



HAL
open science

Characterization of the mechanisms of transcription termination by the helicase Sen1

Zhong Han

► **To cite this version:**

Zhong Han. Characterization of the mechanisms of transcription termination by the helicase Sen1. Molecular biology. Université Paris Saclay (COMUE), 2017. English. NNT : 2017SACLS202 . tel-01596139

HAL Id: tel-01596139

<https://theses.hal.science/tel-01596139>

Submitted on 27 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2017SACLS202

THÈSE DE DOCTORAT
DE
L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE À
L'UNIVERSITÉ PARIS-SUD

ECOLE DOCTORALE N° 577
Structure et dynamique des systèmes vivants (SDSV)

Spécialité de doctorat : Sciences de la vie et de la santé

Par

M. Zhong HAN

Characterization of the mechanisms of transcription termination
by the helicase Sen1

Thèse présentée et soutenue à Paris, le 11 septembre 2017 :

Composition du Jury :

M. Herman VAN TILBEURGH	Directeur de recherche (Institut de Biologie Intégrative de la Cellule)	Président
M. Domenico LIBRI	Directeur de recherche (Institut Jacques Monod)	CoDirecteur
Mme Odil PORRUA	Chargée de recherche (Institut Jacques Monod)	CoDirectrice
M. Alain JACQUIER	Directeur de recherche (Institut Pasteur)	Rapporteur
M. N.Kyle TANNER	Directeur de recherche (Institut de Biologie Physico-Chimique)	Rapporteur
M. Marc BOUDVILLAIN	Directeur de recherche (Centre de biophysique moléculaire)	Examineur
M. Hervé LE HIR	Directeur de recherche (Institut de Biologie de l'École Normale Supérieure)	Examineur

To my tutor Odil Porrua and team leader Domenico Libri

To my family and friends

Summary

Pervasive transcription is a common phenomenon both in eukaryotes and prokaryotes that consists in the massive production of non-coding RNAs from non-annotated regions of the genome. Pervasive transcription poses a risk that needs to be controlled since it can interfere with normal transcription of canonical genes. In *S.cerevisiae*, the helicase Sen1 plays a key role in restricting pervasive transcription by eliciting early termination of non-coding transcription. Sen1 is highly conserved across species and mutations in the human Sen1 orthologue, senataxin (SETX), are associated with two neurological disorders. Despite the major biological relevance of Sen1 proteins, little is known about their biochemical properties and precise mechanisms of action. During my PhD I have studied in detail the mechanisms of termination by Sen1.

In a first project, I have characterized the biochemical activities of Sen1 and investigated how these activities partake in termination. To this end I have employed a variety of *in vitro* approaches, including a minimal transcription-termination system containing only purified Sen1, RNA polymerase II (RNAPII) and DNA transcription templates that allows modifying the different elements of the system in a controlled manner to understand their role in termination. First, I have analysed the function of the different domains of Sen1 in termination. Sen1 is a large protein composed of a central catalytic domain flanked by additional domains with proposed roles in protein-protein interactions. I have demonstrated that the central helicase domain is sufficient to elicit transcription termination *in vitro*. Next, I have shown that Sen1 can translocate along single-stranded nucleic acids (both RNA and DNA) from 5' to 3'. Then, I have analysed the role of the different nucleic acid components of the elongation complex (i.e. nascent RNA and DNA transcription templates) in termination. My results indicate that termination does not involve the interaction of Sen1 with the DNA but requires Sen1 translocation on the nascent RNA towards the RNAPII. Importantly, I have shown that upon encountering RNAPII, Sen1 can apply

a mechanical force on the polymerase that results in transcription termination when RNAPII is paused under certain conditions. This indicates that RNAPII pausing is a strict requirement for Sen1-mediated termination.

In a second project, in collaboration with the group of E. Conti we have performed a structure-function analysis of the helicase domain of Sen1. Comparison of Sen1 structure with that of other related helicases has revealed an overall similar organization consisting in two tandem RecA-like domains from which additional accessory subdomains protrude. In general, the core RecA-like domains are very well conserved among related helicases and most variation is found in the accessory subdomains, which often confer specific characteristics to different helicases. Indeed, we have found that Sen1 contains a unique but evolutionary conserved structural feature that we have dubbed the “brace”. In addition, Sen1 is different from other helicases in an auxiliary subdomain that we have named the “prong”. Importantly, we have shown that the integrity of this subdomain is critical transcription termination by Sen1. We propose that the specific features identified in our structural analyses are important determinants of the transcription termination activity of Sen1.

Finally, we have used Sen1 as a model to investigate the molecular effect of SETX mutations linked to neurodegenerative diseases. We have introduced a set of disease-associated mutations in Sen1 and performed a complete biochemical characterization of the different mutants *in vitro*. Importantly, we found that all mutants were severely affected in transcription termination. Taken together, our results elucidate the key structural determinants of the function of Sen1 and shed light on the molecular origin of the diseases associated with SETX mutations.

Résumé

La transcription cachée est un phénomène répandu aussi bien chez les eucaryotes que chez les procaryotes. Elle se caractérise par une production massive d'ARNs non-codants au niveau de régions non-annotées du génome et est potentiellement dangereuse pour la cellule car elle peut interférer avec l'expression normale des gènes. Chez *S. cerevisiae*, l'hélicase Sen1 induit la terminaison précoce de la transcription non-codante et joue ainsi un rôle clé dans le contrôle de la transcription cachée. Sen1 est très conservée et des mutations dans son homologue humain, senataxin (SETX), ont été associées à des maladies neurodégénératives. Malgré de nombreuses recherches menées sur ces protéines, leurs propriétés biochimiques ainsi que leurs mécanismes d'action restent peu connus. Durant ma thèse, j'ai étudié le mécanisme de terminaison par Sen1.

Premièrement, j'ai caractérisé les activités biochimiques de Sen1 et analysé comment elles permettent d'induire la terminaison. Pour cela, j'ai utilisé un ensemble de techniques *in vitro*, notamment un système de transcription-terminaison qui contient uniquement des composants purifiés : Sen1, l'ARN polymérase II (Pol II) et les ADN matrices. Ce système permet de modifier les différents éléments de façon contrôlée afin de comprendre leur rôle précis dans la terminaison. J'ai tout d'abord analysé la fonction des différents domaines de Sen1 dans la terminaison. Sen1 est une protéine de taille importante qui possède un domaine central catalytique flanqué par deux domaines impliqués dans l'interaction avec d'autres facteurs. J'ai montré que le domaine hélicase est suffisant pour déclencher la terminaison de la transcription *in vitro*. Ensuite, j'ai montré que Sen1 utilise l'énergie de l'hydrolyse de l'ATP pour se déplacer sur des acides nucléiques simple bras (ARN et ADN) dans le sens 5' vers 3'. J'ai alors étudié le rôle des différents acides nucléiques du système dans la terminaison par Sen1 et j'ai montré que l'interaction de Sen1 avec l'ADN n'est pas nécessaire; en revanche Sen1 doit s'associer à l'ARN naissant et se déplacer vers la

polymérase. J'ai aussi montré qu'une fois que Sen1 entre en collision avec la Pol II, elle y exerce une action mécanique qui conduit à la terminaison uniquement quand la Pol II marque une pause. Cela indique que la terminaison est fortement dépendante de la pause transcriptionnelle.

Deuxièmement, en collaboration avec le groupe d'E. Conti, nous avons réalisé une analyse structure-fonction du domaine hélicase de Sen1. Nous avons observé que Sen1 présente une organisation similaire à celle d'autres hélicases proches avec un core composé de deux domaines de type RecA avec plusieurs domaines auxiliaires. En général, le core est très conservé au sein des hélicases proches, alors que les domaines accessoires ont des caractéristiques distinctes qui confèrent des propriétés spécifiques aux différentes hélicases. En effet, nous avons identifié un sous-domaine spécifique à Sen1 mais conservé au cours de l'évolution que nous avons appelé le "brace". Nous avons également détecté des différences notables au niveau d'un autre domaine accessoire que nous avons nommé le "prong". Nous avons pu montrer que le "prong" est essentiel pour la terminaison par Sen1. Nos données suggèrent que les caractéristiques structurales spécifiques de Sen1 que nous avons révélées sont des déterminants majeurs de son activité dans la terminaison de la transcription.

Finalement, nous avons utilisé Sen1 comme modèle pour étudier des mutations dans SETX qui sont associées à des maladies neurodégénératives. Nous avons introduit chez Sen1 une partie des mutations liées à des maladies et nous avons réalisé une caractérisation biochimique complète de chaque mutant. Nous avons ainsi montré que toutes les mutations sont fortement délétères pour la terminaison de la transcription. En conclusion, nos résultats ont permis d'améliorer la compréhension de l'origine des maladies provoquées par des mutations dans SETX.

TABLE OF CONTENTS

INTRODUCTION.....	1
1 TRANSCRIPTION.....	3
1.1 INTRODUCTION OF TRANSCRIPTION.....	3
1.2 THE RNAP II TRANSCRIPTION CYCLE.....	5
1.2.1 Initiation.....	5
1.2.2 Elongation.....	7
1.2.2.1 The role of the CTD of RNAPII.....	7
1.2.2.2 Elongation factors.....	10
1.2.3 Termination.....	13
1.2.3.1 Termination at protein-coding genes: the CPF-CF pathway.....	13
1.2.3.2 Termination at non-coding genes: the NNS-dependent pathway.....	15
1.3 PERVASIVE TRANSCRIPTION.....	19
1.3.1 Occurrence and significance of pervasive transcription.....	19
1.3.2 Control of pervasive transcription by the NNS pathway.....	21
1.4 STRUCTURAL BASIS OF TRANSCRIPTION.....	23
1.4.1 Structure of RNAP II and elongation complex.....	23
1.4.2 Nucleotide addition cycle.....	25
1.4.3 Pausing, proofreading and backtracking.....	27
2 HELICASES.....	30
2.1 CLASSIFICATION.....	30
2.2 SUPERFAMILY 1 HELICASES.....	34
2.2.1 Domain architecture.....	34
2.2.2 NTP hydrolysis and nucleic acid binding.....	35
2.2.3 Translocation and translocation polarity.....	36
2.2.4 Coupling of translocation and strand separation.....	37
2.2.5 Modulation of the helicase activity and function.....	38

3	HELICASES AS TRANSCRIPTION TERMINATION FACTORS....	40
3.1	RHO.....	40
3.2	MFD (MUTATION FREQUENCY DECLINE)	46
3.3	SEN1	51
	OBJECTIVES.....	57
	RESULTS.....	61
	DISCUSSION	137
1	THE MECHANISM OF SEN1-DEPENDENT TRANSCRIPTION	
	TERMINATION	139
1.1	SEN1 HELICASE DOMAIN IS SUFFICIENT FOR TRANSCRIPTION TERMINATION <i>IN VITRO</i>	139
1.2	DISTINCTIVE STRUCTURAL FEATURES OF SEN1 HELICASE DOMAIN INVOLVED IN TRANSCRIPTION TERMINATION	140
1.3	FEATURES OF THE ELONGATION COMPLEX WITH A POTENTIAL IMPACT ON SEN1-MEDIATED TERMINATION	142
2	REGULATION OF SEN1 FUNCTION.....	144
3	IMPLICATIONS FOR THE FUNCTION OF SENATAXIN	147
	BIBLIOGRAPHY.....	151
	RESUMÉ EN FRANÇAIS.....	179

INTRODUCTION

1 TRANSCRIPTION

1.1 Introduction of transcription

Genes encode proteins and proteins dictate cellular functions. The first step in gene expression is the transfer of the genetic information encoded in the DNA into RNA molecules. Although in a different chemical form, the genetic instructions are still written in the same sort of code as they are in the DNA, hence the name transcription.

Depending on the functions of RNA transcripts, they can be divided into protein-coding RNA and non-coding RNA. Protein-coding RNAs are also called messenger RNAs because they convey genetic information from DNA to protein products. The majority of genes carried in the DNA are transcribed into mRNAs; however mRNAs comprise only 1-5% of total cellular RNA. Most of the RNA in cells is non-coding RNA. Some of them accomplish important cellular functions, while many others do not have any assigned role. For example, ribosomal RNAs (rRNAs) are involved in forming ribosomes and catalyzing protein synthesis (1). Transfer RNAs (tRNAs) play important roles in protein synthesis as adaptors between mRNA and amino acids (2). Small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs) function in processing of pre-messenger RNA and rRNAs respectively (3). MicroRNAs (miRNAs) and small interfering RNAs (siRNAs) play important functions in RNA silencing and post-transcriptional regulation of gene expression (4). Cryptic unstable transcripts (CUTs) from yeast and promoter upstream transcripts (PROMPTs) from human, on the other hand, are rapidly degraded in wild-type cells and are considered to be non-functional (5).

The enzymes responsible for transcription of DNA into RNA are called RNA polymerases. In bacteria there is one RNA polymerase (RNAP), which catalyzes the synthesis of both coding and non-coding RNAs. In contrast, eukaryotes possess three RNA polymerases, which are structurally and mechanistically similar and share a

common set of subunits, but they also harbor specific subunits (Table 1) (6) and are in charge of producing different sets of RNAs (Table 2).

Table 1 Subunit composition of different eukaryotic RNA polymerases

RNAP I	RNAP II	RNAP III
A190	Rpb1	C160
A135	Rpb2	C128
AC40	Rpb3	AC40
Rpb5	Rpb5	Rpb5
Rpb6	Rpb6	Rpb6
Rpb8	Rpb8	Rpb8
A12.2a	Rpb9	C11a
Rpb10	Rpb10	Rpb10
AC19	Rpb11	AC19
Rpb12	Rpb12	Rpb12
A14	Rpb4	C17
A43	Rpb7	C25
A49		C37c
A34.5		C53c
		C31
		C34d
		C82d

Adapted from (6).

Table 2 Functions of different RNA polymerases in eukaryotic cells

Type of polymerase	Genes transcribed
RNAP I	5.8S, 18S, 28S rRNA genes
RNAP II	Protein-coding genes, a variety of ncRNAs, snoRNA genes and some snRNA genes
RNAP III	tRNA genes, 5S rRNA genes, some snRNA genes

Because my thesis work exclusively concerns RNAP II transcription, in the next sections I will mainly focus on describing the relevant aspects of RNAP II transcription in eukaryotes.

1.2 The RNAP II transcription cycle

Transcription of genes is a highly complicated and tightly regulated process. The cycle of RNAP II transcription can be divided into three stages: initiation, when polymerase is recruited to a promoter and begins to synthesize RNA, elongation, when polymerase adds nucleotides to the growing RNA chain and termination when the polymerase and the RNA product are released from the DNA template.

1.2.1 Initiation

Initiation involves promoter binding, DNA opening and initial RNA synthesis and is finished when the polymerase escapes from the promoter and forms a stable elongation complex (Figure 1) (7).

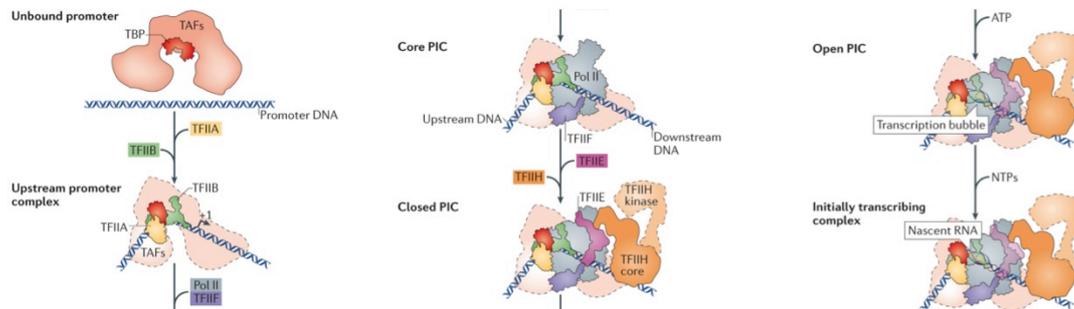


Figure 1: Schematic of RNAP II transcription initiation

Canonical model for stepwise assembly of a pre-initiation complex (PIC) on promoter DNA by RNAP II (grey) and general transcription factors (various colors). Adapted from (8).

The first step of a canonical promoter binding process is the binding of the TATA box binding protein (TBP) to the consensus sequence TATA(A/T)A(A/T)(A/G) (TATA box) in the promoter region (9). TBP and TBP-associated factors (TAFs) make up TFIID, which has been shown to be involved in expression of most RNAP II-transcribed genes in yeast (10). TBP has a saddle-shaped structure so that, upon binding the TATA box, it induces a 90-degree bend in the DNA (11, 12). Surprisingly, only around 20% of genes in yeast possess a TATA box (13), but TBP is somehow

located genome-wide on most yeast promoters (14). Only few base-specific interactions between TBP and DNA are detected; instead the hydrophobic surface formed by A/T-rich DNA is important for binding (12). In fact, most of the TATA-less promoters possess a sequence containing two or less mismatches to the consensus TATA box, referred to as TATA-like elements (14). Analyses of promoter sequences resulted in additional promoter elements that can be recognized by TAFs (8). For example, the initiator element, which overlaps TSS, can be recognized by TAF1-TAF2 (15).

The presence of the auxiliary factor TFIIA can stabilize the TBP-DNA complex, albeit it is not necessary for constitutive transcription. On the contrary, TFIIB is essential for RNAP II recruitment and improves TBP-DNA association. The C-terminal domain of TFIIB recognizes sequences flanking the TATA box and binds TBP, thus facilitating TBP binding to the TATA box (16). The N-terminal domain of TFIIB, on the other hand, can interact with RNAP II and recruit it (17). Approximately half of the RNAP II is associated with TFIIF in yeast (18). Similar to TFIIB, TFIIF can interact with the flanking regions of the TATA box. It can also bind to TFIIB, stabilizing its association with the promoter and therefore facilitates the recruitment of the polymerase to the promoter (19). Subsequently, TFIIE binds to the polymerase (20) and promotes the recruitment of TFIIH to the PIC via its C-terminal region that strongly interacts with TFIIH (21). TFIIH is composed of 10 subunits and mediates the ATP-dependent opening of the PIC. Subunit ssl2 functions as an ATP-dependent, double-strand DNA translocase. It translocates along the downstream DNA and threads DNA into RNAP II (22), causing strain that aids DNA melting. TFIIB also facilitates this process by binding the melted DNA (23) and stabilizing the transcription bubble (24).

After the formation of the open complex, the first ribonucleotide is brought to the active site to initiate polymerization without any primer. The process of initial synthesis can be divided into 3 stages depending on whether the length of the

transcript is 5, 10 or 25 nucleotides (nt). Transcripts of less than 5 nt are prone to dissociate and release, leading to frequent abortive initiation. After transcribing the first 5 nt, the transcript enters the space occupied by TFIIB, which may result in a partial displacement of it. When the RNA is around 10-nt long, it can form a much stronger hybrid with the template strand and the extensive transcript contacts with the polymerase lead to stabilization of the elongation complex. The separation of 5' RNA from the template then favors promoter escape (25). At around 25 nt, transcription initiation is achieved and productive elongation commences (8, 26).

1.2.2 Elongation

After escaping the PIC, RNAP II enters the phase of elongation. For many years, studies of transcription regulation have been focused on transcription initiation, and elongation has been considered as the trivial addition of ribonucleoside triphosphates to the growing RNA chain. However, it has become evident that transcription elongation by RNAP II is a highly regulated process (27). Elongation can be divided into two stages: early elongation and productive elongation. Regulation happens both during early steps of elongation and after RNAP II is released to enter a phase of productive elongation.

1.2.2.1 The role of the CTD of RNAPII

The C-terminal domain (CTD) extends from the largest subunit of RNAP II (Rpb1) as a long, repetitive and largely unstructured polypeptide chain. The CTD is necessary for the regulation of multiple transcription steps and transcription-associated processes, though it is not required for RNAP II catalytic activity (28, 29). This domain is composed of repeats of the heptapeptide YSPTSPS. There are 26 repeats in budding yeast and 52 in humans (Figure 2) (28).

The CTD is subject to a plethora of post-translational modifications, including

peptidyl prolyl isomerization (Pro), glycosylation (Ser and Thr), and phosphorylation. So far the most studied CTD modification is phosphorylation. Generally phosphorylation of the CTD modulates the interaction with different factors, thus enabling the recruitment of specific proteins to the polymerase during different steps of transcription.

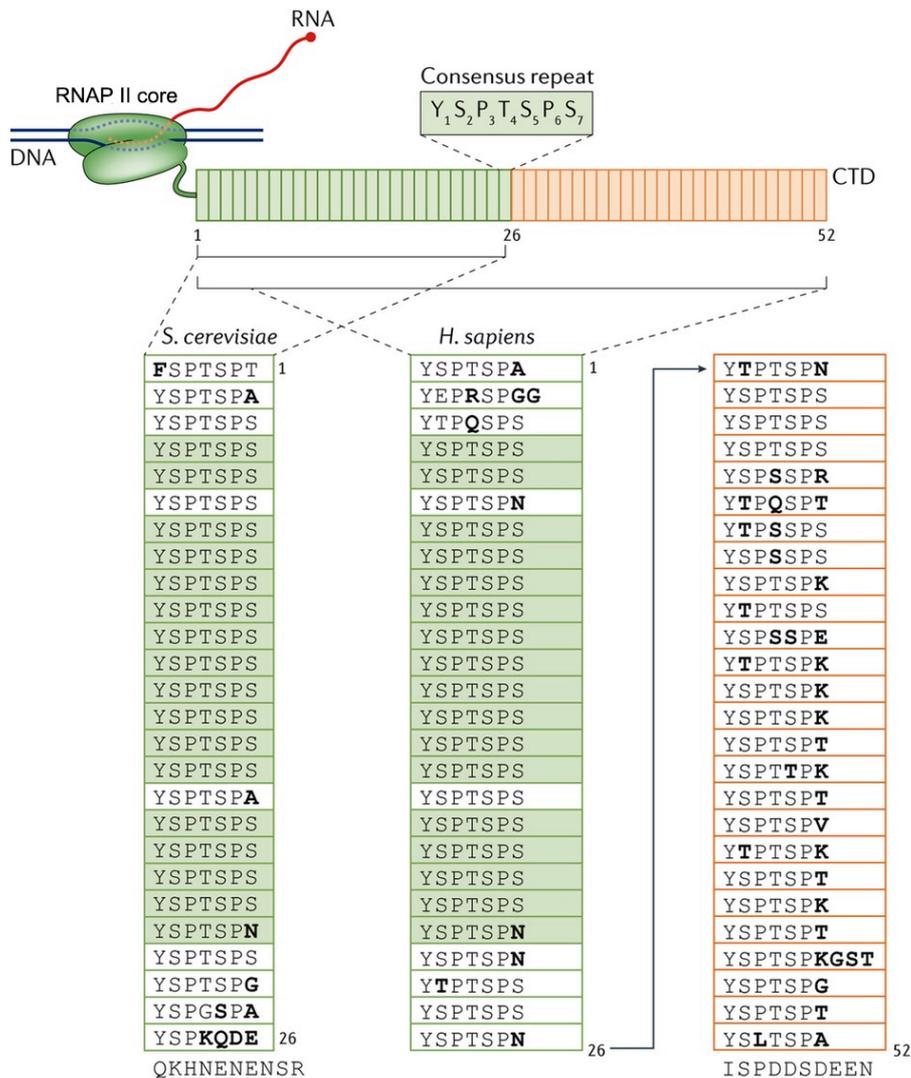


Figure 2: The composition of the CTD

Comparison of the CTD in budding yeast and human. Each rectangle represents one CTD repeat. Repeats conserved in yeast and human are highlighted in green. Adapted from (28)

RNAP II is recruited to promoters with an unphosphorylated CTD (30). Phosphorylation of the Ser5 of the CTD by TFIIF aids promoter escaping (31) and

favors the recruitment of the RNA capping complex (32) during early elongation (before 25 nt). During the transition between early and productive elongation, it exists a key regulatory event in some metazoans called promoter proximal pausing, whereby RNAP II pauses at 30-60 nt downstream of the transcription start site (TSS). Release of the polymerase from this stage into productive transcription requires CTD Ser2 phosphorylation (33). The Ser2 phosphorylated CTD also promotes the recruitment of transcription elongation factors, such as the PAF1 complex (34) and Spt6 (35), ensuring an efficient transition to productive elongation, and transcription termination factors (see next section).

The phosphorylation pattern of the CTD in yeast has been extensively characterized (Figure 3) (36). The first and most studied modifications are the phosphorylation of Ser5 and Ser2. Several studies have revealed that the Ser5 phosphorylation (Ser5P) peaks near the TSS and gradually decreases along the gene, while Ser2P accumulates across the gene and decreases after the polyadenylation site (PAS) (37, 38), consistent with their distinct roles in the transcription cycle. The Thr4P level increases across the gene and peaks downstream of the PAS. The Thr4P CTD has been shown to interact with Rtt103, a protein involved in transcription termination (see the next section), and it has been proposed that this mark regulates the process of transcription termination and cleavage and polyadenylation (28, 39-41). The Tyr1P level also rises downstream of the TSS, but it decreases sharply before the PAS. Importantly, it has been shown that the presence of Tyr1P prevents binding of termination factors (e.g. Nrd1, Rtt103 and Pcf11) to the CTD and therefore defines the window of termination (42-44).

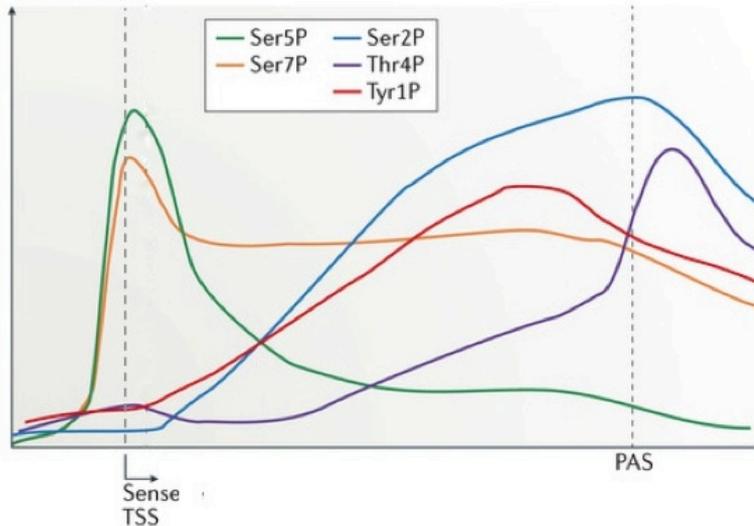


Figure 3: CTD code of *S. cerevisiae*

The phosphorylation pattern of the CTD across protein-coding genes in budding yeast. See the main text for details. Adapted from (28)

1.2.2.2 Elongation factors

Elongation factors can be defined as any proteins that interact with RNAP II to increase (positive elongation factor) or reduce (negative elongation factor) the rate of transcription elongation. They play fundamental roles in the regulation of gene transcription. There are numerous elongation factors belonging to different classes. Some factors can modulate RNAPII catalytic activity, others can help RNAP II transcribe through chromatin (27, 45). I will describe a few examples to gain some insights into the mechanisms of action of elongation factors.

Spt5 (NusG in bacteria) is the only elongation factor that is universally conserved in all domains of life. It binds to Spt4 in eukaryotes, forming the Spt4-Spt5 complex (also named DSIF or DBR sensitivity factor in metazoans). Genome-wide chromatin immunoprecipitation (ChIP) data show that the distribution of Spt4-Spt5 complex along genes mostly mirrors that of RNAP II. The Spt4-Spt5 complex can enhance both the translocation rate and the processivity of the elongating RNAP II *in vivo* (46). Structural and biochemical data (47, 48) show that Spt5 binds to RNAP II in a way

that it can stabilize the association of Rpb4/7 with the core RNAP II (see the structure of RNAP II in following sections), thereby stabilizing RNAP II in the closed conformation and enclosing the DNA in the central cleft of the polymerase to increase transcription processivity (Figure 4).

TFIIS is another well-studied elongation factor that is able to reactivate arrested RNAP II. During gene transcription, RNAP II can move backwards, at certain sequences, causing misalignment of the RNA 3'-OH with the polymerase active site and leading to transcription arrest (49, 50). This phenomenon is called backtracking. In most instances, a backtracked polymerase resumes productive elongation via an evolutionarily conserved mechanism that requires the cleavage of the extruded RNA 3' end, and this process can barely be accomplished by RNAP II itself. TFIIS can stimulate this process by extending from the surface of RNAP II through a pore into the internal active center, positioning a metal ion and a water molecule to enhance the intrinsic nuclease activity of RNAP II (51).

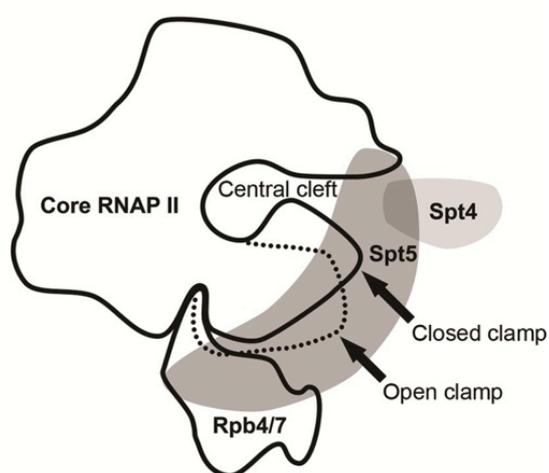


Figure 4: Schematic of the architecture of an elongation complex.

The dashed line indicates the open clamp position observed in the absence of Rpb4/7. See the main text for details. Adapted from (47).

The FACT complex (Facilitates Chromatin Transcription) regulates transcription by promoting RNAP II elongation through nucleosomes. This complex functions as a histone chaperone that facilitates transcription by removing the H2A-H2B dimer from

nucleosomes and reassembling nucleosomes after RNAP II passage (52).

