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 identify a motif in an intrinsically disordered region of Sen1 that mimics the phosphorylated carboxy terminal domain (CTD) of RNA polymerase II and characterize structurally its recognition by the CTD-interacting domain of Nrd1, an RNA-binding protein that binds specific sequences in ncRNAs. In addition, we show that Sen1-dependent termination strictly requires the recognition of the Ser5-phosphorylated form of the CTD by the N-terminal domain of Sen1. Furthermore, we find that the N-terminal and the C-terminal domains of Sen1 can mediate intra-molecular interactions. Our results shed light onto the network of protein-protein interactions that control termination of non-coding transcription by Sen1.

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).

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 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). While snoRNAs are important for the correct modification of rRNA, CUTs are generally considered as non-functional molecules (Arigo et al., 2006; Schulz et al., 2013; Thiebaut et al., 2006; Wyers et al., 2005).

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). 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 snoRNAs, or complete degradation, in the case of CUTs (LaCava et al., 2005; Vanacova et al., 2005; Wyers et al., 2005). 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, 2017a).

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 (Lunde et al., 2010). 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 (Xiang et al., 2014). Similarly, the CID of Nrd1 within the NNS complex interacts with the CTD phosphorylated at S5, which is a mark of early elongation (Kubicek et al., 2012; Vasiljeva et al., 2008), and both Nrd1 and Nab3 recognize specific sequence motifs that are enriched at the target ncRNAs (Hobor et al., 2011; Lunde et al., 2011; Porrua et al., 2012; Schulz et al., 2013; Wlotzka et al., 2011). 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 (Gudipati et al., 2008; Tudek et al., 2014; Vasiljeva et al., 2008). 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, Porrua and Libri, 2013). 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 (Tudek et al., 2014; Fasken et al., 2015; Kim et al., 2016).

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 (Han et al., 2017; Martin-Tumasz and Brow, 2015). 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 (Han et al., 2017; Leonaitė et al., 2017; Porrua and Libri, 2013). Neither *in vitro* nor *in vivo* Sen1 exhibits any sequence-specific RNA-binding capability (Creamer et al., 2011; Porrua and Libri, 2013); 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 Chong et al., 2015; Ghaemmaghami et al., 2003; Kulak et al., 2014; Newman et al., 2006); ii) its low-processivity as a translocase, which makes termination highly dependent on Sen1 efficient recruitment to the elongation complex and RNAPII pausing (Han et al., 2017) 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* (Han et al., 2017; Leonaitė et al., 2017), 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 phosphorylated 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 interaction of Sen1 with its partners Nrd1 and Nab3 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. Moreover, we find that the N-terminal and the C-terminal domains of Sen1 can interact with each other *in vitro*, which might play a role in modulating the interaction of Sen1 with RNAPII CTD and Nrd1. Our findings allow us to propose a detailed molecular model on how inter- and intra-molecular interactions control the specific function of the transcription termination factor Sen1 on ncRNAs.

RESULTS

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

In a previous report (Tudek et al., 2014), we showed that Nrd1 CID domain can recognize a short sequence in Trf4 that mimics 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 (Kim et al., 2016). 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 also reported in an independent study (Heo et al., 2013). 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 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 (figure EV1). Conversely, we did not detect any putative NIM in Sen1 *S. pombe* and human orthologues, in agreement with previous data showing that Nrd1 and Sen1 orthologues do not interact with each other in *S. pombe* (Lemay et al., 2016; Wittmann et al., 2017) and with the fact that no Nrd1 homologue could be identified in association with human Sen1 (Yüce and West, 2013).

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 roughly 100-fold smaller than the K_D of the Nrd1 CID for the S5P-CTD (Tudek et al., 2014). In addition, we observed 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 (Tudek et al., 2014), which indicates that the interaction of Nrd1 CID with Sen1 NIM is quite stable. 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 (figure EV2).

In order to understand the 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 (Kubicek et al., 2012; Tudek et al., 2014; Vasiljeva et al., 2008). 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 (Becker et al., 2008; Kubicek et al., 2012; Lunde et al., 2010; Meinhart and Cramer, 2004; Tudek et al., 2014). 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 (Singh et al., 2015). 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 (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 major difference in the expression profile of snoRNAs in *sen1ΔNIM,* indicating that deletion of the NIM does not have a general impact on transcription termination at this class of NNS-targets (figure 3C). We only detected a small increase in the RNAseq signal downstream of mature *SNR82* in the Sen1 mutant, consistent with a mild termination defect at this locus (figure 3E),

which was further confirmed by northern blot assays (figure EV3A). Metagene analysis of CUTs did not uncover any substantial change in the average RNA signal downstream of the annotated termination site in the *sen1ΔNIM* (figure 3D), indicating that deletion of the NIM does not result in a global decrease in transcription termination efficiency. Inspection of individual cases revealed detectable termination defects only at a minority of CUTs (see examples in figure 3F-I and figure EV3A). Because in our coimmunoprecipitation experiments we observed that some minor interaction between Sen1 and Nrd1 persisted after the deletion of Sen1 NIM (figure 1C), we considered the possibility that this remaining interaction could support important levels of Sen1 recruitment, therefore explaining the weak phenotype of the *sen1ΔNIM* mutant. To investigate this possibility, we deleted the whole C-terminal domain (Cter) of Sen1 downstream of the previously identified nuclear localization signal (Nedea et al., 2008) and analysed the capacity of this mutant (Sen1ΔCter) to interact with Nrd1. Indeed, deletion of the Sen1 Cter did not decrease the protein expression levels but fully abolished the interaction between Sen1 and Nrd1, indicating that this domain of Sen1 possesses additional surfaces that weakly contribute to the interaction between Sen1 and its partners (figure EV3B). However, northern blot analyses of several typical NNS-targets revealed only minor transcription termination defects in the *sen1ΔCter* mutant compared to the wild-type (figure EV3C).

Taken together, our results indicate that, although the interaction between Sen1 and Nrd1-Nab3 might be important for fully efficient termination at a subset of non-coding genes, it is not a strict 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 (Chinchilla et al., 2012; Ursic et al., 2004). 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 (Chen et al., 2014; Ursic et al., 2004) 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 (figure EV4A). 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, Nishimura et al., 2009) strain that allowed us to rapidly target Sen1 for degradation by the proteasome upon addition of auxin (figure EV4B). We supplemented the Sen1-AID strain with a centromeric plasmid expressing either the wt, *ΔNIM, ΔNter* or the *ΔNterΔNIM* version of *SEN1* or an empty vector, as a control, 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 (figure EV4C). 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 (Han et al., 2017), 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 (figure EV4A). Metagene analyses of 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 observed a similar trend at CUTs, although the effect was more moderate (figure 5E). Inspection of individual cases confirmed the transcription termination defects suggested by the

metagene analyses (figure 5F-H). These results are consistent with our observations on isolated cases with the mutant proteins expressed under their own promoter (figure 4B).

We set out to analyse in more detail the termination defects in each mutant. We restricted our analyses to snoRNAs, because their higher expression levels relative to CUTs together with their similar size facilitates more robust and accurate analyses. We quantified the RNA signal for each snoRNA over a window of 300 bp downstream of the 3' end of the mature form and we computed the increase in this signal in each mutant relative to the wt (readthrough ratio, figure EV5B and supplementary table S3). This method likely underestimates the extent of termination defects because the readthrough region often extends to more than 1 Kb downstream of the 3' end of the mature snoRNA, but allows a quantitative comparison of the different mutants. We detected a readthrough increase above 1.5 fold in 50% of snoRNAs in the *ΔNter* mutant (p-value < 0.04). Interestingly, deletion of both the Nter and the NIM provoked substantially stronger termination defects at a fraction of snoRNAs (12 out of 51), suggesting that the presence of the NIM is more critical for the function of Sen1 at a subset of targets.

Because decreased NNS-dependent transcription termination efficiency leads to deregulation of protein-coding genes downstream of and antisense to CUTs and snoRNAs (Schulz et al., 2013), to gain a more global view of the termination defects of either mutant, we compared the sets of protein-coding genes whose expression was significantly altered (p-value < 0.05) in each 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(Colin et al., 2011)). 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 5I-J). 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 the production of higher 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 (figure EV5C-E).

Taken together, these results indicate that deletion of the NIM enhances the termination defects associated with deletion of the Nter at a large fraction of NNS-targets, suggesting that both protein regions have 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 of RNAPII

As evoked above, Sen1 Nter has been proposed to interact with RNAPII, more specifically with the S2P form of the CTD (Chinchilla et al., 2012), 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 6B). 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 (Chinchilla et al., 2012) (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 (Kubicek et al., 2012), 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.

The N-terminal and the C-terminal domains of Sen1 can mediate intra-molecular interactions

Because the NIM mimics the phosphorylated CTD and our results strongly suggest that the Nter of Sen1 recognizes the S5P-CTD, we considered the possibility that the Nter of Sen1 could interact with the NIM. In order to test this hypothesis we conducted *in vitro* pull-down experiments with recombinant Sen1 C-terminal domain (aa 1930-2231) and the Nter expressed in yeast (figure 6E). Indeed, we observe substantial and reproducible interaction between both domains of Sen1, indicating that these protein regions have the potential to mediate intra-molecular interactions. Strikingly, the Nter could bind equally well the wild-type and the ΔNIM version of Sen1 Cter, indicating that the Nter recognizes regions other than the NIM in the C-terminal domain.

