) and a C-terminal intrinsically disordered region (1930-2231). Here we show that the C-terminal end of Sen1 contains a short motif that mimics the S5P-CTD and is recognized by Nrd1 CID. We prove that this motif is the main determinant of the interaction between Sen1 and Nrd1-Nab3 heterodimer and provide the structural details of this interaction. Strikingly, we find that the CTD-mimic in Sen1 is not a strict requirement for Sen1 function, although it contributes to

fully efficient termination at some targets. Instead, we show that the N-terminal domain of Sen1 promotes its recruitment to the S5P-CTD of RNAPII and is a global requirement for non-coding transcription termination. Our results unveil an unique arrangement for Sen1 protein where both a CTD-interacting domain and a CTD-mimic coexist and cooperate to ensure efficient termination. These findings provide important insights into the molecular mechanisms underlying the specific function of Sen1 on ncRNAs but also, more generally, might help understanding how the different transcription-coupled processes can be regulated by protein interactions.

RESULTS

Sen1 possesses a CTD-mimic that is recognized by the CID domain of Nrd1

In a previous report [19], we showed that Nrd1 CID domain can recognize a short sequence in Trf4 that mimic the S5P-CTD of RNAPII and that we dubbed NIM for Nrd1-Interaction Motif. A subsequent report described a second NIM in Mpp6, an exosome cofactor [22]. During the course of our previous work, we discovered that the CID is also required for the interaction between Nrd1 and Sen1 (figure 1a), an observation that was reported in an independent study [31]. This prompted us to search for a putative NIM in Sen1 protein. The S5P-CTD and Trf4 NIM share three important features: i) they contain one or several negatively charged amino acids (aa) at the N-terminal portion that interact with a positively charged surface of the CID; ii) they contain a Y residue followed by several aa at the C-terminal part that adopt a β-turn conformation and interact with a hydrophobic pocket of the CID; and iii) they are placed in protein regions that are predicted to be intrinsically disordered and therefore are fully accessible for the interaction with the CID. We identified a sequence in Sen1 C-terminal domain that fulfils the three characteristics and closely resembles the NIM in Trf4 (figure 1b). Therefore, we tested the role of this motif by comparing the ability of wild-type (wt) or ΔNIM versions of Sen1 to interact with Nrd1 by *in vivo* coimmunoprecipitation experiments using Nrd1-TAP as the bait (figure 1c). Importantly, deletion of the putative NIM did not significantly alter the levels of Sen1 protein but dramatically reduced its interaction with Nrd1. Similar experiments using Sen1 as the bait confirmed these results and showed that deletion of the NIM also strongly affects effect on the association of Sen1 with Nab3 (figure 1d). These results indicate that Sen1 NIM is the main determinant of the interaction of Sen1 with the Nrd1-Nab3 heterodimer. They also strongly suggest that Nab3 interacts with Sen1 *via* Nrd1.

The NIM is one of the very few sequence regions of the C-terminal domain of Sen1 that are conserved in the closest *S. cerevisiae* relatives, suggesting that this mode of interaction between Sen1 and Nrd1 is conserved in these yeast species (supplementary figure 1). Conversely, in agreement with previous data showing that Nrd1 and Sen1 orthologues do not interact with each other in *S. pombe* [32,33] and with the fact that no Nrd1 homologue could be

identified in association with human Sen1[34], we did not detect any putative NIM in Sen1 orthologues from these organisms.

Structural analyses of the Nrd1 CID–Sen1 NIM interaction

To compare the interaction of the newly identified Sen1 NIM (fragment harboring amino acids 2052-2063) and the previously identified Trf4 NIM with Nrd1 CID, we performed a quantitative solution-binding assay using fluorescence anisotropy (FA) with purified recombinant Nrd1 CID and synthetic NIM peptides. We found that Nrd1 CID binds Sen1 NIM with a K_D of 1.2 ±0.02 μM, which is comparable to the dissociation constant of the Trf4 NIM–Nrd1 CID complex (K_D 0.9 ±0.02 μM), and that the binding is similarly affected by previously described mutations in Nrd1 CID (L20D, K21D, S25D, R28D, M126A, I130K, R133A, see table S1). The similar binding strength was also confirmed by [^{1}H,^{15}N] heteronuclear single quantum coherence (HSQC) titration experiment of Nrd1 CID. In this experiment, the protein amide resonances of Nrd1 CID were in slow or intermediate exchange regimes between their free and bound forms relative to the NMR timescale, when titrated with Sen1 NIM, as previously shown for the Trf4-NIM interaction [19]. The analysis of changes in chemical shift perturbations (CSP) of Nrd1 CID in the presence of Sen1 NIM and Trf4 NIM peptides suggests that this domain employs the same interaction surface for both NIMs, with only a minor difference in the area of α4-helix and tip of α7-helix (supplementary figure 2).

In order to understand structural basis of the Nrd1–Sen1 interaction, we solved the solution structure of Nrd1 CID (1-153 aa) in complex with Sen1 NIM (figure 2; table S2). The structure of Nrd1 CID consists of eight α helices in a right-handed superhelical arrangement as previously reported [13,19,35]. The Sen1 NIM peptide is accommodated in the binding pocket of Nrd1 CID in a similar manner to that of Trf4 NIM. The upstream negatively charged part of the Sen1 NIM (D2052-D2053-D2054-E2055-D2056-D2057) interacts with a positively charged region of Nrd1 CID on tips of α1-α2 helices (figure 2b). Charge-swapping mutations of in this region (K21D, S25D and S28D) resulted in a significant decrease in the binding affinity (table S1), confirming the importance of this region for the interaction. Furthermore, the L20D mutant also diminishes the binding affinity as it perturbs the overall geometry of the α1-α2 loop and thus the positioning of the positively charged residues. The downstream hydrophobic part of Sen1 NIM (Y2058, T2059 and P2060) docks into a hydrophobic pocket of the CID formed by L127, M126 and I130. Y2058 makes a putative H-bond with D70 and R74 of the CID. Sen1 I2062 shows multiple intermolecular contacts in NMR spectra with the aliphatic groups of I130, R133 and S134 side chains. Furthermore, the neighboring residue S2061 makes a putative H-bond with R74. As a result, these interactions induce the extended conformation of the downstream region of the Sen1 NIM, which contrasts with the formation of the canonical β-turn that was observed in the structures of CTD and NIM peptides bound to CIDs [10,19,35–37]. In

these complexes, the peptides contain S/NPXX motifs that have a high propensity to form β-turns in which the peptides are locked upon binding to CIDs. In the case of Sen1 NIM, TPSI sequence is predicted as a non-β-turn motif[38]. Indeed, we found that in the extended conformation the positioning of this downstream motif inside the hydrophobic area is energetically the most favorable. Our structural data suggests that Nrd1 CID is able to accommodate not only peptides with motifs that form β-turns but also peptides in the extended conformation that matches hydrophobicity and H-bonding partners inside the binding groove of the CID.

A strong interaction between Sen1 and Nrd1-Nab3 is not essential for non-coding transcription termination

Because of the importance of the NIM for the interaction between Sen1 and Nrd1-Nab3 heterodimer and its significant conservation among yeast species, we analysed the impact of the NIM deletion on growth and Sen1-mediated transcription termination *in vivo*. Surprisingly, we found that deletion of the NIM alone does not affect growth at any of the temperatures tested. Deletion of the NIM only aggravated the thermosensitive phenotype of a *Δrrp6* mutant, which lacks an exonuclease that plays a major role in degradation of ncRNAs targeted by the NNS-complex, when that mutation was combined with the NIM deletion (figure 3a).

In order to test for the role of Sen1 NIM in non-coding transcription termination, we performed RNAseq transcriptome analyses of *Δrrp6* strains expressing either the wt or the *ΔNIM version of* Sen1. As expected, metagene analyses did not reveal any significant effect of the NIM deletion at protein-coding genes (figure 3b). Surprisingly, we did not observe any difference in the expression profile of sn- and snoRNAs in *sen1ΔNIM,* indicating that deletion of the NIM does not have a general impact on transcription termination at these NNS-targets (figure 3c). We only detected small increase in the RNAseq signal downstream of mature *SNR82* in the Sen1 mutant, consistent with slight termination defects at this locus (figure 3e and supplementary figure 3a). Metagene analysis of CUTs, unveiled a moderate increase in the RNA signal at a region covering the annotated termination site in *sen1ΔNIM* (figure 3d). Because transcription events that are not terminated by the NNS pathway are typically targeted by the CPF-pathway that generates stable transcripts, both an increase in the signal and a downstream shift in the RNA profile are expected outcomes of NNS dysfunctions [6]. However, inspection of individual cases indicated that termination defects were only detectable at a minority of CUTs (see examples in figure 3f-i and supplementary figure 3a). Our results indicate that, although the interaction between Sen1 and Nrd1-Nab3 via the NIM is important for fully efficient termination at a subset of non-coding genes, it is not a global requirement for NNS-dependent termination.

The integrity of the N-terminal domain of Sen1 is essential for growth and for transcription termination

Our observation that the interaction between Sen1 and Nrd1 and Nab3 is not essential for transcription termination is surprising since Sen1 is a low-abundance protein that binds RNA without any sequence specificity, and therefore it would unlikely be adequately recruited to its targets solely by virtue of its RNA-binding activity. This implicates that additional protein-protein interactions should ensure timely recruitment of Sen1 to the elongation complex. A good candidate to mediate such interactions is the large Sen1 N-terminal domain (aa 1-975), which has been proposed to interact with RNAPII [39,40]. In order to explore the possible role of the N-terminal domain (Nter) in Sen1 recruitment we constructed mutants that carry the deletion of this domain (*sen1ΔNter*) alone or in combination with the NIM deletion (*sen1ΔNterΔNIM).* To our surprise, contrary to reports from two other groups that only observed slow growth [40,41] we found that deletion of Sen1 N-terminal domain was lethal, both when the mutant gene was expressed from a centromeric plasmid (figure 4a) or from the endogenous locus (supplementary figure 3b). In order to assess the role of the N-terminal domain of Sen1 in NNS-dependent transcription termination, we constructed a Sen1 auxin-inducible degron (AID[42]) strain that allowed us to rapidly target Sen1 for degradation by the proteasome upon addition of auxin (supplementary figure 3c). We supplemented the Sen1-AID strain with an empty vector, as a control, or a centromeric plasmid expressing either the wt, *ΔNIM, ΔNter* or the *ΔNterΔNIM* version of *SEN1* and we analysed by northern blot the expression of several typical NNS-targets upon depletion of the chromosomally-encoded copy of *SEN1* (figure 4b). Strikingly, transcription termination was dramatically impaired in the strain expressing *sen1ΔNter*. In addition, these termination defects were exacerbated in the *sen1ΔNterΔNIM* double mutant. These phenotypes were not due to lower levels of the mutant proteins deleted in the N-terminal domain, since these proteins were expressed at similar levels compared to full-length Sen1 (supplementary figure 3d). Taken together, our results indicate that the N-terminal domain of Sen1 plays a critical role in transcription termination and that the NIM is also required for full efficiency.