The P-TEFb (positive transcription elongation factor b) complex plays a central role in releasing RNAP II from promoter proximal pausing. This regulatory activity depends on the kinase activity of its subunit CDK9 (Bur1 in yeast) (33). The P-TEFb complex mediates phosphorylation of the CTD Ser2, as well as the negative elongation factor (NELF) and DSIF (the Spt4-Spt5 complex). Phosphorylation of NELF provokes its eviction from RNAP II, whereas phosphorylation of DSIF turns it from a negative to a positive elongation factor. There are many similar elongation factors that can modify the CTD phosphorylation pattern, thereby regulating transcription (**Figure 5**).

Modification	CTD decorators			
	Enzyme	Mammals	<i>S. cerevisiae</i>	<i>S. pombe</i>
Tyr1 phosphorylation	Kinases	c-Abl?		
	Phosphatases		Rtr1	Glc7
Ser2 phosphorylation	Kinases	Cdk9, Cdk11, Cdk12, Cdk13, Brd4, DYRK1A	Bur1, Ctk1	Cdk9, Lsk1
	Phosphatases	Fcp1, Cdc14	Fcp1	Fcp1
Pro3 isomerization		PIN1	Ess1	
Thr4 phosphorylation	Kinases	PLK3, CDK9		
	Phosphatases	Fcp1	Fcp1	
Ser5 phosphorylation	Kinases	Cdk7, Cdk8, Cdk9, Cdk12, Cdk13, DYRK1A	Kin28, Cdk8 (Srb10)	Mcs6, Cdk8, Cdk9
	Phosphatases	Ssu72, RPAP2, Scp1, Scp4, Cdc14	Rtr1, Ssu72	
Pro6 isomerization		PIN1	Ess1	
Ser7 phosphorylation	Kinases	Cdk7, Cdk9	Kin28, Bur1	Mcs6?
	Phosphatases	Ssu72	Ssu72	

Figure 5: Summary of CTD modifying enzymes in mammals and yeast. Adapted from (40)

1.2.3 Termination

The final stage of transcription is termination, which results in the release of the transcript and the dissociation of the RNA polymerase from the DNA template. Transcription termination is required for the partition of the genetic information by defining the boundaries of transcription units. However, the functional significance of termination extends beyond the mere definition of gene borders. As termination factors interact with RNA processing and degradation enzymes, they are critical to determine the cellular fate and half-life of the transcript. In this section I will describe the two main pathways for termination of RNAP II-dependent transcription in yeast, which are the cleavage and polyadenylation factor -cleavage factor (CPF-CF) and the Nrd1-Nab3-Sen1 (NNS) dependent pathways.

1.2.3.1 Termination at protein-coding genes: the CPF-CF pathway

Transcription termination of protein-coding genes relies mainly on the cleavage and polyadenylation factor (CPF)-cleavage factor (CF) (Figure 6). Several components of CPF-CF complex (Rna15, Cft1, Cft2, Yth1, Mpe1 and Hrp1) can recognize specific termination and processing signals at the 3' untranslated region (UTR) of the mRNA. In addition, the Pcf11 subunit interacts with the Ser2P form of the CTD (53), which culminates at the 3' ends of protein-coding genes. Therefore, the CPF-CF complex is recruited to the 3' end of genes. Subsequently, the CPF subunit Ysh1 cleaves the nascent RNA at the poly(A) site (54) and adenosine nucleotides are added to the free hydroxyl group on the 3' end (3'OH) by the CPF-associated poly(A) polymerase Pap1. The poly(A)-binding protein (Pab1) then binds to the newly synthesized poly(A) tail, thereby protecting the transcript from 3' degradation and promoting nuclear export. RNAs produced by this pathway are rapidly exported to the cytoplasm and their half-lives are generally dictated by the cytoplasmic turnover pathways (55).

After the mRNA is cleaved from the nascent RNA, the RNAP II transcribes the downstream DNA for less than 150 nucleotides on average (56). The RNAP II is subsequently dissociated from the DNA by mechanisms that are still under debate.

Two models have been proposed for termination by this pathway (Figure 6). The allosteric model posits that after transcribing through the poly(A) site, binding of termination factors leads to a conformational change of the elongation complex, possibly accompanied by the loss of anti-termination factors, which causes termination (57). Consistent with this model, it has been shown that Pcf11 can destabilize an elongation complex *in vitro* by simultaneous binding to RNAP II CTD and to the nascent RNA (58-61).

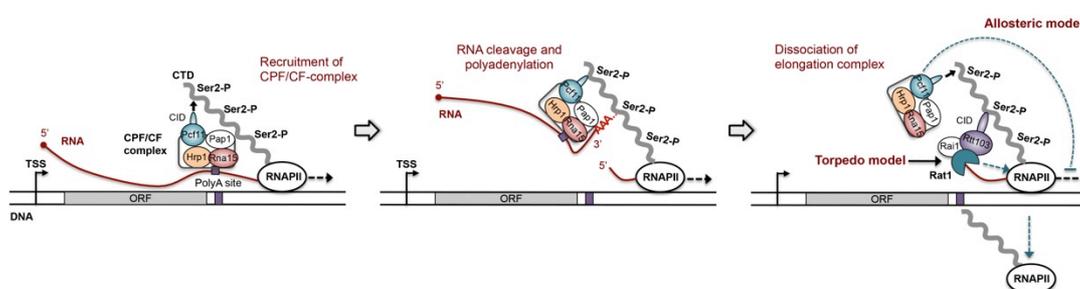


Figure 6: Transcription termination at mRNA-coding genes in yeast. See the main text for details. Adapted from (62).

The torpedo model postulates that the cleavage of the nascent RNA at 3' end of protein-coding genes by the CPF-CF complex provides an entry point for the 5'-3' exonuclease Rat1, which degrades the downstream portion of the transcript up to the transcribing RNAP II, leading to the dissociation of the elongation complex (63, 64). This would explain the coupling between cleavage and termination, and the occurrence of readthrough transcription when the 5'-3' exonuclease is defective (63, 64). In support of this model, Rat1-Rai1 has been shown to be able to dismantle stalled elongation complex *in vitro* (65, 66). Recent genome-wide experiments also shed light on the function of Rat1 (Xrn2 in humans) in termination of protein-coding genes and provided *in vivo* support for this model (67, 68). However, there is also *in vitro* data showing that termination can take place in a poly(A) signal

dependent-manner but without RNA cleavage (69), suggesting that different mechanisms of termination might coexist.

It is apparently not contradictory to combine the two models and propose a unified allosteric-torpedo one according to which the recruitment of both Pcf11 and Rat1 at sites of cleavage and termination ensures efficient and accurate transcription termination of protein-coding genes *in vivo* (70).

1.2.3.2 Termination at non-coding genes: the NNS-dependent pathway

As mentioned above, RNAP II transcribes a plethora of non-coding RNAs (ncRNAs) besides protein-coding genes. Some of them play important functional roles, such as snRNAs and snoRNAs, but most of the ncRNAs in yeast are actually non-functional. In other words, they are considered as by-products of transcription that occurs at the wrong place or in the wrong direction.

In *S. cerevisiae*, the Nrd1-Nab3-Sen1 (NNS) complex is responsible for termination of transcription of most ncRNAs, including snRNAs, snoRNAs and ncRNAs without any assigned function (e.g. cryptic unstable transcripts or CUTs, see below). The ncRNAs terminated in this pathway are normally rapidly processed or degraded.

The NNS-complex is composed of three essential proteins: the sequence-specific RNA binding proteins Nrd1 and Nab3 and the RNA and DNA helicase Sen1 (71, 72). Nrd1 and Nab3 form a tight heterodimer that recognize specific motifs that are enriched in the target ncRNAs (73-79). Nrd1 also contains an N-terminal region that interacts with the CTD of the largest subunit of RNAPII, named CID for CTD interacting domain (77, 80). Specifically, Nrd1 recognizes the Ser5-P form of the CTD, which predominates during early elongation (77, 80). This specific interaction is important for early recruitment of the NNS-complex and for efficient transcription

termination (Figure 7).

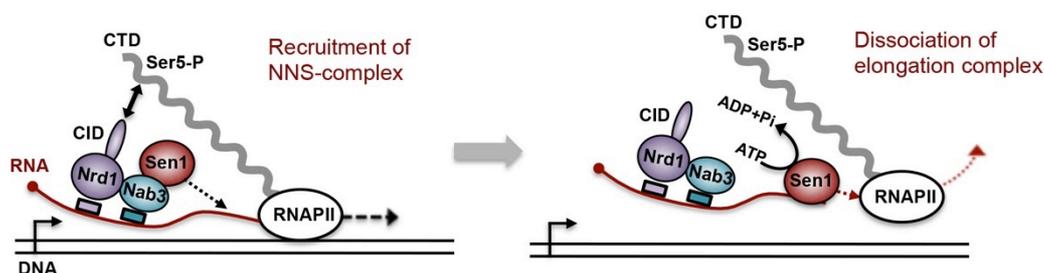


Figure 7: Overview of the NNS-dependent pathway.

The NNS-complex is recruited to the elongation complex by both the recognition of specific motifs at the ncRNA by Nrd1 and Nab3 and by the interaction of Nrd1 CID with the Ser5-P form of the CTD of RNAP II. Dissociation of the elongation complex is then mediated by the Sen1 helicase.

Nrd1 was discovered in a genetic selection in which Nrd1 acts as a trans-acting factor for nuclear pre-mRNA downregulation (NRD) (81). It was subsequently found that Nrd1 was required for termination of snRNAs and snoRNAs. Nrd1 binds specific terminator sequences (GUAA/G) through its RNA recognition motif (RRM) (78, 82, 83). Nuclear magnetic resonance (NMR) and fluorescence anisotropy data show that not only GUAA/G but also several other G-rich and AU-rich sequences are able to bind Nrd1 with good affinity (84). The Nab3 interaction domain (Nab3ID) also plays important roles in recruitment of Nrd1, consistent with the need for cooperativity between Nrd1 and Nab3 for high affinity binding to the RNA (79, 83). Nrd1 also bears a C-terminal low-complexity region enriched in glutamine and proline amino acids (Q/P rich) of unknown function (Figure 8).

Nab3 was found to be able to associate with Nrd1 in a stable heterodimer by genetic and biochemical assays (79, 85). Similar to Nrd1, Nab3 also has a conserved RNA recognition motif (RRM), which is indispensable for cell viability and binds to a consensus target of UCUUG (76, 78, 79, 85) (Figure 8). Flanking the conserved RRM, Nab3 contains two low-complexity domains: a N-terminal aspartic/glutamic acid-rich region (D/E-rich) and a C-terminal glutamine/proline-rich region (Q/P-rich). The N-terminal D/E-rich region is not required for cell viability and its function is

unknown, whereas the C-terminal Q/P-rich domain is essential and might play important roles in Nab3 self-association (86-88).

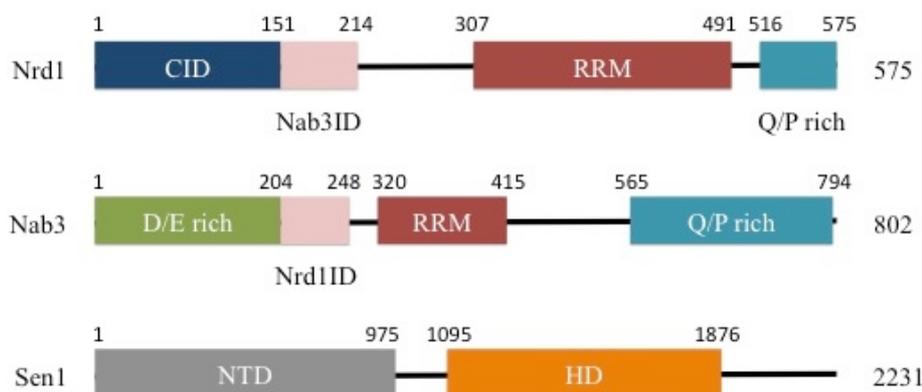


Figure 8: The domain structure of the Nrd1, Nab3, and Sen1.

The lengths of the proteins are indicated on the right (in amino acids). The boxes indicating domains are approximately to scale. The CTD interaction domain (CID), Nab3 interaction domain of Nrd1 (Nab3ID) and Nrd1 interaction domain of Nab3 (NrdIID) are defined by Vasiljeva et al (77). The RNA recognition motifs (RRM) of Nrd1 and Nab3 are defined by Bacikova et al and Hobor et al (76, 89). Q/P-rich, prion-like domains of Nrd1 and Nab3 are described by Alberti et al (90). The length of the helicase domain (HD) and N-terminal domain (NTD) of Sen1 are indicated according to Tumas and Brow (91) and DeMarini (92).

Sen1 was recognized as of importance in short transcript termination in the same screen that yielded Nrd1 (81). It is the only subunit of the NNS-complex that possesses a catalytic activity. It contains a Superfamily I helicase domain flanked by N-terminal and C-terminal extensions involved in protein-protein interactions (71) (Figure 8). Specifically, the N-terminal domain of Sen1 was proposed to mediate interaction with Rnt1, Rad52 and the largest subunit of RNAP II, Rpb1 (93, 94). The C-terminal region of Sen1 is involved in the interaction with Nab3 and Glc7 (71). Helicases are enzymes that utilize the energy from hydrolysis of ATP to remodel nucleic acid molecules or nucleo-protein complexes and are involved in most processes related to RNA and DNA metabolism. A previous *in vitro* work performed in my group has shown that Sen1 is the key protein in the transcription termination reaction. Specifically, Sen1 by itself can elicit the dissociation of an elongation complex through a mechanism involving the interaction of Sen1 with the nascent

RNA and ATP hydrolysis (95). It has been proposed that upon recruitment to the nascent transcript, Sen1 translocates along the RNA until it reaches the RNAP II and dissociates the elongation complex (95, 96). In addition, a mutation in the RNAP II that increases the elongation rate results in increased read-through at Sen1-mediated terminators, and termination defects in a Sen1 mutant can be partially suppressed by a slowly transcribing RNAP II mutant. These connections between the transcription rate and Sen1 activity support the model that Sen1 needs to translocate along the RNA to catch up with the RNAP II and terminate transcription (96).

1.3 Pervasive transcription

1.3.1 Occurrence and significance of pervasive transcription

The development of new sequencing technologies has unveiled an unexpected level of complexity in the eukaryotic and prokaryotic transcription landscapes. High-resolution techniques such as tiling arrays and, more recently, RNA-seq have revealed that a large proportion of transcripts are not associated with any annotated feature, giving rise to the concepts of “pervasive” and “hidden” transcription (5). These transcripts are often rapidly degraded, and therefore remain invisible unless RNA degradation is prevented, for example, by inactivation of the degradation machinery. In *Saccharomyces cerevisiae*, one of the main products of pervasive transcription constitutes a class of ncRNAs dubbed cryptic unstable transcripts (CUTs). CUTs are transcribed by the RNA polymerase II (RNAPII), they are capped, relatively small, with an average length of 200 to 500 bp and contain heterogeneous 3' ends (97). As mentioned above, their transcription is terminated by the NNS-complex, which also promotes their degradation by the nuclear exosome (98, 99).

Work from several laboratories has provided a detailed picture of the genomic distribution of CUTs (73, 100, 101), showing that the vast majority of these transcripts originate from nucleosome-free regions (NFRs) often correspond to promoter regions of *bona fide* genes. This strongly suggests that most yeast promoters are intrinsically bidirectional. Another abundant class of ncRNAs has been named SUTs for stable unannotated transcripts (102). Their origin is the same as for the CUTs, but they likely differ in their transcription termination mode since they are stable and often longer than CUTs. Finally, a third category of ncRNAs includes mainly antisense transcripts that are stabilized upon mutation of the major cytoplasmic 5' to 3' exoribonuclease Xrn1 (103) and have therefore been designed as

XUTs for *xrn1*-sensitive unannotated transcripts. However, there is considerable overlap between the three classes of transcripts, as their classification is based on the relative contribution of different exonucleases to their turnover rather than on functional criteria.

The biological significance of pervasive transcription is still a matter of debate. Some authors propose an evolutionary role for the production of ncRNAs in the generation of new protein-coding genes (104). In addition, in a growing number of cases, expression of an ncRNA has proved to play a role in the regulation of gene expression. This is for instance the case of several genes of amino acid and nucleotide biosynthetic pathways, genes involved in phosphate metabolism and genes involved in meiosis (105). Although the precise mechanisms of regulation are not always fully understood, ncRNAs whose transcription affects the expression of an associated gene are normally either transcribed upstream of a tandem gene or antisense to the cognate gene. In both cases, non-coding transcription that progresses through the promoter region of the downstream or antisense gene induces repression by mechanisms that involve the deposition of repressive chromatin modifications at the promoter and/or chromatin remodeling that prevents the association of activator proteins (106).

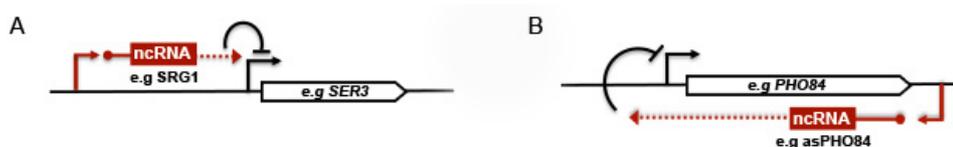


Figure 9: Schematic representation of examples in which non-coding transcription plays a role in regulation of gene expression.

(A) Transcription of a ncRNA represses expression of a downstream gene. In the example showed, non-coding transcription induces chromatin remodelling at the SER3 promoter that prevents the binding of the specific activation factors. (B) Antisense non-coding transcription represses expression of the associated gene at the transcriptional level. In the case of PHO84 and others, chromatin repressive marks are deposited at the promoter of the repressed gene.

1.3.2 Control of pervasive transcription by the NNS pathway

Despite the possible biological roles mentioned above, pervasive transcription poses a risk that is controlled at different levels. Translation of aberrant RNAs that could result in toxic proteins is restricted by the nonsense-mediated decay (NMD) pathway and the Xrn1 exonuclease that degrade the pervasive transcripts in the cytoplasm (103, 107). However, these quality control pathways act after export of the RNA to the cytoplasm and cannot preclude the possible interference of cryptic transcription with transcription of canonical genes for instance by impeding the binding of activator proteins to the promoter region of a downstream gene. At the transcriptional level the cell employs two kinds of mechanisms to limit pervasive transcription. The first one operates at the level of chromatin structure and involves the action of histone modification factors and chromatin remodelers that prevent transcription from initiating divergently at promoters and internally to ORFs (108). This mechanism is, however, only partially efficient so a second pathway plays an additional and major role in the control of pervasive transcription at a post-initiation level. This pathway relies on the NNS-complex, described above, which both elicits early transcription termination and promotes polyadenylation and degradation of the non-coding transcripts by the TRAMP and the exosome complexes, respectively (72, 97, 98).

The exosome is a conserved multisubunit complex that functions in degradation of defective transcripts as well as in processing of the 3' ends of stable ncRNAs (snRNAs, snoRNAs and the 5.8S rRNA). It possesses a cytoplasmic and a nuclear form that differ in their associated factors. The cytoplasmic form contains only one catalytic subunit (Rrp44/Dis3) endowed with both endonuclease and 3'-5' exonuclease activities while the nuclear form contains an additional 3'-5' exonuclease, Rrp6 (109-113). The TRAMP complex is considered as a cofactor of the nuclear exosome. It has two alternative forms with a common structure, containing a polyA

polymerase (either Trf4 or Trf5), a zinc-knuckle RNA binding protein (either Air1 or Air2) and the DExH-box helicase Mtr4. Both the addition of poly(A) tails by Trf4/5-Air2/1 and the unwinding activity of Mtr4 aid the exosome in degradation of structured RNA substrates (97, 113-115)

The fast removal of CUTs relies not only on the independent function of NNS, TRAMP and the exosome but also on the coordinated interactions between these complexes. Specifically, Nrd1-CID also recognizes a CTD-mimic in the Trf4 component of TRAMP, a sequence that has been dubbed NIM for Nrd1-interaction motif (116). The Nrd1-Trf4 interaction mediated by the CID and the NIM is important for the stimulation of RNA polyadenylation by the TRAMP complex *in vitro*, and for efficient RNA degradation/processing by the exosome *in vivo* (117). In addition, crystallographic analyses reveal that the N-terminus of Mtr4 binds to a composite surface groove formed by N-terminal domains of Rrp6 and Rrp47, which might be important for efficient recruitment of Rrp6-containing exosome (118-120). Similarly, another exosome cofactor Mpp6 also contains a NIM sequence, which binds Nrd1-CID and provides an additional connection between the NNS complex and the exosome (121, 122). Finally, *in vitro* data support the idea that both Nrd1 and Nab3 can interact directly with Rrp6 (117, 123).

1.4 Structural basis of transcription

1.4.1 Structure of RNAP II and elongation complex

Yeast and human RNAP II are both comprised of 12 subunits, Rpb1 to Rpb12 with a total mass of more than 500KD. It can dissociate into a 10 subunit core and a heterodimer composed of subunits Rpb4 and Rpb7 (Table 1). Structural data shows that Rpb1 and Rpb2 form the central scaffold, dubbed the “cleft”, that contains the active site (Figure 10). Subunits Rpb3, Rpb10, Rpb11 and Rpb12 form a subcomplex that bridges the two largest subunits. Rpb5, Rpb6 and Rpb8 assemble on Rpb1, and Rpb9 binds to both Rpb1 and Rpb2. The Rpb1 side of the cleft is named the “clamp”. The clamp is mobile and together with the Rpb2 side of the cleft can adopt two distinct conformations, the closed and the open conformation (124). The Rpb4/7 complex forms a wedge structure, inserts into the pocket generated by Rpb1 clamp Rpb2 and Rpb6 via the N-terminal tip of Rpb7, thereby restricting the clamp to the closed conformation (125, 126).

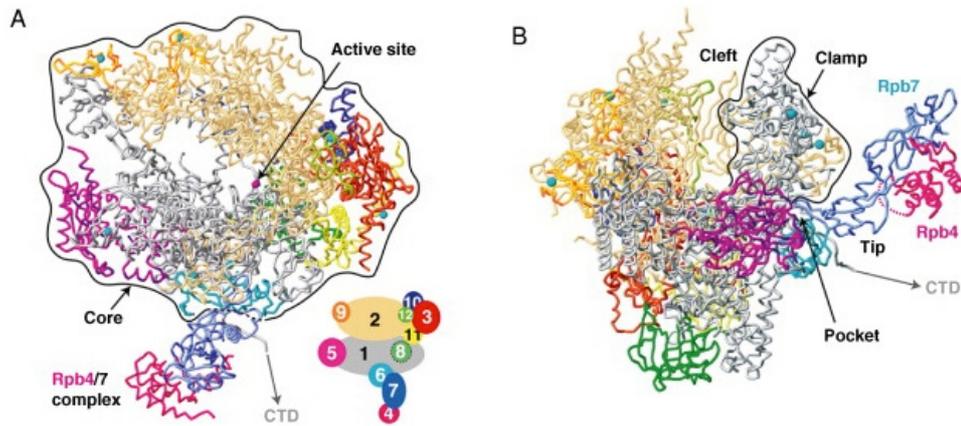


Figure 10: Two views of the complete yeast RNAP II.

The 12 subunits are shown as ribbon diagrams in different colors, as indicated in the schematic diagram. The active site manganese ion is depicted as a pink sphere. Zinc ions are shown as cyan spheres. See the main text for details. Adapted from (127).

During elongation, RNAP II associates with the DNA template and harbors an

RNA-DNA hybrid at the active site, forming the elongation complex. The catalytic center of the RNA polymerase includes the binding sites for the RNA 3'-terminus (i site) and the insertion site for the incoming NTP (i+1 site) (128). The structure of the elongation complex also exhibits a closed conformation. The downstream DNA duplex is unwound before the active center, allowing the template strand to reach the active site. The upstream DNA rewinds as the transcription bubble progresses along the DNA. Within the transcription bubble, the RNA is attached to the catalytic site with its 3' end and forms a 8-9 bp hybrid with the DNA template (**Figure 11**) (7, 127, 129). The DNA-RNA hybrid is the distinguishing feature of the elongation complex. Most interactions between RNAP II and nucleic acids occur in the region of the DNA-RNA hybrid (129, 130). The tight binding of RNAP II to DNA-RNA hybrid is one of the main determinants of the stability of the elongation complex and transcription processivity (131). In addition, extensive interactions between RNAP II and the DNA duplex regions also play an important role in stabilizing the elongation complex: Rpb1 and Rpb5 residues interact with the downstream duplex, while Rpb2 residues interact with the upstream duplex (130, 132).

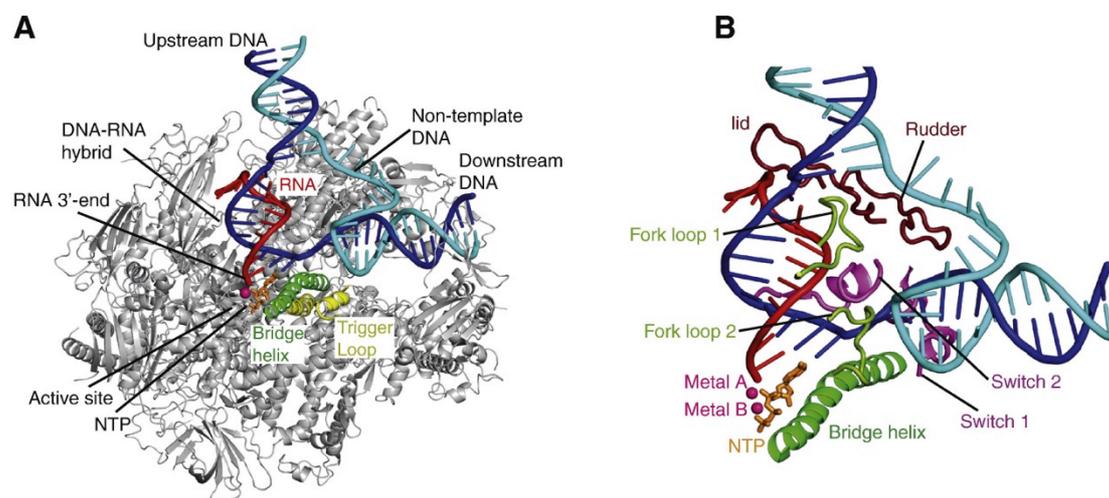


Figure 11: Structure of the RNAP II elongation complex

(A) Overview of the RNAP II elongation complex with bound NTP. The bridge helix (green), trigger loop (yellow), and metal A (pink) are highlighted. DNA template, DNA non-template, RNA, and NTP are shown in blue, cyan, red, and orange, respectively. (B) Nucleic acids and RNAP II structural elements involved in transcription bubble maintenance. Refer to the text for details. Adapted from (7)

During elongation, the maintenance of the transcription bubble requires unwinding of downstream DNA duplex and separation of the RNA product from the DNA template at the 5' end of hybrid. The positively charged residues in the switch 2 region are important for the downstream DNA unwinding by pulling the DNA template strand away from the non-template strand (133). The RNA strand separation involves 3 polymerase loops, named lid, rudder and fork loop1 (129, 134). They interact with either DNA or RNA to prevent formation of extended DNA-RNA hybrids, contributing to the maintenance of hybrid length (**Figure 11**).

1.4.2 Nucleotide addition cycle

A complete nucleotide addition cycle includes several steps: selection of the correct NTP, addition of a nucleotide to the 3' end of the RNA, release of pyrophosphate (PPi) and forward translocation of RNAP II along the DNA and the RNA to free the new nucleotide binding site (i+1 site) (Figure 12). The NTP substrate base-pairs with the DNA template at the i+1 site, located between the RNA 3'-end, the RNAP II bridge helix and the trigger loop (**Figure 11**). The RNAP II active center can be in an open or closed conformation. The NTP substrate can bind transiently to the mobile trigger loop in the open active center conformation. If the substrate is the cognate NTP, there will be a conformational change that will lead to closing of the active center and subsequent catalysis. Otherwise, binding of a non-complementary NTP keeps the active center opened, favoring the release of the incorrect NTP rather than its incorporation, therefore preventing misincorporation (135). In addition, yeast RNAP II has a 500- to 5000-fold preference for NTPs over dNTPs (136). A conserved asparagine residue in the active center has been shown to play a role in the NTP/dNTP discrimination by contacting the ribose 2'-OH group (137, 138).

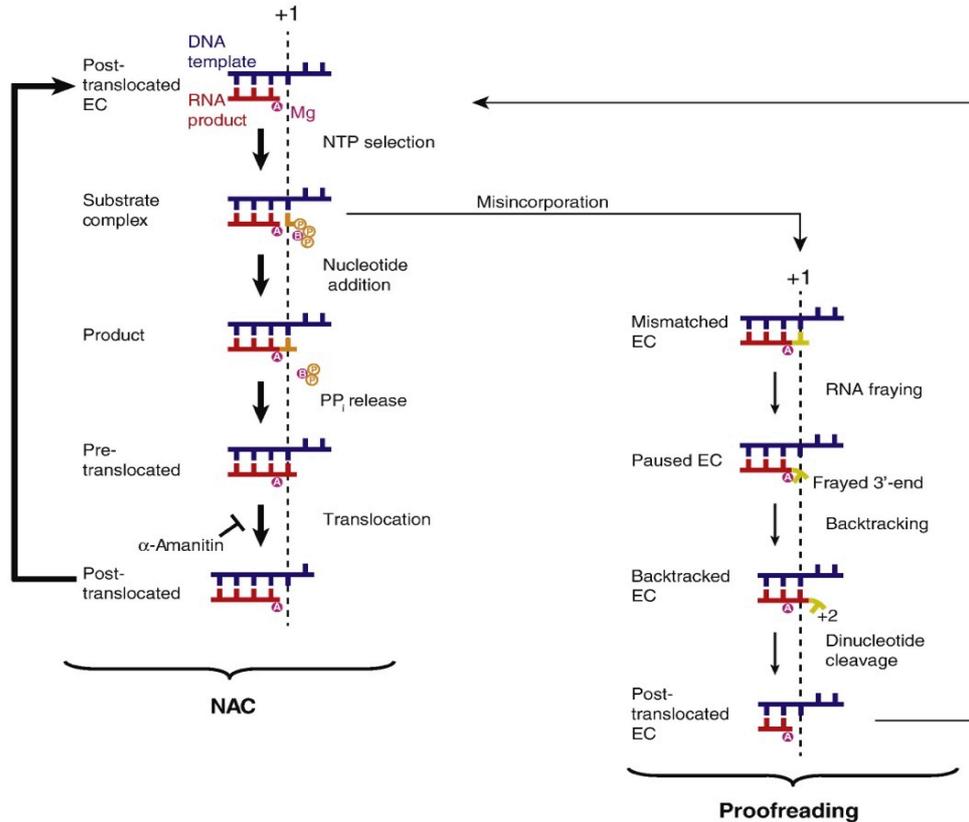


Figure 12: Schematic of functional states of the elongation complex.