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 snoRNAs, which in turn have important functions in 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 (Schulz et al., 2013). 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 (Harlen and Churchman, 2017b, 2017a). One of these factors is Nrd1, that preferentially binds the S5P-CTD, which contributes to its recruitment to the transcribing RNAPII during early elongation (Kubicek et al., 2012; Vasiljeva et al., 2008). 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 (Kim et al., 2016). 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 several major structural elements: one or several negatively charged residues at the N-terminal part and several hydrophobic amino acids that adopt a β-turn at the C-terminal part, 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 (Kim et al., 2016) as its DLDK C-terminal motif (figure 2E) has no propensity to form a β-turn (Singh et al., 2015).

This relaxed sequence requirement implies that other 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. Of course, such protein regions should be present in the appropriate protein context (i.e. intrinsically disorder regions and not covered by other protein-protein interaction regions) and in the appropriate cellular compartment (i.e. in the nucleus, in the case of Nrd1-interacting factors). 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. In addition, full deletion of the C-terminal domain of Sen1 completely abolishes the interaction between Sen1 and its partners but has only a minor impact on transcription termination efficiency. Therefore, the recruitment of Sen1 to the transcription elongation complex does not seem to depend to a large extent on the interaction with Nrd1 and Nab3. In contrast, deletion of the Nter of Sen1 is lethal and results in massive transcription termination defects. This result would be, in principle, compatible with an autoregulatory role for this domain. For instance, the Nter could be an activation domain or could mediate intra-molecular interactions that relieve autoinhibition. However, in a previous *in vitro* study using purified proteins we have shown that deletion of the Nter does not impair Sen1 catalytic activity or capacity to induce transcription termination (Han et al., 2017). This result, together with 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 aggravates the transcription termination defects observed in a *sen1ΔNter* mutant. It is therefore possible that under certain physiological conditions the protein interactions mediated by the Nter are altered and the binding to Nrd1 and Nab3 becomes more relevant for Sen1 recruitment. Indeed, the remarkable conservation of the NIM in close yeast species indicates that there is some selective pressure to maintain this motif.

The essential character of the Nter of Sen1 was overlooked in former studies (Chen et al., 2014; Ursic et al., 2004), 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 (Chinchilla et al., 2012). 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. In our case, we detect the association of Sen1 with RNAPII harbouring S2P-CTD, but this association does not depend on the presence of the Nter. Bearing in mind that Sen1 plays additional different transcription-related roles (e.g. resolution of transcription-replication conflicts, transcription-coupled DNA repair), it is conceivable that Sen1 interacts with RNAPIIs that simply happen to be phosphorylated at S2 because the levels of this mark are quite high along most of the transcription cycle. In contrast, the levels of S5P are high only during early elongation. A former study has shown a strong correlation between the levels of S5 phosphorylation and the efficiency of NNS-dependent termination (Gudipati et al., 2008). 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 moderate, albeit global, transcription termination defects (Kubicek et al., 2012; Tudek et al., 2014). 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.

Multiple protein-protein interactions control the dynamics of termination factors during transcription

The fact that the Nter of Sen1 behaves as a “reader” of S5P-CTD and that the NIM functions as a “mimic” of S5P-CTD for Nrd1 lead us to test whether Sen1 Nter would recognize the NIM in the Cter. Our *in vitro* results (figure 6E) do not support such idea, however, we found that the Nter can bind the Cter *in trans*, strongly suggesting that these two domains of Sen1 can mediate intra-molecular interactions. It is important to note that there are several structurally dissimilar families of protein domains that can recognize the CTD of RNAPII. It is certainly possible that the CTD-interaction domain of Sen1 does not possess the same S5P-CTD recognition determinants than Nrd1 CID and, therefore, does not bind the NIMs. Structural analyses of the Nter should shed light on this matter. Concerning the function of these intra-molecular interactions, as mentioned above, our former *in vitro* study disfavour the hypothesis of an autoregulatory role. Therefore, we propose that these interactions would play a role in modulating the interaction of the Nter and the Cter with additional factors (see below).

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 would be in close proximity to Sen1, the NIM of Sen1, being a much stronger binder for Nrd1

CID, would displace Nrd1 from the CTD. Indeed, we have shown that the interactions of the CID with the CTD of RNAPII and the NIM of Sen1 are mutually exclusive and that the affinity of Sen1 NIM for Nrd1 CID is ~100 fold higher than that of the S5P-CTD. The interaction of Nrd1 with Sen1 might stabilize Sen1 on the elongation complex or eventually bring it in close proximity to the nascent RNA. Although the Nter of Sen1 does not seem to recognize the NIM, it is possible that the intra-molecular interactions between the Nter and the Cter change the conformation of Sen1 in a way that destabilizes its association with RNAPII and Nrd1. This would, thus, facilitate the release of Sen1 onto the RNA to subsequently translocate towards RNAPII and induce termination. In addition, dissociating Sen1 from Nrd1 would be necessary for the CID to interact with Trf4 NIM to subsequently promote polyadenylation and degradation of the released ncRNA.

Our former and present data are consistent with the idea that competitive protein interactions play an important role in regulating the dynamics of termination and RNA quality factors in ncRNA metabolism. An additional mechanism that might partake in the modulation of these protein-protein interactions could rely on post-translational modification. Indeed, Nrd1 and Sen1, being associated with the CTD during transcription, might also be targeted by CTD enzymes and modified in the regions we have characterized in this study.

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* (Porrua and Libri, 2013), 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* (Carroll et al., 2007; Conrad et al., 2000; Steinmetz and Brow, 1998; Steinmetz et al., 2001, 2006). 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 1). 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 (Schaughency et al., 2014). 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 (Han et al., 2017). 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 (Longtine et al., 1998; Rigaut et al., 1999) 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, 200 mM sodium acetate, 0.25% NP-40, 2 mM EDTA, 1 mM 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 (10 mM Tris pH 8, 150 mM NaCl, 0.1 % NP-40, 0.5 mM EDTA and 1 mM DTT) for 2h at 21°C or by boiling the beads in 2x Laemmli buffer (100 mM Tris pH 6.8, 4% SDS, 15% glycerol, 25 mM 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.

For pull-down experiments each recombinant version of His_6-GST-tagged Sen1 Cter or the His_6-GST control was overexpressed by growing BL21 (DE3) CodonPlus (Stratagene) cells harboring the appropriate plasmid (see table S10) on auto-inducing medium (Studier, 2005) at 20°C overnight. Protein extracts were prepared in GST binding buffer (50 mM Tris-HCl, pH

7.5, 500 mM NaCl, 5% glycerol, 0.1% NP40, 1 mM DTT) by sonication and subsequent centrifugation at 13000 rpm for 30 minutes at 4°C. Approximately 5 mg of extract containing the corresponding recombinant protein was incubated with 25 µl of glutathione sepharose (SIGMA) and subsequently mixed with yeast extracts expressing Sen1 TAP-Nter (typically 0.5 mg of extract per binding reaction). Beads were washed with lysis buffer (10 mM sodium phosphate pH 7, 200 mM sodium acetate, 0.25% NP-40, 2 mM EDTA, 1 mM EGTA, 5% glycerol) and the proteins were eluted by incubation for 15 minutes in 80 µl GST elution buffer containing 50 mM Tris-HCl pH 8, 20 mM reduced glutathione, 150 mM NaCl, 0.1% NP40 and 10% glycerol. Protein extracts were treated with 50 µg/ml of RNase A for 20 minutes at 20ºC prior to incubation with beads.

Fluorescence anisotropy analyses

Nrd1 CID and its mutants were produced and purified as described previously (Kubicek et al., 2012). 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 (Kuzmic, 2009).

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% H_2O/10% D_2O). 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 (Kay et al., 1993) was measured with a non-uniform sampling; acquired data were processed and analysed analogously as described previously (Nováček et al., 2011, 2012). 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 (Peterson et al., 2004; Zwahlen et al., 1997) 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 (Güntert and Buchner, 2015). 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 (Case, 2002). These calculations employed a modified version (AMBER ff14SB) of the force field (Maier et al., 2015), using a protocol described previously (Hobor et al., 2011; Stefl et al., 2010). 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(Bolger et al., 2014) and mapped to the R64 genome(Cherry et al., 2012) with bowtie2 using the default options (Langmead and Salzberg, 2012).

Annotations

For protein coding genes, snoRNAs and CUTs we used the annotations in Xu et al, 2008 (Xu et al., 2009). To facilitate snoRNA analyses we excluded intronic and polycistronic snoRNAs. 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 400 bp upstream of any gene (protein-coding genes, sn- and snoRNAs and CUTs). From this group of CUTs we only retained for our analyses the 50% more highly expressed. The final subset of CUTs included 357 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 (Ramírez et al., 2016)**.** 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. Quantification of readthrough signal for calculation of the readthrough index was performed using the multiBigwigSummary tool from the deepTools2 package using the normalized bigwig files for the different samples as inputs and limiting the quantifications to a window of 300 bp downstream of the 3' end of the mature snoRNAs.

Differential expression analyses were performed by counting the reads from two independent biological replicates that mapped at protein-coding genes using the htseq-count tool(Anders et al., 2015) with the union mode. The count files thus generated were used as input files for the SARTools DESeq2 package (Varet et al., 2016). 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 Mara Barucco for her help with bioinformatic analyses. 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.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

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

Arigo, J.T., Eyler, D.E., Carroll, K.L., and Corden, J.L. (2006). Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. Mol Cell *23*, 841–851.

Becker, R., Loll, B., and Meinhart, A. (2008). 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.