Overlapping functions for the N-terminal domain and the NIM of Sen1

The requirement of Sen1 Nter for transcription termination could be explained either by a role of this domain in activating Sen1 catalytic activity or by a function in mediating its recruitment to the elongation complex. Because we have previously shown that deletion of the Nter does not reduce the efficiency of transcription termination by Sen1 *in vitro* nor does it affect any of

Sen1 measurable catalytic activities[24], we favour the second possibility. In agreement with this hypothesis, we found that overexpression of *sen1ΔNter* from the *GAL1* promoter (pGAL) restored cell growth (figure 4c) and suppressed to a large extent the termination defects associated with deletion of Sen1 Nter (figure 4d). We took advantage of this genetic context in which the different mutants are viable and their termination phenotypes are less extreme to explore at the genome-wide level the function of the Nter and the NIM in termination at the different NNS targets. To this end we performed RNA-seq transcriptome analyses on two independent biological replicates of *Δrrp6* strains overexpressing either the wt, the *ΔNter* or the *ΔNterΔNIM* version of Sen1. As expected we did not detect any effect of these mutations on transcription termination at protein-coding genes (supplementary figure 4a). Metagene analyses of sn- and snoRNAs unveiled a substantial increase of the RNA signal downstream of the 3'-end of the mature form in the *ΔNter* that is more pronounced in the double *ΔNterΔNIM* mutant, suggesting global transcription termination defects in those mutants (figure 5a-d). We also observed a downstream shift in the RNA profile of the mutants relative to the wt at CUTs, which tended to be more pronounced in the double *ΔNterΔNIM* than in the *ΔNter* mutant (figure 5e). Inspection of individual cases confirmed the transcription termination defects suggested by the metagene analyses (figure 5f-h). However, the effect of these Sen1 mutations was less prominent at CUTs than at sn- and snoRNAs. These results are consistent with our observations on isolated cases with the mutant proteins expressed under their own promoter (figure 4b).

Clustering of sn- and snoRNAs according to the increase in the RNA signal downstream of the 3'-end actually distinguished two groups of genes (supplementary figure 4b). Cluster 1 is composed of sn- and snoRNAs with major accumulation of read-through transcripts, whereas cluster 2 includes genes with very moderate or no accumulation (see supplementary table S3). Clusters 1 obtained from the analysis of *ΔNterΔNIM* and *ΔNter* mutants largely overlapped, with 29 sn- and snoRNAs being present in both and 8 only present in the *ΔNterΔNIM* cluster 1 (figure 5l). A more detailed examination of these 8 cases revealed that they actually exhibited termination defects also in the *ΔNter* mutant, but to a lesser extent (supplementary table S3). Cluster 2 contained mainly intronic snoRNAs, multicistronic snoRNAs, among which only the last one in the multicistron could harbour a terminator at its 3'-end, and sn- and snoRNAs that were previously shown to be poorly sensitive to Nrd1 and Nab3 mutations, so they are unlikely to be *bona fide* targets of the NNS termination pathway [43]. Therefore, the integrity of the Nter is a general requirement for termination by the NNS-complex at snoRNAs.

Because decreased NNS-dependent transcription termination efficiency leads to deregulation of protein-coding genes downstream of and antisense to CUTs and snoRNAs [6], we compared the sets of protein-coding genes whose expression was significantly altered (p-value < 0.05)

in either mutant. We identified 209 and 544 genes with a fold change > 1.5 in the *ΔNter* and the *ΔNterΔNIM* mutant, respectively (supplementary tables S4-7), among which many well-characterized examples of genes regulated by NNS-dependent termination (e.g. *URA2*, *SER3* and *IMD2*, for review see Colin et al, 2011[44]). Whereas we observed substantial overlap between the sets of genes that are upregulated in each mutant, only a few genes were downregulated in both mutants (figure 5j). Assessment of individual cases suggested that the poor overlap between the sets of genes downregulated in each mutant could be mostly due to the differences in termination efficiency. While mild termination defects at a non-coding gene in *sen1ΔNter* would induce downregulation of a downstream gene by transcriptional interference, the more severe defects in the *sen1ΔNterΔNIM* mutant could lead to high levels of a chimeric RNA (covering the non-coding gene and the downstream protein-coding gene) that would be interpreted in our differential expression analysis as a case of no change in expression or even upregulation (supplementary figure 4c-e).

Taken together, these results indicate that deletion of the NIM enhances the termination defects associated with deletion of the Nter at most NNS-targets, suggesting that both protein regions have largely overlapping functions in mediating the recruitment of Sen1. In addition, the interactions mediated by the NIM might be particularly important for the efficiency of termination at a subset of non-coding genes.

The N-terminal domain of Sen1 mediates its recruitment to the S5P-CTD form of RNAPII

As evoked above, Sen1 Nter has been proposed to interact with RNAPII, more specifically with the S2P form of the CTD[39], which could be the main way to recruit Sen1 to elongation complexes. We set out to further test this possibility by performing *in vivo* coimmunoprecipitation experiments with *sen1ΔNter* either expressed from its own promoter in a centromeric plasmid using the Sen1-AID system described above or overexpressed from pGAL (figure 6a and b). Strikingly, we did not detect any significant decrease in the capacity of Sen1 to interact with total RNAPII (shown by Rpb1 and/or Rpb3 subunits) upon deletion of the Nter, regardless the expression levels of the mutant protein. In both kind of experiments, protein extracts were treated with RNase to detect only non-RNA-mediated protein interactions. This result strongly suggests that regions other than the Nter can mediate the interaction between Sen1 and RNAPII. We considered the possibility that Sen1 Nter would be important for the recognition of a specific phosphorylated form of RNAPII. Indeed, analyses of Sen1 coimmunoprecipitates with antibodies against different CTD phospho-marks revealed a substantial decrease in the association of S5P-CTD RNAPII with Sen1ΔNter, whereas the levels of S2P-CTD RNAPII remained unchanged (figure 6a). This implies that the distribution

of Sen1 among the different forms of RNAPII is altered in the absence of the Nter. This result is in contrast with previous two-hybrid assays suggesting that the Sen1 Nter would interact with the S2P form of the CTD [39](see discussion for possible explanations). In order to further substantiate the notion that Sen1 Nter is critical for the recognition of S5P-CTD, we tested whether replacement of this protein region by another domain that can also interact with S5P-CTD could suppress the lethality and termination defects associated with deletion of Sen1 Nter. To this end we constructed a chimera in which the Nrd1 CID (amino acids 1-150), which recognizes preferentially the S5P form of RNAPII CTD[35], was fused to Sen1 amino acids 976-2231 (i.e. equivalent to Sen1ΔNter). Indeed, we found that the strain producing the chimeric Nrd1 CID- Sen1ΔNter protein become viable (figure 6c) and northern blot analysis of a typical NNS-targets showed that the efficiency of transcription termination was also restored to a large extent in this strain (figure 6d). Strikingly, the presence of Nrd1 CID at the place of Sen1 Nter rescued the growth of a double mutant *sen1ΔNterΔNIM* to a lesser extent and did not suppress the transcription termination defects (figure 6c and d). This indicates that in this particular genetic context (i.e. with Nrd1 CID at the place of the Nter in Sen1) the NIM becomes particularly relevant for Sen1 function. Taken together, these results strongly support the idea that the essential role of Sen1 Nter is to ensure early recruitment of Sen1 by recognizing the S5P-CTD.

DISCUSSION

In budding yeast, the NNS-complex emerges as a safeguard for gene expression. On one hand, it is required for transcription termination and maturation of sn- and snoRNAs, which in turn have important functions in splicing and rRNA modification. On the other hand, via its transcription termination activity coupled to RNA degradation, it prevents massive genomic deregulation that results from uncontrolled pervasive transcription [6]. Nevertheless, the NNS-complex needs to be tightly regulated in order to restrict its activity to the right targets and avoid premature transcription termination at protein-coding genes and/or degradation of mRNAs. The final step of transcription termination (i.e. dissociation of the elongation complex) exclusively depends on the helicase Sen1, which cannot discriminate non-coding from protein-coding RNAs on its own. The fact that Sen1 forms a complex with Nrd1 and Nab3, together with the capacity of Nrd1 and Nab3 to recognize motifs that are enriched in the target ncRNAs, led several authors, including us, to propose that Nrd1 and Nab3 would play a critical role in the recruitment of Sen1, therefore conferring the necessary specificity to Sen1 activity.

In the present study, we characterize molecularly and functionally the interaction between Sen1 and its partners Nrd1 and Nab3 and provide data that challenge the former model. Furthermore, we obtain evidence that other protein-protein interactions that do not concern Nrd1 and Nab3 play a more prominent role in Sen1 recruitment. Our results allow redefining the rules that govern the specific function of the NNS-complex in non-coding transcription termination.

Molecular mimicry to coordinate transcriptional and post-transcriptional processes

As mentioned above, the CTD of RNAPII is considered a master regulator of the transcription cycle as well as of co-transcriptional processes as mRNA capping and splicing. This is due to the capacity of the CTD to be dynamically modified by kinases and phosphatases resulting in complex phosphorylation patterns that are differentially recognized by a plethora of factors with key roles in the aforementioned processes[9,45]. One of these factors is Nrd1, that preferentially binds the S5P-CTD, which contributes to its recruitment to the transcribing RNAPII during early elongation [35,46]. Surprisingly, in a previous study, we have discovered that the same domain of Nrd1 recognizes a sequence that mimics the S5P-CTD in the non-canonical poly(A)-polymerase Trf4 . In a subsequent report, a second CTD-mimic was identified in another cofactor of the exosome, Mpp6[22]. Here we reveal the presence of a third functional CTD-mimic, designated NIM for Nrd1-Interaction Motif, in the essential helicase Sen1. This mechanism of mediating mutually-exclusive interactions with multiple partners that should act sequentially in the same RNA molecule seems an efficient manner to temporally coordinate the different steps of the NNS pathway: recruitment of Nrd1 to RNAPII, then recruitment/stabilization of Sen1 to

facilitate dissociation of the transcription elongation complex and finally polyadenylation of the released RNAs by Trf4 and subsequent degradation/processing by the exosome.