Models of the nucleotide addition cycle (left) and the proofreading cycle (right). Adapted from (7)

Nucleotide incorporation catalysis follows a nucleophilic substitution mechanism, with the RNA 3'-OH group acting as a nucleophile that attacks the NTP α -phosphate. After nucleotide addition, release of the pyrophosphate is proposed to cause a trigger loop conformational change and to open the active center (139, 140). Just after addition of a nucleotide to the growing RNA chain, the polymerase is in a pre-translocation state in which the active site is occupied by the extended RNA 3'-end. In order to free the active site for binding the next NTP, the polymerase needs to translocate along the DNA and the RNA by one nucleotide to reach the post-translocation state. Alternative models have been proposed to explain the mechanism of translocation: the power-stroke model and the Brownian ratchet model (141). In the power-stroke model, the conversion of chemical energy into mechanical work by RNAP II is supplied during transfer of an NMP moiety from pyrophosphate (PPi) to the 3' end of RNA. Hence, the forward translocation is coupled by either

phosphodiester-bond formation or PPi release, with a transition state existing between the pre- and post-translocated states (142). In the Brownian ratchet model, there is no transition state with significant activation energy. The forward translocation is not synchronized to the chemical steps (143). The ratchet model postulates that during the nucleotide addition cycle, the elongation complex is in an equilibrium between the pre- and post-translocation states, and the binding of complementary incoming NTP stabilizes the post-translocation state (135, 139).

The mechanism of translocation has also been elucidated with structures of RNAP II bound by the mushroom toxin α -amanitin (136, 144). The toxin contacts the bridge helix and the trigger loop, traps the elongation complex in a new conformation that is intermediary between the pre- and the post-translocation states and thereby interfering with the translocation process.

1.4.3 Pausing, proofreading and backtracking

Various factors can induce transcriptional pausing during elongation, such as certain DNA sequences, a mismatched NTP incorporation and DNA lesions (7, 141, 145). Pausing can be divided into two stages. The first phase is referred to as the elemental pause (145). At elemental pausing, an initial rearrangement of the active site, including a conformational change of the trigger loop, interrupts elongation (146). Further rearrangements then lead to a long-lived pause or polymerase backtracking along the DNA and RNA. It was proposed that in the elemental pause step, the RNA 3'-end frays away from the DNA template due to the aforementioned rearrangement of the active center (146).

A misincorporated nucleotide can be removed by proofreading (Figure 12). After nucleotide misincorporation, the 3'-nucleotide frays away from the DNA template, inducing pausing. The polymerase then backtracks by one nucleotide, allowing the alignment of the phosphodiester bond with the catalytic site (147). The RNA 3'-end

then stimulates the intrinsic endonucleolytic activity of the polymerase, cleaving a dinucleotide containing the mismatched nucleotide (148), thereby allowing resumption of transcription. The proofreading process, together with the NTP selection as described above, accounts for the fidelity of transcription (149).

It has been shown that both bacterial RNAP and eukaryotic RNAP II undergo backtracking at some sequence-specific pausing sites (150-153). During backtracking, the polymerase moves back along the DNA and the RNA, the RNA 3'-end disengages from the active site and extrudes into the secondary pore and funnel, where NTP substrates enter into the catalytic center during active transcription (154, 155) (Figure 13), rendering the elongation complex inactive (156). Fraying of the RNA 3'-end is proposed to trigger both backtracked and non-backtracked pausing (149). The first backtracked nucleotide stacks between the $i+1$ site and the Rpb2 gating tyrosine (49, 147) (Figure 13). Backtracking by one position may thus be facilitated, but further backtracking is probably disfavored because RNA base stacking must be disrupted at the gating tyrosine. This also explains the preference of the polymerase for cleaving dinucleotides (147). However, backtracking often happens more extensively and the RNA extrudes beyond the gating tyrosine and into the secondary pore because the free energy of the elongation complex is lower at an upstream position. Backtracked RNAP II can resume transcription with the help of the transcription factor TFIIS as described before. In short, TFIIS interacts with the polymerase surface and inserts its C-terminal domain (a zinc ribbon) into the pore, reaches the active site and thereby stimulates the RNAP II intrinsic endonucleolytic cleavage activity (Figure 13).

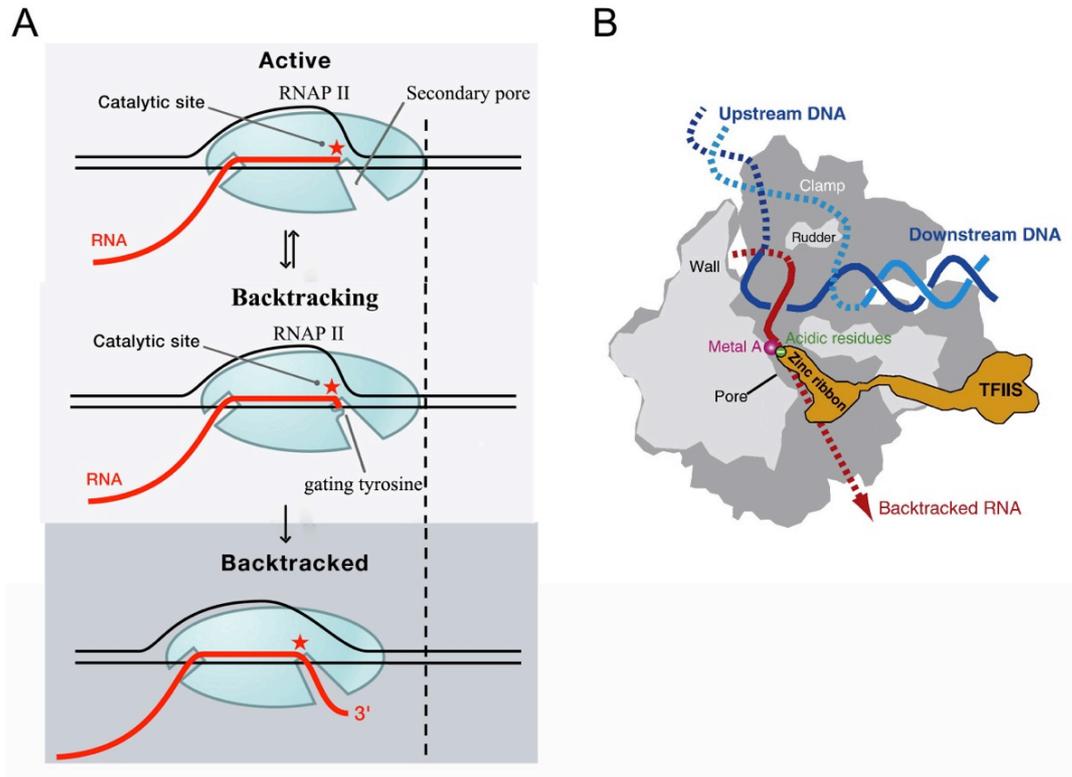


Figure 13: RNAP II backtracking and reactivation

(A) Schematic of RNAP II backtracking, see the main text for details. (B) Schematic cutaway view of yeast RNAP II-TFIIS. Adapted from (7, 157).

2 Helicases

Helicases use NTP to bind or remodel nucleic acids, nucleic acid-protein complexes, or both (158). They play important roles in almost every cellular process that involves nucleic acids, including DNA replication and repair, transcription, splicing and translation. Mutations in genes coding for helicases have been linked to numerous human diseases, such as xeroderma pigmentosum, Cockayne Syndrome, Werner's syndrome, Bloom's syndrome (159-161). Mutations in SETX, the human orthologue of Sen1, involved in transcription, cause juvenile amyotrophic lateral sclerosis type 4 (ALS4) and ataxia ocular apraxia type 2 (AOA2) (162-165), while mutations in IGHMBP2, involved in translation, give rise to distal muscular atrophy (166).

2.1 Classification

According to sequence and structural and functional analyses, all helicases are classified into six superfamilies (SFs). SFs 3 to 6 helicases are hexamers, and most of them can form a ring-shape structure, while helicases from SF1 and 2 do not form the toroidal shape (158, 167). There are several structural and mechanistic features that are conserved across these diverse superfamilies of enzymes. In all cases, the core domains adopt a RecA or an AAA⁺-like phosphate-loop (P-loop) NTPase fold and can be located either within the same polypeptide chain or between different subunits. These motifs convert chemical into mechanical energy by coupling the cycle of NTP hydrolysis to protein conformational changes (167, 168).

Most of SFs 3 to 6 helicases have similar ring-shape structural features (167) (Figure 14). Although all of them contain a P-loop NTPase core, individual evolution happens in different superfamilies (169). The nucleotide-binding sites are located at the interface between monomers and include amino-acid residues coming from neighbouring subunits. The ring structure has a central channel which encircles substrate oligonucleotides. The topological link between the protein and the nucleic

acid substrate might increase the stability and processivity of the enzyme. This is probably the reason why unwinding of dsDNA ahead of the replication fork seems to be almost carried out by a toroidal helicase in all kingdoms of life (169).

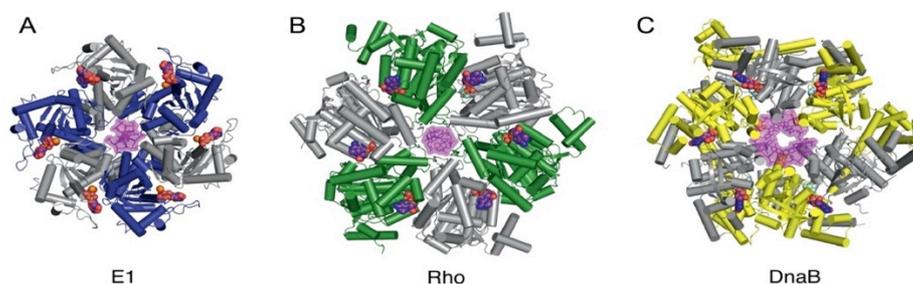


Figure 14: Ring helicases share similar structural features

(A) The papilloma virus E1 helicase bound to DNA (PDB: 2GXA). (B) The *E. coli* Rho transcription termination factor bound to RNA (PDB: 3ICE). (C) The *G. stearothermophilus* DnaB helicase bound to DNA (PDB: 4ESV). DNA and RNA molecules are located in the central hole and are shown in violet, the nucleotides are indicated in magenta, the magnesium ions are indicated in cyan. Adapted from (170).

SF1 and SF2 helicases share a catalytic core with high structural similarity, consisting of two RecA domains. Based on phylogenetic analyses of the sequence of these helicases, they are further divided into families or groups (Figure 15). Proteins within each of the identified families display similar sequence characteristics. In addition, some mechanistic features are also shared by proteins within the same family, such as the preference for a particular NTP, the ability to unwind nucleic acid duplexes and the translocation polarity. Several characteristic sequence motifs can be identified (Figure 16), although not all motifs are present in each family (158). These motifs were originally considered to be equivalent between the two groups, however structural studies showed later that this was not always the case (159). The level of conservation in the characteristic motifs is high within each family, but only limited conservation remains across both SFs. The highest level of sequence conservation across both SFs is located in motifs responsible for NTP binding and hydrolysis (motif I, II and VI Figure 16) (171).

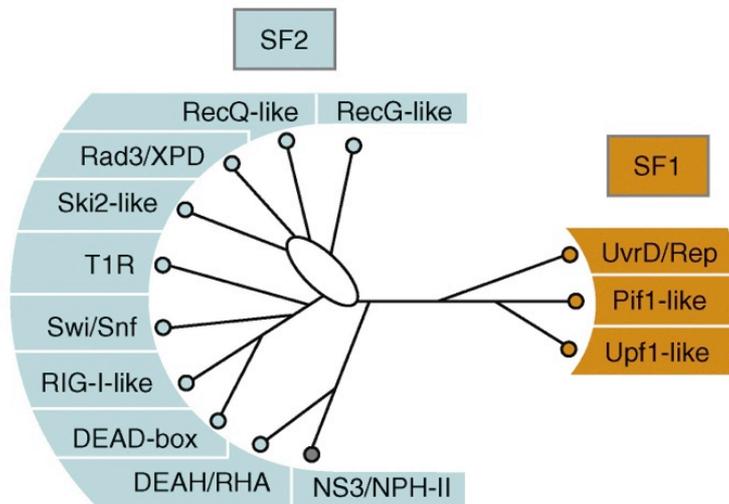


Figure 15: The families of the SF1 and SF2 helicases.

Schematic, cladogram showing the three identified families of SF1 (right), and the ten families of SF2 (left). Adapted from (158).

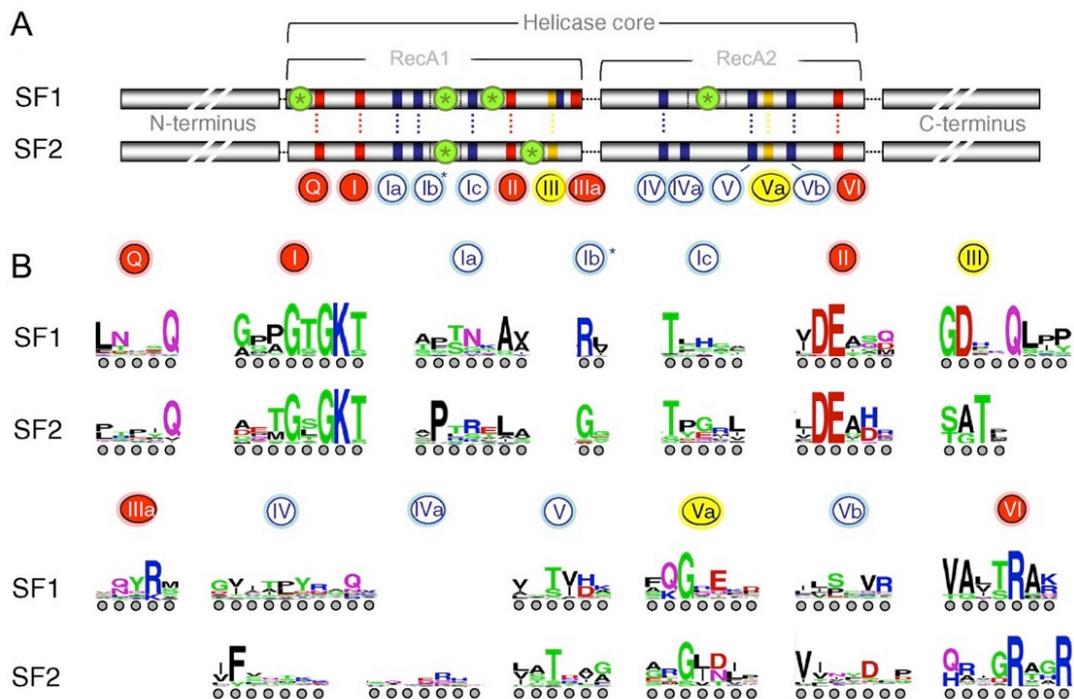


Figure 16: Sequence of the helicase core of SF1 and SF2 proteins.

(A) Sequence organization of SF1 and SF2 helicase cores. Characteristic sequence motifs are colored according to their predominant biochemical function: red, ATP binding and hydrolysis; yellow, coordination between nucleic acid and ATP binding sites; blue, nucleic acid binding. Green circled asterisks mark insertions of additional subdomains. The lengths of the blocks and the distance between the conserved domains are not to scale. (B) Sequence

conservation within the core helicase motifs. The height of the amino acids reflects the level of conservation. Coloring marks the chemical properties of a given amino acid: green and purple - polar, blue - basic, red - acidic, and black - hydrophobic. Adapted from (158).

Because my PhD work is focused on the mechanism of action of the SF1 helicase Sen1, in the next sections I will describe the main features of SF1 helicases.

2.2 Superfamily 1 helicases

Members of the SF1 superfamily of helicases are defined by several specific sequence motifs, for example the highly conserved motif III sequence GDxxQ is a useful hallmark and diagnostic of SF1 proteins. Several distinct families have been defined within SF1, among them three families are better studied: UvrD/Rep-like helicases, Pif1-like helicases and Upf1-like helicases. Upf1-like helicases are one of the main families in SF1 that can work on RNAs, and Sen1 is one of the important members of this family. Unlike some special helicases, which can translocate along nucleic acids without unwinding activity (eg. dsDNA translocase EcoR124I) (172) or can unwind duplexes without translocation (eg. DEAD-box helicases) (173), SF1 enzymes are mostly canonical helicases (174). In other words, SF1 helicases can translocate on either single-strand DNA (ssDNA) or RNA (ssRNA) and are capable of unwinding duplexes. Based on the direction of translocation on the single-strand nucleic acid, they are further divided into two groups: SF1A helicases can translocate from 3' to 5', whereas SF1B helicases translocate 5' to 3'.

2.2.1 Domain architecture

All SF1 helicases share a common domain architecture comprising two tandem RecA-fold domains constituting the motor domains of the helicase. The NTP binding pocket is located at the interface of the two core domains and motifs contacting the nucleic acids are located across the top surface of both core domains (Figure 17). SF1 helicases normally contain accessory domains either flanking or inserted within the core domains. The position, primary sequence and structural conformation of the accessory domains are quite variable between different SF1 helicases, which endows them with specific functions. The accessory domains can play important roles in targeting helicases to specific substrates or directing them to specific cellular locations, in modulating the helicase activity and in mediating the interaction with other proteins (159, 174, 175).

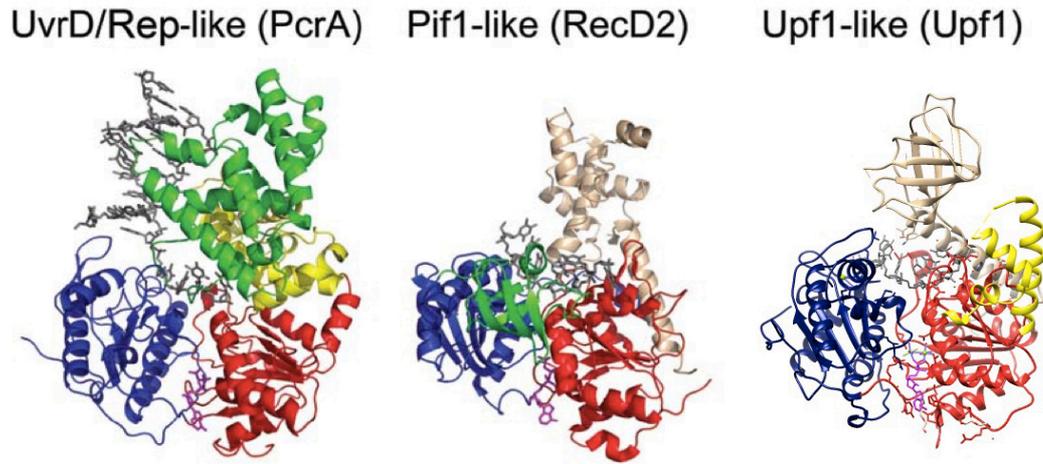


Figure 17: Representative structures of SF1 helicases

The structures shown are *B. stearothermophilus* PcrA (UvrD/Rep-like family), *Deinococcus radiodurans* RecD2 (Pif1-like family) and the core of human Upf1 (Upf1-like family). The helicase core domains are shown in red and blue respectively for all three proteins and a non-hydrolysable ATP analogue bound at their interface appears in magenta. Where present, DNA is shown in gray. Accessory domains are shown in different colors and are specific to each helicase. Adapted from (174)

2.2.2 NTP hydrolysis and nucleic acid binding

As mentioned above, the binding sites for the nucleotide are conserved across all six superfamilies of helicases. The NTP binding pocket in SF1 helicases is located at the interface between the RecA1 and RecA2 domains (Figure 17), with only subtle differences observed in available structures. Some specific roles have been elucidated for the amino acids in the binding pockets. For example, the glutamine residue in the Q-motif (in both SF1 and SF2) forms hydrogen bonds with the adenine resulting in a preference for ATP over other NTPs (176). Binding to the NTP leads to closing of the core domains because conserved residues from motif III (RecA1 domain) and motif VI (RecA2 domain) approach to coordinate the γ -phosphate (177). Catalysis is promoted by a divalent cation that is coordinated between the β - and γ -phosphates via a conserved threonine and an aspartate in motifs I and II, respectively (177).

As illustrated above (Figure 16), numerous conserved motifs within the core domains are involved in nucleic acid binding. Regardless the polarity of the helicase, the

single-strand nucleic acid is always bound to the core domains with the same orientation: the 3'-end associates with the RecA1 domain and the 5'-end associates with the RecA2 domain (158, 174). Accessory domains in SF1 helicases are also able to associate with nucleic acid. Importantly, for some helicases these interactions have been shown to be critical for translocation (177, 178).

SF1 enzymes typically display a very low or even undetectable basal ATPase activity that is remarkably increased by binding to single-strand nucleic acids (159, 174). Evidence obtained with the PcrA helicase suggests that nucleic acid binding to the helicase leads to a conformational change in the ATP binding pocket that allows the coordination of Mg^{2+} required for ATP hydrolysis (174, 179).

2.2.3 Translocation and translocation polarity

The interaction with the single-strand nucleic acid stimulates the helicase ATPase activity, and ATP hydrolysis in turn triggers conformational changes of the two core domains, leading to a movement of the helicase on the nucleic acid. The translocation polarity is a consequence of the alteration of the relative grip of the RecA1 and RecA2 domains on the single-strand nucleic acid during the ATP hydrolysis cycle. It has been shown for SF1 helicases that the accessory domains are important for translocation. The mechanism of translocation has been nicely illustrated for the SF1B model helicase RecD2, member of the Pif1-like family. Without ATP binding, the RecA1 domain of RecD2 binds tightly to the DNA, and both RecA domains are in an open, more extended, conformation. Binding of an ATP analog results in the closing of the two RecA domains, enabling the RecA2 to slide along the DNA and to move toward the RecA1. This conformational change also leads to increased interactions between the RecA2 with the DNA (180) (Figure 18). Upon ATP hydrolysis and nucleotide release, the RecA1 domain, that is weakly bound to the DNA, moves 1 nt forward along the ssDNA (in the 5' to 3' direction), so that the two RecA domains adopt again an open conformation. A similar mechanism also applies

to translocation of the SF1A model helicase PcrA that belongs to the UvrD/Rep family helicase (177, 180). The mechanism of translocation of Upf1-like family helicases is less clear. Nonetheless it is very possible that they use an analogous mechanism for translocation since they adopt the same domain organization when bound to nucleic acids (Figure 17) (181).

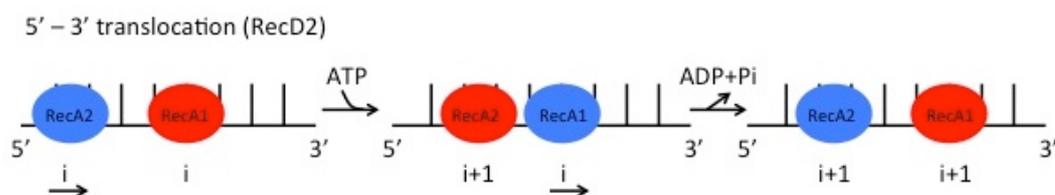


Figure 18: Translocation mechanism of RecD2.

Schematic of the translocation mechanism for the RecD2 helicase in a cycle of ATP binding and hydrolysis. Red circles represent domains that have a tight grip on the ssDNA, and blue circles represent domains that have a weaker grip and can slide along the DNA. In the absence of ATP, both RecA1 and RecA2 are at position “i”, and RecA1 binds tightly to the DNA while RecA2 binds loosely. When the enzyme binds the ATP, the cleft between RecA1 and RecA2 motor domains closes, causing RecA2 to slide along the DNA backbone (black) by one nucleotide, at position “i+1” site. The contacts between RecA1 and the DNA remain tight to anchor the DNA as RecA2 slides along it. When the conformational change is complete, the grip of RecA1 on the DNA is loosened and that of RecA2 is tightened. Then, ATP hydrolysis takes place and allows the cleft to relax to the open conformation, thereby enabling RecA1 to slide forward along the DNA to the position “i+1”. The result is translocation by one base in the 5’–3’ direction during a single round of ATP binding and hydrolysis. Adapted from (159, 180)

2.2.4 Coupling of translocation and strand separation

One common structural feature involved in duplex unwinding in SF1 helicases is a “pin” or “wedge” which locates at ss-ds nucleic acid junction and helps to prise the incoming strands apart during translocation (Figure 19).

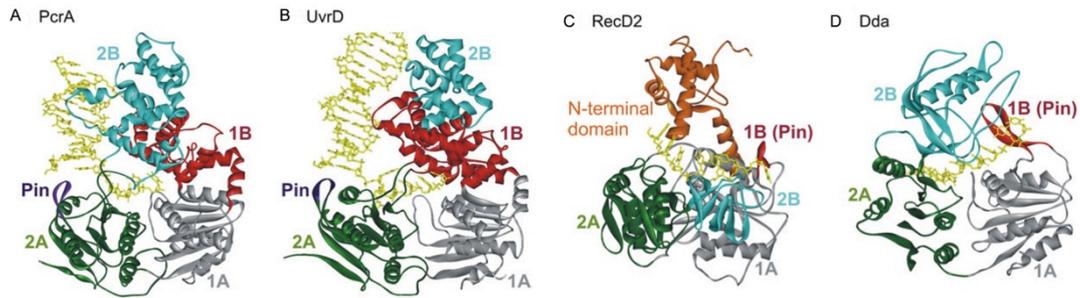


Figure 19: Crystal structures of SF1A (PcrA, UvrD) and SF1B (RecD2, Dda) helicases

(A) The PcrA helicase from *Bacillus stearothermophilus* (PDB code 3PJR). The RecA-like domains 1A and 2A are shown in grey and green, respectively. The structure shows the pin (purple) separating the strands of duplex DNA. (B) Structure of *Escherichia coli* UvrD helicase (PDB code 2IS1). (C) Structure of the RecD2 helicase from *Deinococcus radiodurans* (PDB code 3GPL). The beta-hairpin pin (1B) is indicated in red. (D) Structure of the bacteriophage T4 Dda helicase (PDB code 3UPU), indicating the pin in red. Adapted from (159).

The location of the pin varies according to the direction of translocation. For SF1A helicases, the RecA2 domain leads during translocation. The structures of PcrA and UvrD show that the pin emanates from the leading RecA2 domain (Figure 19). For some SF1B helicases, on the contrary, the pin is located on the RecA1 domain as an insertion (Figure 19). Whether there is also a pin structure involved in strand separation in Upf1-like helicases remains unclear.

2.2.5 Modulation of the helicase activity and function

Functional specificity of SF1 helicases arises from the presence of distinctive accessory domains and/ or the interaction with partner proteins.

Accessory domains can regulate the helicase activity. It has been shown that monomers of Rep, a SF1A helicase from *E. coli*, cannot unwind duplex DNA, whereas monomers carrying the deletion of its domain 2B results in proficient, albeit poorly processive, unwinding activity (182). It was suggested that the auto-inhibitory

effect of the 2B domain could be relieved by dimerization of Rep proteins or by the interaction with partner proteins (183). A similar auto-inhibition mechanism operates on Upf1 protein as well. The N-terminal domain of Upf1, named the CH domain (rich in cysteine and histidine domain), exerts a *cis* inhibitory action on both the ATPase and unwinding activity of the helicase core. The CH domain forces Upf1 to bind RNA in a clamping conformation that prevents its function (181). The inhibition can be relieved by binding of Upf2 to the CH domain. This interaction triggers a conformational change on the CH domain that promotes the activation of the helicase core. The C-terminal domain of Upf1, conserved in higher eukaryotes, can also lock the Upf1 helicase core in a conformation that is not productive for ATP hydrolysis or duplex unwinding (175).

Partner proteins help helicases to target their specific substrates. For example, *E. coli* UvrD is involved in both mismatch DNA repair and nucleotide excision DNA repair. The function of UvrD in the distinct pathways is directed by the interaction with different protein partners. MutL recruits UvrD to the mismatch repair pathway, whereas UvrB directs UvrD to the nucleotide excision repair pathway (184, 185). In the case of yeast Sen1, the aforementioned Upf1-like helicase that plays a key role in non-coding transcription termination, it has been proposed that its recruitment to specific ncRNA targets is enabled by the interaction with Nrd1 and Nab3 (62). As detailed in a previous section (see section 1.2.3.2), Nrd1 and Nab3 are RNA binding proteins that can recognize specific motifs that are highly enriched on certain classes of ncRNAs.

3 Helicases as transcription termination factors

In this section I will summarize the most recent advances on the function and mechanisms of action of helicases, both prokaryotic and eukaryotic, that specifically work as transcription termination factors.