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

Carroll, K.L., Ghirlando, R., Ames, J.M., and Corden, J.L. (2007). Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. Rna *13*, 361–373.

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

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

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, D700-705.

Chinchilla, K., Rodriguez-Molina, J.B., Ursic, D., Finkel, J.S., Ansari, A.Z., and Culbertson, M.R. (2012). Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae. Eukaryotic Cell *11*, 417–429.

Chong, Y.T., Koh, J.L.Y., Friesen, H., Duffy, S.K., Duffy, K., Cox, M.J., Moses, A., Moffat, J., Boone, C., and Andrews, B.J. (2015). Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. Cell *161*, 1413–1424.

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

Conrad, N.K., Wilson, S.M., Steinmetz, E.J., Patturajan, M., Brow, D.A., Swanson, M.S., and Corden, J.L. (2000). A yeast heterogeneous nuclear ribonucleoprotein complex associated with RNA polymerase II. Genetics *154*, 557–571.

Creamer, T.J., Darby, M.M., Jamonnak, N., Schaughency, P., Hao, H., Wheelan, S.J., and Corden, J.L. (2011). Transcriptome-wide binding sites for components of the Saccharomyces cerevisiae non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. PLoS Genet *7*, e1002329.

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

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

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

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

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

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

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

Harlen, K.M., and Churchman, L.S. (2017b). Subgenic Pol II interactomes identify region-specific transcription elongation regulators. Mol Syst Biol *13*.

Heo, D., Yoo, I., Kong, J., Lidschreiber, M., Mayer, A., Choi, B.-Y., Hahn, Y., Cramer, P., Buratowski, S., and Kim, M. (2013). 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.

Hobor, F., Pergoli, R., Kubicek, K., Hrossova, D., Bacikova, V., Zimmermann, M., Pasulka, J., Hofr, C., Vanacova, S., and Stefl, R. (2011). Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. J. Biol. Chem. *286*, 3645–3657.

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

Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R., and Forman-Kay, J.D. (1993). 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.

Kim, K., Heo, D.-H., Kim, I., Suh, J.-Y., and Kim, M. (2016). 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.

Kubicek, K., Cerna, H., Holub, P., Pasulka, J., Hrossova, D., Loehr, F., Hofr, C., Vanacova, S., and Stefl, R. (2012). Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. Genes Dev. *26*, 1891–1896.

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

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

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A., and Tollervey, D. (2005). RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell *121*, 713–724.

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

Lemay, J.-F., Marguerat, S., Larochelle, M., Liu, X., van Nues, R., Hunyadkürti, J., Hoque, M., Tian, B., Granneman, S., Bähler, J., et al. (2016). The Nrd1-like protein Seb1 coordinates cotranscriptional 3' end processing and polyadenylation site selection. Genes Dev. *30*, 1558–1572.

Leonaitė, B., Han, Z., Basquin, J., Bonneau, F., Libri, D., Porrua, O., and Conti, E. (2017). Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. EMBO J. *36*, 1590–1604.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Lunde, B.M., Reichow, S.L., Kim, M., Suh, H., Leeper, T.C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., and Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. Nat. Struct. Mol. Biol. *17*, 1195–1201.

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

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. *11*, 3696–3713.

Martin-Tumasz, S., and Brow, D.A. (2015). 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.

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

Nedea, E., Nalbant, D., Xia, D., Theoharis, N.T., Suter, B., Richardson, C.J., Tatchell, K., Kislinger, T., Greenblatt, J.F., and Nagy, P.L. (2008). The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. Mol. Cell *29*, 577–587.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

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

Nováček, J., Zawadzka-Kazimierczuk, A., Papoušková, V., Žídek, L., Šanderová, H., Krásný, L., Koźmiński, W., and Sklenář, V. (2011). 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR *50*, 1–11.

Nováček, J., Haba, N.Y., Chill, J.H., Žídek, L., and Sklenář, V. (2012). 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.

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

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

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

Porrua, O., Hobor, F., Boulay, J., Kubicek, K., D'Aubenton-Carafa, Y., Gudipati, R.K., Stefl, R., and Libri, D. (2012). In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. EMBO J. *31*, 3935–3948.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160-165.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Séraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. *17*, 1030–1032.

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

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. (2013). Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell *155*, 1075–1087.

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

Stefl, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B., et al. (2010). The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. Cell *143*, 225–237.

Steinmetz, E.J., and Brow, D.A. (1998). 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.

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

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

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J., and Libri, D. (2006). Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. Mol Cell *23*, 853–864.

Tudek, A., Porrua, O., Kabzinski, T., Lidschreiber, M., Kubicek, K., Fortova, A., Lacroute, F., Vanacova, S., Cramer, P., Stefl, R., et al. (2014). Molecular basis for coordinating transcription termination with noncoding RNA degradation. Mol. Cell *55*, 467–481.

Ursic, D., Chinchilla, K., Finkel, J.S., and Culbertson, M.R. (2004). 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.

Vanacova, S., Wolf, J., Martin, G., Blank, D., Dettwiler, S., Friedlein, A., Langen, H., Keith, G., and Keller, W. (2005). A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol *3*, e189.

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

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

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191.

Wittmann, S., Renner, M., Watts, B.R., Adams, O., Huseyin, M., Baejen, C., El Omari, K., Kilchert, C., Heo, D.-H., Kecman, T., et al. (2017). The conserved protein Seb1 drives transcription termination by binding RNA polymerase II and nascent RNA. Nat Commun *8*, 14861.

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

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725–737.

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

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-M|[uuml]|nster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Yüce, Ö., and West, S.C. (2013). 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.

Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. (1997). 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.

FIGURES

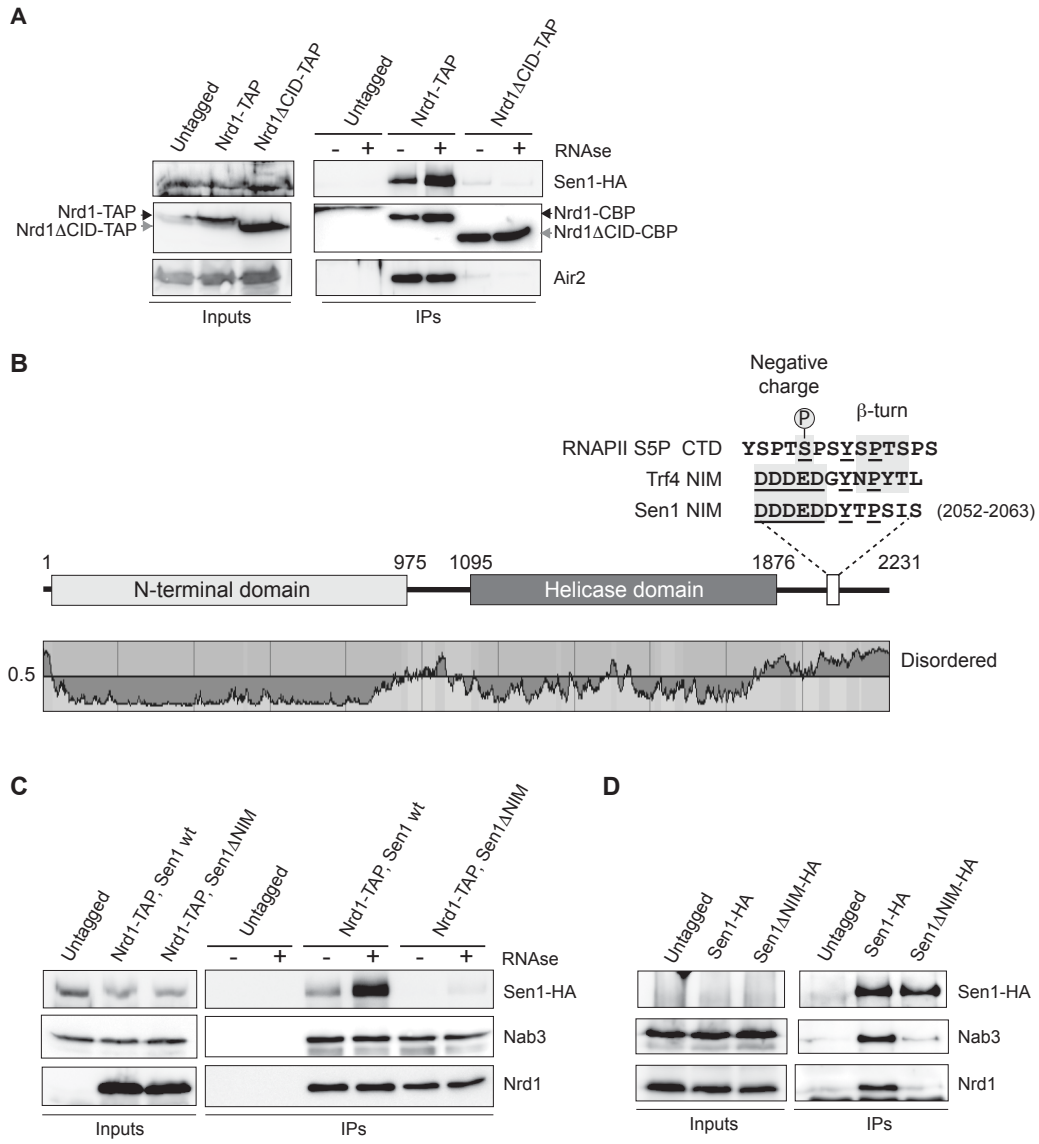


Figure 1: Identification of a Nrd1-Interaction Motif (NIM) in Sen1 that is critical for the integrity of the NNS-complex. A) Deletion of the CID domain dramatically reduces the interaction of Nrd1 with Sen1. Coimmunoprecipitation (CoIP) experiments using TAP-tagged Nrd1 (either wt or ΔCID) as the bait. Representative gel of one out of two independent experiments. **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 (Dosztányi et al., 2005). 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. Representative gel of one out of two

independent experiments. **D**) CoIP experiments using HA-tagged Sen1, either wt or ΔNIM, as the bait. Representative gel of one out of two independent experiments. 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.