Our previous structural analyses have shown that the NIM in Trf4 share with the S5P-CTD two major structural elements: one or several negatively charged residues at the N-terminal part and a β-turn at the C-terminal flanked by a conserved tyrosine. Strikingly, in the present structure of Nrd1 CID in complex with Sen1 NIM we have not observed the characteristic β-turn, yet several important H-bonds and hydrophobic interactions between the CID and Sen1 NIM are maintained and the affinity of the CID for the NIMs of Trf4 and Sen1 is almost identical. This is due to an alternative conformation of the C-terminal region of Sen1 NIM that can be accommodated in the binding pocket of Nrd1 CID. We suggest that similar extended conformation exists also in the case of the CTD mimic found in Mpp6[22] as its DLDK C-terminal motif (figure 2e) has no propensity to form a β-turn[38].

This relaxed sequence requirement implies that a large number of protein regions could potentially behave as *bona fide* CTD-mimics mediating the interaction with Nrd1 or with other factors with structurally related CTD-Interaction domains. We anticipate that other proteins that function both in association with the CTD of RNAPII and in a separate complex might employ a similar mechanism to either coordinate transcriptional and post-transcriptional steps of the same pathway, as in the case of Nrd1, or to mediate independent functions of the same protein (e.g. docking of a CTD kinase to a different substrate).

Redundant ways to recruit Sen1 to transcribing RNAPII

Despite the major role of the NIM in mediating the interaction of Sen1 with its partners Nrd1 and Nab3, we have shown here that this motif plays only an accessory role in transcription termination by Sen1, since only few NNS-targets exhibited detectable termination defects upon deletion of the NIM. Therefore, the recruitment of Sen1 to the transcription elongation complex does not seem to depend on the interaction with Nrd1 and Nab3. We cannot exclude that additional weak interactions between regions of the NNS-complex other than the CID and the NIM also contribute to Sen1 function, but overall it is unlikely that they play a critical role in Sen1 recruitment. In contrast, deletion of the Nter of Sen1 is lethal and results in massive transcription termination defects. The fact that overexpression of *sen1ΔNter* suppresses lethality and partially restores the termination efficiency strongly suggests that the Nter plays a critical role in the recruitment of Sen1. Indeed, a higher concentration of the mutant protein might increase its association with the elongation complex either via unspecific interactions between the helicase domain and the nascent RNA or via protein-protein interactions with the RNAPII itself or associated factors. Unfortunately, a more direct assessment of the role of the Nter in Sen1 recruitment is hampered by the fact that upon crosslinking (an obligate step in

chromatin immunoprecipitation experiments), the efficiency of immunoprecipitation of Sen1ΔNter is several folds lower than that of full-length Sen1 (data not shown). Our data suggest that the reason of the weak phenotype associated with deletion of the NIM is the fact that the interactions mediated by the NIM and by the Nter of Sen1 are at least partially redundant. This idea is supported by the fact that deletion of the NIM generally aggravates the transcription termination defects observed in a *sen1ΔNter* mutant. It is therefore possible that under certain conditions the protein interactions mediated by the Nter are altered and the binding to Nrd1 and Nab3 become more relevant for Sen1 recruitment.

The essential character of the Nter of Sen1 was overlooked in former studies[40,41], most likely because the *sen1ΔNter* gene was not expressed under its own promoter at the endogenous locus, and therefore possibly higher levels of the truncated protein produced in those genetic contexts would be sufficient for viability. In addition, the role of this domain in transcription termination has not systematically been analysed before. Here we present genome-wide data showing that the Nter is a global requirement for non-coding transcription termination. The results of our biochemical and genetic analyses strongly suggest that the essential function of the Nter is to promote the interaction of Sen1 with the S5P-CTD. A previous work reported that Sen1 preferentially binds the S2P form of the CTD in two-hybrid assays [39]. It is possible that in that artificial context, because of a different conformation of the CTD separated from the rest of RNAPII and/or a different repertoire of factors bound to the CTD, Sen1 exhibits a more efficient interaction with S2P-CTD. On the other hand, a former study has shown a strong correlation between the levels of CTD phosphorylation at S5 and the efficiency of NNS-dependent termination[18]. This dependency could not be explained solely by the capacity of Nrd1 to interact with the S5P-CTD, since the deletion of Nrd1 CID provoked very moderate transcription termination defects[19,35]. Our present data indicating that Sen1 can recognize the same form of the CTD on its own allows better understanding the former pieces of evidence.

Taking into account our past and present results, we propose a model according to which both Nrd1-Nab3 and Sen1 would be independently recruited to the S5P-CTD (figure 7). Once Nrd1-Nab3 bind their cognate recognition sequences at the nascent RNA, they would dissociate from the CTD and interact with Sen1 via the NIM, therefore stabilizing Sen1 on the elongation complex or eventually bringing it in close proximity to the nascent RNA. Because the Nter of Sen1 can recognize the S5P-CTD and the NIM mimics this motif, it is very tempting to speculate that these two regions of Sen1 could mediate intra-molecular interactions that might modulate the dynamics of association and dissociation of Sen1 with the nascent RNA. For instance, these interactions might promote the release of Sen1 from RNAPII and Nrd1, facilitating Sen1 translocation along the RNA to induce termination.

This configuration of Sen1, with both a CTD-reader and a CTD-mimic facilitating the function of its catalytic domain, is peculiar but likely not restricted to budding yeast. Indeed, the Nter is

relatively well-conserved from yeast to human and we have shown that the NIM is also present in other yeast species (supplementary figure 1).

The role of Nrd1 and Nab3 in non-coding transcription termination

We have previously shown that Nrd1 and Nab3 are irrelevant for transcription termination by Sen1 *in vitro*[20], indicating that they are not necessary for Sen1 catalytic activity. However, Nrd1 and Nab3 are essential and their interaction with specific sequences in the target ncRNAs is critical for termination *in vivo* [47–51]. This led us to propose that Nrd1 and Nab3 are required for the recruitment of Sen1 *in vivo*. The data presented in this study are in seemingly contrast with this idea. On the other hand, we have observed that, whereas overexpression of Sen1 bypasses the requirement of the N-terminal domain, that seems critical for recruitment, in the context of Sen1 overexpression, Nrd1 and Nab3 are still essential (see supplementary figure 5). This genetic evidence would support the idea that these proteins play a role other than Sen1 recruitment in non-coding transcription termination. It has previously been proposed that Nrd1 and Nab3 would induce RNAPII pausing[52]. In a former study, we have provided *in vitro* evidence that Sen1 is a poorly-processive translocase and that transcription termination by Sen1 strictly requires RNAPII to be paused[24]. Therefore, a model in which Nrd1 and Nab3 would promote termination by Sen1 by modifying RNAPII elongation properties would be consistent with present and previous data. How the interaction of Nrd1 and Nab3 with the nascent RNAs and/or with RNAPII-associated factors would induce polymerase pausing is an interesting subject for future studies.

METHODS

Construction of yeast strains and plasmids

Yeast strains used in this paper are listed in table S9. Gene deletions, tagging and insertion of the *GAL1* promoter were performed with standard procedures [53,54] using plasmids described in table S10. Strain DLY2769 expressing untagged *sen1ΔNIM* was constructed by transforming a *Δsen1 strain* harbouring the *URA3*-containing plasmid pFL38-*SEN1* (DLY2767) with the product of cleavage of pFL39- *sen1ΔNIM* (pDL703) with restriction enzymes MluI, BstZ17I and Bsu36I. Cells capable of growing on 5-FOA were then screened by PCR for the presence of the NIM deletion and the absence of plasmids pFL38-*SEN1* and pFL39-*sen1ΔNIM.*

Plasmids expressing different *SEN1* variants were constructed by homologous recombination in yeast. Briefly, a wt yeast strain was transformed with the corresponding linearized vector and a PCR fragment harbouring the region of interest flanked by 40-45 bp sequences allowing recombination with the vector. Clones were screened by PCR and the positive ones were verified by sequencing.

Plamids for overexpression of the Nrd1 CID variants R133G and R133D were obtained using QuikChange site-directed mutagenesis kit (Stratagene).

Coimmunoprecipitation experiments

Yeast extracts were prepared by standard methods. Briefly, cell pellets were resuspended in lysis buffer (10 mM sodium phosphate pH 7, 200mM sodium acetate, 0.25% NP-40, 2mM EDTA, 1mM EGTA, 5% glycerol) containing protease inhibitors, frozen in liquid nitrogen and lysed using a Retsch MM301 Ball Mill. For TAP-tagged proteins, the protein extract was incubated with IgG Fast Flow Sepharose (GE Healthcare) and beads were washed with lysis buffer. Tagged and associated proteins were eluted either by cleaving the protein A moiety of the tag with TEV protease in cleavage buffer (10mM Tris pH 8, 150mM NaCl, 0.1 % NP-40, 0.5mM EDTA and 1mM DTT) for 2h at 21°C or by boiling the beads in 2x Laemmli buffer (100mM Tris pH 6.8, 4% SDS, 15% glycerol, 25mM EDTA, 100 mM DTT, 0.2% bromophenol blue) for 5 min. For HA-tagged proteins, protein extracts were first incubated with 25 µg of anti-HA antibody (12CA5) for 2 hours at 4°C. Then, 30 µl Protein A-coupled beads (Dynabeads, Thermo Fisher, 30mg/ml) were added to each sample and incubated for 2 hours at 4°C. After incubation, the beads were washed with lysis buffer and proteins were eluted by incubating with 2x Laemmli buffer (without DTT) for 10 minutes at 37°C.