3.1 Rho

Rho is a bacterial homo-hexameric toroidal motor protein, belonging to the SF5 of helicases according to its sequence. It utilizes the energy from ATP hydrolysis to translocate along the RNA from 5' to 3' and dissociate the transcription elongation complex (EC) (6, 186-189). Rho is well conserved across the different bacteria kingdom, and more than 90% of the species contain at least one copy of the *rho* gene (190). In the gram-negative bacterium *E.coli* it is an essential and abundant protein that is responsible for more than 20% of transcription termination events (191, 192). The number of genes terminated by Rho in the gram-positive model organism *B. subtilis*, where Rho is dispensable and low abundance, is believed to be substantially lower (187, 193). Rho expression is auto-regulated by attenuation, which means that Rho recognizes termination signals located at the 5' end of its own gene and induces premature termination of part of transcription events (194). Rho performs a number of functions that are important for a variety of biological processes. For example, genome-wide analyses suggest that Rho-dependent termination plays a role in repressing pervasive antisense transcription in both *E. coli* and *B. subtilis* (192, 195). Rho can terminate transcription of untranslated mRNAs via transcription-translation coupling, a phenomenon also known as polarity (186, 196). Rho has also been shown to play important roles in removing R-loops in *E. coli*, which otherwise can be processed into deleterious double-stranded breaks that lead to genomic instability (197-199).

The first step of transcription termination is the recruitment of Rho to the nascent RNA. Rho can recognize specific sequences on the nascent transcript, called *rut* (Rho utilization) site, via its N-terminal RNA binding domain (RBD). At each Rho monomer, the RBD forms a cleft that is only big enough to fit two pyrimidine bases, preferentially YC dinucleotides (where Y represents a pyrimidine) (200). A spacer of at least 12 nucleotides is required between two consecutive YC dinucleotides (201, 202). Therefore a *rut* site usually comprises 60-90 nucleotides (203), including the 6 YC dinucleotides bound by the six Rho monomers and the respective spacers, which are typically C-rich and G-poor (204). The depletion of G within a *rut* site decreases the probability of forming secondary structures, which are obstacles for Rho binding (191, 204, 205). There are two distinct types of *rut* binding sites on a Rho hexamer, called primary and secondary (206). The RBD mentioned above is the primary binding site (PBS), which supports the specific interaction of Rho to the *rut* site without ATP hydrolysis. In addition to the PBS, the Rho N-terminal domain (NTD) contains a positively charged N-terminal helix bundle (NHB) that may help binding of RNA (Figure 20) (207). After initial binding to the *rut* site, the inward tilt conformation of the PBS helps directing the RNA into the central pore of the hexameric ring, where the secondary binding site (SBS) locates (206, 208) (Figure 20). There are several key sequence motifs located in the C-terminal RecA-like ATPase domain (AD) (Figure 20). One is the P-loop, which is required for ATP binding and hydrolysis (208). A second important motif in Rho's CTD is the Q-loop, which makes intimate contacts with the RNA. For example, a backbone carbonyl from V284 forms a hydrogen bond with the 2'-OH of the RNA ribose, explaining the specificity of Rho for RNA substrates (209, 210). A third motif, the R-loop, makes a single contact with the nucleic acid substrate. The location of the R-loop is consistent with biochemical data showing that this motif is involved in coupling RNA binding with ATP hydrolysis (209, 211, 212). The ATP-binding pocket of Rho is formed at the interface between two adjacent CTDs and contains the conserved P-loop motif (Figure 20). The binding of the RNA to the SBS stimulates Rho ATPase activity, which in turn

promotes ring closure (213) and allow Rho to translocate along the RNA (209, 214). Rho has been crystalized as both a closed-ring and a notched-shape (201, 208, 211, 215), suggesting that the Rho hexamer can transiently open to allow the RNA to get into the central channel. There are also recent single-molecule experiments showing two RNA binding patterns for Rho: one consisting in Rho bound to around 57 nt, corresponding to the PBS alone, and another in which Rho binds about 85 nt corresponding to both the primary and the secondary binding sites (216), confirming previous structural and biochemical data.

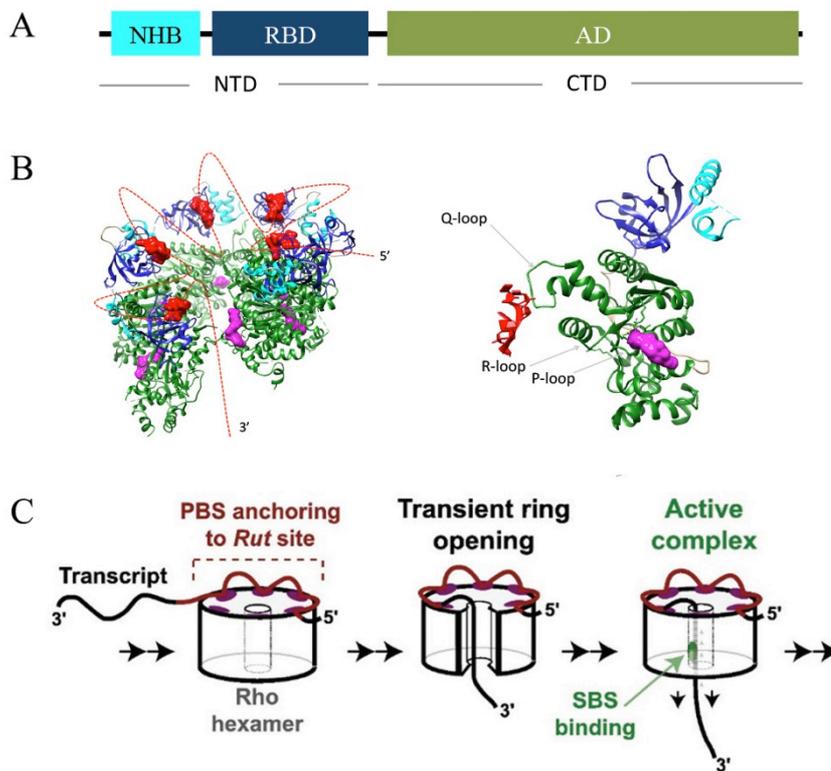


Figure 20: Structure of Rho and RNA loading

(A) Schematic representation of Rho domain architecture. Structured domains are represented as colored blocks, while black lines denote flexible connecting linkers (B) Left: Open-ring structure of a Rho hexamer (PDB: 1PVO). Right: Rho protomer from a closed-ring structure containing RNA bound to the SBS (PDB: 3ICE). Rho domain colors are labeled as in (A). The RNA is indicated in red. The ATP is shown in magenta. (C) Schematic of RNA loading. Sequence specific contacts happen between the RNA substrate and the Rho hexamer at its PBS. Transient opening of Rho helps threading the RNA into the central channel of the hexamer. The RNA then binds to the SBS, promoting ring closure. Adapted from (217).

After recruitment to the *rut* site, Rho couples ATP hydrolysis to translocation via its

secondary binding site. It has been shown that the *rut* site sticks to the PBS during translocation, a process named “tethered tracking”, leading to formation of an RNA loop between the PBS and the SBS upon translocation (206, 216, 218, 219). Structural blocks on the nascent RNA, such as RNA hairpins, RNA binding proteins or ribosomes, can impede Rho translocation, thus affecting Rho-dependent termination (220). It has been shown that, by translocating on the RNA, Rho can displace a streptavidin molecule bound to biotinylated RNA (221, 222) and can unwind up to 80 base-pairs RNA-DNA hybrids, indicating that Rho is a strong but moderately processive motor (6).

The last step of transcription termination (i.e. dissociation of the EC) is triggered when the moving Rho catches up with the polymerase. There is a considerable amount of studies showing a correlation between RNAP pausing and transcription termination (203, 223-226), and polymerase pausing is considered as a prerequisite for Rho-dependent termination (188, 220). However transcriptional pausing does not necessarily result in termination as backtracked and arrested ECs have been shown to be poor substrates for Rho termination (227). It remains to be determined whether most sites of Rho termination are elemental pauses (see section 1.4.3).

The mechanism by which Rho dislodges the RNAP from the DNA template remains relatively obscure. Several models have been proposed (Figure 21). The hybrid shearing model posits that Rho can generate a pulling force on the nascent RNA that can disrupt the RNA-DNA hybrid in the elongation complex, giving rise to EC collapse (228, 229). This proposal is based on the fact that double-stranded nucleic acids are known to separate briefly due to thermal motions (called breathing), so that Rho can capture the RNA in a form that will not be able to reform a duplex with the DNA template (229). The hyper-translocation model postulates that Rho exerts a pushing force on the polymerase, rather than pulling the nascent RNA, causing RNAP to translocate forward without RNA synthesis and leading to EC destabilization (230). This model is supported by evidence showing that rewinding of the DNA duplex

upstream of the transcription bubble, which can also promote forward translocation of RNAP, facilitates Rho-dependent transcription termination (230). In addition, it has also been shown that Rho can promote forward translocation of RNAP stalled by DNA-binding protein serving as a roadblock (230). Finally, the allosteric model proposes that Rho triggers a rapid rearrangement of the RNAP active center that induces EC inactivation and is followed by a relatively slow step of EC dissociation. Several experiments support this model (231). For example, crosslinking data shows that the RNA 3' terminus remains in the RNAP catalytic center upon Rho treatment, but exhibits a different crosslinking pattern, suggesting that a conformational change rather than RNAP hyper-translocation is induced by Rho during termination. Furthermore, the antibiotic tagetitoxin, which binds RNAP near the active center and traps RNAP in an inactive conformation (232, 233), can abolish Rho-mediated EC dissociation (231), consistent with the idea that a conformational change in RNAP is the key element in the termination process. Finally, Rho cannot promote dissociation of an EC composed of T7 RNAP, even though it is less stable than the *E. coli* EC, suggesting that a specific Rho-RNAP interaction is required for termination (231). Strikingly, there is also evidence showing that *in vitro* *E. coli* Rho can efficiently terminate transcription of *Saccharomyces cerevisiae* RNAP II, but not RNAP I and III (234, 235), suggesting that Rho might require features that are shared by *E. coli* RNAP and *S. cerevisiae* RNAPII.

Genome-wide ChIP-chip analyses suggest that Rho might associate with RNAP throughout transcription cycle (236). Consistently, one study showed that Rho is able to bind to the RNAP in the absence of DNA, RNA or other proteins (231). However, another group claims that the interaction with nascent RNA is the primary event enabling the recruitment of Rho to the EC (237). Further studies are required to elucidate these discrepancies.

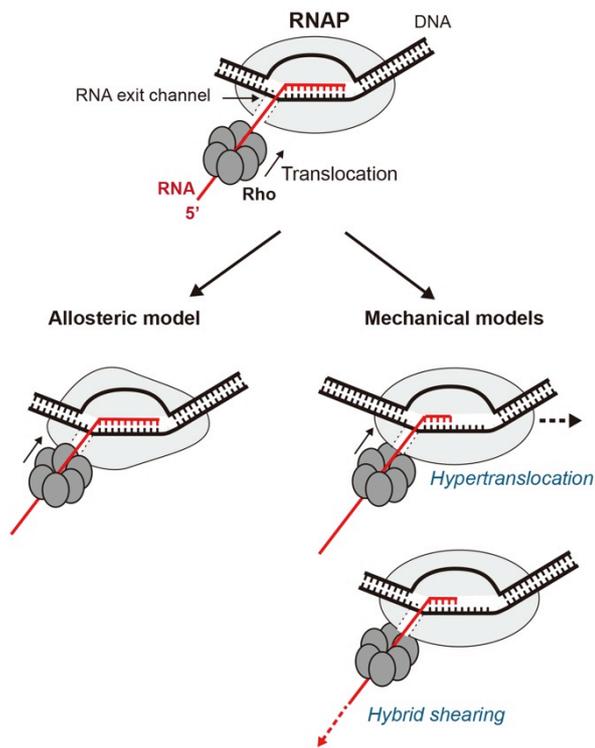


Figure 21: Models for Rho-dependent EC dissociation

Black lines indicate the DNA template. Red lines indicate the RNA. Proteins are indicated in the figure. See main text for more details.

3.2 Mfd (mutation frequency decline)

E. coli Mfd is a monomeric, multimodular protein, belonging to the RecG-like family of SF2 helicases according to its sequence (158). In terms of catalytic activity, it is actually a double-stranded DNA (dsDNA) translocase rather than a helicase because it translocates along dsDNA without unwinding the DNA duplex (238, 239). Mfd plays a pivotal role in DNA repair, specifically in transcription-coupled repair (TCR), one of the two main pathways in nucleotide excision repair (NER).

The NER pathway is responsible for removing bulky DNA lesions, such as thymine dimers and 6,4-photoproducts induced by ultraviolet light (UV). NER can be divided into two subpathways: global genomic repair (GGR) and transcription-coupled repair (TCR). The main difference between the two subpathways is the damage recognition step. In GGR, the UvrA-UvrB complex scans the DNA and binds to the damage site with high affinity (Figure 22). After damage recognition, UvrA dissociates and UvrC is recruited to the damage site, producing DNA cleavage at both sides around the site. Following incision, the UvrD helicase is recruited, removing UvrB and UvrC as well as the damage-containing oligonucleotides. Finally, the resultant gap is filled in by the action of the DNA Polymerase I and DNA ligase (240). In TCR, RNAP plays the main role in DNA damage scanning via transcription. When RNAP encounters a DNA damage, such as a thymine dimer, it stalls at the damage site. Then Mfd is recruited through the direct interaction with the polymerase and elicits dissociation of the stalled EC. Finally Mfd recruits UvrA-UvrB to the damage site and thus triggers the DNA repair process (241). In such way, damages on the template strand can be removed more efficiently than those on the non-template strand (239, 242-244).

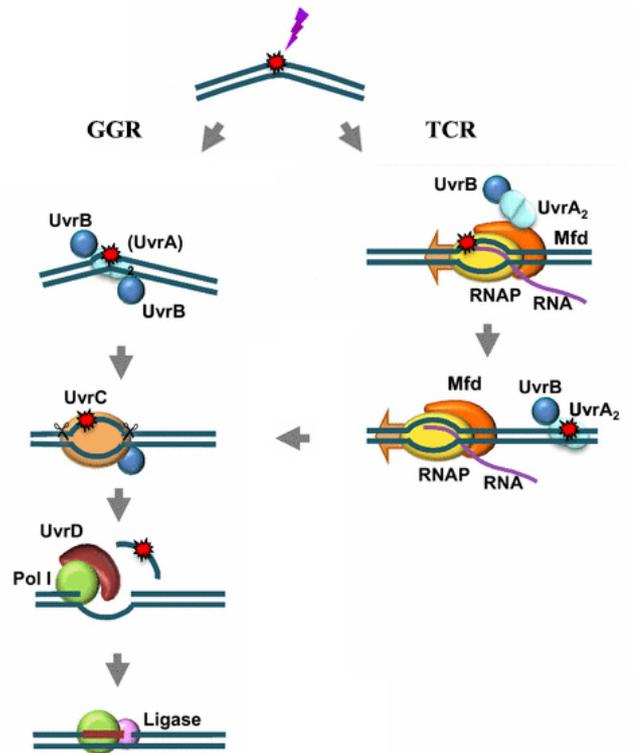


Figure 22: Model for nucleotide excision repair (NER) in *E. coli*.

See the main text for more details. Adapted from (245).

The structure of Mfd consists of a compact arrangement of eight domains referred to as D1a, D1b, and D2-D7 (Figure 23) (246). The RecA-like catalytic cores, composed of the D5 and D6 modules, are strikingly similar in sequence and structure to the corresponding RecG. However, unlike RecG, Mfd requires ATP (or a non-hydrolyzable ATP-analog) for dsDNA binding (247, 248). A particularly interesting feature located in helicase core is the TRG (translocation in RecG) motif. Mutations in this motif completely abolished Mfd-mediated EC dissociation, while the DNA binding (in the presence of the analog ATP γ S) and nucleotide hydrolysis activities remained intact. It has been proposed that the TRG motif couples ATP hydrolysis and DNA translocation (246). The D4 module flanking the helicase core at the N-terminal side is the RNAP interacting domain (RID). The structure of the Mfd D4 module in complex with the RNAP β 1 subunit has been determined (Figure 23) (249). Disruption of this specific interaction impairs the proper function of Mfd in EC dissociation while Mfd maintains the capability of DNA binding and ATP hydrolysis

(246, 249, 250). The N-terminal region of Mfd is composed of three modules, with sequence and structure similarity with the UvrB component of the NER machinery. Both proteins interact with the same region of UvrA as revealed by biochemical and structural evidences (251-253). An N-terminal truncated version of Mfd missing the UvrB homology region is active for dissociation of stalled ECs but defective in UvrA recruitment (248, 254). Module D7 interacts with module D2 and occupies the UvrA interaction sites (246, 252), therefore it is inhibitory for UvrA binding. This domain is dispensable for RNAP release (246).

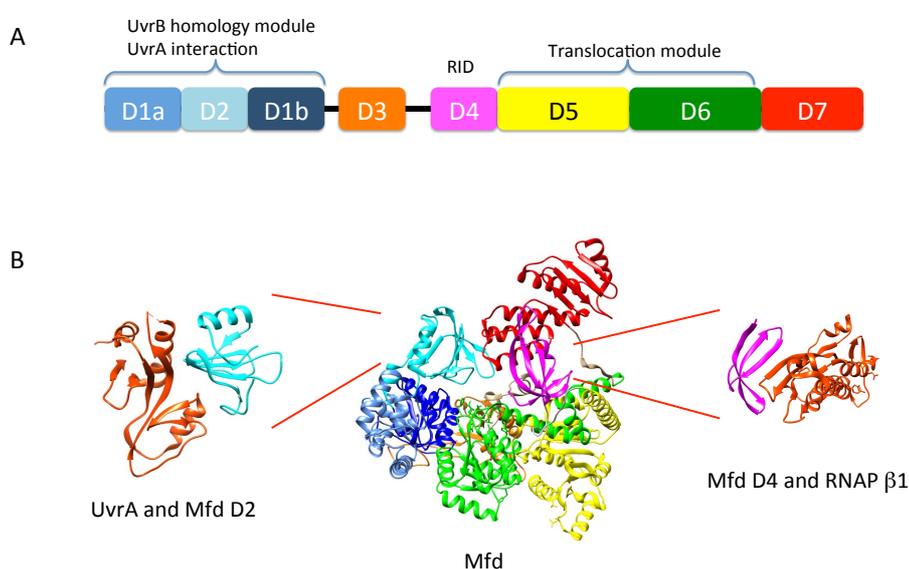


Figure 23: Structural features of Mfd.

(A) Schematic representation of Mfd domain architecture. Structured domains are represented as colored blocks, while black lines denote flexible connecting linkers. (B) Middle: structure of *E. coli* Mfd (PDB ID 2EYQ), left: structure of *E. coli* Mfd D2 module in complex with UvrA (PDB ID 4DFC), right: structure of *Thermus thermophilus* Mfd D4 module in complex with *Thermus aquaticus* RNAP β 1 subunit (PDB ID 3MLQ). Mfd domain colors are labeled as in (A), UvrA and RNAP are labeled in orange.

Similar to most SF1 and SF2 helicases, the N-terminal and C-terminal domains of Mfd are important for the regulation of Mfd activities (158, 255). The full-length Mfd protein by itself is not capable to translocate along dsDNA. Translocation by full-length Mfd can be detected only in the presence of stalled RNAP (239, 256). Interestingly, deletion of either the N-terminal D1a/D2/D1b region or the C-terminal

D7 domain generates Mfd proteins that are constitutively active for translocation (256-258). The D7 module is mobile during the catalytic cycle (252), and it has been proposed that binding to RNAP via module D4 switches Mfd from a “closed” and autoinhibited form, resulting from the interaction of D7 with D2, to a more “open” and active conformation, promoting Mfd translocation to displace the stalled RNAP and allow the recruitment of UvrA/B (252, 254, 256, 257).

The process of resolving damage-stalled TEC (Figure 24) begins with the recruitment of Mfd to the blocked RNAP via the interaction between Mfd RID and the RNAP $\beta 1$ subunit. Whether Mfd is only recruited to the stalled RNAP or it can be brought to the transcribing RNAP as well remains unclear. Structures of eukaryotic RNAP II encountering TCR-related damage sites, such as cyclobutane pyrimidine dimers and cisplatin lesions, show a polymerase conformation that is nearly identical to that in a normal EC in the absence of DNA lesions (129, 259, 260), suggesting non-allosteric recruitment of repair factors. However the structure of Mfd RID in complex with RNAP $\beta 1$ reveals a conformational change in RNAP, suggesting this novel conformation might be involved in the recruitment of Mfd (249), although it might also be a consequence of the interaction between both proteins. Once bound to RNAP, Mfd undergoes a structural rearrangement and reveals otherwise repressed activities. Specifically, Mfd binds to dsDNA upstream of RNAP, and translocates toward the stalled RNAP exerting a pushing force that ultimately induces forward translocation of RNAP and the collapse of the transcription bubble (239, 261). Approximately 25 bp of dsDNA upstream of RNAP, but not downstream, is required for stalled EC dissociation (239). Interestingly, Mfd can also push forward backtracked and arrested RNAP *in vitro*, thereby rescuing an arrested EC (239). Recent single-molecule experiments have shown that after transcription termination by Mfd, the displaced RNAP remains associated with Mfd, which translocates slowly and processively along the dsDNA with RNAP attached to it (261, 262). This complex is then released from the DNA template upon recruitment of the UvrA/B complex and ATP hydrolysis (244). In this way, TCR can take place much more efficiently than GGR

(244, 253, 263), probably because the recruitment of UvrA/B by Mfd is more efficient than the direct recognition of the damage by UvrA/B (244).

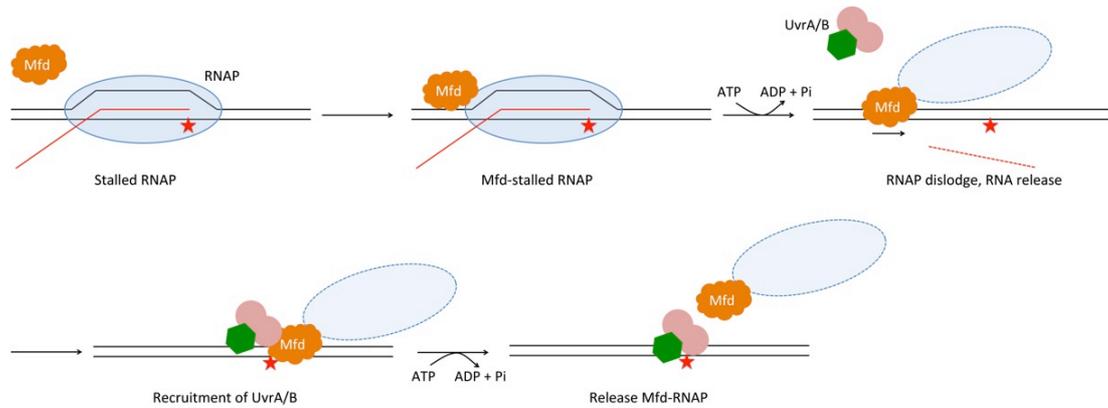


Figure 24: Process of Mfd-mediated EC dissociation.

Mfd binds the stalled RNAP and begins translocating toward the RNAP, applying a pushing force on the polymerase that eventually dismantles the EC. The nascent RNA is released but the RNAP remains associated with Mfd, which translocates slowly along the dsDNA until the recruitment of the UvrA/B complex to the DNA lesion site.

3.3 Sen1

Sen1 is a SF1, Upf1-like helicase. In *S. cerevisiae*, it is one of the components of the Nrd1-Nab3-Sen1 (NNS) complex, which plays important roles in termination of transcription of most ncRNAs, like snRNAs, snoRNAs and CUTs, as well as some small protein-coding RNAs (62, 264-266). Sen1 is the only subunit in the NNS complex that performs enzymatic functions. It is a 252-KD protein harboring a central helicase domain that is essential for cell viability (Figure 25), N-terminal and C-terminal regions that have been shown to mediate protein-protein interactions (see section 1.2.3.2 for details) (71, 93, 94, 267). Yeast Sen1 is an exclusively nuclear protein that contains two nuclear localization signals flanking the helicase domain (Figure 25) (268).

The helicase domain of Sen1 is composed of two tandem RecA-like domains (RecA1 and RecA2), which contains all the conserved motifs of SF1 helicases (Figure 25) (158). According to the structure of Upf1, the RecA1 domain is predicted to contain two insertion domains (1B and 1C) (181, 269), which in Upf1 are involved in the interaction with the RNA. The helicase domain of Sen1, produced in *E. coli*, has been proved to be able to translocate on single-stranded nucleic acids (both RNA and DNA) and unwind double-stranded nucleic acids (91). Translocation occurs in the 5' to 3' direction, as for the Upf1 helicase (91, 269). A single amino acid mutation in the helicase domain of Sen1 (*Sen1-E1597*) is sufficient to generate transcription termination defects at both ncRNAs and some short protein-coding genes (265).

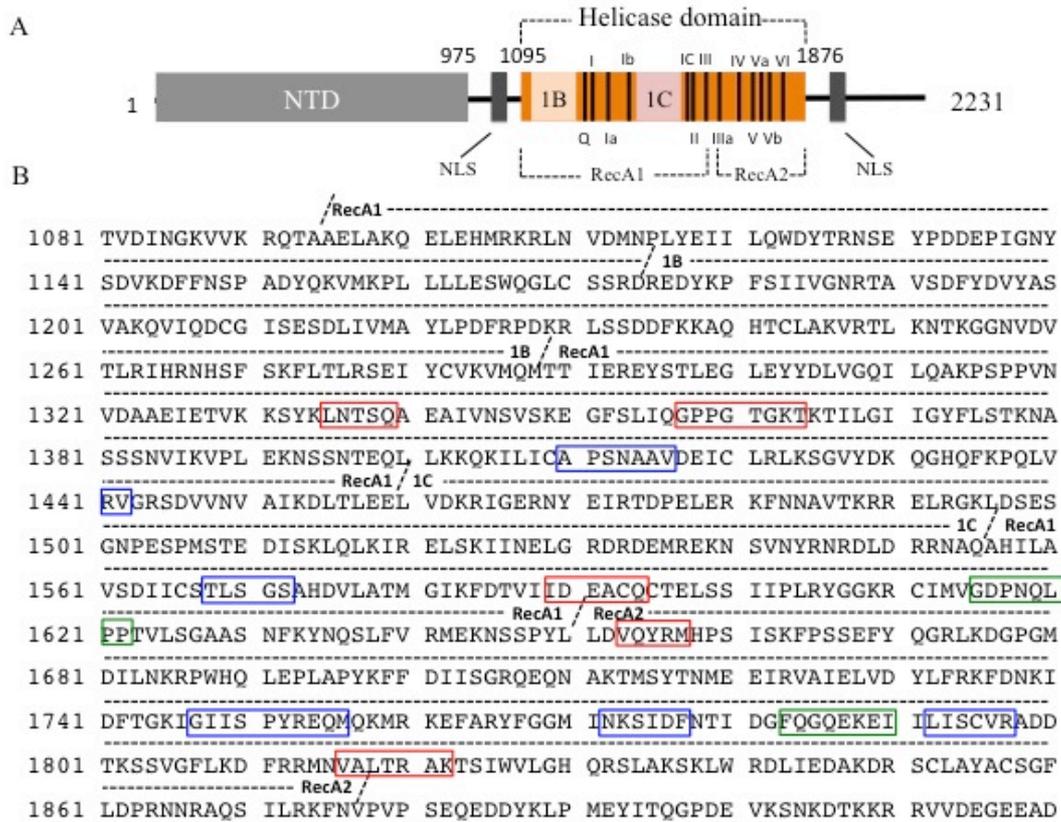


Figure 25: Scheme and sequence of Sen1

(A). Schematic representation of Sen1. The conserved helicase domain is indicated by an orange bar, and it is composed of RecA1 and RecA2 domains, with 1B and 1C subdomains inserted within RecA1 (91, 181, 269). The black thin bars within the helicase domain represent motifs conserved across SF1 helicases (158). The N-terminal domain (NTD) and the nuclear localization signals (NLS) are shown in gray (268). (B). Sequence and organization of the helicase domain of Sen1. The two RecA domains (RecA1 and RecA2) and the two insertion domains (1B and 1C) are indicated above the respective sequences. Motifs involved in ATP binding and hydrolysis are shown in red. Motifs responsible for nucleic acid binding are indicated in blue and sequences necessary for the coordination between ATP binding and nucleic acid binding are shown in green (158).

Unlike Rho, Sen1 cannot recognize specific sequences on the nascent RNA for its appropriate function in transcription termination (56). However, Nrd1 and Nab3, the other two components of the NNS complex, can bind to short sequence motifs that contribute to the specificity and efficiency of termination (56, 74, 78, 79, 266). In addition, Nrd1 interacts with the Ser5-phosphorylated C-terminal domain (CTD) of RNAP II (77), which peaks over the 5' end of genes; therefore, the NNS complex usually promotes transcription termination at short non-coding genes (73, 78, 270,

271).

After recruitment of the NNS complex to the nascent RNA, Sen1 applies a mechanism of termination similar to *E. coli* Rho-dependent termination (62, 95, 272). By using highly purified *in vitro* transcription termination system, it has been demonstrated that Sen1 alone can dissociate the EC, preferentially at transcription pausing sites (95). Similar to Rho, Sen1-dependent termination *in vitro* requires interaction with the nascent RNA and ATP hydrolysis. Contrary to Rho, which can release elongation complex of both *E. coli* RNAP and *S. cerevisiae* RNAP II *in vitro*, Sen1-dependent termination is specific to *S. cerevisiae* RNAP II (95). Experiments showing that the RNAP II transcription rate could affect the position of Sen1-dependent termination sites *in vivo* (96) together with data showing that Sen1 helicase domain can translocate along the RNA from 5' to 3' *in vitro* (91), support a model whereby Sen1 translocate along the nascent RNA to catch up with the polymerase to provoke transcription termination.