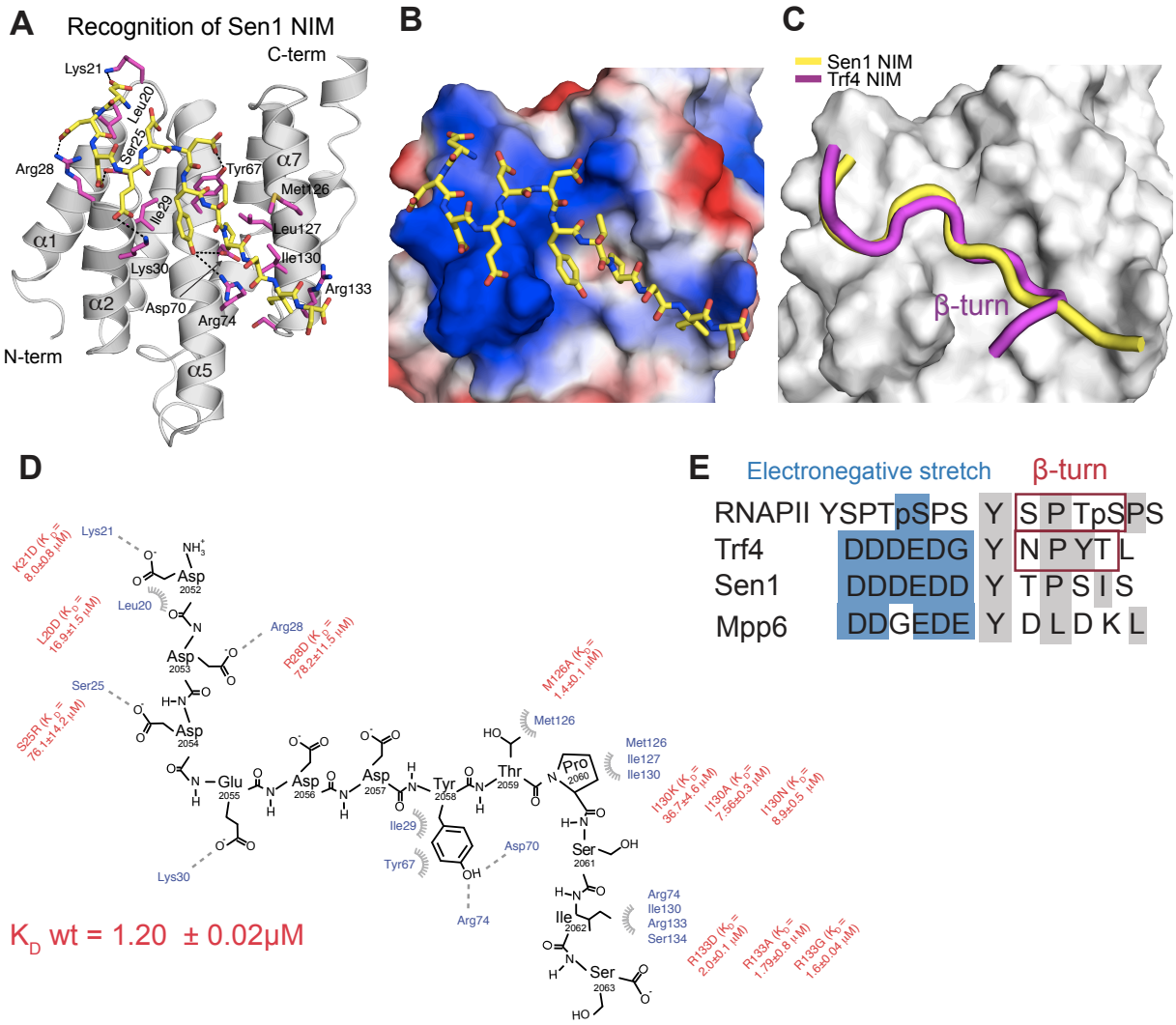


Figure 2: Recognition of Sen1 NIM by Nrd1 CID. A) NMR structure of Nrd1 CID bound to Sen1 NIM. The NIM peptide is represented in yellow sticks (only non-hydrogen atoms are shown) and Nrd1 CID is shown as a grey ribbon model. Nrd1 CID residues that form hydrophobic contacts and putative hydrogen bonds 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 (Kubicek et al., 2012; Tudek et al., 2014).

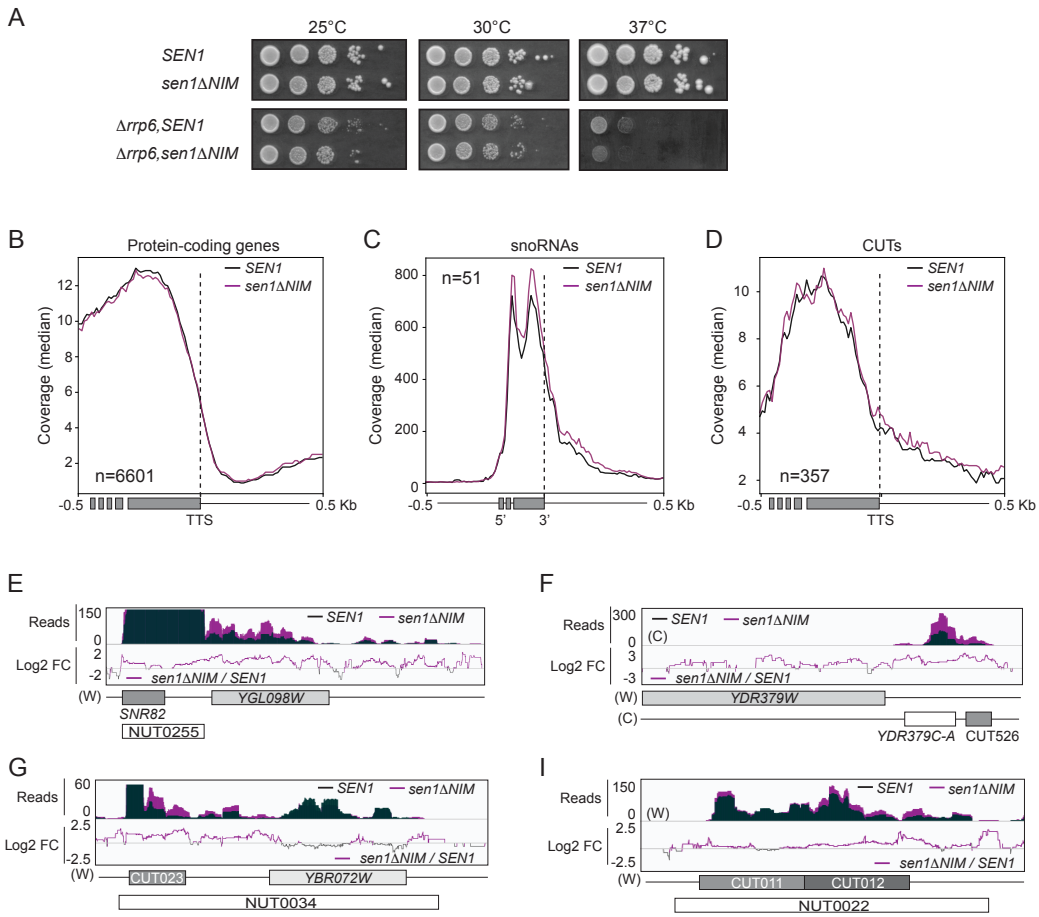


Figure 3: Deletion of the Sen1 NIM has only a moderate effect on growth and on non-coding transcription termination. A) Growth tests of the *sen1ΔNIM* mutant in either a wt or a *Δrrp6* background. **B-D)** Metagene analyses of RNAseq experiments performed in a *Δrrp6* background in the presence of the 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 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 *sen1ΔNIM* 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(Schulz et al., 2013) 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.