Fluorescence anisotropy analyses

Nrd1 CID and its mutants were produced and purified as described previously[35]. The equilibrium binding of the different versions of Nrd1 CID to Trf4 NIM and Sen1 NIM was analyzed by fluorescence anisotropy. The NIM peptides were N-terminally labelled with the 5,6-carboxyfluorescein (FAM). The measurements were conducted on a FluoroLog-3 spectrofluorometer (Horiba Jobin-Yvon Edison, NJ). The instrument was equipped with a thermostatted cell holder with a Neslab RTE7 water bath (Thermo Scientific). Samples were excited with vertically polarized light at 467 nm, and both vertical and horizontal emissions were recorded at 516 nm. All measurements were conducted at 10°C in 50 mM Na_2HPO_4 100 mM NaCl pH=8. Each data point is an average of three measurements. The experimental binding isotherms were analyzed by DynaFit using 1:1 model with non-specific binding [55].

NMR analyses

All NMR spectra were recorded on Bruker AVANCE III HD 950, 850, 700, and 600 MHz spectrometers equipped with cryoprobes at a sample temperature of 20°C using 1 mM uniformly ^{15}N,^{13}C-labelled Nrd1-CID in 50 mM Na_2HPO_4 100 mM NaCl pH=8 (20°C) (90% H2O/10% D2O). The initial nuclei assignment was transferred from BMRB entry 19954 and confirmed by HNCA, HNCACB, HCCCONH, HBHACONH and 4D-HCCH TOCSY spectra. The spectra were processed using TOPSPIN 3.2 (Bruker Biospin) and the protein resonances were assigned manually using Sparky software (Goddard T.G. and Kellner D.G., University of California, San Francisco). 4D version of HCCH TOCSY[56] was measured with a non-uniform sampling; acquired data were processed and analysed analogously as described previously[57,58]. All distance constraints were derived from the three-dimensional 15N- and 13C-edited NOESYs collected on a 950 MHz spectrometer. Additionally, intermolecular distance constraints were obtained from the three-dimensional F1-13C/15N-filtered NOESY-[13C,1H]-HSQC experiment [59,60] with a mixing time of 150 ms on a 950 MHz spectrometer. The NOEs were semi-quantitatively classified based on their intensities in the 3D NOESY spectra. The initial structure determinations of the Nrd1 CID–Sen1 NIM complex were performed with the automated NOE assignment module implemented in the CYANA 3.97 program [61]. Then, the CYANA-generated restraints along with manually assigned Nrd1 CID–Sen1 NIM intermolecular restraints were used for further refinement of the preliminary structures with AMBER16 software [62]. These calculations employed a modified version (AMBER ff14SB) of the force field [63], using a protocol described previously[14,64]. The 20 lowest-energy conformers were selected (out of 50 calculated) to form the final ensemble of structures.

Northern blot assays

Unless otherwise indicated, cells used for northern blot assays were grown on YPD medium at 30°C to OD_{600} 0.3 to 0.6 and harvested by centrifugation. RNAs were prepared using standard methods. Samples were separated by electrophoresis on 1.2% agarose gels, and then transferred to nitrocellulose membranes and UV-crosslinked. Radiolabeled probes were prepared by random priming of PCR products covering the regions of interest with Megaprime kit (GE Healthcare) in the presence of α-^{32}P dCTP (3000 Ci/mmol). Oligonucleotides used to generate the PCR probes are listed in table S11. Hybridizations were performed using a commercial buffer (Ultrahyb, Ambion) and after washes, membranes were analysed by phosphorimaging.

RNA-seq library preparation and deep-sequencing

RNA samples were treated with RiboZero to deplete rRNA and RNA libraries were prepared by the Imagif sequencing platform using a NextSeq 500/550 High Output Kit v2. Samples were sequenced on a NextSeq 500 sequencer. Reads were demultiplexed with bcl2fastq2-2.18.12; adaptor trimming of standard Illumina adaptors was performed with cutadapt 1.15. and reads were subsequently quality-trimmed with trimmomatic[65] and mapped to the R64 genome[66] with bowtie2 using the default options [67].

Annotations

For protein coding genes, sn- and snoRNAs and CUTs we used the annotations in Xu et al, 2008[68]. We noticed that a relatively high number the annotated CUTs were located immediately downstream of the mature form of sn- and snoRNAs and therefore they correspond most likely to the precursor of sn- and snoRNAs and not to *bonafide* CUTs. In order to have a more reliable annotation of CUTs we excluded those falling within a window of 400 bp downstream of sn- ad snoRNAs. To facilitate metagene analyses, we also excluded CUTs located just upstream of highly expressed genes. The final subset of CUTs included 442 out of the 925 originally annotated ones.

Bioinformatic analyses

Bioinformatic analyses of RNAseq data were performed using the Galaxy framework (http://galaxy.sb-roscoff.fr and http://deeptools.ie-freiburg.mpg.de). In order to represent the average distribution of reads over a defined region (-/+ 500 bp) relative to the transcription termination site (i.e. metagene analyses), we used a set of tools from the deepTools2 package[69]**.** Reads were normalized to the library size and strand-specific coverage bigwig files

were obtained with bamCoverage using a bin size of 1 and default parameters. The coverage files for each strand were used as inputs for the computeMatrix tool together with separate annotations for each strand and for each feature (e.g. protein-coding gene, CUTs, etc), using a bin size of 10 and the transcription termination site as the reference point. Matrices constructed that way for each strand were subsequently combined using the rbind option of the computeMatrixOperations tool and used as the input for the plotProfile tool. We represented the median instead of the mean values to minimize the bias towards the very highly expressed features.

For heatmaps and clustering we computed the log2 of the signal ratio (mutant/wt) for each strand using the aforementioned normalized bigwig files as the input and the bigwigCompare tool from the deepTools2 package with the following parameters: bin size of 10 and pseudocount of 1 for non-covered regions. With the output files, we run the computeMatrix tool using the sn- and snoRNAs annotations, a bin size of 10 and scaling the body of the genes to 0.2 Kb. Matrices for each strand were subsequently combined using the rbind option of the computeMatrixOperations tool and used for the plotHeatmap tool, selecting the mean as the statistics to represent. We performed clustering using the K-means clustering option of plotHeatmap and setting the number of clusters to 2 (a larger number of clusters did not allow distinguishing any group with a different profile).

Differential expression analyses were performed by counting the reads from two independent biological replicates that mapped at each feature (sn- and snoRNAs or protein-coding genes) using the htseq-count tool[70] with the union mode. The count files thus generated were used as input files for the SARTools DESeq2 package[71]. We set default parameters but we added a blocking factor to correct the batch effect due to the library preparation and sequencing of the two replicates on different days.

ACCESSION NUMBERS

The atomic coordinates for the NMR ensemble of the Nrd1CID–Sen1 NIM complex have been deposited in the Protein Data Bank under ID code 6GC3. RNAseq data have been deposited in GEO with accession number GSE117604.

ACKNOWLEDGEMENTS

We thank M. J. Martin-Niclos for technical assistance and other members of D. L. lab for fruitful discussions. We thank the Roscoff Bioinformatics platform ABiMS (http://abims.sb-roscoff.fr) for providing computational resources and support. This work has benefited from the facilities and expertise of the high throughput sequencing core facility of I2BC (http://www.i2bc.paris-saclay.fr/).

Z.H. was supported by PhD fellowships from the China Scholarship Council and La Ligue contre le Cancer and by a post-doctoral fellowship from the labex “Who am I?”. D.L. was supported by the CNRS, the Agence National pour la Recherche (ANR-12-BSV8-0014-01) and by the labex “Who am I?” (Idex ANR-11-IDEX-0005-02 and ANR-11-LABX-0071). O.P ; was supported by the CNRS and the Agence National pour la Recherche (ANR-16-CE12-0001-01). O.J. was supported by 2017 FEBS Long-Term Fellowship. K.K. was supported by the Czech Science Foundation (P305/12/G034).R.S. was supported by the Czech Science Foundation (15-17670S) and Ministry of Education, Youths and Sports of the Czech Republic (CEITEC 2020 project (LQ1601)). This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 649030 to R.S.). This publication reflects only the author’s view and the Research Executive Agency is not responsible for any use that may be made of the information it contains.

AUTHORS CONTRIBUTIONS

Z.H. performed biochemical and genetic experiments and prepared RNA samples for deep sequencing. A.T. performed biochemical experiments. O.J. designed experiments, prepared protein samples, performed FA measurements and analysis, analyzed NMR spectra, performed structure calculation; K.K. collected and processed NMR spectra, assisted with structure calculation; R.S. designed experiments and assisted with structure calculation. D.L. advised on experimental design and contributed to manuscript writing. O.P. conceived the project and performed biochemical experiments and bioinformatic analyses. O.P. wrote the manuscript with input from all authors.

REFERENCES

1. Jensen, T. H., Jacquier, A. & Libri, D. Dealing with pervasive transcription. *Mol. Cell* **52,** 473–484 (2013).

2. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16,** 190–202 (2015).

3. Wyers, F. *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121,** 725–737 (2005).

4. Arigo, J. T., Eyler, D. E., Carroll, K. L. & Corden, J. L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol Cell* **23,** 841–51 (2006).

5. Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J. & Libri, D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol Cell* **23,** 853–64 (2006).

6. Schulz, D. *et al.* Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* **155,** 1075–1087 (2013).

7. LaCava, J. *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121,** 713–724 (2005).

8. Vanacova, S. *et al.* A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol* **3,** e189 (2005).

9. Harlen, K. M. & Churchman, L. S. The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat. Rev. Mol. Cell Biol.* **18,** 263–273 (2017).

10. Lunde, B. M. *et al.* Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **17,** 1195–1201 (2010).

11. Xiang, K., Tong, L. & Manley, J. L. Delineating the Structural Blueprint of the Pre-mRNA 3′-End Processing Machinery. *Mol Cell Biol* **34,** 1894–1910 (2014).

12. Kubicek, K. *et al.* Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* **26,** 1891–1896 (2012).

13. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhart, A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat Struct Mol Biol* **15,** 795–804 (2008).

14. Hobor, F. *et al.* Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J. Biol. Chem.* **286,** 3645–3657 (2011).

15. Porrua, O. *et al.* In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **31,** 3935–3948 (2012).

16. Lunde, B. M., Hörner, M. & Meinhart, A. Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic Acids Res.* **39,** 337–346 (2011).

17. Wlotzka, W., Kudla, G., Granneman, S. & Tollervey, D. The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *EMBO J.* **30,** 1790–1803 (2011).

18. Gudipati, R. K., Villa, T., Boulay, J. & Libri, D. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nat. Struct. Mol. Biol.* **15,** 786–794 (2008).

19. Tudek, A. *et al.* Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Mol. Cell* **55,** 467–481 (2014).

20. Porrua, O. & Libri, D. A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nat. Struct. Mol. Biol.* **20,** 884–891 (2013).