Aside from its role in terminating non-coding transcription, additional functions have also been proposed for Sen1. Evidence from a few candidate genes suggests a role for Sen1 in transcription termination of some long protein-coding genes (273, 274). However, in contrast to these studies, Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) analyses detected very little, if any, differences in RNAP II localization at the 3' end of most protein-coding genes in Sen1-depleted strains (275). The function of Sen1 in transcription termination at protein-coding genes requires further investigation. Sen1 was also proposed to be involved in termination of RNAP I transcription, since inactivation of the temperature sensitive mutant *sen1-1* leads to accumulation of readthrough transcripts at the 35S pre-rRNA gene (274). Consistently, Nrd1 and Nab3 have been shown to colocalize at the rDNA locus by chromatin immunoprecipitation (ChIP) experiments (276), suggesting that the NNS complex might play a role in termination of RNAP I-dependent transcription or in rRNA processing.

In addition, it has been proposed that Sen1, through its helicase activity, can help resolving RNA/DNA hybrids (R-loops) that are formed when the nascent RNA invades the DNA and pairs with the template strand during transcription (277). Sen1 also plays a role in resolving conflicts between replication and transcription (278), which can cause genome instability (279). Specifically, it was proposed that a fraction of Sen1 associates with replication forks and counteracts R-loop formation at sites of collision between the replisome and the RNAP II transcription machinery (278). In addition, the Sen1 N-terminal domain, rather than the helicase domain, has been shown to play an important role in transcription-coupled repair (280). It should be noted that Sen1 dysfunction may cause indirect effects on other cellular processes because inefficient non-coding transcription termination by the NNS-complex can impact the expression of numerous protein-coding genes via transcriptional interference (see section 1.3 about pervasive transcription). For example, Sen1 was originally identified in a screening for mutants that affect tRNA splicing endonuclease activity (281). However it was shown subsequently that this phenotype is likely due to transcriptional interference on the *SEN2* gene encoding a subunit of the tRNA splicing endonuclease, upon impaired termination of upstream *SNR79* (92).

Sen1 is the only protein of the NNS-complex that is conserved in most eukaryotes. Its human ortholog, senataxin (SETX), is a 302-kDa protein and localizes to both the nucleus and the cytoplasm (282, 283). Mutations in the most conserved regions of senataxin, the N-terminal and the helicase domains, are linked to amyotrophic lateral sclerosis type 4 (ALS4) and ataxia-ocular apraxia type 2 (AOA2) (Figure 26) (162, 163, 284). Introducing some of the AOA2-associated mutations in the equivalent residues of Sen1 provokes transcription termination defects *in vivo* (268). Human senataxin has also been shown to play a role in transcription termination at several model genes (285-287). It was proposed that senataxin could resolve R-loops, which are formed at termination sites and facilitate RNAP II pausing, so that the nascent RNA becomes accessible to the 5'-3' "torpedo" exonuclease Xrn2 (285). Consequently, Xrn2 degrades the 3' portion of the cleaved transcripts and induces

RNAP II release. This role of senataxin in resolving R-loops has also been proposed to be important for the protection of the genome against transcription-associated DNA damages since depletion of SETX results in accumulation of DNA breaks and hyperactivation of the DNA damage response (164, 278, 282, 287-289).

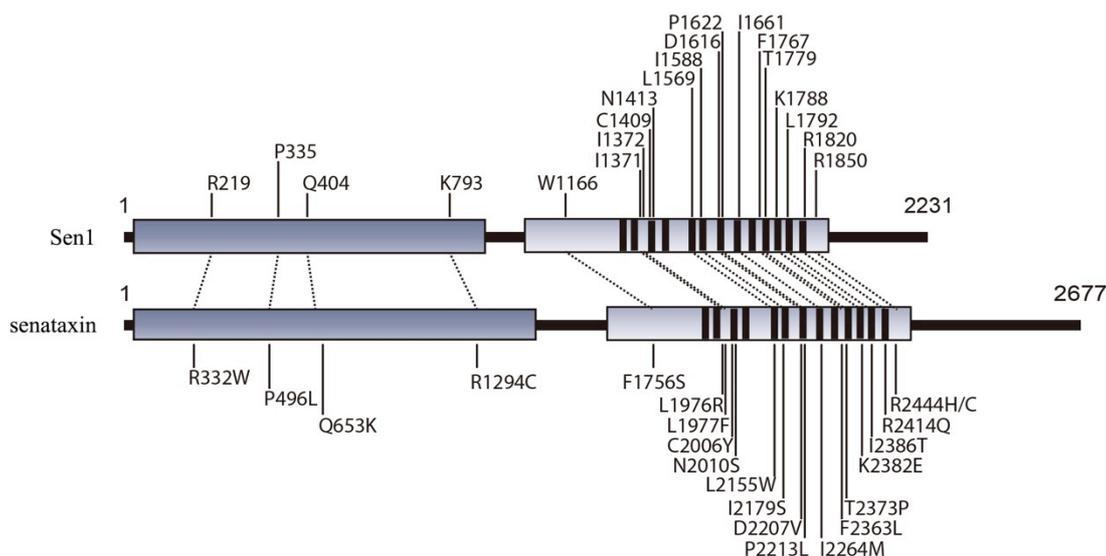


Figure 26: Comparison of Sen1 and senataxin architectures.

Scheme of Sen1 and senataxin (SETX) proteins. Both share a similar domain organization. Several disease mutations conserved in the two proteins are indicated.

Senataxin has been shown to interact with various proteins, leading to different cellular functions (164). For example, the CTD of human RNAP II can be symmetrically dimethylated at the conserved Arg residue (R1810) and this modification recruits the survival of motor neuron protein (SMN), which interacts with senataxin and recruits it to RNAP II, to resolve R-loops and mediate transcription termination (287). Breast cancer type1 susceptibility protein (BRCA1) has also been reported to localize to R-loops and mediate the recruitment of senataxin, which is important for resolving the R-loops and preventing DNA breaks (288). The N-terminal region of senataxin is covalently modified by the addition of SUMO2/3, and the SUMO2/3 modification is required for the interaction of senataxin with the Rrp45 subunit of exosome, which might play a role in the resolution of transcription-replication conflicts (290). The microprocessor complex that consists of

at least two subunits, the RNase III Droscha and the double-stranded RNA (dsRNA) binding protein Dgcr8, orchestrates the recruitment of senataxin and Xrn2 to initiate RNAP II premature termination at HIV-1 promoter region and a subset of cellular genes (291). Senataxin also promotes premature termination at virus-induced genes via the interaction with transcriptional cofactor TAF4 (TBP associated factor 4) (292). In *SETX*-deficient cells, the overactivation of virus-induced genes can cause abnormally high inflammation and cause tissue damage and contribute to AOA2/ALS4 clinical symptoms (292). *SETX* performs a function in the circadian rhythms as well. Specifically, a negative feedback loop is established by autonomous regulation of circadian clock genes PERIOD (PER) and CRYPTOCHROME (CRY). PER and CRY proteins bind to the elongation complex at *Per* and *Cry* termination sites, inhibiting the function of *SETX* and impeding RNAP II transcription termination, which leads to downregulation of gene expression (286).

OBJECTIVES

Transcription termination by Sen1 is crucial for the control pervasive transcription and plays a role in regulation of gene expression (5, 62, 164, 264). The high conservation of Sen1 across species together with evidence associating mutations in Sen1 human orthologue SETX with two several neurodegenerative diseases suggests that Sen1 proteins have maintained important biological roles during evolution (164). However, despite intense investigations on the function of Sen1 *in vivo* during the last few years, little is known about the biochemical properties and precise mechanisms of action. In a previous work from our lab some of the key features of Sen1-dependent transcription termination were revealed using a minimal *in vitro* system (95). However, many relevant aspects remain obscure. The objective of my thesis work is to characterize the mechanism of termination by Sen1. To this end, I have taken a major part in two projects that have generated the following publications:

1) Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination (*Han et al, 2017*).

In this article, we have characterized the biochemical activities of Sen1 and we have studied how these activities partake in transcription termination *in vitro*.

2) Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family (*Leonaite*, Han* et al, 2017*).

In this study, we have performed a structure-function analysis of the helicase domain of Sen1, which is sufficient for termination *in vitro*, and we have revealed important determinants of Sen1 function. We have also employed Sen1 as a model to understand the molecular origin of diseases associated with SETX mutations.

RESULTS

Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination

Zhong Han^{1,2}, Domenico Libri¹ and Odil Porrua^{1,*}

¹Institut Jacques Monod, UMR7592, Centre Nationale pour la Recherche Scientifique (CNRS), Université Paris-Diderot, Sorbonne Paris Cité, F-75205 Paris, France and ²Université Paris-Saclay, 91190 Gif sur Yvette, France

Received July 21, 2016; Revised November 10, 2016; Editorial Decision November 22, 2016; Accepted November 28, 2016

ABSTRACT

Pervasive transcription is widespread and needs to be controlled in order to avoid interference with gene expression. In *Saccharomyces cerevisiae*, the highly conserved helicase Sen1 plays a key role in restricting pervasive transcription by eliciting early termination of non-coding transcription. However, many aspects of the mechanism of termination remain unclear. In this study we characterize the biochemical activities of Sen1 and their role in termination. First, we demonstrate that the helicase domain (HD) is sufficient to dissociate the elongation complex (EC) *in vitro*. Both full-length Sen1 and its HD can translocate along single-stranded RNA and DNA in the 5' to 3' direction. Surprisingly, however, we show that Sen1 is a relatively poorly processive enzyme, implying that it must be recruited in close proximity to the RNA polymerase II (RNAPII) for efficient termination. We present evidence that Sen1 can promote forward translocation of stalled polymerases by acting on the nascent transcript. In addition, we find that dissociation of the EC by Sen1 is favoured by the reannealing of the DNA upstream of RNAPII. Taken together, our results provide new clues to understand the mechanism of Sen1-dependent transcription termination and a rationale for the kinetic competition between elongation and termination.

INTRODUCTION

Pervasive transcription is a common phenomenon both in eukaryotes and prokaryotes that consists in the massive production of non-coding RNAs (ncRNAs) from non-annotated regions of the genome (1). Pervasive transcription poses a risk that needs to be controlled since it can interfere with normal transcription of canonical genes. In *Saccharomyces cerevisiae*, a major actor in the control of

pervasive transcription generated by the RNA polymerase II (RNAPII) is the Nrd1–Nab3–Sen1 (NNS) complex, which elicits early termination of non-coding transcription and promotes degradation of the RNAs thus produced by the nuclear exosome (2–4). The NNS-complex plays two additional important roles in gene expression. One is through its function in the biogenesis of sn- and snoRNAs, which are important for splicing and rRNA modification, respectively. Both transcription termination and 3' end maturation of most sn- and snoRNAs by the exosome depend on the NNS-complex (5). In addition, in a growing number of cases the NNS-complex mediates regulation of gene expression by a mechanism of premature termination or attenuation (for review, see 6).

The NNS-complex is composed of three essential proteins: the sequence-specific RNA-binding proteins Nrd1 and Nab3 and the RNA and DNA helicase Sen1 (7,8). Nrd1 and Nab3 recognize specific motifs on the target ncRNAs, GUAA/G and UCUUG being the optimal respective binding sites (9–12). In addition, Nrd1 interacts with the C-terminal domain (CTD) of the largest subunit of RNAPII (13,14). Both the recognition of the ncRNA and the interaction with the RNAPII CTD are necessary for the recruitment of the NNS-complex to the elongation complex (EC) and for efficient transcription termination (3,15–17).

Sen1 is the only subunit of the NNS-complex that possesses a catalytic activity. It contains a conserved central domain with homology to the Superfamily 1 (SF1) helicases, flanked by N-terminal and C-terminal extensions involved in protein–protein interactions. Specifically, the N-terminal domain has been proposed to mediate the interaction with RNAPII, while the C-terminal domain contains sequences important for the nuclear localisation of Sen1 and for the interaction with the phosphatase Glc7 and Nab3 (7,18,19). Deletion of the N-terminal domain or mutation of the helicase domain provoke transcription termination defects *in vivo* (19–21). Sen1 is also the only protein of the NNS-complex that is conserved in most eukaryotes. Its human ortholog, senataxin, has also been shown to play a role in transcription termination at several model genes

*To whom correspondence should be addressed. Tel: +33 157278034; Fax: +33 157278135; Email: odil.porrua@ijm.fr

(22–25) and mutations in the most conserved regions of senataxin, the N-terminal and the helicase domains, are linked to amyotrophic lateral sclerosis type 4 (ALS4) and ataxia-ocular apraxia type 2 (AOA2) (26). In addition, introducing AOA2 mutations in the equivalent residues of Sen1 provokes growth defects (19) suggesting that these neurological disorders could be due to termination defects.

Transcription termination can be envisioned as a multi-step process involving the recruitment of termination factors, pausing of the transcribing polymerases and finally the action of the termination factors on the nucleic acid and/or the protein components of the EC to elicit its dissociation (for review, see 6). In a previous study, we have used a highly purified *in vitro* transcription termination (IVTer) system to address the role of Sen1 in transcription termination. More precisely, we analysed the capacity of Sen1 to achieve the final steps of termination (i.e. dissociation of the EC). Importantly, we have found that Sen1 alone can dismantle a paused EC *in vitro*. In addition, we have observed that, similarly to the bacterial termination factor Rho, Sen1 needs to interact with the nascent RNA and hydrolyse ATP to promote termination (27). However, many details of the termination reaction remain to be elucidated. For instance, it is unclear whether termination requires the translocation of Sen1 along the RNA, as it is the case for its bacterial counterpart. In addition, although the integrity of the helicase domain is essential for termination, the possible contribution of additional domains of Sen1 to the step of dissociation of the EC has so far not been studied. In this work we employ a variety of *in vitro* approaches to explore these aspects. First, we have performed a functional dissection of Sen1 protein to understand the role of the different domains of Sen1 in the final step of termination. Importantly, we have found that the helicase domain is sufficient to dissociate the EC and that the presence of the N-terminal domain partially inhibits termination *in vitro*, suggesting a possible function for this domain in modulating the termination activity of Sen1. Second, we have found that both full-length Sen1 and its helicase domain can translocate along single-strand DNA and RNA with 5' to 3' polarity. Our data indicate that Sen1 is a low-processivity enzyme that displays a significant preference for DNA over RNA for translocation. However, IVTer assays performed under particular conditions indicate that dissociation of the EC does not involve the interaction of Sen1 with the DNA component of the EC. Nevertheless, substituting the nascent RNA by ssDNA in IVTer reactions substantially increases the efficiency of EC dismantling. Because Sen1 translocation seems significantly more processive on ssDNA than on ssRNA, this result strongly suggests that termination requires Sen1 translocation on the RNA. Furthermore, we provide evidence that Sen1 can promote forward translocation of stalled RNAPII, similar to the bacterial termination factor Rho. Taken together, our data provide considerable advances in the understanding of the mechanisms of action of Sen1 in non-coding transcription termination.

MATERIALS AND METHODS

Strain and plasmid construction

Oligonucleotides used for the constructs are detailed in Supplementary Table S1 and plasmids are described in Supplementary Table S2. Strain YDL2556 for overexpression of wild-type *SEN1* at the endogenous locus was obtained by inserting the *GALI* promoter just upstream of *SEN1* in a protease-deficient strain (BJ2168, kind gift of B. Seraphin) using standard procedures (28,29). Plasmids for overexpression of *SEN1* variants are derived from pCM185 and were constructed by homologous recombination in yeast using previously described methods (28).

Protein purification

Saccharomyces cerevisiae RNAPII was purified from strain BJ5464 (30) by Nickel-affinity chromatography followed by anion exchange essentially as previously described (31), except for a few modifications. For this study, the ammonium sulfate precipitation step was omitted and all the buffers used in Ni²⁺-affinity chromatography were supplemented with 5 mM MgCl₂ and 2 mM β-mercaptoethanol. Recombinant His₆-tagged Rpb4/7 heterodimer was purified from *Escherichia coli* by Ni²⁺-affinity chromatography followed by gel filtration as previously described (31).

N-terminally TAP-tagged Sen1 proteins were purified from yeast, either from strain YDL2556, in the case of wild-type Sen1, or from strain BJ2168 harboring the appropriate plasmid (see Supplementary Table S2), in the case of Sen1 variants. Overexpression was induced by growing cells in YPA (for wild-type Sen1) or minimal synthetic media (for the mutant versions) containing 20 g/l of galactose for 14–24 h at 30°C. Proteins were purified using a standard TAP protocol as described before (31), with the following modifications: the concentration of NaCl in elution buffers was elevated to 500 mM to increase the elution yield and a treatment with 20 μg/ml of RNase A was included during elution from the IgG-beads at 4°C overnight. The purity of the protein preparation was monitored by SDS-PAGE followed by silver staining and by mass spectrometry (MS) analyses. No traces of Nrd1 and Nab3 or any known partners were detected and only minor amounts of unspecific contaminants (e.g. ribosomal proteins) were revealed by MS (see Supplementary Table S3). To accurately quantify the different Sen1 preparations, we loaded 5–10 μl of each preparation onto a protein gel next to four different concentrations of highly pure BSA (typically 50, 100, 200, 500 and 1000 ng). We stained the gel with SYPRO Ruby (Thermo Scientific), scanned it with a Typhoon scanner (GE Healthcare) and quantified the different protein bands using the Image Quant software (GE Healthcare). The concentration of Sen1 proteins was calculated by comparison to the BSA standard curve (signal versus the protein concentration), correcting the values according to the molecular weight of the different proteins.

ATP hydrolysis assay

ATPase assays were performed in 10 μl reactions containing 10 nM of purified Sen1 proteins in 10 mM Tris-HCl pH 7.5,

100 mM NaCl, 1 mM MgCl₂, 1 μM ZnCl₂, 1 mM DTT and 15% glycerol in the presence of 50 ng/μl of polyU. The reaction was initiated by the addition of an ATP solution (250 μM of cold ATP and 0.25 μM of 800 Ci/mmol α³²P-ATP as the final concentrations in the reaction) and allowed to proceed for a total of 12 min at 28°C. Aliquots (1.5 μl) were removed at various times and mixed with 1.5 μl of quench buffer (10 mM EDTA, 0.5% SDS). The hydrolysis products were separated by thin layer chromatography on PEI cellulose plates (Merck) in 0.35 M potassium phosphate (pH 7.5) and analysed by phosphorimaging using a Typhoon scanner (GE healthcare).

***In vitro* termination (IVTer) assay**

Termination assays were performed basically as previously described (31). Briefly, ternary ECs were assembled in a promoter-independent manner by first annealing a fluorescently labeled RNA with the template DNA and subsequently incubating the RNA:DNA hybrid with purified RNAPII. Next, the non-template strand and recombinant Rpb4/7 heterodimer were sequentially added to the mixture. The ternary ECs were then immobilized on streptavidin beads (Dynabeads MyOne Streptavidin T1 from Invitrogen) and washed with transcription buffer (TB) containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM MgCl₂, 10 μM ZnCl₂, 10% glycerol and 2 mM DTT; then with TB/0.1% Triton, TB/0.5 M NaCl and finally TB. The termination reactions were performed in TB in the presence of RNase inhibitors in a final volume of 20 μl. The amount of EC-beads used for each assay was optimized experimentally, depending on the efficiency of EC assembly for each template. For standard IVTer assays as presented in Figure 2, transcription was initiated after addition of a mixture of ATP, UTP and CTP at 1 mM each as the final concentration in the reaction, to allow transcription through the G-less cassette up to the first G of a G-stretch in the non-template strand. For assays performed in the absence of transcription as in Figures 5 and 6, the ECs contained a sufficiently long RNA to allow Sen1 loading (i.e. 44-nt long) and termination was assessed in the presence of 1 mM ATP. The reactions were incubated for 15 min at 28°C and then stopped by the addition of 1 μl of 0.5 M EDTA and the mixtures were separated into beads and supernatant fractions. The bead fractions were resuspended in 8 μl of loading buffer (1× Tris-borate-EDTA, 8 M urea) and boiled for 5 min at 95°C and RNAs in the supernatant fractions were ethanol-precipitated and resuspended in 8 μl of loading buffer. Transcripts were subjected to 10% (w/v) denaturing PAGE (8 M urea) and the gels were scanned using with a Typhoon scanner. The efficiency of termination by the different Sen1 variants was inferred from the % of nascent RNA released into the supernatant.

Duplex unwinding assay

The DNA:DNA and RNA:DNA substrates for the unwinding assays were formed by annealing a 5'-end radiolabeled short DNA oligonucleotide to either the 5' or 3' end of a longer DNA or RNA oligonucleotide (Supplementary Table S1). The DNA oligonucleotides were purchased from

Sigma and PAGE-purified before use. The RNA was obtained by *in vitro* transcription with the appropriate templates (see Supplementary Table S1) using the MEGAscript T7 kit (Ambion). Unwinding assays were performed in unwinding buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 7.5 μM ZnCl₂, 0.5 mM DTT, 10% glycerol, 0.1 mg/ml BSA) in 20 μl reactions. Sen1 proteins were preincubated with the corresponding duplex substrate and the reaction was initiated by adding a mixture containing ATP and MgCl₂ at 2 mM as the final concentrations, and an excess of unlabeled competitor oligonucleotide (0.1 μM) to trap the unwound unlabeled oligonucleotide. Aliquots were withdrawn at the indicated times and mixed with 1 volume of stop/loading buffer containing 50 mM EDTA, 1% SDS and 20% glycerol. The samples were separated by electrophoresis on a 20% native PAGE. Gels were directly exposed on phosphorimager screens overnight at -80°C and subsequently analysed using a Typhoon scanner.

Processivity assays were performed on a substrate composed of a 76-nt 5' radiolabeled DNA oligonucleotide annealed to two tandem shorter DNA oligos (20-nt and 19-nt long, respectively, see Supplementary Table S1). A 1.5-fold excess of the short oligonucleotides was used to ensure complete substrate formation. Briefly, the tandem duplex substrate (0.25 nM final concentration) was mixed with an excess of Sen1 (30 nM) in unwinding buffer and incubated for 10 min at 28°C. The reaction was initiated by adding a mixture containing ATP and MgCl₂ (2 mM final concentrations), an excess of competitor oligonucleotides (25 nM final concentration) to trap the unwound unlabeled oligonucleotides and heparin (2.5 μg/ml final concentration) to trap free/released Sen1 molecules. Reaction aliquots were withdrawn at various times and mixed with 1 volume of stop/loading buffer. Samples were separated by electrophoresis on a native 8% PAGE. Gels were dried and analyzed by phosphorimaging as before.

RESULTS

The helicase domain of Sen1 is sufficient for transcription termination *in vitro*

Sen1 contains three different structural modules: a conserved helicase domain and two adjacent regions with proposed roles in protein-protein interactions. In order to understand which domains of Sen1 are actually involved in the step of dissociation of the EC, we purified several truncated versions of Sen1 from yeast and we assessed their capacity to terminate transcription *in vitro* (Figures 1 and 2). As a control we produced a catalytic mutant (Sen1 K1363A) harboring a substitution in an essential lysine at the conserved motif I (also known as Walker A motif, Figure 1A). Because Sen1-mediated transcription termination strictly requires ATP hydrolysis both *in vivo* and *in vitro* (21,27), we first monitored the ATPase activity of the different Sen1 variants (Figure 1). As expected, the Sen1 K1363A mutant and the Sen1 N-terminal domain did not display any detectable ATPase activity. Deletion of the N-terminal domain alone or together with the C-terminal region did not significantly affect the efficiency of ATP hydrolysis, suggesting that these domains do not substantially contribute to or regulate Sen1 ATPase activity. We next analysed the capability of the

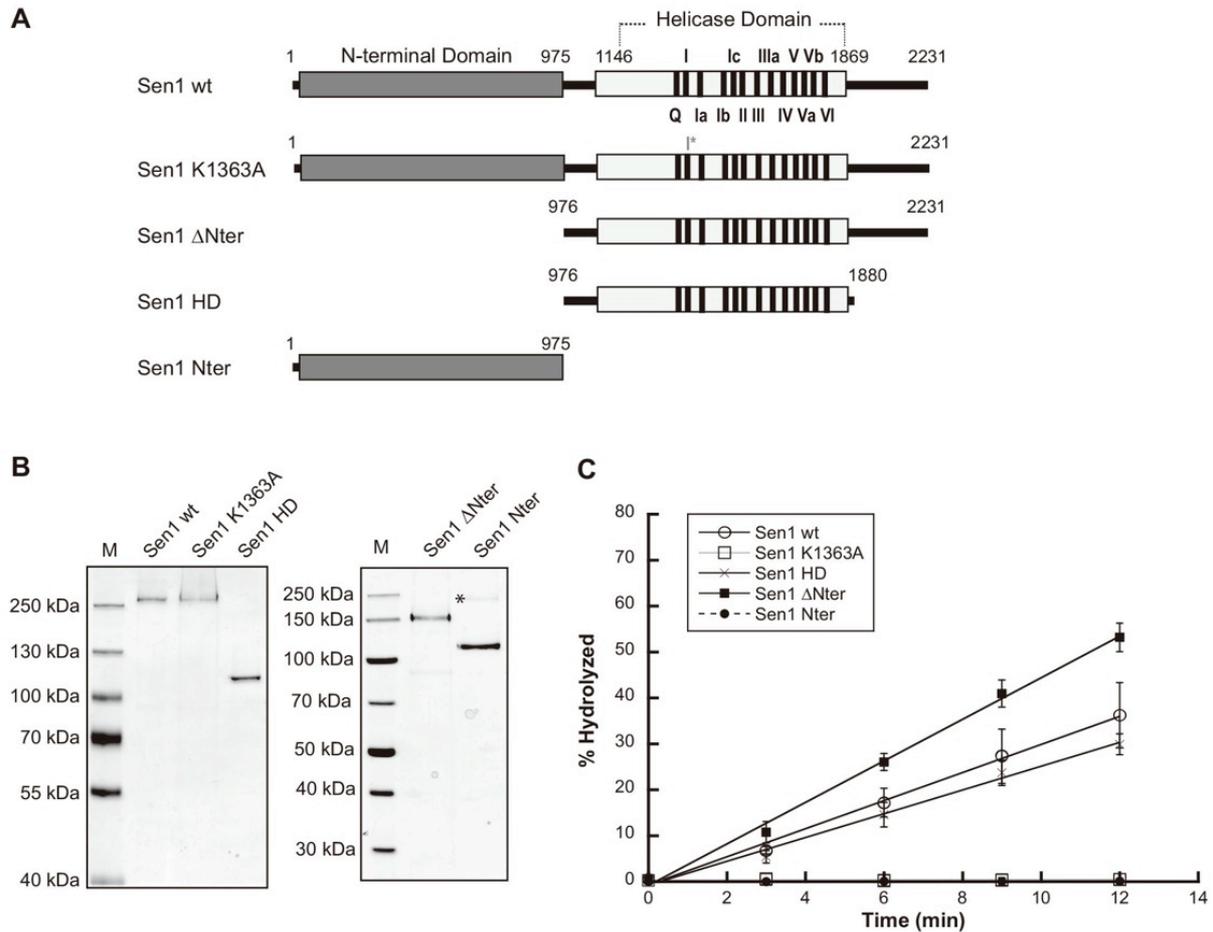


Figure 1. Analysis of the ATPase activity of different Sen1 variants. **(A)** Schematic of the Sen1 proteins analyzed. The two regions predicted to form globular domains (the N-terminal and the helicase domains) are indicated by gray bars. A black line represents the regions that are predicted to be disordered. Highly conserved motifs in SF1 helicases are indicated. The K1363A mutant harbouring a mutation at the universally conserved motif I serves as control. **(B)** Coomassie-stained protein gel showing the Sen1 variants purified from yeast (M, molecular size marker). An asterisk denotes a fraction of Sen1 Nter that exhibits anomalous migration on a SDS-PAGE (as identified by MS). **(C)** Graphical representation of the ATP hydrolyzed by the different Sen1 variants as a function of time. The values correspond to the average and standard deviation (SD) of three independent experiments.

different Sen1 versions to terminate transcription using a highly purified *in vitro* transcription-termination system containing solely purified RNAPII and Sen1 proteins (27 and Figure 2A). Briefly, in this system we assemble, in a promoter-independent manner, ternary ECs composed of RNAPII (12 subunits), the transcription templates and a short nascent RNA. ECs are immobilized on streptavidin beads via a biotin moiety present at the 5'-end of the non-template strand. The transcription templates contain a G-less cassette followed by a G-stretch in the non-template strand so that upon addition of a nucleotide mix devoid of GTP, the RNAPII transcribes until the G-stretch and is retained in the beads. The capacity of the different Sen1 variants to elicit EC dissociation is inferred from the fraction of nascent RNA released into the supernatant. Consistent with the notion that transcription termination is ATP-dependent, the Sen1 K1363A mutant and the Sen1 N-terminal domain failed to terminate transcription in our IVTer assay (Figure 2B). Importantly, the helicase domain alone was sufficient to terminate transcription *in vitro*, indi-

cating that this region retains all the activities and properties that are required for dissociation of the EC. Both Sen1 ΔNter and HD actually exhibited a somewhat more efficient termination activity, suggesting that the N-terminal region might play an inhibitory role in termination. This inhibition was only observed *in cis*, since addition of increasing concentrations of purified N-terminal domain did not affect the efficiency of termination by Sen1 ΔNter or HD *in vitro* (Supplementary Figure S1).