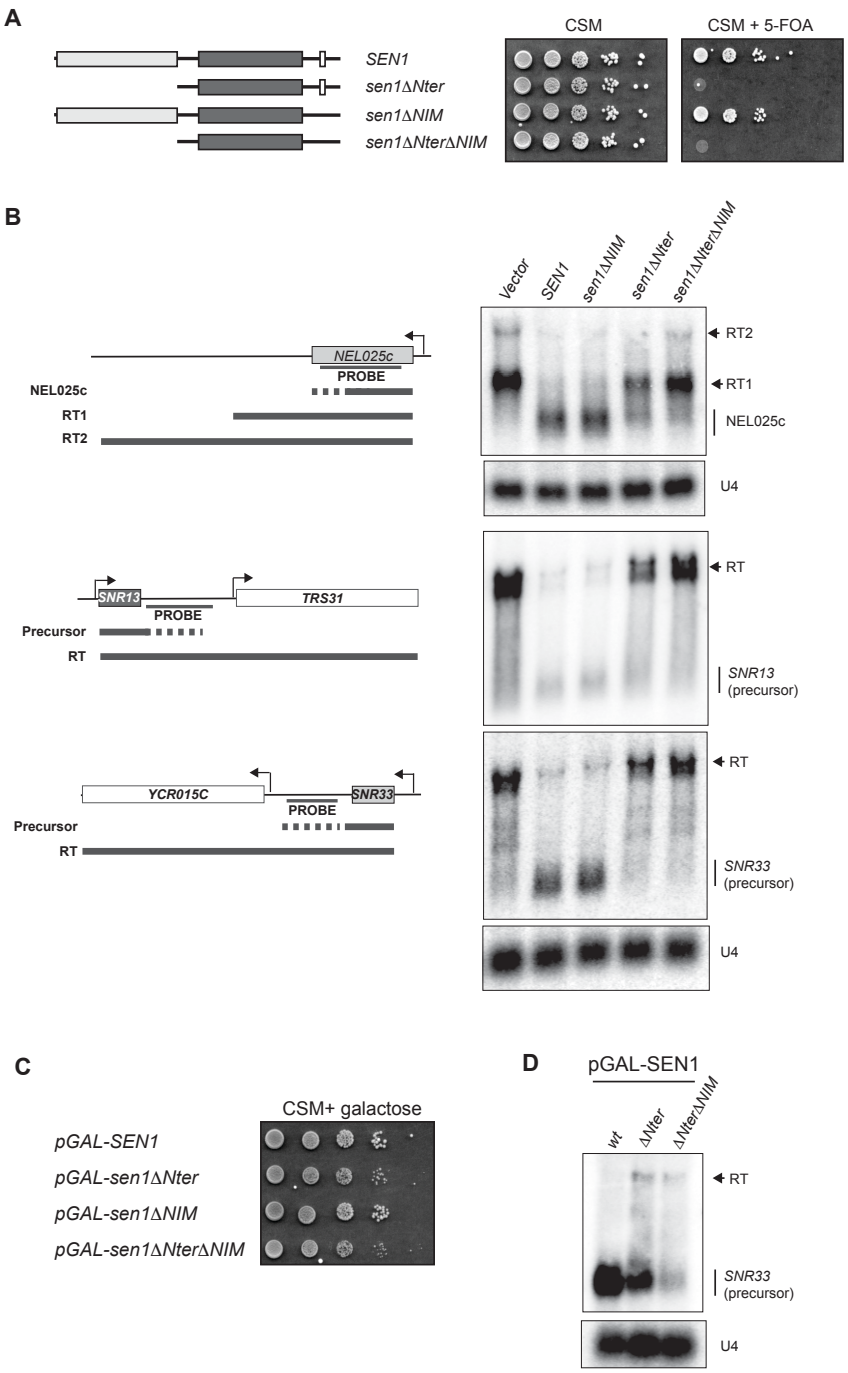


Figure 4: The N-terminal domain of Sen1 plays an essential role in non-coding transcription termination *in vivo*. A) Deletion of the N-terminal of Sen1 is lethal. A *Δsen1* strain (YDL2767) covered by an *URA3*-containing plasmid (pFL38) expressing wt Sen1 was transformed with a *TRP1*-plasmid (pFL39) 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. All results shown were reproduced in at least one additional independent experiment. Probes used for RNA detection are described in table S11.

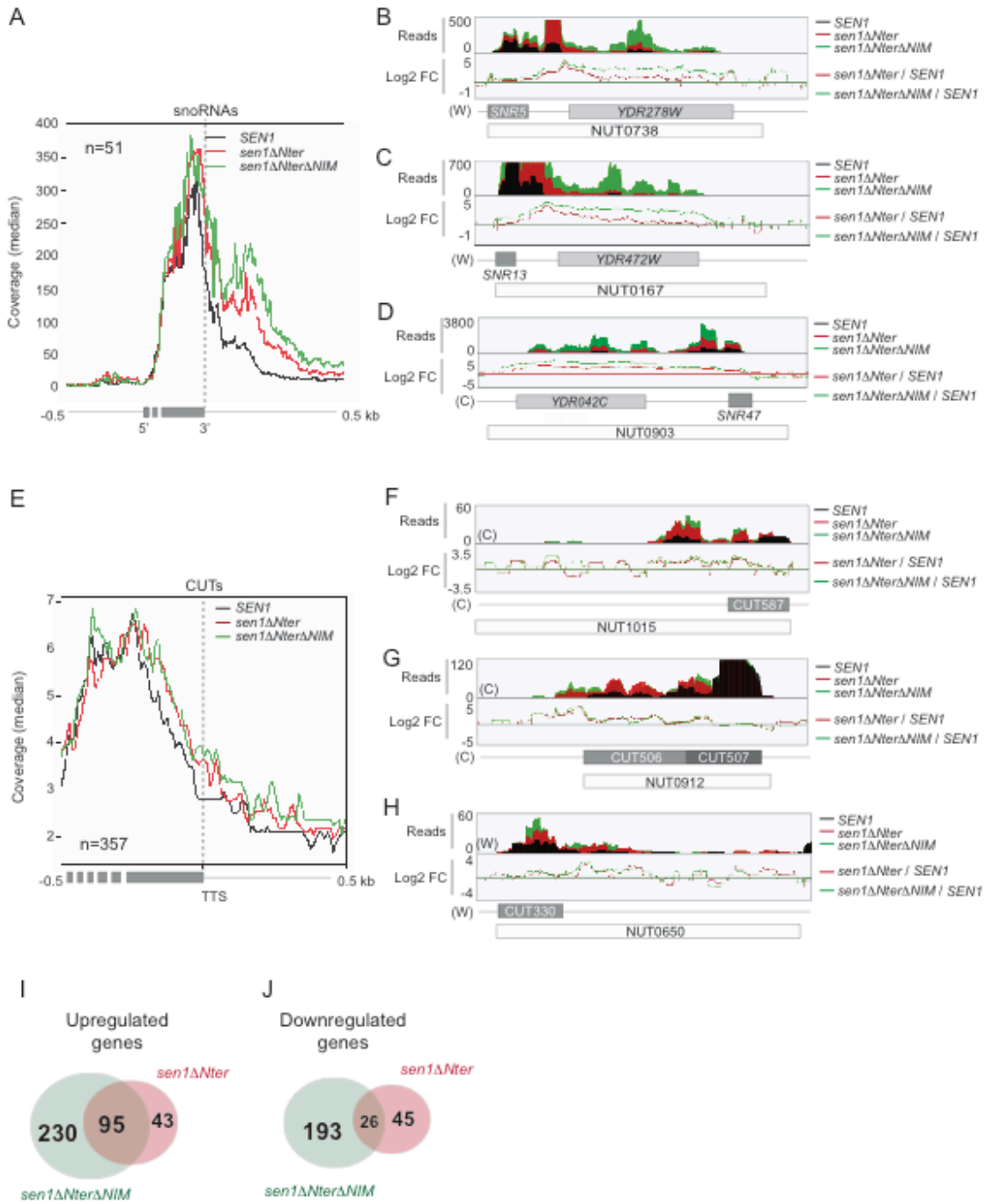


Figure 5: The absence of a NIM in Sen1 generally aggravates the transcription termination defects associated with deletion of the N-terminal domain. A) Metagene analyses of RNAseq experiments performed in a *Δrrp6* background in the presence of the wt, the *ΔNter* or the *ΔNterΔNIM* version of *SEN1* 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 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. Venn diagrams illustrating the overlap between the protein-coding genes that are up- (**I**) or downregulated (**J**) 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.