21. Fasken, M. B., Laribee, R. N. & Corbett, A. H. Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1. *PLoS Genet.* **11,** e1005044 (2015).

22. Kim, K., Heo, D.-H., Kim, I., Suh, J.-Y. & Kim, M. Exosome Cofactors Connect Transcription Termination to RNA Processing by Guiding Terminated Transcripts to the Appropriate Exonuclease within the Nuclear Exosome. *J. Biol. Chem.* **291,** 13229–13242 (2016).

23. Martin-Tumasz, S. & Brow, D. A. Saccharomyces cerevisiae Sen1 Helicase Domain Exhibits 5'- to 3'-Helicase Activity with a Preference for Translocation on DNA Rather than RNA. *J. Biol. Chem.* **290,** 22880–22889 (2015).

24. Han, Z., Libri, D. & Porrua, O. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res.* **45,** 1355–1370 (2017).

25. Leonaitė, B. *et al.* Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. *EMBO J.* **36,** 1590–1604 (2017).

26. Creamer, T. J. *et al.* Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS Genet* **7,** e1002329 (2011).

27. Chong, Y. T. *et al.* Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell* **161,** 1413–1424 (2015).

28. Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441,** 840–846 (2006).

29. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11,** 319–324 (2014).

30. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425,** 737–741 (2003).

31. Heo, D. *et al.* The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. *J. Biol. Chem.* **288,** 36676–36690 (2013).

32. Wittmann, S. *et al.* The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. *Nat Commun* **8,** 14861 (2017).

33. Lemay, J.-F. *et al.* The Nrd1-like protein Seb1 coordinates cotranscriptional 3' end processing and polyadenylation site selection. *Genes Dev.* **30,** 1558–1572 (2016).

34. Yüce, Ö. & West, S. C. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol. Cell. Biol.* **33,** 406–417 (2013).

35. Kubicek, K. *et al.* Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes Dev.* **26,** 1891–1896 (2012).

36. Becker, R., Loll, B. & Meinhart, A. Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* **283,** 22659–22669 (2008).

37. Meinhart, A. & Cramer, P. Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature* **430,** 223–226 (2004).

38. Singh, H., Singh, S. & Raghava, G. P. S. In silico platform for predicting and initiating β-turns in a protein at desired locations. *Proteins* **83,** 910–921 (2015).

39. Chinchilla, K. *et al.* Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae. *Eukaryotic Cell* **11,** 417–429 (2012).

40. Ursic, D., Chinchilla, K., Finkel, J. S. & Culbertson, M. R. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic Acids Res.* **32,** 2441–2452 (2004).

41. Chen, X., Müller, U., Sundling, K. E. & Brow, D. A. Saccharomyces cerevisiae Sen1 as a model for the study of mutations in human Senataxin that elicit cerebellar ataxia. *Genetics* **198,** 577–590 (2014).

42. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6,** 917–922 (2009).

43. Chen, X. *et al.* Transcriptomes of six mutants in the Sen1 pathway reveal combinatorial control of transcription termination across the Saccharomyces cerevisiae genome. *PLoS Genet.* **13,** e1006863 (2017).

44. Colin, J., Libri, D. & Porrua, O. Cryptic transcription and early termination in the control of gene expression. *Genet Res Int* **2011,** 653494 (2011).

45. Harlen, K. M. *et al.* Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* **15,** 2147–2158 (2016).

46. Vasiljeva, L., Kim, M., Mutschler, H., Buratowski, S. & Meinhart, A. The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **15,** 795–804 (2008).

47. Steinmetz, E. J. & Brow, D. A. Control of pre-mRNA accumulation by the essential yeast protein Nrd1 requires high-affinity transcript binding and a domain implicated in RNA polymerase II association. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 6699–6704 (1998).

48. Steinmetz, E. J., Conrad, N. K., Brow, D. A. & Corden, J. L. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* **413,** 327–331 (2001).

49. Steinmetz, E. J., Ng, S. B. H., Cloute, J. P. & Brow, D. A. cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Mol. Cell. Biol.* **26,** 2688–2696 (2006).

50. Carroll, K. L., Ghirlando, R., Ames, J. M. & Corden, J. L. Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *Rna* **13,** 361–73 (2007).

51. Conrad, N. K. *et al.* A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. *Genetics* **154,** 557–571 (2000).

52. Schaughency, P., Merran, J. & Corden, J. L. Genome-wide mapping of yeast RNA polymerase II termination. *PLoS Genet.* **10,** e1004632 (2014).

53. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast* **14,** 953–961 (1998).

54. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17,** 1030–1032 (1999).

55. Kuzmic, P. DynaFit--a software package for enzymology. *Meth. Enzymol.* **467,** 247–280 (2009).

56. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R. & Forman-Kay, J. D. A gradient-enhanced HCCH TOCSY experiment for recording side-chain H-1 and C-13 correlations in H2O samples of proteins. *J. Magn. Reson. Ser.* **3,** 333–337 (1993).

57. Nováček, J., Haba, N. Y., Chill, J. H., Žídek, L. & Sklenář, V. 4D Non-uniformly sampled HCBCACON and 1J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J Biomol NMR* **53,** 139–148 (2012).

58. Nováček, J. *et al.* 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR* **50,** 1–11 (2011).

59. Zwahlen, C. *et al.* Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage λ N-Peptide/boxB RNA Complex. *J. Am. Chem. Soc.* **119,** 6711–6721 (1997).

60. Peterson, R. D., Theimer, C. A., Wu, H. & Feigon, J. New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes. *J Biomol NMR* **28,** 59–67 (2004).

61. Güntert, P. & Buchner, L. Combined automated NOE assignment and structure calculation with CYANA. *J Biomol NMR* **62,** 453–471 (2015).

62. Case, D. A. Molecular Dynamics and NMR Spin Relaxation in Proteins. *Acc. Chem. Res.* **35,** 325–331 (2002).

63. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **11,** 3696–3713 (2015).

64. Stefl, R. *et al.* The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* **143,** 225–237 (2010).

65. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

66. Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40,** D700-705 (2012).

67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

68. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457,** 1033–1037 (2009).

69. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44,** W160-165 (2016).

70. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

71. Varet, H., Brillet-Guéguen, L., Coppée, J.-Y. & Dillies, M.-A. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS ONE* **11,** e0157022 (2016).

72. Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21,** 3433–3434 (2005).

FIGURES

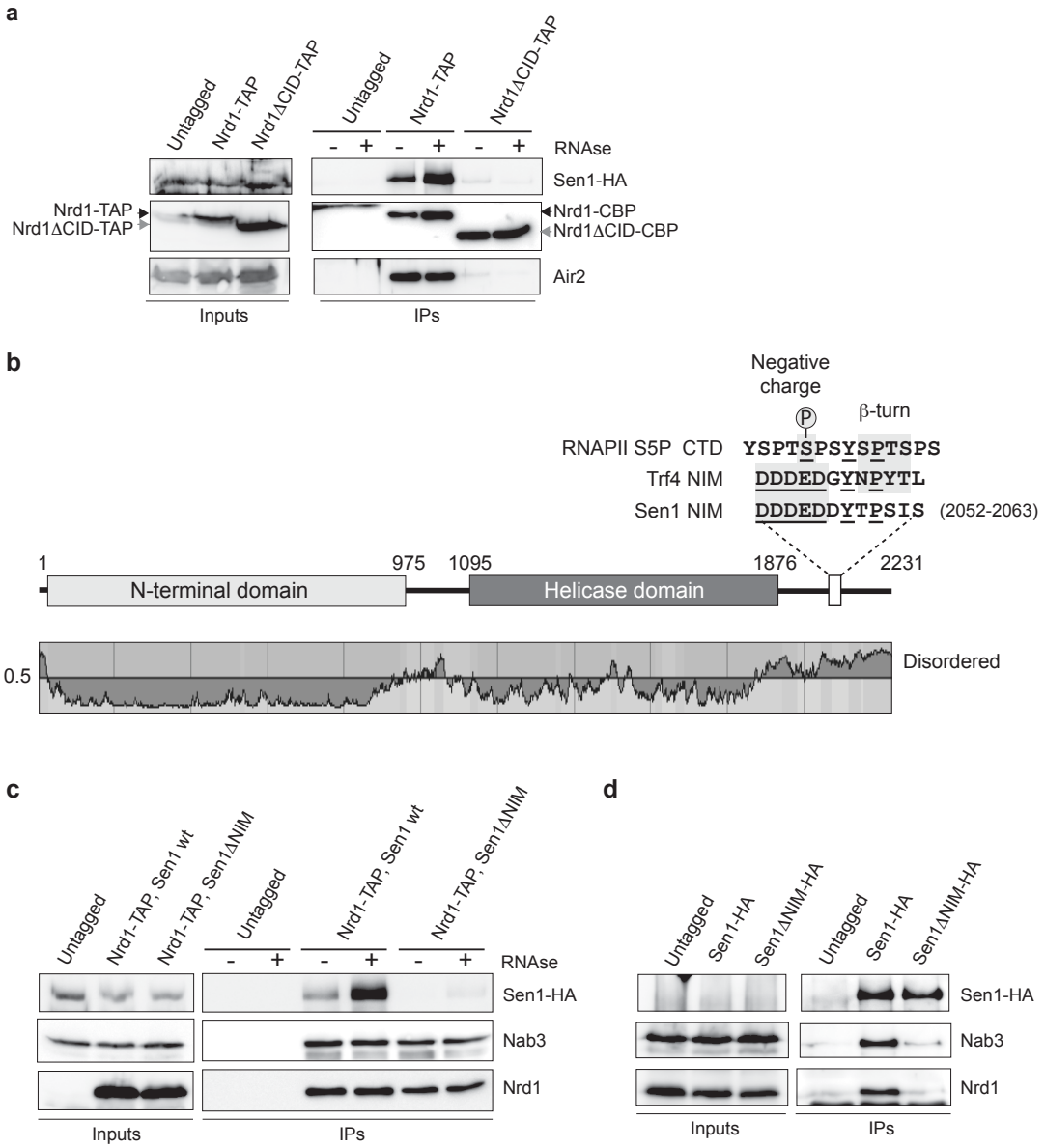

Figure 1: Identification of a Nrd1-Interaction Motif (NIM) in Sen1 that is critical for the integrity of the NNS-complex. a) Deletion of the CID domain dramatically reduces the interaction of Nrd1 with Sen1. Coimmunoprecipitation (CoIP) experiments using TAP-tagged Nrd1 (either wt or ΔCID) as the bait. **b**) Scheme of Sen1 protein. Globular domains are denoted by solid bars whereas intrinsically disordered regions are shown by a line. The disorder prediction was obtained using IUPred [72]. The sequence of the RNAPII S5P-CTD and Trf4 and Sen1 NIMs is shown on the top. Structural elements that are important for the interaction with Nrd1 CID are indicated. Conserved positions are underlined. **c**) Deletion of the NIM decreases substantially the association of Nrd1 with Sen1. CoIP experiments using Nrd1-TAP as the bait in a *SEN1* or *sen1ΔNIM* background. **d**) CoIP experiments using HA-tagged Sen1, either wt or ΔNIM, as the bait. Protein extracts were treated with RNaseA prior to immunoprecipitation.