Sen1 can translocate on both ssRNA and ssDNA in the 5' to 3' direction

Our previous observation that Sen1-dependent termination *in vitro* requires both the interaction of Sen1 with the nascent RNA and ATP hydrolysis (27) suggests that, akin to Rho-dependent termination in bacteria, disruption of the EC by Sen1 might require Sen1 translocation along the RNA. To investigate this possibility, we decided to analyze the capacity of Sen1 proteins to unwind different nu-

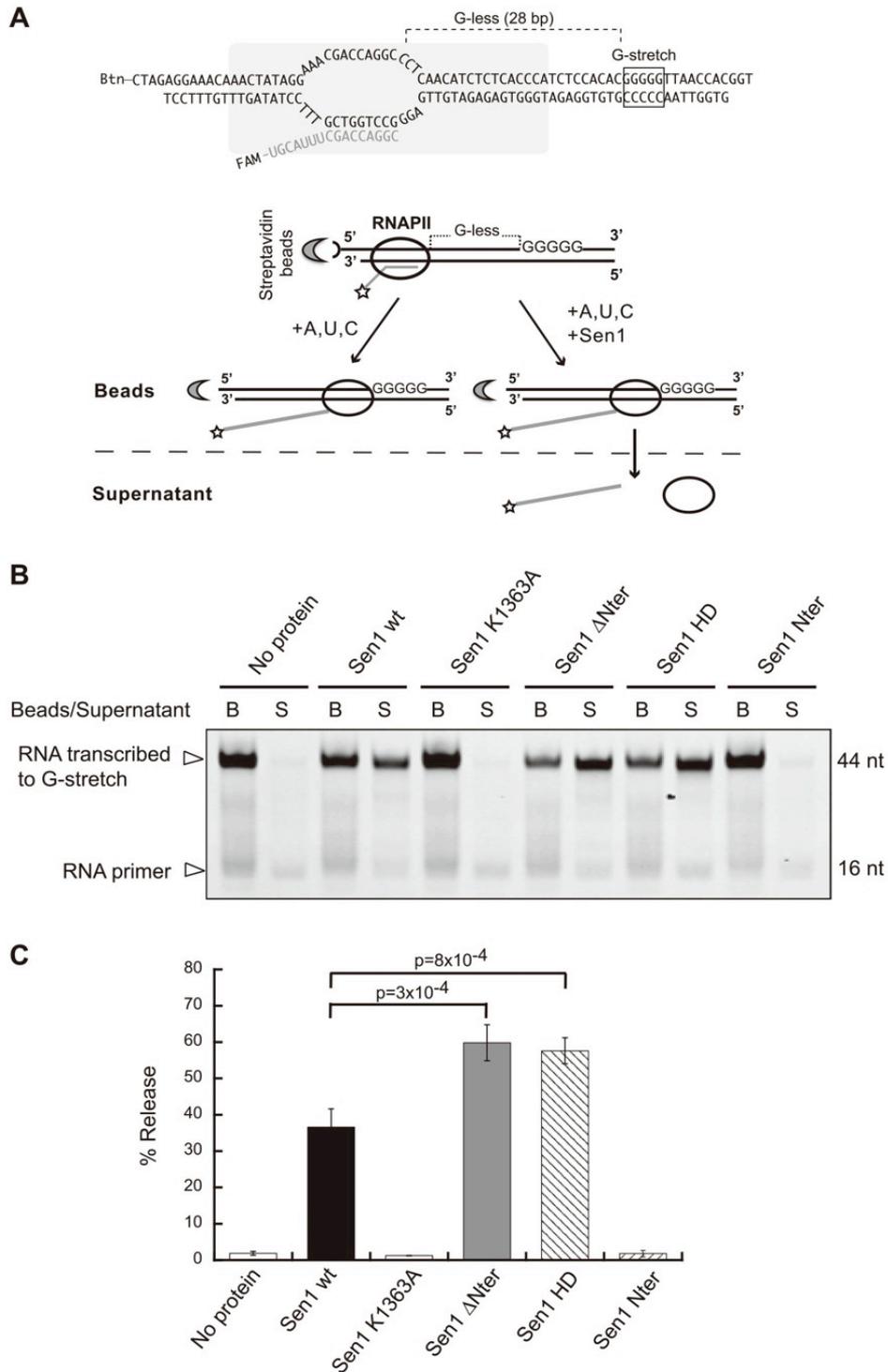


Figure 2. Analysis of the capability of Sen1 variants to terminate transcription *in vitro*. (A) Scheme of an *in vitro* transcription termination assay. A schematic of a ternary EC based on previous biochemical and structural analyses (47,48) is shown on the top. Promoter-independent assembly of ECs is performed using a 9 nt RNA:DNA hybrid that occupies the RNAPII catalytic center. Ternary ECs are attached to streptavidin beads via the 5' biotin of the non-template strand allowing subsequent separation of bead-associated (B) and supernatant (S) fractions. The RNA is fluorescently labeled with FAM at the 5'-end. The transcription template contains a G-less cassette followed by a G-stretch in the non-template strand. After adding an ATP, UTP, CTP mix, the RNAPII transcribes until it encounters the G-rich sequence. Sen1 provokes dissociation of ECs paused at the G-rich stretch and therefore the release of RNAPII and associated transcripts to the supernatant. (B) Denaturing PAGE analysis of RNAs from a representative IVTer assay in the absence and in the presence of Sen1 proteins. (C) Quantification of the fraction of transcripts released from ECs stalled at the G-stretch as a measure of the termination efficiency. Values represent the average and standard deviation of three independent experiments. The p-value associated with a *t*-test (*p*) is indicated.

cleic acid duplexes, which is an indirect way to monitor translocation. We tested both DNA:DNA and RNA:DNA duplexes containing a single-strand overhang at either the 5' or the 3' end (Figure 3). Similarly to its *Schizosaccharomyces pombe* ortholog, both full-length Sen1 and the helicase domain could dissociate DNA:DNA and RNA:DNA duplexes containing a 51–55 nt single-strand 5' overhang but were essentially inactive on duplexes with a 3' overhang, suggesting that these proteins can translocate on ssDNA or ssRNA in the 5'-3' direction (Figure 3A and B). We tested several lengths of overhang and we found that a 5' overhang as short as 5 nt can actually support unwinding (although less efficiently, data not shown). However, the full-length protein and the helicase domain exhibited a somewhat different behaviour. Whereas the helicase domain was slightly more efficient for DNA:DNA duplex unwinding, it was significantly less active than the full-length protein for RNA:DNA duplex unwinding, suggesting that additional regions in the full-length protein might be important for optimal activity on RNA:DNA duplexes. The Sen1 Δ Nter variant exhibited levels of RNA:DNA duplex unwinding activity that were similar to those of full-length Sen1, or even higher, suggesting that the C-terminal domain contains sequences that improve the activity on RNA:DNA duplexes (Supplementary Figure S2). Strikingly, both the full-length protein and the helicase domain worked much more efficiently on a DNA:DNA duplex than on an RNA:DNA duplex, since both the amplitude and the rate of the unwinding reaction were substantially higher for the former compared to the latter duplex (Figure 3C). Because we observed a similar affinity for the ssDNA and for the ssRNA or even a higher affinity for the RNA, (Supplementary Figure S3), these results suggest that Sen1 is significantly more processive on ssDNA than on ssRNA and/or very sensitive to the stability of the duplex (i.e. unwinding is less efficient for more stable duplexes). While this work was in progress, an independent study using the recombinant Sen1 helicase domain obtained similar results regarding the preference of Sen1 for ssDNA over ssRNA (32). The authors of that study performed complementary analyses and concluded that the stability of the duplex only partially accounts for the differences in unwinding efficiency, strongly suggesting that Sen1 is substantially more processive on ssDNA than on ssRNA.

Sen1 is a relatively low-processivity helicase

The former unwinding assays do not allow to directly evaluate Sen1 processivity, defined as the average number of nucleotides that Sen1 can translocate before dissociating from the nucleic acid substrate. The reasons are that: (i) these assays were performed in multiple-round conditions, such that the overall level of duplex dissociation might be the result of several proteins acting subsequently on the same substrate and (ii) a single duplex substrate does not allow to detect translocation distances shorter or longer than the length of this duplex. Because the processivity of translocation might be a critical parameter of the termination reaction mediated by Sen1, we set out to assess directly Sen1 processivity on nucleic acids. To this end, we performed single-round unwinding experiments using a substrate composed of a 76-nt ssDNA molecule annealed to two short oligos (20-nt and

19-nt) in tandem and containing a 35-nt single-strand overhang at the 5' end (Figure 4A). Because the second oligo can only be dissociated after translocation of Sen1 throughout the first duplex region, comparing the efficiency of release of the first and the second oligo allows the estimation of the processivity of translocation. In order to perform single-round assays, heparin was added together with ATP to initiate the unwinding reactions so that Sen1 reloading on the substrate was prevented. Heparin effectively prevented Sen1 reassociation with the duplex because when added before Sen1, the reaction was fully inhibited (Figure 4B, control 'H'). We observed a dramatic decrease of the overall unwinding efficiency in single-round experiments compared to multiple-round assays performed with the same substrate in the absence of heparin (Figure 4C), suggesting that heparin might outcompete the binding of a fraction of Sen1 proteins to the substrate and/or that part of Sen1 molecules do not even translocate long enough to dissociate the first duplex. We found that the reaction is relatively fast, since already after 30 sec the maximal level of unwinding was reached, (for technical reasons we could not monitor the activity at times shorter than 30 s). The fact that the fraction of unwound duplex did not increase after 30 s further validates our conditions to reproduce single-round reactions. Strikingly, we observed a clear accumulation of the product of dissociation of the first oligo and only 30% of the substrates from which the first oligo was released were fully unwound by Sen1. This indicates that >70% of Sen1 molecules dissociate from the substrate DNA after unwinding the first duplex and before reaching the 3' end, which implies that Sen1 processivity is \approx 20–40 nt. We attempted similar experiments using two tandem RNA:DNA duplexes but the resulting unwinding activity was too low to be accurately quantified (data not shown). These results indicate that Sen1 is a relatively poorly-processive translocase, which has important implications for the mechanisms of termination (see Discussion).

Termination does not require the interaction of Sen1 with the ssDNA in the transcription bubble

The striking preference of Sen1 for ssDNA over ssRNA led us to ask whether this property of Sen1 would be part of the mechanism of termination. Because the only portion of ssDNA that is present in our transcription system is the unwound DNA in the transcription bubble, we considered the possibility that, in addition to interacting with the nascent RNA, Sen1 recognizes the single stranded non-template DNA in the transcription bubble. Pulling the non-template DNA by virtue of its helicase activity might help Sen1 destabilize the EC. To explore this possibility, we set out to analyse whether Sen1 can dismantle an EC in which the non-template strand just upstream of stalled RNAPII has been substituted by a non-nucleic acid spacer. Specifically, we used two molecules of hexa-ethyleneglycol (EG) in tandem to replace the 6 nucleotides at the edge of the bubble (Figure 5). We first assessed whether the presence of the EG spacer could block the progression of Sen1 along the DNA by monitoring the capability of Sen1 to dissociate a duplex containing the EG spacer immediately upstream of the double strand region. We tested a duplex composed of the non-

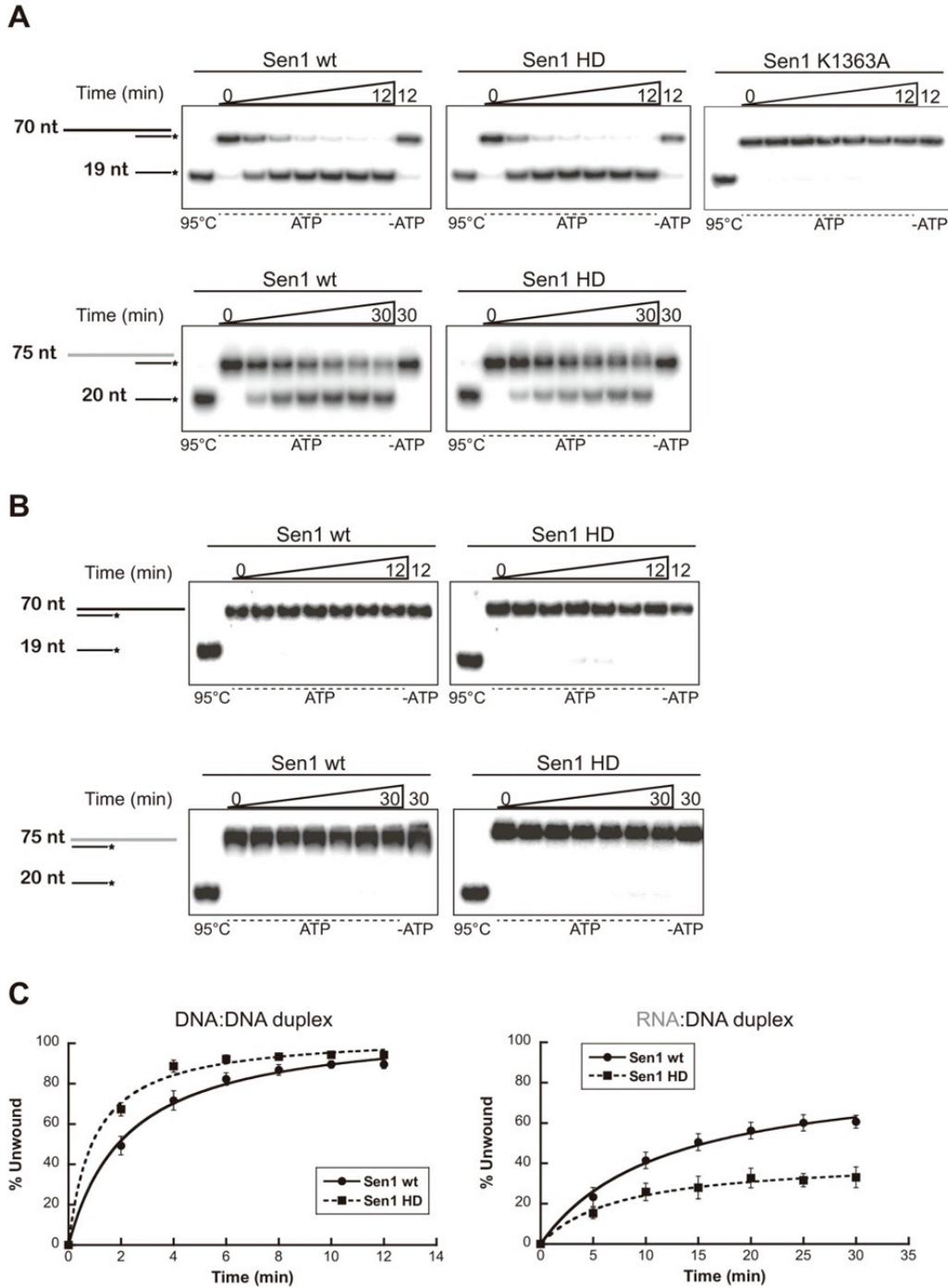


Figure 3. Nucleic acid duplex unwinding assays with full-length Sen1 (wt) and its helicase domain. (A) Representative gels showing ATP-dependent unwinding of DNA:DNA versus RNA:DNA duplexes containing a 5' single stranded overhang by Sen1 variants. The RNA molecule is shown in grey. An asterisk indicates the radioactive label at the 5' end of the short DNA molecule. The first lanes correspond to heat-denatured (95°C) samples and the last lanes are control reactions incubated with Sen1 proteins in the absence of ATP. DNA:DNA duplex unwinding reactions were performed with 0.5 nM of the substrate and 3 nM of Sen1 proteins whereas RNA:DNA duplex reactions contained 0.5 nM of the substrate and 30 nM of Sen1 proteins. Sen1 K1363A was used as a negative control. (B) Same assays as in (A) but using duplexes containing a 3' single stranded overhang. (C) Graphs showing the fraction of duplex unwound by Sen1 proteins as a function of time. Values represent the average and SD from three independent experiments. Data were fitted with Kaleidagraph to the Michaelis–Menten equation.

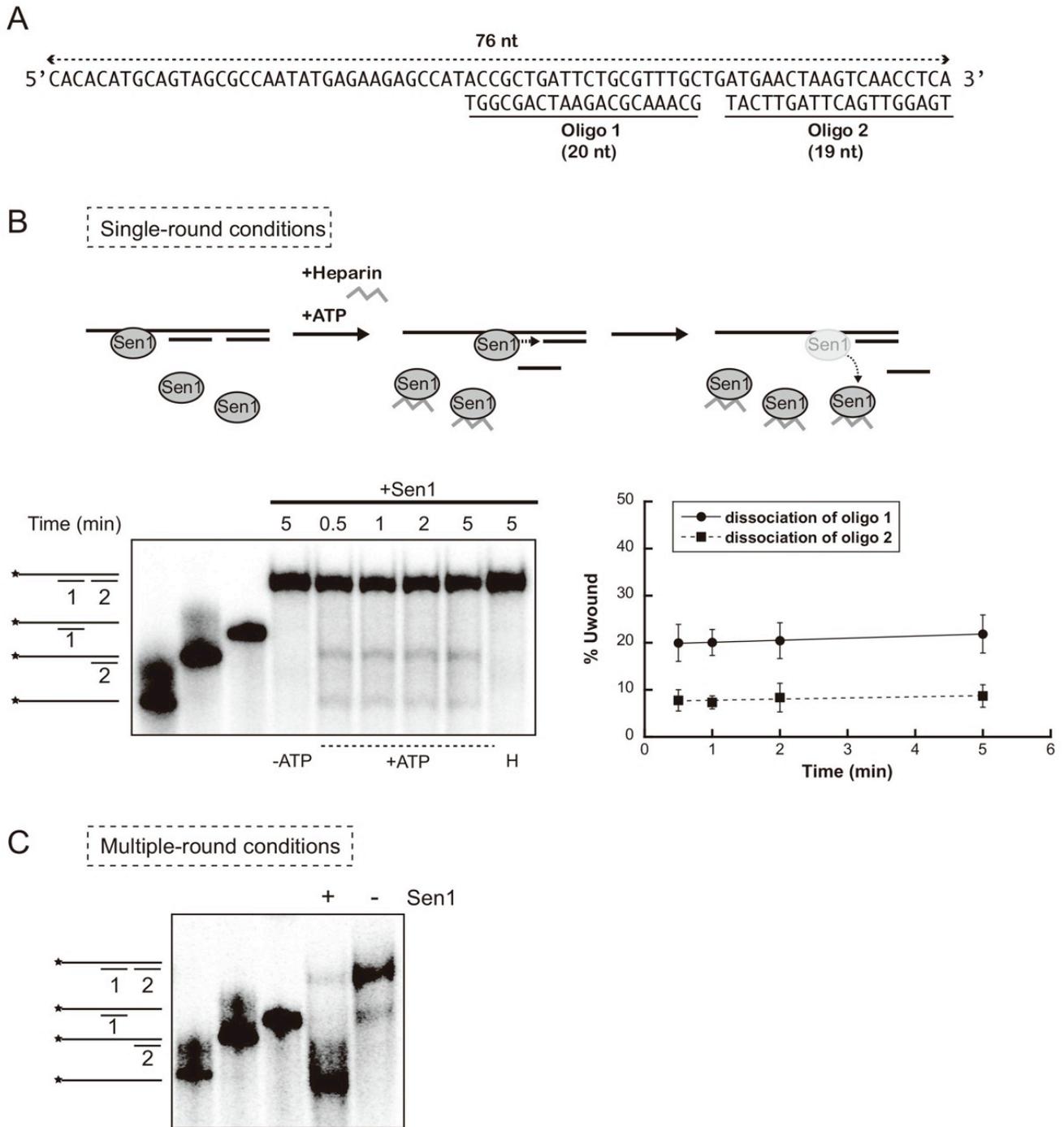


Figure 4. Analysis of the processivity of Sen1 translocation on ssDNA. (A) Sequence and structure of the substrate used in these experiments. (B) Unwinding experiments performed in single-round conditions with Sen1 HD. An experiment performed with full-length Sen1 gave identical results (data not shown). Top: Schematic description of the assay. The presence of heparin in the reaction prevents re-association of Sen1 molecules that have dissociated from the substrate, thus precluding new rounds of unwinding. Bottom-left: Representative gels of unwinding experiments. An asterisk indicates the radioactive label at the 5' end of the long DNA molecule. The first three lanes show different nucleic acid molecules that could be detected in these experiments. -ATP, control reaction incubated with Sen1 HD in the absence of ATP. H, control reaction in which heparin was added before challenging the substrates with Sen1 HD. Bottom-right: Graphs showing the fraction of oligonucleotide dissociated by Sen1 HD as a function of time. Values represent the average and SD from three independent experiments. (C) Control multiple-round unwinding assay. The reaction proceeded for 5 min. In the absence of heparin, Sen1 can undertake multiple cycles of association-translocation-dissociation that allow carrying the unwinding reaction to completion (89% of the substrate fully unwound).

template strand containing the spacer annealed to a short complementary DNA oligonucleotide. An identical duplex without the spacer served as control and was efficiently unwound by Sen1 (Figure 5A). However, the presence of the EG spacer completely abolished unwinding, indicating that translocation was effectively prevented. We next assessed whether Sen1 could dissociate a stalled EC containing the EG as described above (Figure 5B-C). In order to place the EG at the desired position, we assembled ECs with a long RNA primer (44-mer) so that 28 nt of nascent transcript are exposed to interact with Sen1 and we assessed termination in the absence of transcription. As shown in Figure 5C, in these conditions Sen1 could induce dissociation of the EC, indicating that termination does not strictly require ongoing transcription. Importantly, the presence of the spacer in the transcription bubble did not significantly affect the efficiency of EC dismantling, strongly suggesting that Sen1 does not interact with or translocate on the non-template DNA to elicit termination.

Sen1 can promote forward translocation of stalled RNAPII

The presence of the EG spacer in the non-template strand might alter the catalytic properties of the EC that, under certain conditions, might respond differently to Sen1. We assessed whether ECs containing the EG spacer were catalytically competent, by monitoring the capacity of RNAPIIs to transcribe in the presence of nucleotides (Figure 5D). Interestingly, we observed that the EG-containing ECs stalled after addition of 3 nt. This was due to the inability of the upstream DNA to rewind, because substituting the same portion of non-template strand by a sequence non-complementary to the template produced the same effect. Strikingly, the addition of Sen1 helped these stalled RNAPIIs to resume elongation and transcribe up to the G-stretch where they stalled because of the absence of GTP. Sen1 required both the nascent RNA and ATP hydrolysis to promote elongation, because neither a short, non-exposed nascent RNA nor the addition of a non-hydrolyzable ATP analogue supported EC rescue (Supplementary Figure S4). This result indicates that, by acting on the nascent RNA, Sen1 can apply a mechanical force on the EC that promotes forward translocation of RNAPII. Importantly, because in these conditions Sen1 promotes elongation rather than dissociation of RNAPII from the templates, these data also suggest that termination requires a persistent paused state of RNAPII or a particular conformation that RNAPII does not adopt when it stalls due to the lack of upstream DNA rewinding. Finally, we performed additional experiments to assess whether Sen1 could induce termination instead of promoting elongation of the former stalled complexes if elongation were precluded. To this end, we assembled ECs with transcription templates containing a sequence that in the absence of cytidine provokes RNAPII stalling after transcribing 3 nt (Figure 5B). We then analysed the capacity of Sen1 to dismantle ECs harboring normal complementary, EG-containing or non-annealing non-template strand in the absence of cytidine (Figure 5E). Interestingly, termination occurred in these conditions, but with reduced efficiency in the presence of the EG spacer or a non-complementary non-template DNA. This result sug-

gests that reannealing of the DNA at the upstream portion of the transcription bubble is necessary for fully efficient termination.

A long portion of ssDNA at either the non-template or the template strand cannot replace the nascent RNA in termination

The above results strongly suggest that the nascent RNA is the only nucleic acid molecule required for Sen1-mediated termination. We decided to explore in more detail the role of the nascent transcript in termination. The RNA could be a mere way to bring Sen1 in contact with specific surfaces of RNAPII. Alternatively, because Sen1 can promote forward translocation of stalled RNAPII, the RNA could support translocation of Sen1 towards the paused polymerase, which in particular conditions might ultimately prompt RNAPII forward movement without nucleotide addition and therefore EC destabilization. This would be analogous to the so-called 'hypertranslocation' model proposed for the bacterial termination factor Rho (for review, see 33). If Sen1 employs such a mechanism of termination, given that it can translocate along the ssDNA, it would be conceivable that a sufficiently long portion of ssDNA at the non-template strand upstream of RNAPII could support termination in the absence of exposed nascent RNA. In order to test this possibility, we performed IVTer assays under modified conditions in which most of the template strand upstream of RNAPII was omitted leaving 34 nt of single-strand non-template DNA exposed for the interaction with Sen1. As shown in Figure 6A, in the presence of a long nascent RNA Sen1 could elicit dissociation of the EC, although with somewhat reduced efficiency compared to previous assays under normal conditions. Importantly, in the absence of exposed RNA Sen1 failed to dismantle the EC, indicating that the ssDNA at the non-template strand cannot functionally replace the nascent transcript in the termination reaction. We also tested whether a 25 nt single-strand extension of the template strand downstream of the assembled EC could support termination by Sen1 (Supplementary Figure S5). Again, Sen1 could induce EC dissociation only in the presence of an exposed RNA, indicating that the single-strand template DNA even if it should support the translocation of Sen1 toward the RNAPII, cannot substitute the nascent RNA in termination.

Substitution of the extruded RNA by ssDNA increases the efficiency of termination

The above results indicate that Sen1-mediated termination specifically requires the interaction with the nascent RNA. We envisaged two possible explanations for this. Because the behaviour of Sen1 on DNA and on RNA is different (Figure 3), the first possibility is that it is the nature of the nucleic acid that matters (i.e. only the interaction with RNA promotes the termination activity of Sen1). The second possibility is that the position of the nucleic acid is crucial, for instance because it brings Sen1 to a specific region of RNAPII. In order to distinguish between these two possibilities we performed modified IVTer assays in which we compared the efficiency of termination on normal ECs

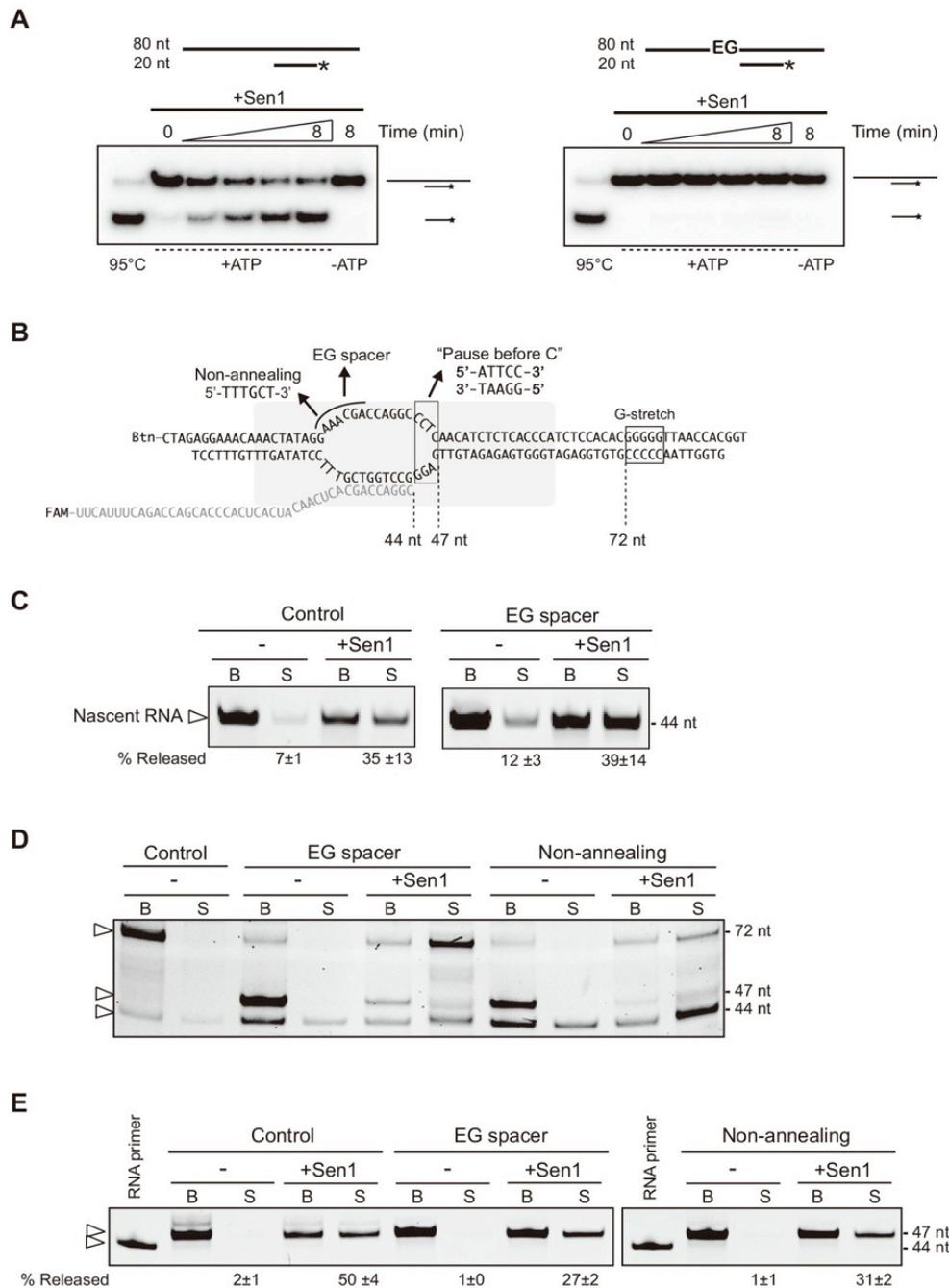


Figure 5. Analysis of the role of the ssDNA in the transcription bubble in Sen1-dependent termination. **(A)** Unwinding assays showing that Sen1 cannot translocate through ethyleneglycol (EG) spacers. Top: Scheme of the substrates used for the unwinding assays. The duplex consists in the non-template strand from the IVTer assays annealed to a 20-nt DNA oligo, leaving a 34-nt single-strand overhang at the 5' end, either in the presence or in the absence of two hexa-ethyleneglycol spacers in tandem that replace 6 nt of the non-template strand. An asterisk denotes the radioactive label at the 5' end of the short DNA oligonucleotide. Bottom: Representative gels of unwinding assays. The experiments were performed twice, with the same results. **(B)** Schematic of the different ternary ECs used as termination substrates in the experiments shown in panels **(C)**, **(D)** and **(E)**. **(C)** IVTer assays comparing the efficiency of EC dissociation in the absence and in the presence of the EG. ECs were assembled on both types of templates using a 44-mer RNA, so that the nascent RNA is sufficiently long to support termination in the absence of transcription. Values represent the average and SD from two independent experiments. **(D)** IVTer assays performed in the presence of ATP, CTP and UTP to allow transcription up to the G-stretch. The ECs harboured either a normal complementary (control), an EG-containing or a non-annealing portion of non-template DNA. Sen1 can rescue stalled ECs but can only induce EC dissociation when transcription is prevented by the absence of an incoming nucleotide (i.e. at the G-stretch). The major RNA species detected are indicated by an arrowhead. **(E)** Experiments performed with ECs containing a different sequence at the downstream portion of the transcription bubble as indicated in panel **(B)** as 'Pause before C' templates. Assays were conducted in the presence of ATP and UTP to induce RNAPII pausing after addition of 3 nt. The 44-mer RNA primer used for the assembly of ECs was loaded next to the IVTer reactions. Values correspond to the release of the 47 nt RNA species (average and SD of two independent experiments).

and on complexes assembled with a chimeric DNA-RNA in which the nucleic acid portion that is exposed for interaction with Sen1 is ssDNA instead of ssRNA (Figure 6B). We found that Sen1 can elicit dissociation of an EC with emerging ssDNA, indicating that Sen1-dependent termination specifically requires the interaction with the nascent RNA due to its position relative to RNAPII as any single-strand nucleic acid at the place of the nascent RNA can support termination. Strikingly, we reproducibly observed substantially higher efficiency of termination on ECs harbouring the chimeric DNA-RNA relative to a normal RNA transcript. Because Sen1 seems more processive on ssDNA than on ssRNA, this result suggests that the translocation processivity is a limiting parameter in the termination reaction. Importantly, this strongly suggests that termination requires the translocation of Sen1 along the RNA.