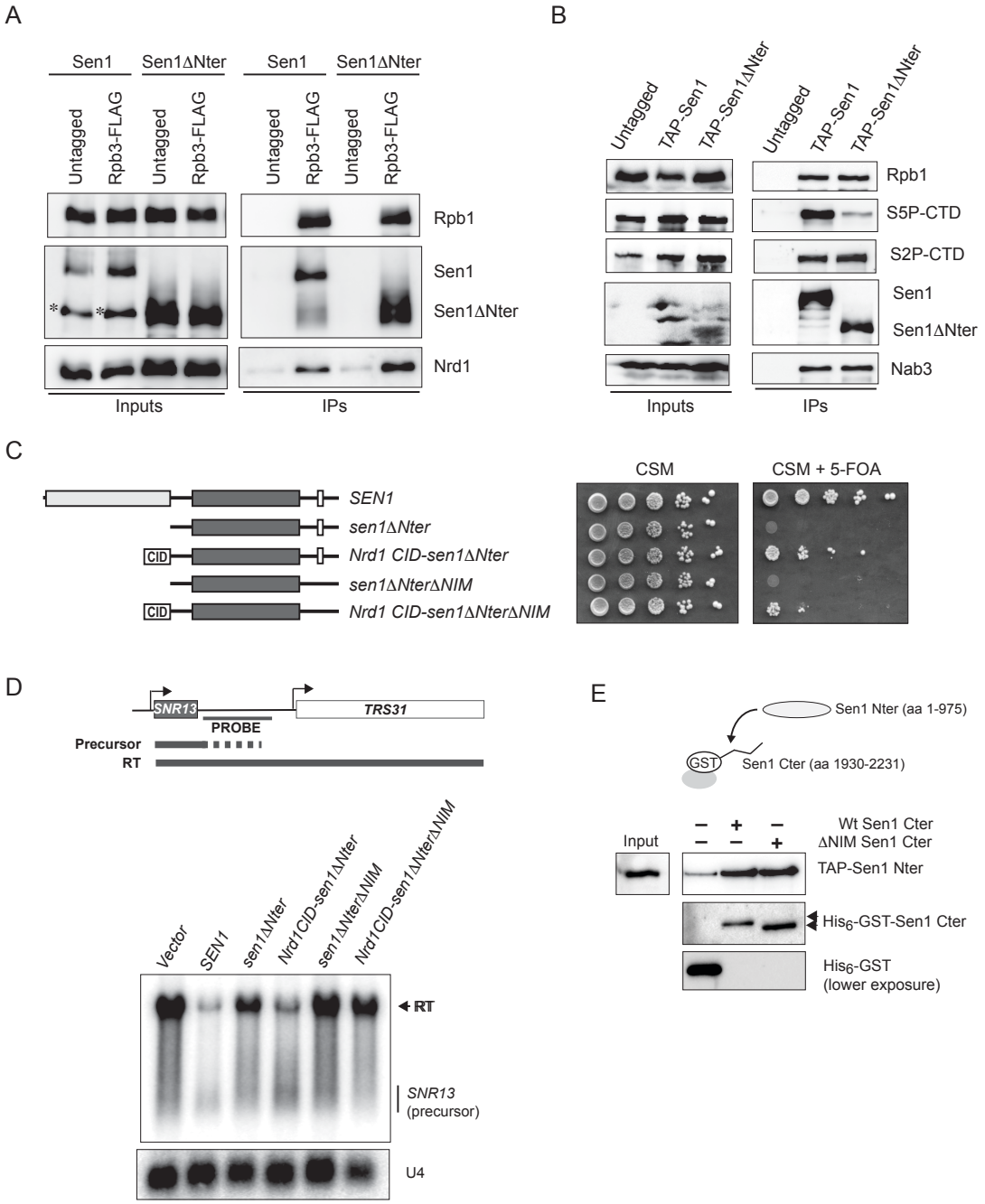


Figure 6: The N-terminal domain of Sen1 recognizes the S5-phosphorylated form of RNAPII CTD and Sen1 C-terminal domain. A) Deletion of Sen1 N-terminal domain does not prevent the interaction of Sen1 with RNAPII. CoIP experiments using Rbp3-FLAG as the bait.

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. Representative gel of one out of two independent experiments. **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. Representative gel of one out of three independent experiments. In both **A**) and **B**) protein extracts were treated with RNaseA before immunoprecipitation. **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. Representative gel of one out of two independent experiments.The U4 RNA is used as a loading control. Probes used for RNA detection are described in table S11. **E**) Sen1 Nter interacts with the C-terminal domain (Cter) of Sen1 both in the presence and in the absence of the NIM *in vitro*. Pull-down experiments using either a wt or a ΔNIM version of recombinant Sen1 Cter immobilized on glutathione-sepharose beads and a TAP-tagged version of Sen1 Nter expressed in yeast. Representative gel of one out of three independent experiments. Antibodies used for protein detection are listed in table S8.

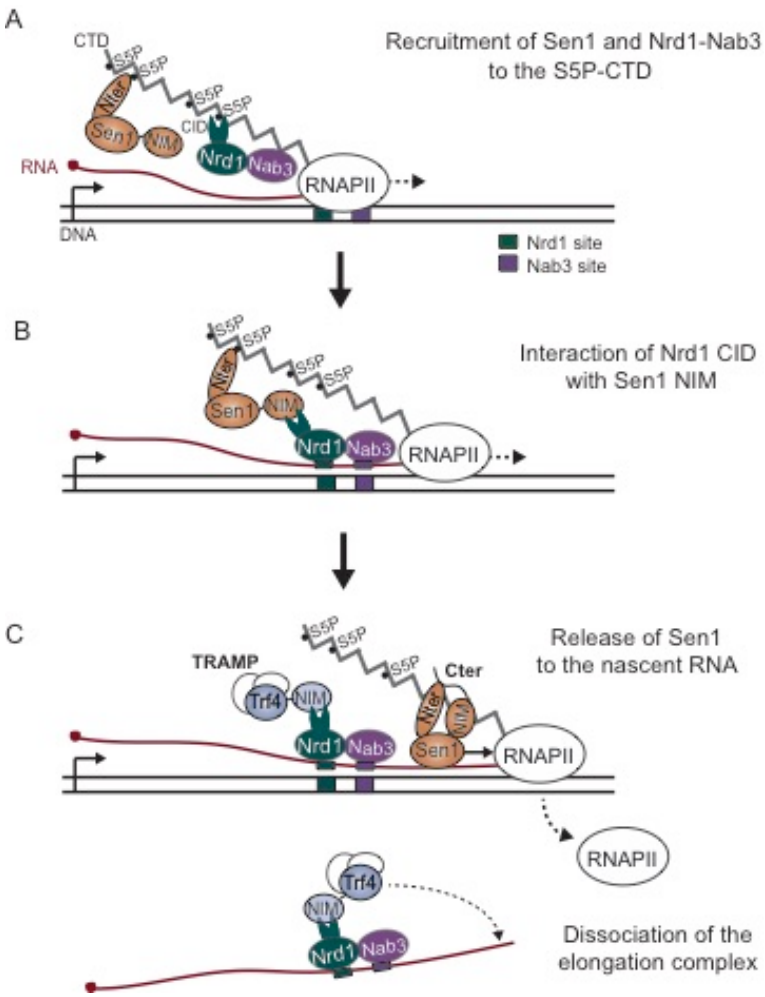
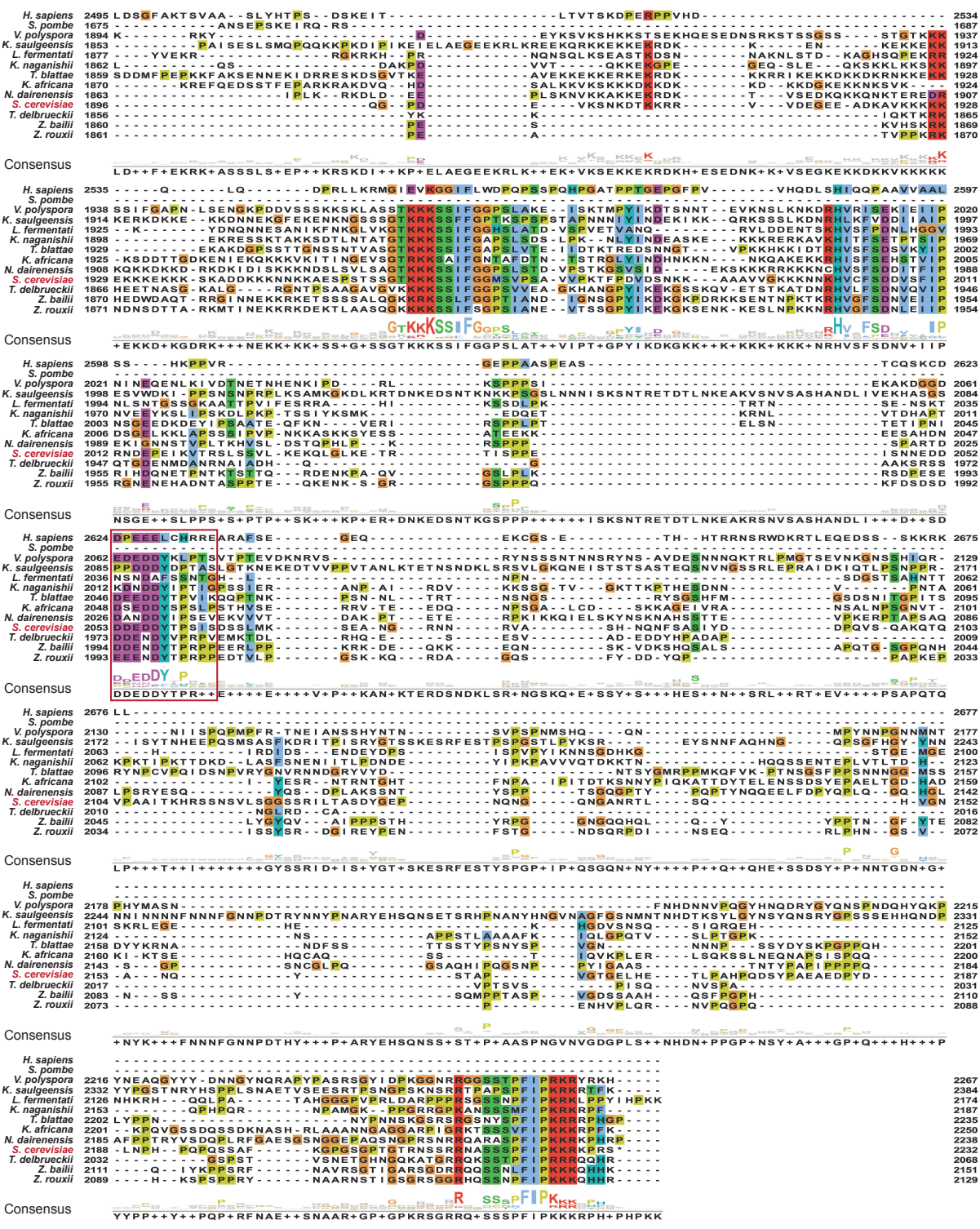