In these experiment, Sen1 could not be detected in the input extracts. Antibodies used for protein detection are listed in table S8.

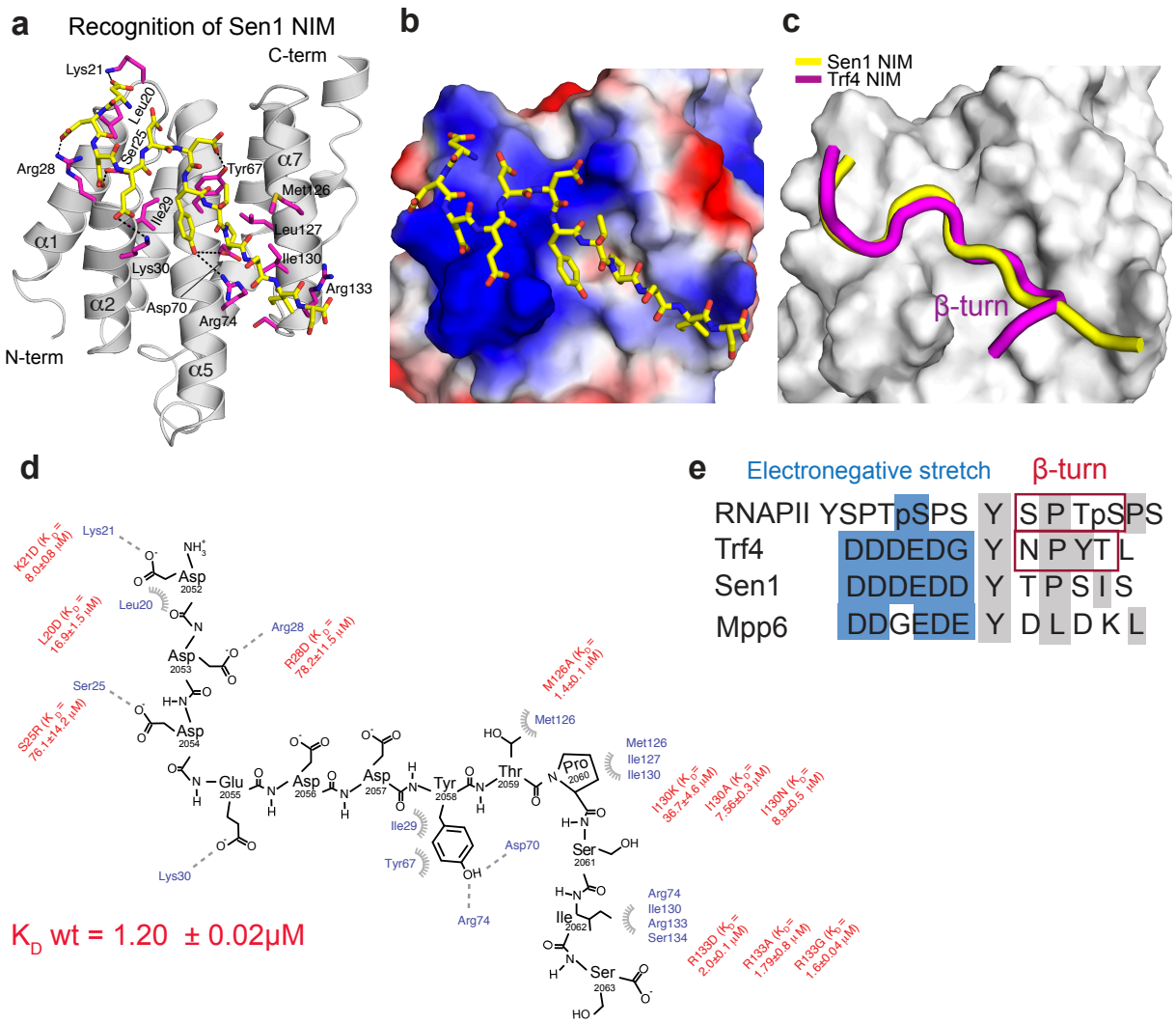

Figure 2: Recognition of Sen1 NIM by Nrd1 CID. a) NMR structure of Nrd1 CID bound to Sen1 NIM. The NIM peptide is represented in yellow sticks (only non-hydrogen atoms are shown) and Nrd1 CID is shown as a grey ribbon model. Nrd1 CID residues that form hydrophobic contacts and putative hydrogen bonds to Sen1 NIM peptide are shown in magenta sticks. **b)** Electrostatic surface representation of Nrd1 CID (electropositive in blue; electronegative in red; neutral in white) with the Sen1 NIM peptide (represented in yellow sticks; only non-hydrogen atoms are shown). The upstream electronegative stretch of Sen1 NIM interacts with an electropositive pocket of Nrd1 CID, while the C-terminal part of the peptide adopts an extended conformation that docks in a hydrophobic pocket of Nrd1 CID. **c)** Superposition of Nrd1 CID–Sen1 NIM (yellow) and Nrd1 CID–Trf4 NIM (magenta) complexes, displaying only peptide ribbons on the surface of Nrd1 CID. The comparison highlights the

extended and β-turn conformations of Sen1 NIM and Trf4 NIM, respectively. **d)** Scheme showing contacts (putative H-bonds and hydrophobic contacts) and energetics between the Sen1 NIM peptide and Nrd1 CID. Equilibrium binding experiments with the protein mutants were monitored by FA. L20D mutant disrupts the hydrophobic contact with F17 and impairs the overall geometry of the α1-α2 loop that contributes to the interaction with the upstream electronegative stretch of NIM. K_D (wild type Nrd1 CID-Sen1 NIM) equals 1.20 ± 0.02 μM. **e)** Alignment of RNAPII CTD and CTD mimics (Trf4 NIM, Sen1 NIM, and Mpp6 CTD mimic). Blue and grey boxes highlight the upstream electronegative stretches and hydrophobic regions, respectively. Previous structural works reported that S/NPXX motifs form the β-turn conformation[19,35].

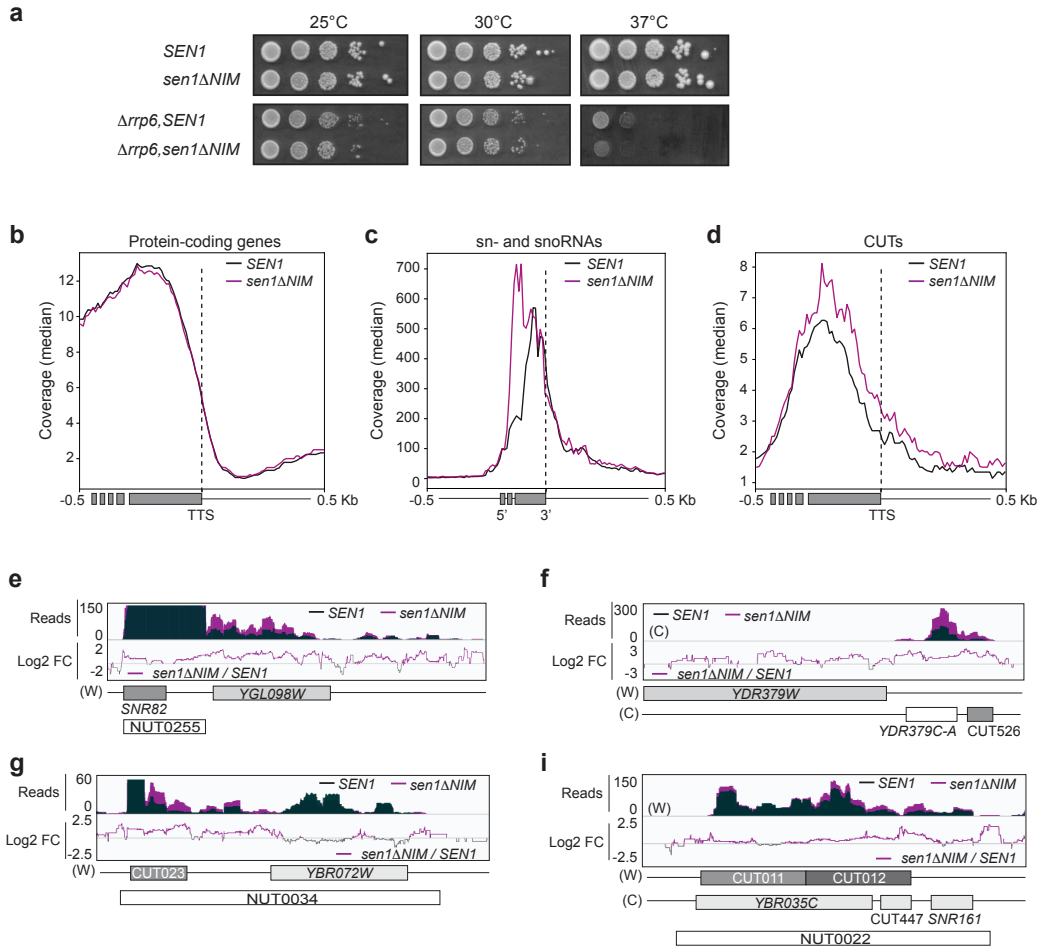

Figure 3: Deletion of the Sen1 NIM has only a moderate effect on growth and on non-coding transcription termination. a) Growth tests of the *sen1ΔNIM* mutant in either a wt or a *Δrrp6* background. **b-d)** Metagene analyses of RNAseq experiments performed in a *Δrrp6* background in the presence of the wt or the *ΔNIM* version of Sen1. The profile corresponds to the median coverage (reads per 10^7 reads mapping at each genomic position) from 0.5 kb

upstream to 0.5 kb downstream of the annotated transcription termination site (TTS) of protein-coding genes (**b**) and CUTs (**d**) or the 3'end of the mature sn- and snoRNAs (**c**). Experiments were performed in biological duplicates. **e-i**) Integrative Genomics viewer (IGV) screenshots of selected snoRNAs and CUTs exhibiting transcription termination defects in the *sen1⊿NIM* mutant. For each example the plot on the top corresponds to the coverage (reads per 10^7 reads, in short, "reads") obtained for each strain, whereas the plot on the bottom represents the log2 of the fold-change (FC) of the signal of the mutant relative to the wt. Features annotated at the Watson (w) and/or the Crick strand (c) are indicated. Extended forms of these snoRNAs and CUTs detected upon depletion of Nrd1 (Nrd1-Unterminated Transcripts or NUTs) in a former study[6] are shown to indicate the expected products of inefficient termination. The images in **e**) and **g**) correspond to zoom in views to better visualize the differences in the signal at the termination region.