DISCUSSION

Transcription termination by Sen1 plays several crucial roles in maintaining the integrity of the functional transcriptome, from preventing genome deregulation by uncontrolled pervasive transcription to regulating gene expression by premature termination. The high conservation of Sen1 across species together with evidence linking mutations in the human Sen1 orthologue to several neurological disorders argues for evolutionarily conserved functions of major biological relevance for Sen1 proteins. However, despite intense investigations on the function of Sen1 *in vivo* during the last few years, little is known about the Sen1 biochemical properties and precise mechanisms of action. The large size of Sen1 (252 kDa) makes its purification very challenging and our attempts to produce a recombinant version of full-length Sen1 using several overexpression systems have failed. Yet, we have succeeded in purifying functional Sen1 from its natural host, *S. cerevisiae*. In a previous work, we reconstituted Sen1-dependent transcription termination using a minimal *in vitro* system and thus unveiled some of the key features of Sen1 mechanisms of action. However, many relevant questions about the mechanisms of termination remain unanswered. In this study, we characterize the helicase activity of Sen1 and investigate its role in transcription termination *in vitro*. Our results lead us to propose a model according to which, akin to Rho-dependent termination in bacteria, translocation of Sen1 on the nascent RNA is a critical step in the termination reaction.

The functional architecture of Sen1

Helicases are enzymes that often have little substrate specificity and that require additional factors to specify their biological function and regulate their activity. In the case of Sen1, previous *in vivo* studies have assigned functions to both the N-terminal and the C-terminal domain in mediating the interaction with RNAPII and Nab3, respectively (7,18). Because Sen1 itself does not exhibit any clear sequence preference in RNA binding (27,34), these interactions might be critical for Sen1 recruitment to the EC and for loading of Sen1 on non-coding transcripts harbouring the characteristic motifs of NNS-dependent targets (i.e. Nrd1 and Nab3 recognition sequences). However, whether

the N- and C-terminal domains of Sen1 participate in the step of EC dissociation or whether they play a direct role in modulating Sen1 activity have remained open questions.

In this study, we have performed a functional dissection of Sen1 to explore these aspects. Importantly, we have found that the helicase domain is sufficient for the step of dissociation of the EC, indicating that this region contains all the properties that are essential for the final step of termination. In a previous report, we showed that the capacity to dismantle an EC is not an unspecific property of any RNA helicase (27). Thus, it would be interesting to undertake structural analyses of Sen1 helicase domain to try to identify the determinants of Sen1 termination activity.

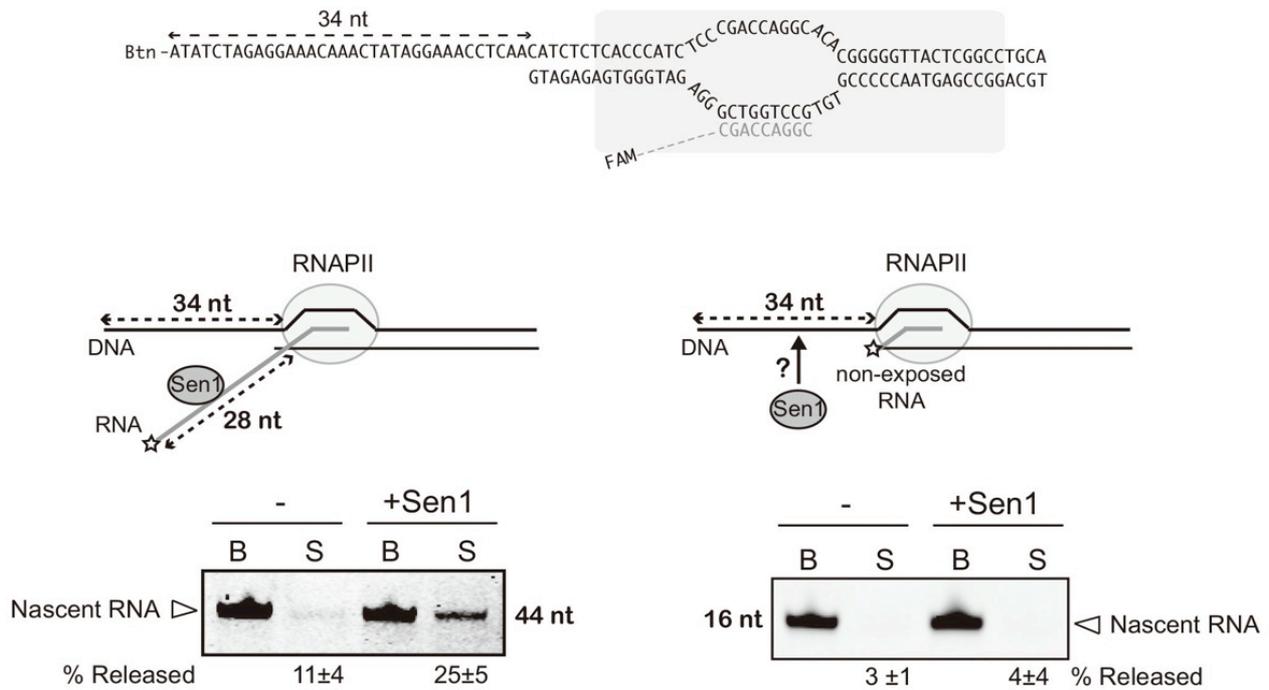
Our results therefore suggest that the N-terminal and C-terminal domains of Sen1 are implicated in earlier steps of the termination process that are only limiting *in vivo*, possibly related to the timely and specific recruitment of Sen1 to its targets as discussed above. In addition, we have found that *in vitro*, the presence of the N-terminal domain decreases the termination efficiency. This behaviour is reminiscent of that of the closest homologue of Sen1, the helicase Upf1, a key factor in the nonsense-mediated decay pathway for RNA quality control. In the case of Upf1, intra-molecular interactions mediated by N-terminal and C-terminal extensions of the helicase domain induce autorepression. The autoinhibition exerted by the N-terminal domain, also called CH domain, is relieved upon interaction with the Upf1 partner Upf2 (35,36). It remains to be tested whether the N-terminal domain of Sen1 also mediates inhibitory intra-molecular interactions. Alternatively, its mild inhibitory role might be due to its large size and its position towards the 3' end of the RNA, and therefore towards the RNAPII. It is possible that the N-terminal domain partially occludes regions of interaction with RNAPII that might be important for termination.

We have also observed that the presence of the C-terminal domain improves Sen1 helicase activity, especially on RNA:DNA duplexes. Helicases often possess additional nucleic-acid binding domains at the extensions of their core helicase domains (37). It is therefore possible that the C-terminal domain contains an additional RNA-binding domain, a hypothesis that we are currently exploring. Such possible RNA-binding activity could be particularly relevant for termination *in vivo*, where the presence of a multitude of competing RNA-binding proteins (e.g. RNP factors) might limit the access to the nascent RNA.

An intrinsically low-processivity helicase

In a previous report we did not succeed in detecting Sen1 duplex unwinding activity (27). Later, we discovered that minor contaminations of RNA in our protein preparations prevented Sen1 from exhibiting nucleic-acid binding and unwinding activity (unpublished observations). Thus, we have modified our purification protocol (see the *Methods* for details) and we have managed to observe that Sen1 can unwind both RNA:DNA and DNA:DNA duplexes in an ATP-dependent manner, but only in the presence of a 5' overhang, indicating that Sen1 can translocate along nucleic acids in the 5' to 3' direction (Figure 3). Surprisingly, our tests using tandem duplexes have revealed that Sen1 is

A



B

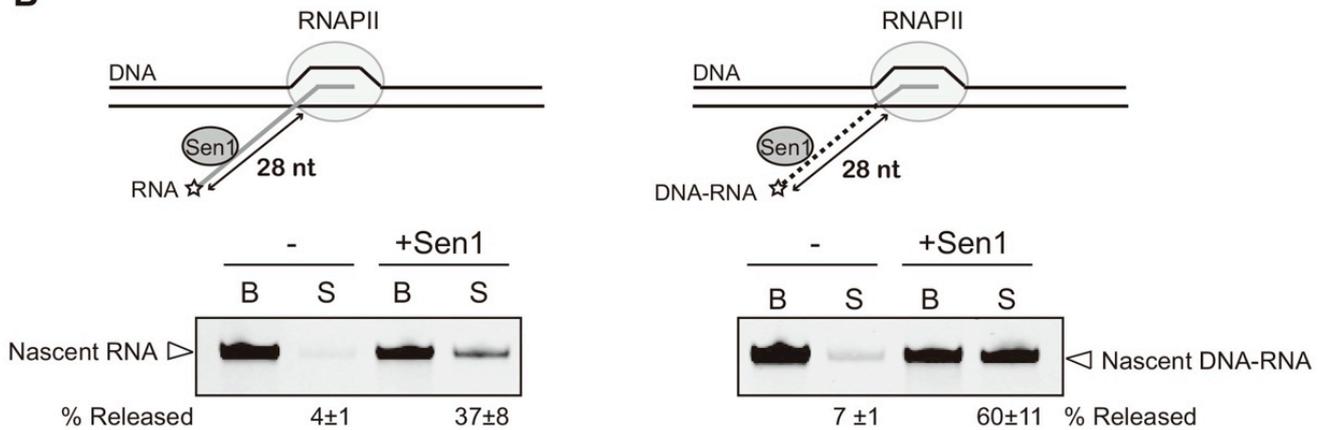


Figure 6. Analysis of the role of the nascent RNA in Sen1-mediated termination. **(A)** IVTer assays showing that a single-strand non-template DNA cannot replace the nascent RNA for termination. A schematic of the modified EC is shown on the top. The ECs were assembled using a shortened template strand so that the non-template strand upstream of the transcription bubble is single-stranded. A 15-bp region of upstream duplex DNA was maintained to avoid perturbations of the RNAPII-DNA interaction network that might lead to destabilization of the EC. **(B)** IVTer assays showing that substituting the nascent RNA by ssDNA increases the efficiency of termination. ECs were assembled using either a 44-mer RNA, as in Figure 5, or a chimeric DNA-RNA molecule in which the 16 nt at the 3' end (embedded in RNAPII) are RNA and the additional 28 nt (exposed for Sen1 interaction) are DNA (indicated by a dotted line). The sequence of the DNA templates and the structure of the EC is the same as in Figure 2A. Values represent the average and SD from three independent experiments. The *P*-value for the comparison of the ssRNA and the ssDNA in termination is 0.03.

a relatively low-processivity enzyme since more than half of Sen1 molecules dissociate before translocating ≈ 40 nt (Figure 4). This behaviour is in contrast with that of Upf1, which can translocate over more than 10 kb (38).

This feature of the Sen1 helicase activity could be important for its regulation. Given the specificity of Sen1 function, which contrasts with its low inherent capacity to discriminate between target and non-target RNAs, several mechanisms must exist to ensure timely termination at the appropriate substrates. On one hand, Sen1 protein levels are kept low (around 100 molecules/cell; 39). On the other hand, the interaction of Sen1 with Nrd1 and Nab3, which recognize specific motifs on ncRNAs, likely plays an important role in attracting Sen1 to the right targets. Low processivity might have been selected for Sen1 proteins throughout evolution to provide an additional layer of control as it leaves a relatively narrow window for the action of Sen1 and makes termination highly dependent on slowing down elongation or polymerase pausing, an event that might naturally occur or be induced by specific factors at termination regions.

Indeed, a previous *in vivo* study has revealed a kinetic competition between Sen1-mediated termination and RNAPII elongation as fast RNAPII mutants exhibit delayed termination (40). Our observation that *in vitro* Sen1 can only operate on paused polymerases (27; and Figure 5D) together with our results showing that Sen1 is an intrinsically low-processivity enzyme provide mechanistic data to better understand the former *in vivo* evidences. It remains possible that the processivity of translocation on unpaired RNA is significantly higher than on a duplex, yet *in vivo* the RNA is unlikely to exist devoid of secondary structures and bound proteins. A more direct assessment of translocation with higher resolution techniques (e.g. single-molecule systems) in the absence and in the presence of a barrier (a nucleic acid duplex or RNA-bound proteins) would be required to fully understand what is the contribution of Sen1 translocation rate to the kinetics of termination and to estimate the impact of competing processes such as elongation and RNP assembly.

Our results demonstrating the relatively low processivity of Sen1 on RNA are also relevant in the light of a previously-proposed role for Sen1 and senataxin in removing R-loops that form during transcription when the nascent RNA hybridizes with the DNA template (22,25,41). In humans R-loops that form *in vivo* are very long (even more than 1 kb in humans, see 42). Sen1, given its low processivity, could not in principle unwind such long structures. However, it is possible that in yeast R-loops are much shorter. Alternatively, the accumulation of R-loops observed upon Sen1 mutation could be an indirect consequence of impaired transcription termination. Finally, an interesting possibility is that *in vivo* the processivity of Sen1 is significantly enhanced by the interaction with other factors, for instance with the Sen1 partners Nrd1 and Nab3.

The mechanism of transcription termination

One of the most relevant but also most elusive questions in the field is what is the precise mechanism of termination by Sen1. Our previous results showing that ATP-dependent

termination by Sen1 requires its interaction with the nascent RNA together with *in vivo* evidence of a kinetic competition between termination and transcription elongation suggested a mechanism of termination involving Sen1 translocation on the nascent RNA (27,40). However, a formal proof for this model has been missing. Here, using classical duplex unwinding assays, we have proved that Sen1 can translocate on single stranded nucleic acids. Our observation that Sen1 is at least one order of magnitude more efficient on DNA:DNA than on RNA:DNA duplexes together with data from another group obtained with a recombinant helicase domain strongly suggest that Sen1 is more processive on ssDNA than on ssRNA.

The prominent preference of Sen1 for ssDNA over ssRNA led us to ask whether Sen1 might need to interact with the ssDNA in the transcription bubble to elicit termination. However, the results of our IVTer assays with modified transcription templates strongly suggest that this is not the case (Figure 5). Nevertheless, we have taken advantage of the different behaviour of Sen1 on ssDNA and ssRNA to obtain further insight into the mechanism of termination. We have shown that the substitution of the extruded nascent RNA by ssDNA increases the efficiency of termination *in vitro* (Figure 6B). Because the affinity of Sen1 for the ssDNA and the ssRNA is similar (Supplementary Figure S3, note that we have measured the affinity for the same sequences used in the experiments in Figure 6B), we exclude the possibility that this enhanced termination is the result of better recruitment to the EC. Instead, because Sen1 seems more processive on ssDNA than ssRNA, the enhancement of termination obtained with ssDNA is most likely due to increased processivity, supporting the idea that Sen1-mediated termination requires Sen1 translocation on the nascent RNA (Figure 7).

The efficient activity of Sen1 on ssDNA remains striking and might be relevant for its previously proposed roles in DNA repair and replication. Indeed, the N-terminal domain of Sen1 has been shown to mediate the interaction with the nucleotide excision repair (NER) factor Rad2 (18) and to be critical for efficient transcription coupled repair (TCR), an NER subpathway (43). Our data showing that Sen1 can promote forward translocation of stalled RNAPIIs *in vitro* (Figure 5D) suggest that, similar to the bacterial transcription-repair coupling factor Mfd (44), *in vivo* Sen1 might facilitate DNA repair by displacing the lesion-stalled polymerases and promoting the recruitment of the NER machinery. In addition, given the efficient DNA:DNA unwinding activity of Sen1, it is tempting to speculate an additional role for Sen1 in subsequent steps of the DNA repair process involving the exposure of ssDNA regions. Furthermore, a previous report suggested that Sen1 associates with replication forks and promotes their progression through highly transcribed regions (45). The DNA helicase activity of Sen1 might be important for this role in replication.

The behavior of Sen1 is reminiscent of the bacterial terminator factor Rho. Three alternative models have been proposed to explain the mechanism of termination by Rho. The 'hypertranslocation' model posits that the powerful helicase activity of Rho would exert a pushing force on the RNAP that would force it to translocate without transcription and

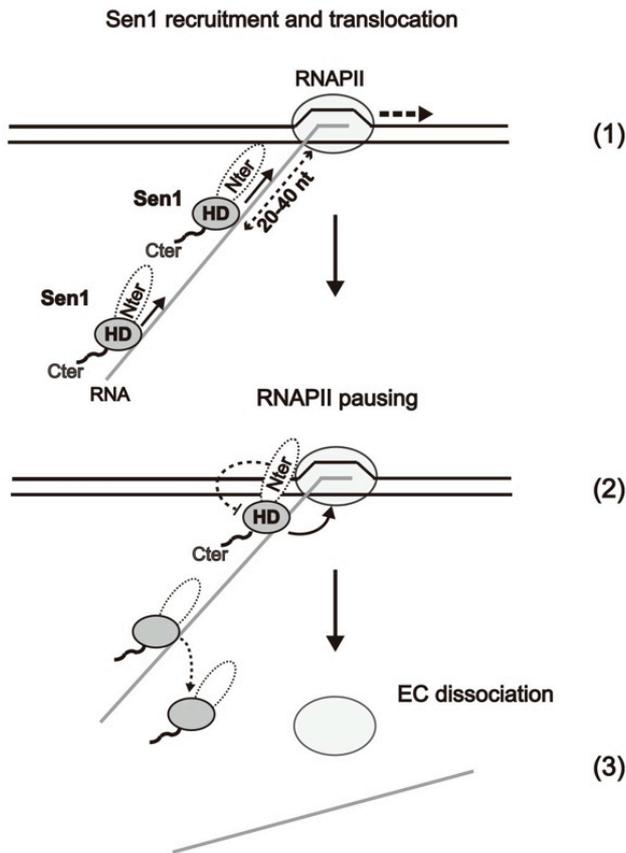


Figure 7. Model for the mechanisms of transcription termination by Sen1. After recruitment to the EC by the other NNS components, Nrd1 and Nab3, and/or the direct interaction with RNAPII, Sen1 must be loaded onto the nascent RNA in close proximity to the RNAPII. Sen1 molecules loaded more than 40 nt upstream of the RNAPII would dissociate before accomplishing termination. Sen1 translocation along the RNA allows dismantling of the EC in a reaction that requires exclusively the action of the helicase domain of Sen1 and likely involves ‘hypertranslocation’ of RNAPII. Termination strictly requires RNAPII pausing. In the absence of additional factors, the Sen1 N-terminal domain partially inhibits the transcription termination activity of the helicase domain.

this would ultimately provoke EC dissociation. The ‘hybrid shearing’ model postulates that Rho would rather disrupt the hybrid in the catalytic centre by dragging the nascent RNA out of the EC with a similar destabilizing effect as in the previous model. Finally, the allosteric model proposes that the simultaneous interaction of Rho with the RNAP and the nascent transcript allows Rho to induce a conformational change in RNAP that ultimately leads to EC disassembly (33 and references therein).

Our finding that Sen1 can apply a mechanical force on a stalled EC allowing the RNAPII to resume elongation would support a ‘hypertranslocation’ model (Figure 5D). In addition, we have observed that disrupting the base-pairing of the duplex region immediately upstream of the transcription bubble decreases the efficiency of termination (Figure 5E). This suggests that bubble rewinding provides additional energy that assists EC dissociation by Sen1, as previously shown for its bacterial counterpart Rho (46). Yet, given the extensive interactions of RNAPII with the

upstream duplex DNA in the EC (47), we cannot exclude the possibility that the modifications introduced in the non-template DNA alter the conformation of the EC making it a less favourable substrate for termination. Finally, additional evidence supports a more complex model for termination with not only mechanical but also allosteric components. Indeed, in our previous study we showed that Upf1 could not terminate RNAPII transcription in our *in vitro* assay, indicating that a processive RNA translocase activity is not sufficient for termination. In addition, we found that Sen1 cannot terminate transcription by the bacterial RNAP, suggesting that specific recognition of RNAPII might be required for termination (27). Understanding whether Sen1 interacts with specific regions of RNAPII to elicit termination and the nature of the molecular transitions leading to EC dissociation remain a major challenge and is the focus of our future research.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Karine dos Santos for her attempts in producing recombinant Sen1 and for helpful discussions. We thank the proteomic platform of Institut Jacques Monod for technical assistance. We thank all other members of the Libri lab for their input and Marc Boudvillain, Terence Strick, Anne-Lise Haenni and Bronislava Leonaite for critical reading of the manuscript.

FUNDING

Centre National de la Recherche Scientifique; Agence National pour la Recherche [ANR-08-Blan-0038-01 and ANR-12-BSV8-0014-01 to D.L.]; Fondation pour la Recherche Medicale (programme Equipes 2013 to D.L.); PhD fellowship from the China Scholarship Council (to Z.H.). Funding for open access charge: Agence National pour la Recherche [ANR-12-BSV8-0014-01]. *Conflict of interest statement.* None declared.

REFERENCES

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

7. Nedeá,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.
8. Vasiljeva,L. and Buratowski,S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell*, **21**, 239–248.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. Finkel,J.S., Chinchilla,K., Ursic,D. and Culbertson,M.R. (2010) Sen1p performs two genetically separable functions in transcription and processing of U5 small nuclear RNA in Saccharomyces cerevisiae. *Genetics*, **184**, 107–118.
21. Steinmetz,E.J., Warren,C.L., Kuehner,J.N., Panbehi,B., Ansari,A.Z. and Brow,D.A. (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell*, **24**, 735–746.
22. Skourti-Stathaki,K., Proudfoot,N.J. and Gromak,N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell*, **42**, 794–805.
23. Suraweera,A., Lim,Y., Woods,R., Birrell,G.W., Nasim,T., Becherel,O.J. and Lavin,M.F. (2009) Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum. Mol. Genet.*, **18**, 3384–3396.
24. Wagschal,A., Rousset,E., Basavarajaiah,P., Contreras,X., Harwig,A., Laurent-Chabalier,S., Nakamura,M., Chen,X., Zhang,K., Meziane,O. *et al.* (2012) Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell*, **150**, 1147–1157.
25. Zhao,D.Y., Gish,G., Braunschweig,U., Li,Y., Ni,Z., Schmitges,F.W., Zhong,G., Liu,K., Li,W., Moffat,J. *et al.* (2016) SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*, **529**, 48–53.
26. Bennett,C.L. and La Spada,A.R. (2015) Unwinding the role of senataxin in neurodegeneration. *Discov. Med.*, **19**, 127–136.
27. 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.
28. 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 Chichester Engl.*, **14**, 953–961.
29. 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.
30. Kireeva,M.L., Lubkowska,L., Komissarova,N. and Kashlev,M. (2003) Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from Saccharomyces cerevisiae. *Methods Enzymol.*, **370**, 138–155.
31. Porrua,O. and Libri,D. (2015) Characterization of the mechanisms of transcription termination by the helicase Sen1. *Methods Mol. Biol. Clifton NJ*, **1259**, 313–331.
32. 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.
33. Peters,J.M., Vangeloff,A.D. and Landick,R. (2011) Bacterial transcription terminators: the RNA 3'-end chronicles. *J. Mol. Biol.*, **412**, 793–813.
34. Creamer,T.J., Darby,M.M., Jamonnak,N., Schaugency,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.
35. Chakrabarti,S., Jayachandran,U., Bonneau,F., Fiorini,F., Basquin,C., Domcke,S., Le Hir,H. and Conti,E. (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell*, **41**, 693–703.
36. Fiorini,F., Boudvillain,M. and Le Hir,H. (2013) Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic Acids Res.*, **41**, 2404–2415.
37. Fairman-Williams,M.E., Guenther,U.P. and Jankowsky,E. (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.*, **20**, 313–324.
38. Fiorini,F., Bagchi,D., Le Hir,H. and Croquette,V. (2015) Human Upf1 is a highly processive RNA helicase and translocate with RNP remodelling activities. *Nat. Commun.*, **6**, 7581.
39. 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.
40. Hazelbaker,D.Z., Marquardt,S., Wlotzka,W. and Buratowski,S. (2012) Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol. Cell*, **49**, 55–66.
41. Mischo,H.E., Gomez-Gonzalez,B., Grzechnik,P., Rondon,A.G., Wei,W., Steinmetz,L., Aguilera,A. and Proudfoot,N.J. (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol Cell*, **41**, 21–32.
42. Yu,K., Chedin,F., Hsieh,C.-L., Wilson,T.E. and Lieber,M.R. (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat. Immunol.*, **4**, 442–451.
43. Li,W., Selvam,K., Rahman,S.A. and Li,S. (2016) Sen1, the yeast homolog of human senataxin, plays a more direct role than Rad26 in transcription coupled DNA repair. *Nucleic Acids Res.*, **44**, 6794–6802.
44. Howan,K., Smith,A.J., Westblade,L.F., Joly,N., Grange,W., Zorman,S., Darst,S.A., Savery,N.J. and Strick,T.R. (2012) Initiation of transcription-coupled repair characterized at single-molecule resolution. *Nature*, **490**, 431–434.
45. Alzu,A., Bermejo,R., Begnis,M., Lucca,C., Piccini,D., Carotenuto,W., Saponaro,M., Brambati,A., Cocito,A., Foiani,M.

- et al.* (2012) Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell*, **151**, 835–846.
46. Park, J.-S. and Roberts, J.W. (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 4870–4875.
47. Barnes, C.O., Calero, M., Malik, I., Graham, B.W., Spahr, H., Lin, G., Cohen, A.E., Brown, I.S., Zhang, Q., Pullara, F. *et al.* (2015) Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Mol. Cell*, **59**, 258–269.
48. Kireeva, M.L., Komissarova, N., Waugh, D.S. and Kashlev, M. (2000) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J. Biol. Chem.*, **275**, 6530–6536.

Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination

Zhong Han, Domenico Libri and Odil Porrua

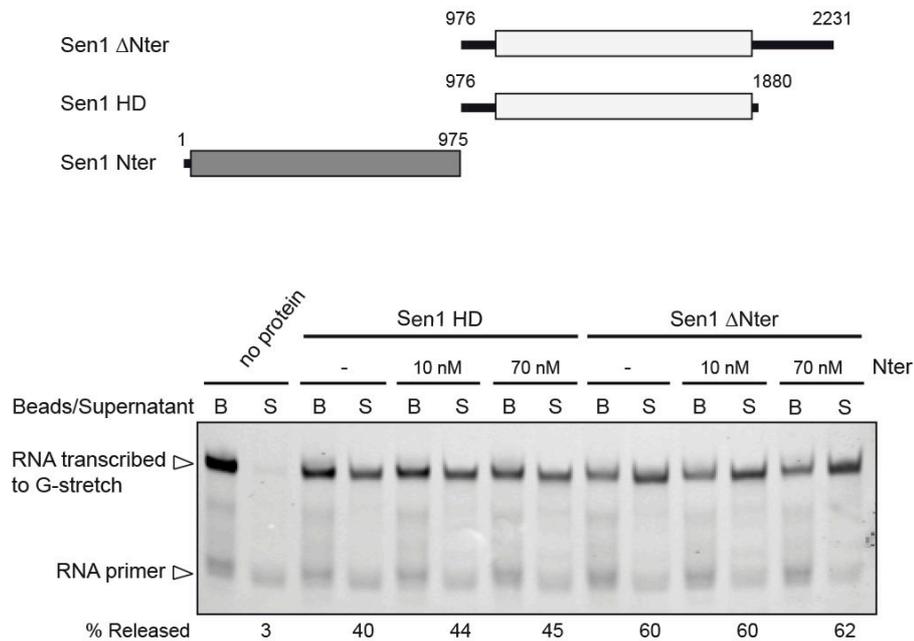
Supplementary material:

Figures S1-S5.

Tables S1-S3.

Supplementary methods.