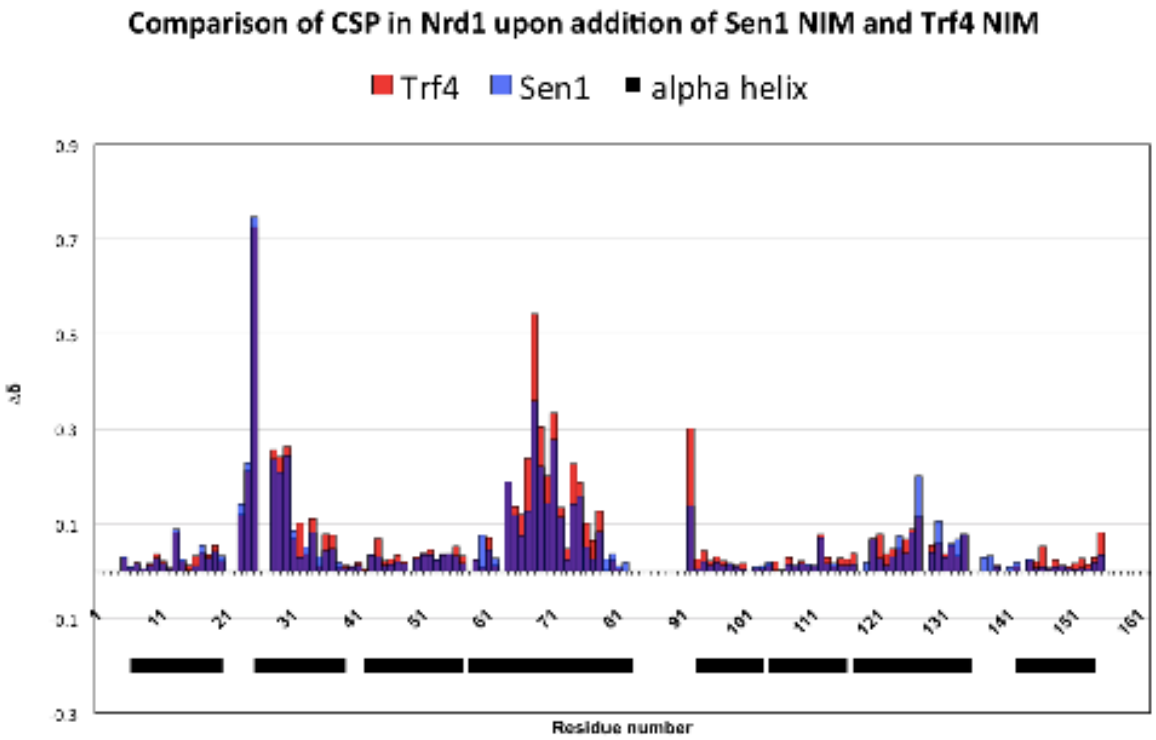
Figure 7: Model for the role of the N-terminal domain and the NIM in the recruitment of Sen1. A) Sen1 and Nrd1 are independently recruited to the S5P-CTD via the Nter and the CID, respectively. **B**) Once Nrd1 and Nab3 RNA recognition sequences emerge from the transcribing RNAPII, Nrd1 and Nab3 would interact with them. Sen1 NIM, being a much stronger binder of Nrd1 CID, would likely displace Nrd1 from the CTD. 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 Cter of Sen1 might be important for the release of Sen1 from RNAPII and Nrd1 to the nascent RNA to subsequently provoke dissociation of the elongation complex.

EXPANDED VIEW FIGURES

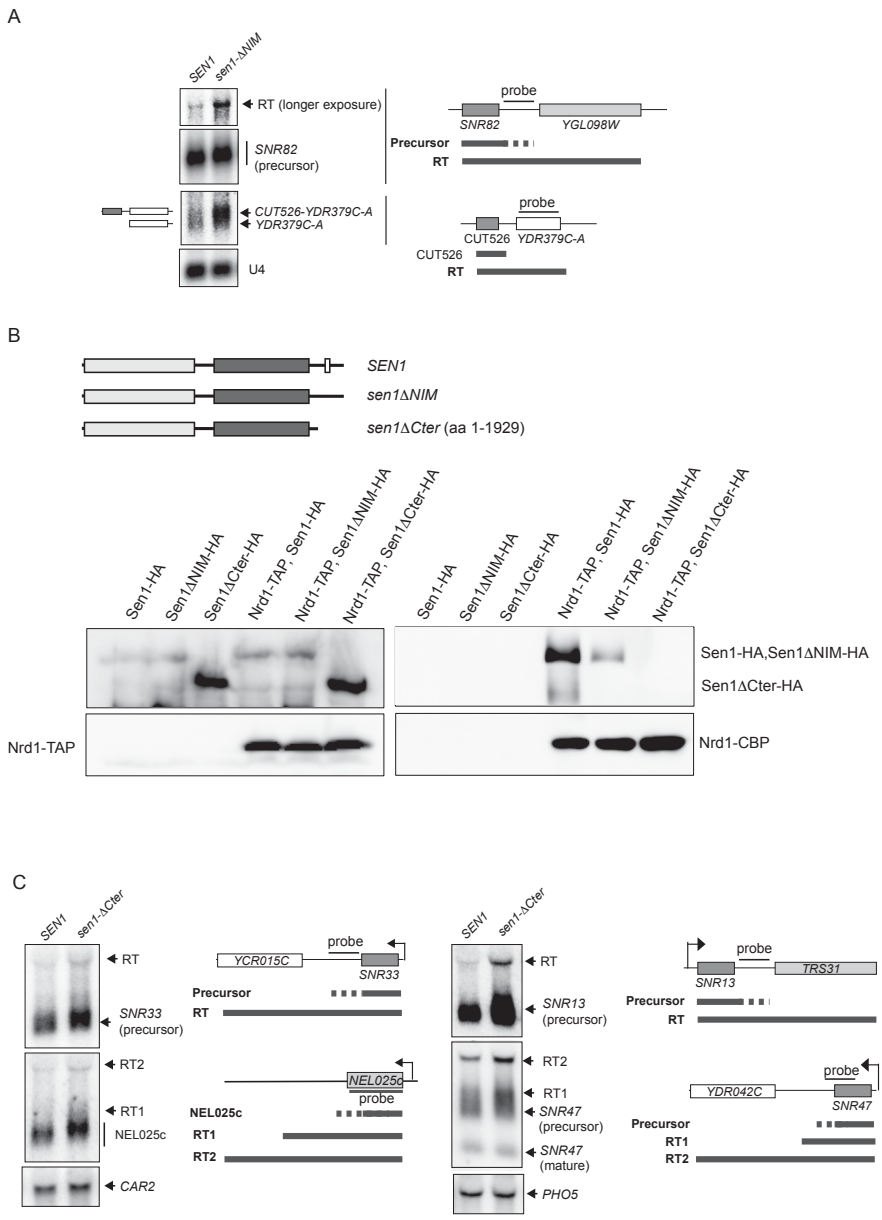


Expanded view figure 1: The Sen1 NIM is conserved in close yeast species. *S. cerevisiae* Sen1 protein sequence was submitted to blastp excluding the genus *Saccharomyces* from the search. The ten most conserved protein sequences together with Sen1 orthologues from *Homo sapiens* (SETX) and *Schizosaccharomyces pombe* were aligned to *S. cerevisiae* Sen1

using clustal omega. Visualization of the alignment and calculation of the consensus sequences ware performed with Jalview(Waterhouse et al., 2009). Amino acids are colored according to clustal color scheme. The protein identifiers for the proteins used in the alignment are the following: Q7Z333 (*H. sapiens*), Q92355 (*S. pombe*), XP_001644478.1 (*V. polyspora*), SMN21961.1 (*K. saulgeensis*), SCV99407.1 (*L. fermentati*), CCK69072.1 (*K. naganishii*), XP_004181671.1 (*T. blattae*), XP_003955130.1 (*K. Africana*), XP_003672383.1 (*N. dairenensis*), Q00416 (*S. cerevisiae*), XP_003680903.1 (*T. delbrueckii*), CDF91445.1 (*Z. bailii*) and GAV52597.1 (*Z. rouxii*). The NIM is indicated by a red box.