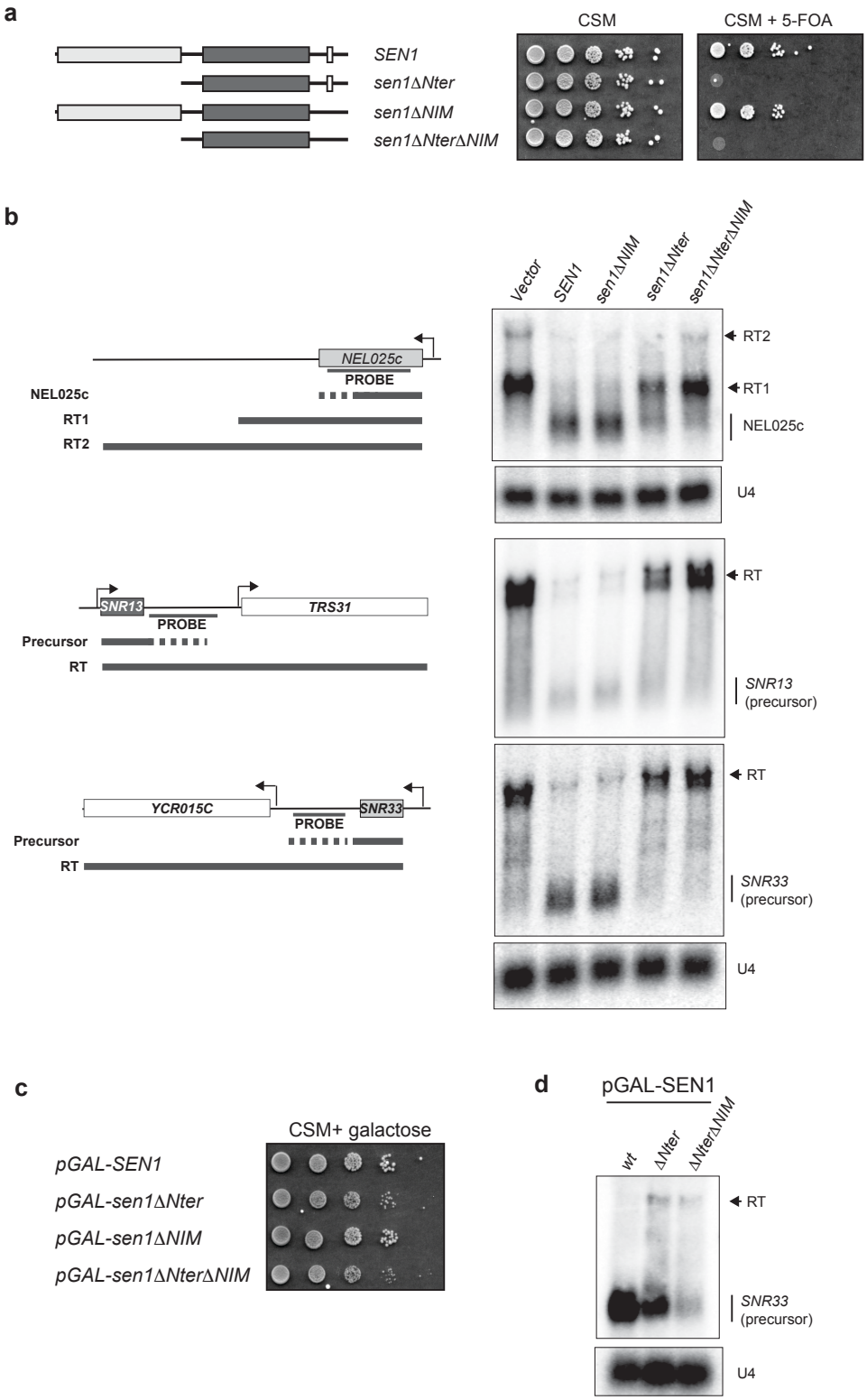

Figure 4: The N-terminal domain of Sen1 plays an essential role in non-coding transcription termination *in vivo*. a) Deletion of the N-terminal of Sen1 is lethal. A *Δsen1* strain (YDL2767) covered by an *URA3*-containing plasmid (pFL38) expressing wt Sen1 was transformed with a *TRP1*-plasmid (pFL39) carrying either the wt or the mutant versions of *SEN1* indicated in the scheme on the left. After over-night growth in non-selective medium,

cells were plated on minimal medium (CSM) containing 5-fluorootic acid (5-FOA) to select those that have lost the *URA3* plasmid. The absence of growth in 5-FOA implies that the *SEN1* version expressed in the *TRP1*-plasmid does not support viability. **b**) Deletion of the N-terminal domain of Sen1 provokes dramatic defects in transcription termination *in vivo*. Northern blot analyses of three well-characterized NNS-targets, the CUT NEL025c and the snoRNAs *SNR13* and *SNR33*, in a Sen1-AID (auxin-induced degron) strain carrying an empty vector or a plasmid expressing either the wt or the indicated versions of *SEN1.* Sen1-AID was depleted for 2h by the addition of 500 mM indole-3-acetic acid (IAA) to monitor the capacity of the plasmid-borne versions of *SEN1* to induce transcription termination. The expected RNA species resulting from inefficient termination (RT for readthrough species) are indicated. **c**) Overexpression of *sen1ΔNter* restores viability. Growth test of strains harboring the indicated version of *SEN1* at the endogenous locus under the control of the *GAL1* promoter (pGAL) in the presence of galactose. **d**) Overexpression of *SEN1* partially suppresses the termination defects associated with deletion of the N-terminal domain. Northern blot analysis of a typical NNS-target in strains expressing the indicated *SEN1* versions from pGAL in the presence of galactose. For **b**) and **d**) assays were performed in a *Δrrp6* background to detect the primary products of NNS-dependent termination and the U4 RNA is used as a loading control. Probes used for RNA detection are described in table S11.

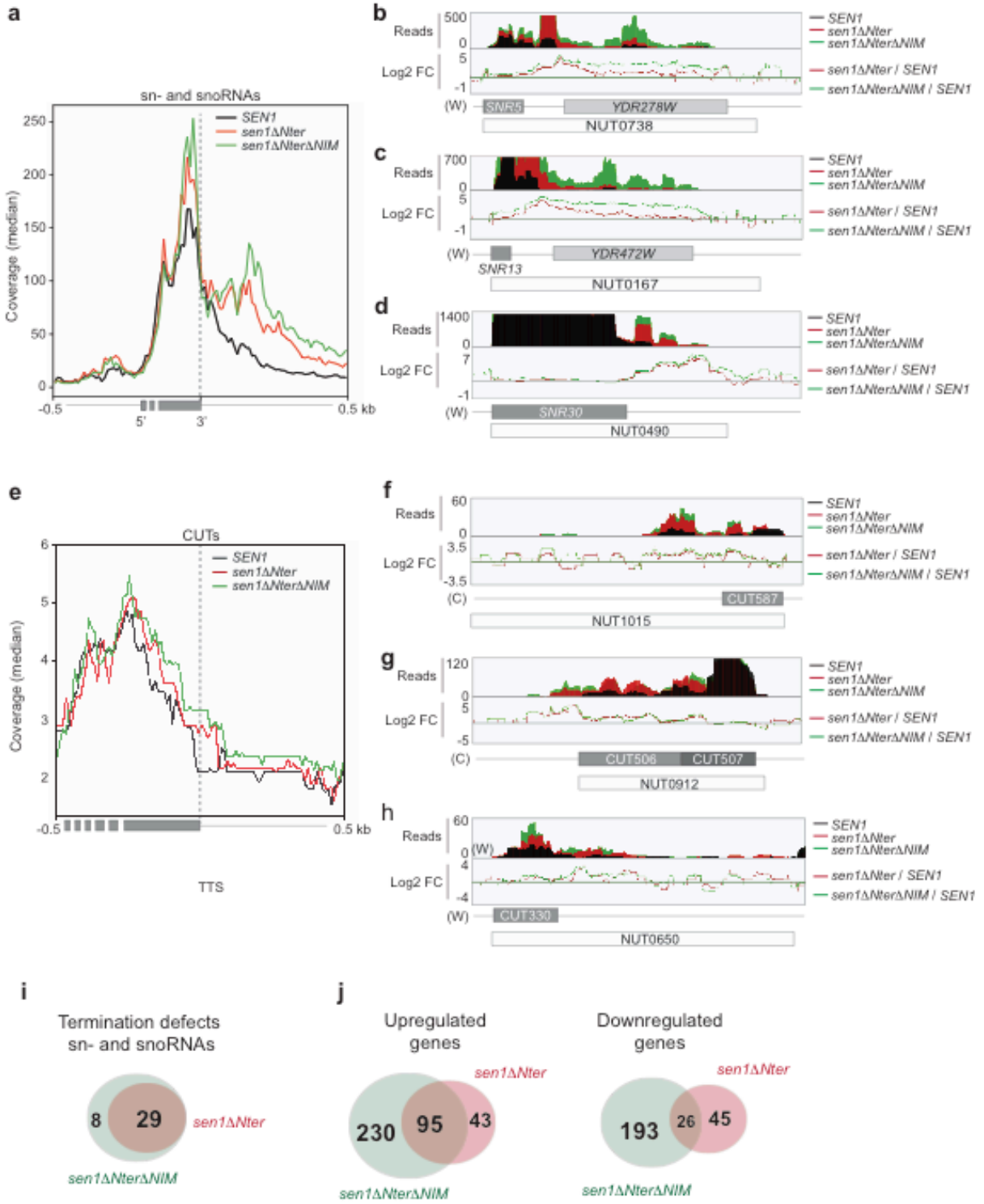

Figure 5: The absence of a NIM in Sen1 generally aggravates the transcription termination defects associated with deletion of the N-terminal domain. a) Metagene analyses of RNAseq experiments performed in a *Δrrp6* background in the presence of the wt, the *ΔNter* or the *ΔNterΔNIM* version of *SEN1* expressed from pGAL in CSM containing galactose. The profile corresponds to the median coverage from 0.5 kb upstream to 0.5 kb downstream of the 3'end of the mature sn- and snoRNAs. Experiments were performed in