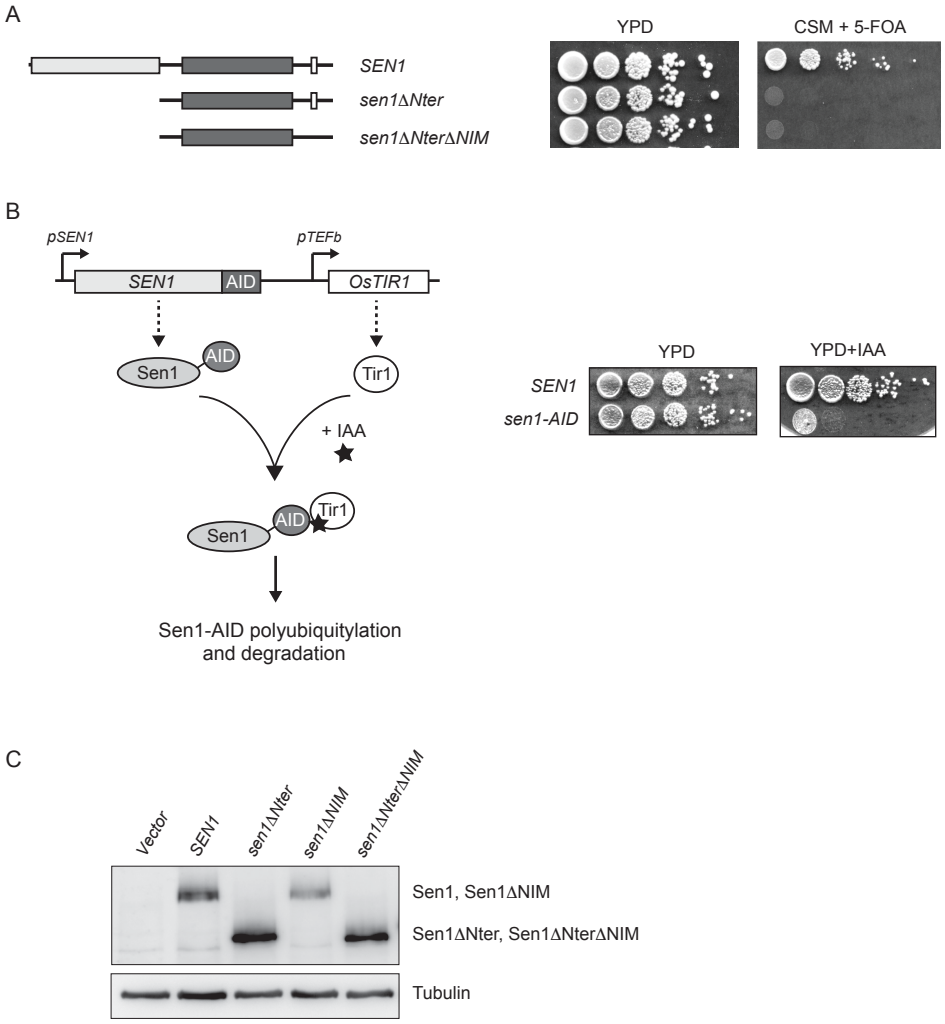


Expanded view figure 2: Quantification of chemical shift perturbations (CSP) of Nrd1 CID upon binding to the Trf4 NIM (in red) and Sen1 NIM peptides (in blue). The combined chemical shift perturbations are plotted versus the amino-acid residue number. The black boxes indicate α-helices of Nrd1 CID.

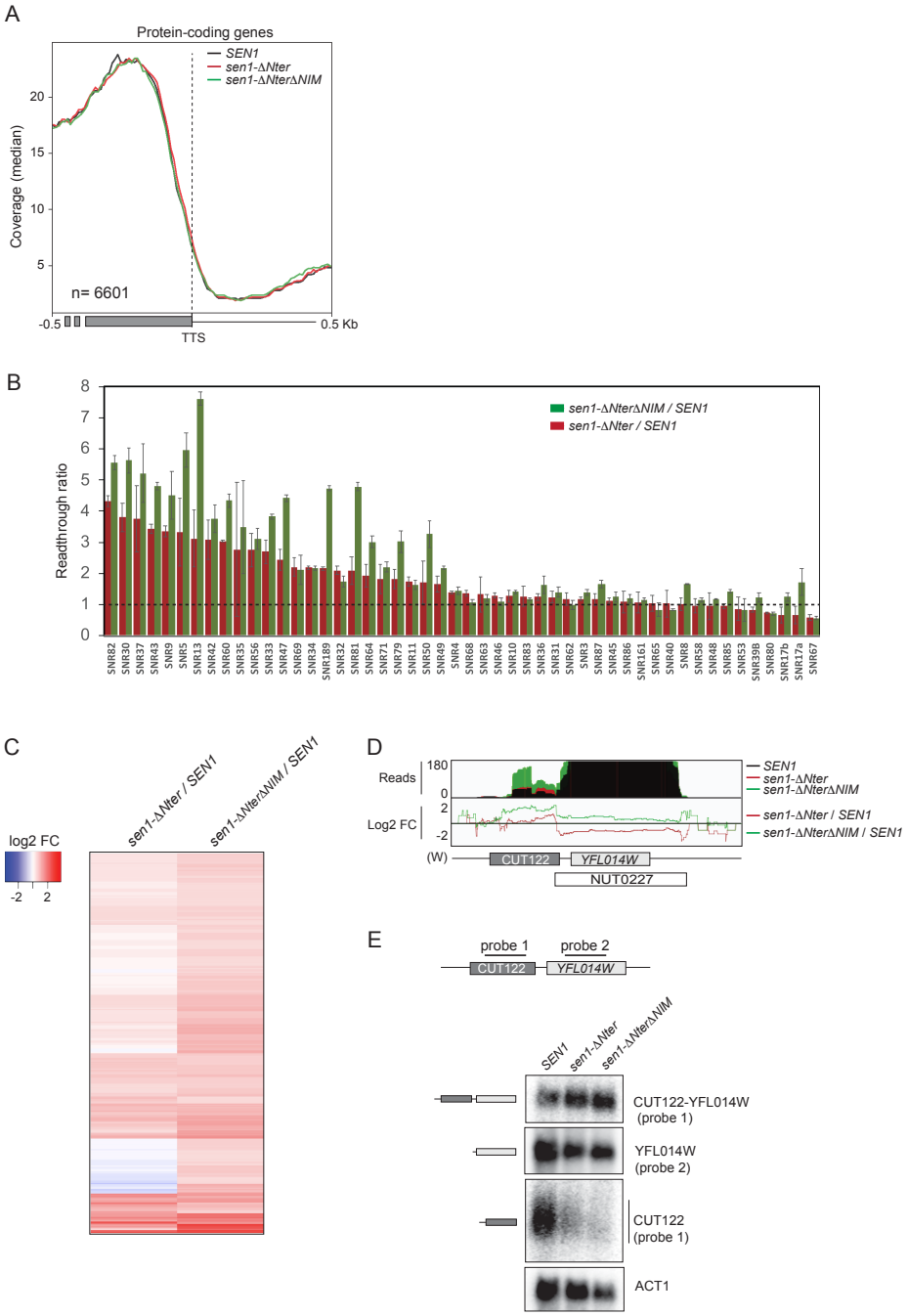


Expanded view figure 3 (related to figure 3): A) Northern blot analyses of two of the examples shown in figure 3 that exhibit termination defects in the *sen1ΔNIM* mutant. The expected RNA species resulting from inefficient termination (RT for readthrough species) are indicated. Experiments performed in a *Δrrp6* background. The U4 RNA is used as a loading control. **B)** Deletion of Sen1 C-terminal domain completely abolishes the interaction of Sen1 with Nrd1. **Top:** scheme of proteins analysed in these experiments**. Bottom:** CoIP assays using Nrd1-TAP as the bait. Representative gel of one out of two independent experiments. Antibodies used for protein detection are detailed in table S8. **C)** Deletion of Sen1 Cter provokes minor transcription termination defects at typical NNS-dependent non-coding genes. Northern blot assays performed in a in a *Δrrp6* background. Results correspond to one out of

two independent biological replicates.The *CAR2* and *PHO5* RNAs are detected as a loading controls. Probes used for RNA detection are described in table S11.



Expanded view figure 4 (related to figure 4): A) Growth tests performed as in figure 4A but using strains harbouring the TAP-tagged versions of *SEN1* indicated on the left at the endogenous locus in the presence of a *URA3* plasmid (pFL38) expressing *SEN1*. **B**) Description of the degron system employed in this study to deplete Sen1. The *SEN1* endogenous locus is modified by inserting the sequence of an AID (auxin inducible degron) tag followed by a cassette for the expression of *Oryza sativa TIR1* (*OsTIR1*), an ubiquitin ligase. In the presence of auxin hormones such as indole-3-acetic acid (IAA), Tir1 recognizes the AID-tag and promotes polyubiquitylation and subsequent degradation of Sen1-AID by the proteasome. A growth test illustrating the inhibition of cell growth upon depletion of Sen1 is shown on the right. **C**) Western blot analysis of HA-tagged versions of Sen1 expressed from a centromeric plasmid in a Sen1-AID strain upon depletion of the chromosomally-encoded copy of Sen1 as in figure 4B. Tubulin is detected as a loading control.



Expanded view figure 5 (related to figure 5): A) Metagene analyses of RNAseq experiments described in figure 5 comparing the signal distribution in the backgrounds indicated upstream and downstream of the TTS of protein-coding genes. **B**) Quantification of termination defects at snoRNAs in the *sen1ΔNterΔNIM* and the *sen1ΔNter* mutants. The readthrough ratio is calculated as the RNA signal over a window of 300 bp downstream of the mature snoRNA in the corresponding mutant divided by the signal in the wt. Bars represent the average of two independent biological replicates. Error bars denote the standard error of the mean. **C**) Heatmap visualization of the changes in expression of protein-coding genes in the different

mutants relative to the wt measured as the log2 FC of the total amount of normalized reads mapping at each gene (see methods for details). The set of genes analysed correspond to those that are upregulated in the *sen1ΔNterΔNIM* mutant (see table S5). Note that several of those genes appear as unchanged or even downregulated in the *sen1ΔNter* mutant. **D**) Screenshot of one example illustrating how differences in the efficiency of transcription termination at CUTs (or snoRNAs) in each mutant can result in apparent differences in deregulation of neighbouring genes. **E**) Northern blot analysis of example shown in **D**). Transcription from the *YFL014W* promoter is decreased as a consequence of inefficient termination at the upstream CUT (CUT122) in the *sen1* mutants. However, high levels of readthrough species (dicistronic CUT122- *YFL014W*) are generated and would contribute to the reads that map to the *YFL014W* locus in RNAseq experiments. The actual expression levels of the *YFL014W* gene in the different mutants cannot, therefore be accurately quantified by RNAseq. The *ACT1* transcript is used as a loading control. Note that the lower amount of *sen1ΔNterΔNIM* RNA loaded likely masks the higher levels of readthrough species that are produced in this mutant.