biological duplicates. **b-d**) IGV screenshots of examples of snoRNAs displaying substantial termination defects in the mutants. The images are zoom in views to better visualize the differences in the signal at the termination region. **e**) Metagene analyses of the same RNAseq experiments to monitor for global termination defects at CUTs. The annotated TTS (see methods for details) was used as the reference point. **f-h**) IGV screenshots of examples of CUTs displaying substantial termination defects in the mutants. **i**) Venn diagram showing the overlap between the sets of snoRNAs that exhibit substantial termination defects in either mutant. These sets of snoRNAs correspond to those in cluster 1 generated by heatmap and clustering analyses (see supplementary figure 4). **j**) Venn diagrams illustrating the overlap between the protein-coding genes that are up- (left) or downregulated (right) in the different mutants relative to the wt at least 1.5 fold. The size of the circles is not proportional to the number of genes.

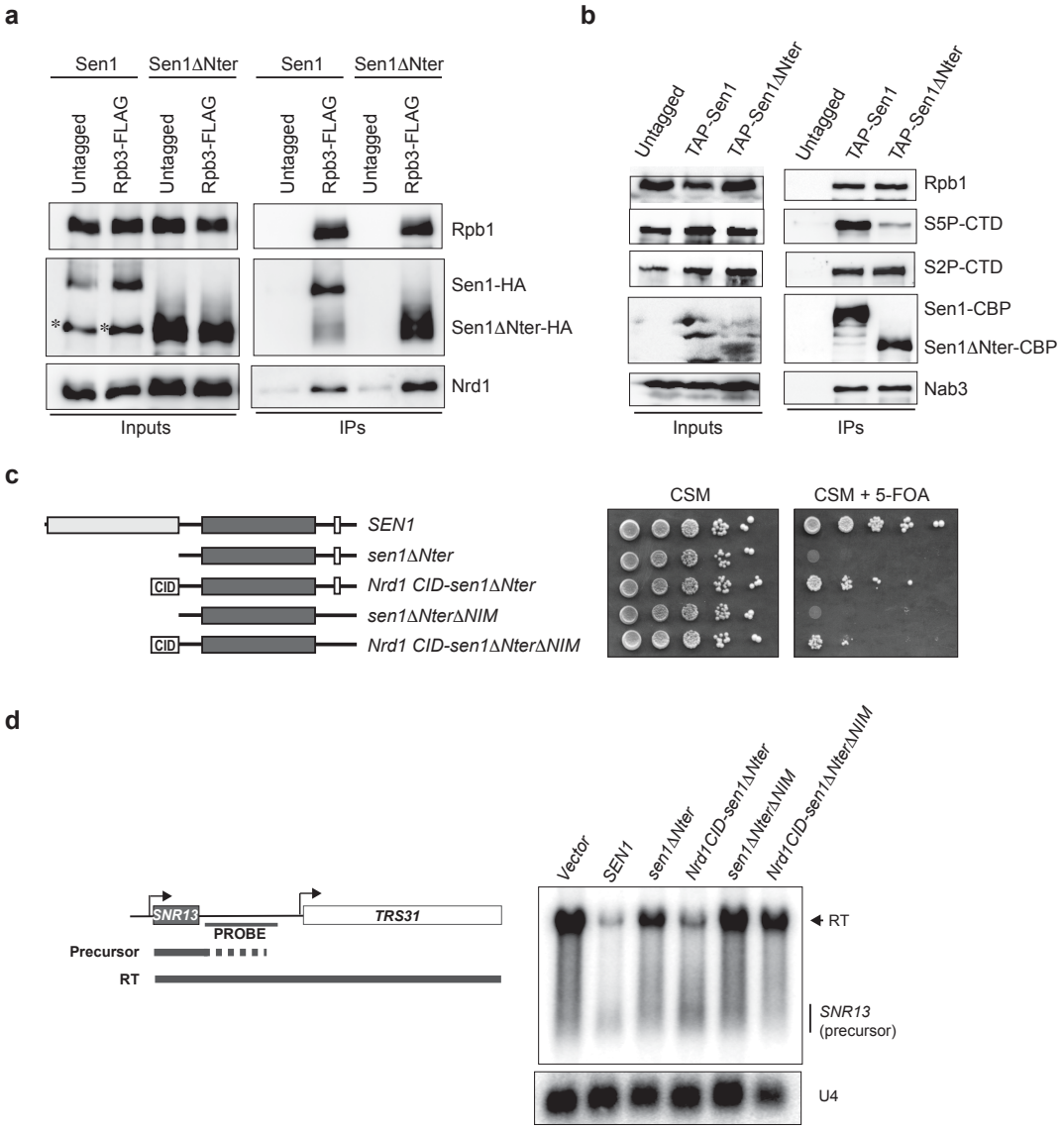

Figure 6: The N-terminal domain of Sen1 is critical for the recognition of the S5-phosphorylated form of RNAPII CTD. a) Deletion of Sen1 N-terminal domain does not prevent the interaction of Sen1 with RNAPII. CoIP experiments using Rbp3-FLAG as the bait. Assays were performed in a Sen1-AID strain harboring a plasmid expressing either *SEN1* or *sen1ΔNter* upon depletion of Sen1-AID in the presence of IAA for 2h. An asterisk denotes a major proteolytic Sen1 fragment detected in the extracts of roughly the size of *sen1ΔNter.* Nrd1 is detected as a positive control. **b**) Deletion of the Sen1 N-terminal domain reduces the interaction of Sen1 with the S5P-CTD. CoIP experiments using TAP-Sen1 as the bait. Sen1 proteins were expressed from pGAL in the presence of galactose. Nab3 is detected as a positive control. In both **a**) and **b**) protein extracts were treated with RNaseA before immunoprecipitation. Antibodies used for protein detection are listed in table S8. **c**) Replacing the Nter of Sen1 by the CID of Nrd1 restores viability. Growth test performed in the same conditions as in figure 4A but in the presence of a *TRP1*-plasmid carrying the *SEN1* versions

indicated in the scheme on the left. The growth of the strain expressing the Nrd1 CID-sen1ΔNter chimera in 5-FOA implies that this gene can support viability. **d**) Substituting the Nter of Sen1 by Nrd1 CID partially suppresses the termination defects detected in the *sen1ΔNter* mutant, but only in the presence of a functional NIM. Northern blot assays performed in a Sen1-AID strain carrying an empty vector or a plasmid expressing the indicated versions of *SEN1* upon depletion of the endogenous Sen1 protein as in figure 4B. Experiments performed in a *Δrrp6* background. The U4 RNA is used as a loading control. Probes used for RNA detection are described in table S11.

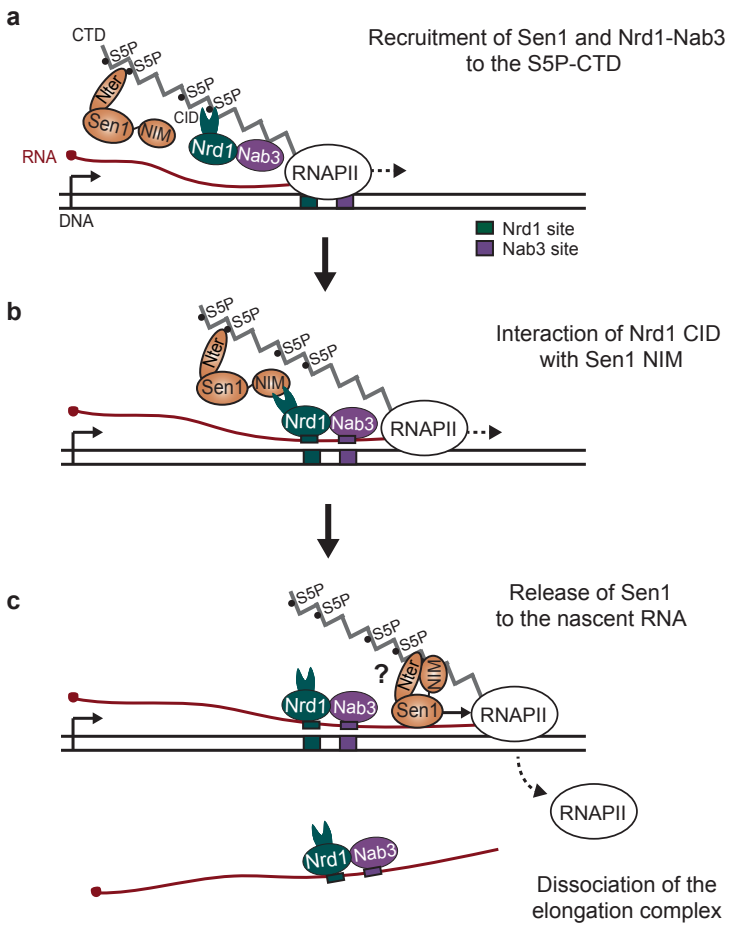

Figure 7: Model for the role of the N-terminal domain and the NIM in the recruitment of Sen1. a) Sen1 and Nrd1 are independently recruited to the S5P-CTD via the Nter and the CID, respectively. **b**) Once Nrd1 and Nab3 RNA recognition sequences emerge from the transcribing RNAPII, Nrd1 and Nab3 would interact with them and Nrd1 would likely release the CTD to bind Sen1 NIM. This interaction between Nrd1 and Sen1 might be important to stabilize Sen1 association with the elongation complex and/or recruit Sen1 when the function

or integrity of the Nter is compromised. **c**) Intra-molecular interactions between the Nter and the NIM might be important for the release of Sen1 from RNAPII and Nrd1 to the nascent RNA to subsequently provoke dissociation of the elongation complex.