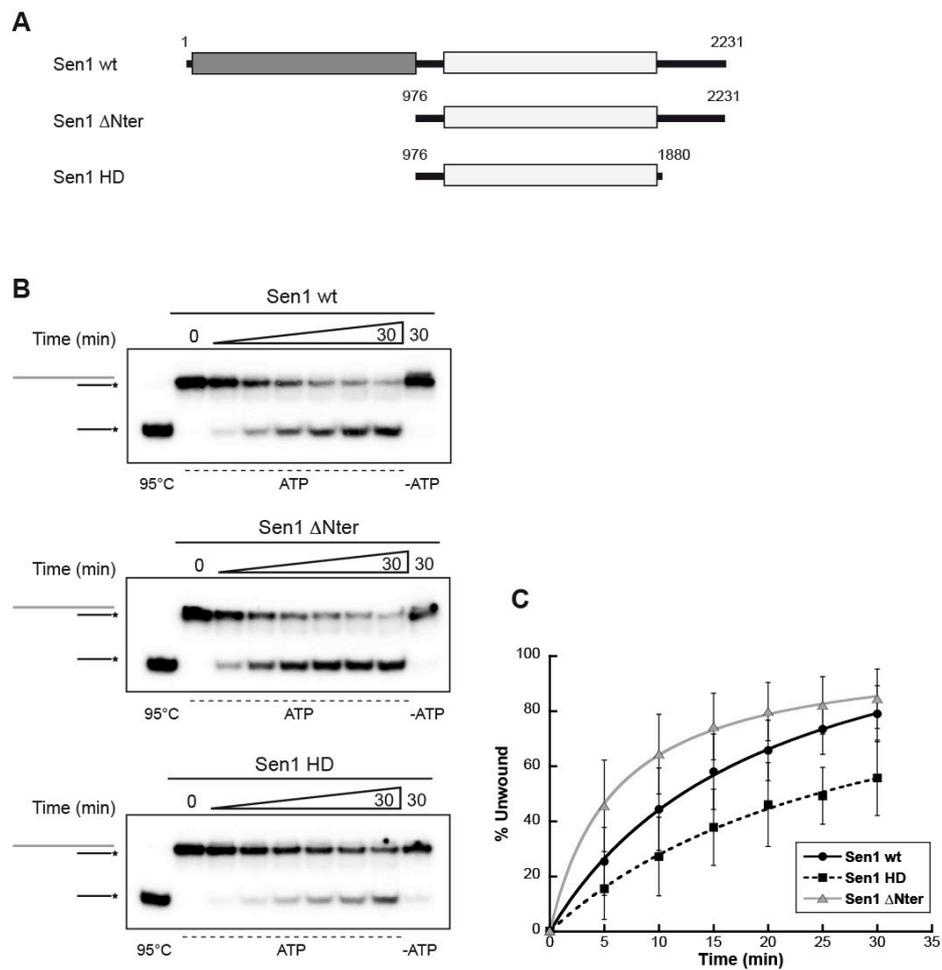
Supplementary references.



Han et al, figure S1

Figure S1: Sen1 N-terminal domain does not inhibit termination by Sen1 HD and Sen1 ΔNter *in trans*.

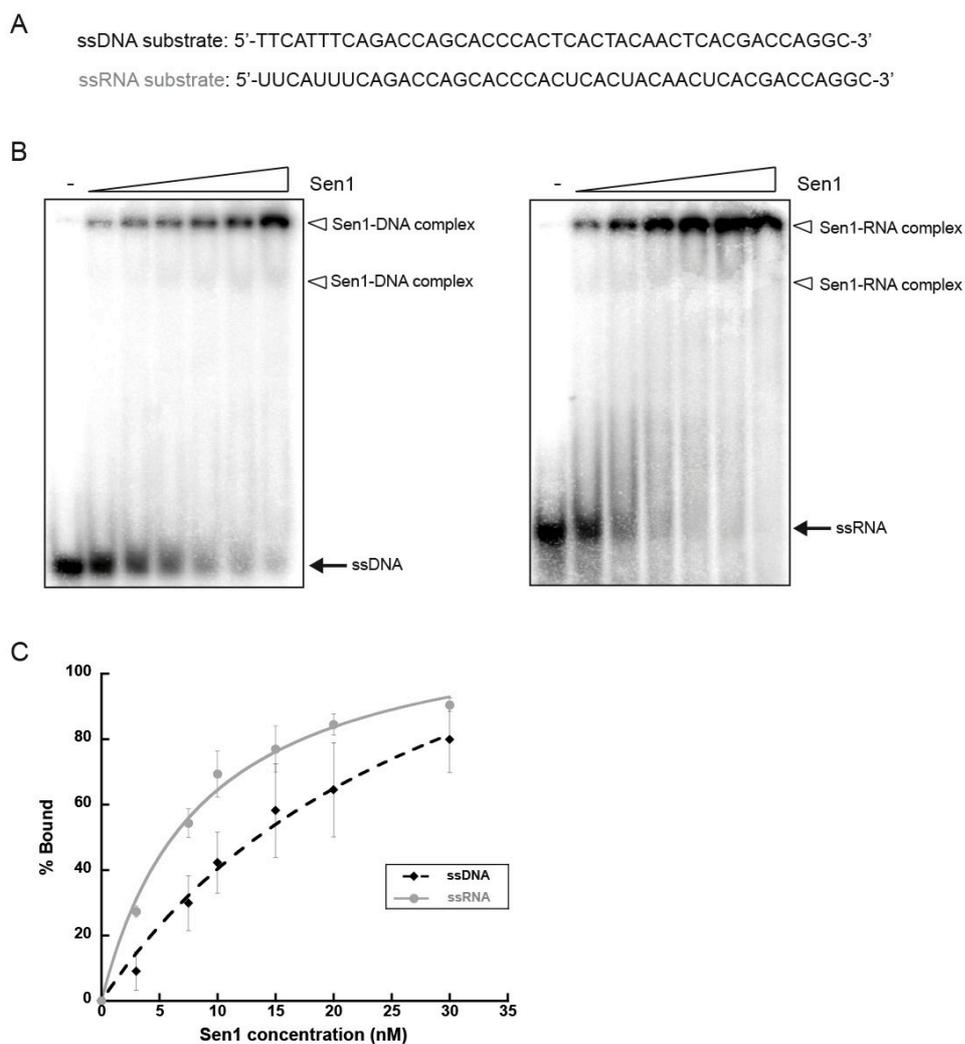
Top: schematic of the Sen1 proteins used. Here we added 10 nM of Sen1 HD or ΔNter (note that IVTer reactions in figure 2 were instead performed in the presence of 20 nM of Sen1 proteins). Bottom: representative gel of IVTer assays with the Sen1 variants indicated in the presence of increasing amounts of Sen1 N-terminal domain in the reaction. The experiment was performed twice with essentially the same result.



Han et al, figure S2

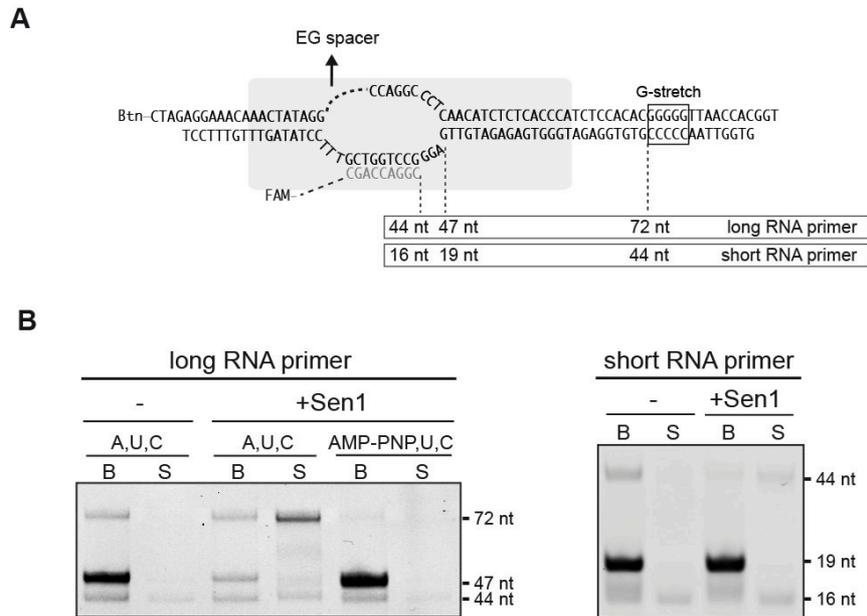
Figure S2: The presence of Sen1 C-terminal domain is necessary for optimal unwinding activity.

A) Schematic of Sen1 proteins used in these experiments. **B)** Representative gels showing ATP-dependent unwinding of RNA:DNA duplexes as in figure 3. The RNA strand is shown in grey. An asterisk indicates the radioactive label at the 5' end of the short DNA molecule. **C)** Graphs showing the fraction of duplex unwound by Sen1 proteins as a function of time. Values represent the average and SD from three independent experiments.



Han et al, figure S3

Figure S3: Analysis of the interaction of Sen1 with single-strand nucleic acids. A) Sequence of the DNA and RNA molecules used as substrates in electrophoretic mobility shift assays with full-length Sen1 . **B)** Native PAGE illustrating the interaction of Sen1 with ssDNA and ssRNA. **C)** Graph showing the percentage of nucleic acids bound as a function of Sen1 concentration. Values represent the average and SD from three independent experiments.



Han et al, figure S4

Figure S4: Rescue of stalled ECs by Sen1 requires ATP hydrolysis and an accessible nascent RNA. **A)** Schematic of the different ECs used as termination substrates indicating the length of the RNAs produced in the different conditions. The non-template strand contains an EG spacer at the indicated position that induces RNAPII stalling after transcription of 3 nt. **B)** Left: IVTer assays performed on ECs harbouring a long portion of extruding RNA that allows the interaction with Sen1. Experiments performed in the presence of ATP or its non-hydrolysable analogue AMP-PNP. This analogue can be used by RNAPII for transcription elongation (1). Right: IVTer tests on ECs carrying a short RNA primer that is embedded in RNAPII and therefore not accessible to Sen1. Experiments were performed in the presence of ATP, UTP and CTP.

Supplementary table S2: List of plasmids used in this work.

Name	Description/use	Source
pBS1761	Apr, oriColE1; plasmid bearing a TRP1 cassette for N-terminally TAP-tagging	(2)
pDL576	pCM185-derived plasmid harbouring the PGAL1-TAP portion of the TAP-tagging cassette from pBS1761	This work
pDL667	pDL576-derived plasmid to express the TAP-tagged Sen1 N-terminal domain (residues 1-975) from the PGAL1 promoter.	This work
pDL668	pDL576-derived plasmid to express the TAP-tagged Sen1 helicase domain (residues 976-1880) from the PGAL1 promoter.	This work
pDL670	pDL576-derived plasmid to express the TAP-tagged Sen1 Δ Nter (residues 976-2231) from the PGAL1 promoter.	This work
pDL761	pDL576-derived plasmid to express the TAP-tagged <i>sen1 K1363A</i> mutant from the PGAL1 promoter.	This work
pFL39-Sen1	pFL39 harbouring <i>SEN1</i> expressed from its own promoter. Used as a template for PCR amplifications.	This work

Supplementary methods

Mass spectrometry analyses

A full-length Sen1 preparation at 0.2 μM (10 μl) was subjected to low migration on a NUPAGE 4-12% acrylamide gel (Invitrogen) and stained with Coomassie blue (Instant blue, Expedeon). Three bands covering the entire protein content of the sample were excised and digested separately by sequencing grade trypsin (12.5 $\mu\text{g/ml}$; Promega, Madison, WI, USA) in 20 μl of 25 mmol/L NH_4HCO_3 overnight at 37°C. The protein digests were pooled and analyzed with an Orbitrap Fusion Tribrid coupled to a Nano-LC Proxeon 1000 equipped with an easy spray ion source (all from Thermo Scientific). Peptides were separated by chromatography with the following parameters: Acclaim PepMap100 C18 pre-column (2 cm, 75 μm i.d., 3 μm , 100 Å), Pepmap-RSLC Proxeon C18 column (50 cm, 75 μm i.d., 2 μm , 100 Å), 300 nl/min flow rate, gradient from 95 % solvent A (water, 0.1% formic acid) to 35% solvent B (100 % acetonitrile, 0.1% formic acid) over a period of 97 minutes, followed by column regeneration for 23 min, giving a total run time of 2 hours. Peptides were analyzed in the Orbitrap cell, in full ion scan mode, at a resolution of 120,000 (at m/z 200), with a mass range of m/z 350-1550 and an AGC target of 4×10^5 . Fragments were obtained by high collision-induced dissociation (HCD) activation with a collisional energy of 30%, and a quadrupole isolation window of 1.6 Da. MS/MS data were acquired in the Orbitrap cell in the top-speed mode, with a total cycle of 3 seconds at a resolution of 30,000, with an AGC target of 1×10^4 , a dynamic exclusion of 50 seconds and an exclusion duration of 60 seconds. Precursor priority was highest charge state, followed by most intense. Peptides with charge states from 2 to 8 were selected for MS/MS acquisition. The maximum ion accumulation times were set to 250 ms for MS acquisition and 60 ms for MS/MS acquisition in parallelization mode. MS/MS data were processed with an in-house Proteome Discoverer 1.4 using MASCOT node and the Swissprot Yeast Database. The mass tolerance was set to 7 ppm for precursor ions and 0.5 Da for fragments.

Electrophoretic mobility shift assay (EMSA)

EMSA were carried out in 10 μl reactions containing 10 mM Tris-HCl pH 7.5, 70 mM NaCl, 2 mM MgCl_2 , 7.5 μM ZnCl_2 , 10% glycerol and 1 mM DTT. The nucleic acid substrates (1 nM final concentration) were incubated with increasing concentrations of purified Sen1 for 30 min on ice prior to electrophoresis on 15 cm-long 5% polyacrylamide gels in TBE at 200V for 1.5 hours. The gels were dried and analyzed by phosphorimaging.

Supplementary references

1. Bunick,D., Zandomeni,R., Ackerman,S. and Weinmann,R. (1982) Mechanism of RNA polymerase II--specific initiation of transcription in vitro: ATP requirement and uncapped runoff transcripts. *Cell*, **29**, 877–886.
2. Puig,O., Caspary,F., Rigaut,G., Rutz,B., Bouveret,E., Bragado-Nilsson,E., Wilm,M. and Séraphin,B. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods San Diego Calif*, **24**, 218–229.

Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family

Bronislava Leonaitė^{1,2,5}, Zhong Han^{3,4,5}, Jérôme Basquin¹, Fabien Bonneau¹, Domenico Libri³, Odil Porrua^{3,*} and Elena Conti^{1,*}

¹ Max Planck Institute of Biochemistry, Munich, Germany

² Graduate School of Quantitative Biosciences, Ludwig-Maximilians-University, Feodor-Lynen-Str. 25, 81337 Munich, Germany

³ Institut Jacques Monod, Centre Nationale pour la Recherche Scientifique (CNRS), UMR 7592, Université Paris Diderot, Sorbonne Paris Cité, F-75205 Paris, France.

⁴ Université Paris-Saclay, 91190 Gif sur Yvette, France.

⁵ These authors contributed equally to the work

*Correspondence: conti@biochem.mpg.de, odil.porrua@ijm.fr

Running title: *S. cerevisiae* Sen1 helicase structure

Key words: RNA helicases, non-coding transcription, transcription termination

Final character counts: 65,788

ABSTRACT

The superfamily 1B (SF1B) helicase Sen1 is an essential protein that plays a key role in the termination of non-coding transcription in yeast. Here, we identified the ~90 kDa helicase core of *S. cerevisiae* Sen1 as sufficient for transcription termination *in vitro* and determined the corresponding structure at 1.8 Å resolution. In addition to the catalytic and auxiliary subdomains characteristic of the SF1B family, Sen1 has a distinct and evolutionary conserved structural feature that “braces” the helicase core. Comparative structural analyses indicate that the “brace” is essential in shaping a favorable conformation for RNA binding and unwinding. We also show that subdomain 1C (the “prong”) is an essential element for 5'-3' unwinding and for Sen1-mediated transcription termination *in vitro*. Finally, yeast Sen1 mutant proteins mimicking the disease forms of the human orthologue, senataxin, show lower capacity of RNA unwinding and impairment of transcription termination *in vitro*. The combined biochemical and structural data thus provide a molecular model for the specificity of Sen1 in transcription termination and more generally for the unwinding mechanism of 5'-3' helicases.

INTRODUCTION

In yeast, there are two major transcription termination pathways. In the case of canonical protein-coding genes, Pol II normally terminates transcription via the cleavage and polyadenylation factor (CPF) complex, yielding stable mature mRNAs that are then exported to the cytoplasm (reviewed in Mischo & Proudfoot, 2013). In the case of non-coding RNAs, such as cryptic unstable transcripts (CUTs) and small nucleolar RNAs (snoRNA), Pol II terminates transcription via a non-canonical pathway that is coupled to RNA degradation (reviewed in Jensen *et al*, 2013). This non-canonical termination pathway depends on the Nrd1-Nab3-Sen1 (NNS) complex (reviewed in Arndt & Reines, 2015; Porrua & Libri, 2015a). Nrd1 and Nab3 form a heterodimer (Carroll *et al*, 2007) that underpins the substrate specificity of the NNS complex (Wlotzka *et al*, 2011; Porrua *et al*, 2012; Schulz *et al*, 2013). Nrd1-Nab3 also recruits the Trf4 subunit of TRAMP (Tudek *et al*, 2014), a major cofactor of the RNA-degrading exosome in the nucleus (LaCava *et al*, 2005; Vanacova *et al*, 2005; Wyers *et al*, 2005). TRAMP polyadenylates the 3' end of its RNA substrates and feeds them to the nuclear exosome, resulting in the complete or partial 3'-5' degradation of CUTs and snoRNAs, respectively (Allmang *et al*, 1999; Wyers *et al*, 2005). While Nrd1-Nab3 couples the NNS complex to RNA degradation, Sen1 is the key enzyme in the transcription termination reaction (Porrua & Libri, 2013). Sen1 has also been shown to be involved in termination of short protein-coding genes (Steinmetz *et al*, 2006) and inactivation of Sen1 leads to the accumulation of R-loops (RNA:DNA hybrids that form during transcription when the nascent RNA invades the DNA template) (Mischo *et al*, 2011).

Sen1 is an RNA/DNA helicase and is the only evolutionary conserved subunit of the NNS complex. The human orthologue of yeast Sen1, senataxin (SETX), is associated with neurological pathologies: recessive mutations in the *SETX* gene cause ataxia with oculomotor apraxia type 2 (AOA2) and dominant mutations provoke amyotrophic lateral sclerosis type 4 (ALS4) (reviewed in Bennett & La Spada, 2015). Disease mutations cluster in the two most conserved regions of SETX, the N-terminal domain and the helicase domain. Like its yeast orthologue, SETX has been assigned functions in transcription termination and in the control of R-loop formation (Suraweera *et al*, 2009; Skourti-Stathaki *et al*, 2011; Zhao *et al*, 2016).

Sen1 belongs to the superfamily 1B (SF1B) Upf1-like family of helicases together with Upf1 and IGHMBP2. Structural studies of Upf1 and IGHMBP2 have shown the presence of a common domain organization (Cheng *et al*, 2007; Clerici *et al*, 2009; Chakrabarti *et al*, 2011; Lim *et al*, 2012). SF1B RNA helicases contain two RecA domains (RecA1 and RecA2) with the classical helicase motifs involved in nucleic acid binding and ATP hydrolysis. Helicases of this family also contain two SF1B-specific subdomains (1B and 1C) that modulate RNA binding (Cheng *et al*, 2007; Clerici *et al*, 2009; Chakrabarti *et al*, 2011; Lim *et al*, 2012). SF1B helicases bind nucleic acids with the same polarity as all other RNA-dependent ATPases, i.e. with the 3' end at RecA1 and the 5' end at RecA2 (reviewed in Pyle, 2008; Ozgur *et al*, 2015). However, the directionality of duplex unwinding of the SF1B superfamily is opposite to that of processive SF2 helicases, which unwind duplexes in the 3'-5' direction (Büttner *et al*, 2007). For example, Upf1 has been shown to be a highly processive 5'-3' RNA helicase (Bhattacharya *et al*, 2000; Fiorini *et al*, 2015). Similarly, Sen1 uses ATP hydrolysis to unwind RNA or DNA duplexes in the 5'-3' direction (Kim *et al*, 1999; Martin-Tumasz & Brow, 2015).

Sen1 is expected to have a similar domain organization as compared to Upf1 and IGHMBP2 and a similar 5'-3' unwinding mechanism. However, Sen1 also has a distinct function, namely the ATPase-dependent ability of promoting transcription termination *in vitro* (Porrua & Libri, 2013). In this work, we used biochemical and structural approaches to dissect the elements that underpin the general 5'-3' unwinding and the distinctive properties of Sen1. We demonstrate the existence of features that are specific to Sen1 and integrate structural knowledge into a refined model for 5'-3' unwinding and transcription termination.

RESULTS AND DISCUSSION

Identification of the active helicase core of *S. cerevisiae* Sen1

S. cerevisiae Sen1 is a multi-domain protein of about 250 kDa (2231 residues). Analysis of the Sen1 amino-acid sequence by secondary structure and fold recognition programs HHpred (Söding *et al*, 2005) and Phyre2 (Kelley *et al*, 2015) predicted the presence of an α -helical region (amino acids 1-975) at the N-terminus

followed by a Upf1-like helicase region and a low-complexity segment of roughly 300 residues at the C-terminus (Figure 1A). To obtain a soluble fragment of Sen1 encompassing the helicase core, we subcloned a fragment of Sen1 cDNA coding for residues 976-1904 into a vector for bacterial expression. After purification, we noticed that the Sen1 fragment was smaller than expected (Figure EV1A). Mass-spectrometry analyses suggested that endogenous protease digestion had likely occurred at the Sen1 N-terminus during purification.

For structural studies, we expressed and purified a fragment of Sen1 encompassing residues 1095-1904 (hereafter referred to as Sen1_{Hel}), with the shorter N-terminus previously identified by the Brow's group (Martin-Tomasz & Brow, 2015). Sen1_{Hel} showed levels of RNA-dependent ATPase activity similar to those of full-length Sen1 purified from yeast (*y*Sen1 FL). A mutant version with the E1591Q substitution in the highly conserved helicase motif II, which is essential for ATP hydrolysis, was inactive (Figure 1A). Next, we analyzed the helicase activity of Sen1_{Hel} by performing duplex unwinding assays. As a substrate we used a 19-nt RNA:DNA duplex harboring a 25-nt single-stranded RNA extension at either the 5'-end or the 3'-end. Similar to the full-length protein, Sen1_{Hel} displayed significant duplex unwinding activity on the substrate containing a 5' single-strand overhang (Figure 1B), while no activity was detected on the substrate containing a single-stranded extension at the 3' end (Figure EV1B), consistent with 5'-3' helicase activity. We note that Sen1_{Hel} has similar unwinding properties as compared to the Sen1 fragment that had been previously characterized by the Brow's group (Martin-Tomasz & Brow, 2015).

Next, we tested whether recombinant Sen1_{Hel} retains the capability of the full-length protein to terminate transcription *in vitro* (Porrua & Libri, 2013). We assembled ternary elongation complexes (ECs) in a promoter-independent manner using purified RNA Pol II, DNA transcription templates and a short RNA oligonucleotide primer that forms a 9-bp duplex with the template strand and occupies the active center of the polymerase (Figure 1C, left panel). We biotinylated the non-template strand to allow the association of ECs with streptavidin beads and the subsequent separation of Pol II molecules (and associated transcripts) engaged in transcription from those that have been released from the DNA templates after transcription termination. In order to assess the capacity of Sen1 proteins to elicit termination, we monitored the efficiency

of release of nascent RNA into the supernatant. Similar to the full-length protein, Sen1_{Hel} promoted the release of a significant fraction of nascent transcripts (Figure 1C, right panel). This activity was dependent on the integrity of the Sen1_{Hel} active site, as no release was observed with the Sen1_{Hel} E1591Q mutant (Figure 1C, right panel). We concluded that the ~90 kDa helicase core is essentially responsible for the transcription termination properties of Sen1 *in vitro*.

Overall structure of the helicase core of yeast Sen1

We crystallized Sen1_{Hel} in the presence of ADP and determined the structure by single-wavelength anomalous dispersion (SAD) phasing using the signal from the sulfur atoms in the native protein. The structure was refined at 1.8 Å resolution with *R*_{free} of 18%, *R*_{factor} of 15% and good stereochemistry (Table 1) (Figure EV1C). Overall, Sen1_{Hel} has a domain organization similar to that of the helicase core of Upf1 (Upf1_{Hel}, also known as Upf1-ΔCH) (Cheng *et al*, 2007; Clerici *et al*, 2009; Chakrabarti *et al*, 2011) as well as IGHMBP2 (IGHMBP2_{Hel}) (Lim *et al*, 2012) (Figure 2). In the Sen1_{Hel} - ADP structure, the two RecA domains are positioned side by side, separated by a cleft about 10 Å wide (Figure 2). In this open conformation, RecA2 is rotated about 30° from the position it acquires in the closed conformation that is typical of helicases in the active RNA-ATP-bound state (Linder & Lasko, 2006; Pyle, 2008). Two mutations shown to affect the function of Sen1 *in vivo* map to the RecA domains and are likely to cause partial unfolding of the protein due to unfavorable electrostatic clashes (G1747D, DeMarini *et al*, 1992 and E1597K, Steinmetz & Brow, 1996).

The ADP nucleotide binds at the bottom of the cleft and interacts directly with the RecA1 domain (Figure 2 and Figure EV2A). Similarly to what has previously been observed in the structures of nucleotide-bound Upf1_{Hel} (Chakrabarti *et al*, 2011), the adenine ring is sandwiched between an apolar surface of RecA1 and an aromatic residue (Tyr1655) that is present in the linker connecting RecA1 to RecA2 and is part of motif IIIa (Fairman-Williams & Jankowsky, 2012). In addition, the conserved side chain of Gln1339 forms a bidentate hydrogen-bond interaction with the N6 and N7 moieties of the adenine ring (Figure EV2A). A similar Gln-based specificity

determinant for adenine nucleotides was originally identified in the so-called Q motif of DEAD-box proteins (Cordin *et al*, 2004). Although at the sequence level the corresponding Q motif of Upf1-like helicases is also present upstream of motif I, at the three-dimensional level it forms part of a different structural element as compared to the Q motif of DEAD-box proteins and is instead similar to the Q motif found in the Ski2-like family of DExH-box proteins (Sengoku *et al*, 2006; Weir *et al*, 2010; Jackson *et al*, 2010).

Sen1_{Hel} also contains two accessory subdomains that extend on the surface of RecA1: a ~160-residue insertion known as subdomain 1B and a 120-residue insertion known as subdomain 1C. Subdomain 1B contains two antiparallel helices that pack against each other and against the side of RecA1 with extensive hydrophobic interactions, forming the so-called “stalk” (Chakrabarti *et al*, 2011) (Figure 2). At the top of the “stalk”, subdomain 1B features a β -barrel fold (the “barrel”), which hovers over RecA1. Subdomain 1C is also formed by α -helices, forming a prong-like feature. As outlined below, the “stalk”, the “barrel” and the “prong” show specific differences when comparing Sen1_{Hel} to Upf1_{Hel} and IGHMBP2_{Hel}. The major difference from other known SF1B helicases, however, is the presence in Sen1_{Hel} of a ~50-residue N-terminal segment that we refer to as the “brace” (Figure 2, left panel).

Sen1 is a SF1B helicase with distinct structural features

Structural comparisons of Sen1_{Hel} with Upf1_{Hel} and IGHMBP2_{Hel} reveal several distinct features in the accessory subdomains of Sen1_{Hel} (Figure 2). First, the ordered portion of the “prong” is shorter with respect to Upf1_{Hel} and IGHMBP2_{Hel}. Second, the “barrel” has a more elaborate topology with respect to Upf1_{Hel} and IGHMBP2_{Hel}, with additional helical turns. Perhaps more importantly, the “barrel” is connected to the “stalk” helices by short linkers as compared to Upf1_{Hel} and IGHMBP2_{Hel}. The short linkers appear to restrict the conformational space that the Sen1_{Hel} “barrel” domain can sample. This spatial restriction is further compounded by the interactions with the N-terminal “brace” described below.

The “brace” (residues 1097-1149) fastens three different structural features of the helicase core, namely RecA1, the “stalk” and the “barrel”. A first short α -helix ($\alpha 1$) inserts aliphatic side chains (Leu1109 and Arg1108) into a hydrophobic surface groove formed between the RecA1 domain (Ala1573, Ala1578 and Tyr1606) and a “stalk” helix (Tyr1303) (Figure 3A). The polypeptide chain then continues with a second α -helix ($\alpha 2$) sandwiched between the “stalk” helices and the “barrel”. Hydrophobic residues on one side of helix $\alpha 2$ (Leu1116, Ile1120 and Trp1123) make apolar interactions with residues of the “stalk” (with Ile1291 and with Leu1162, Trp1166 and Leu1169, respectively) (Figure 3B). Hydrophobic residues on the other side of helix $\alpha 2$ (Tyr1117 and Leu1121) are engaged in van der Waals interactions with aliphatic side chains of the “barrel” (with Leu1216, Leu1244 and Lys1246). After another hydrophobic interaction with the “barrel” (Tyr1125 with Val1218 and Val1283), the “brace” makes an 180° turn via the clustering of Pro1132 with Trp1123 and Trp1166. It then continues to connect to the ascending helix of the “stalk” via van der Waals interactions (Val1143 with Phe1147 with Tyr1153) (Figure 3B). Overall, the “brace” buries 2400 Å² of the Sen1_{Hel} surface area with evolutionary conserved interactions (Figure 3C).

The “brace” appears to stabilize the overall fold of the protein. *In vitro*, deletions of the N-terminal 1128-1149 residues resulted in an insoluble protein, likely because removal of the “brace” led the hydrophobic residues on RecA1, the “stalk” and the “barrel” to be exposed to solvent (unpublished observations). Consistently, a deleted variant of Sen1 lacking the N-terminal 1134 residues does not support yeast viability while deletion of the N-terminal 1088 residues (which leaves the “brace” intact) results in termination defects but does not lead to lethality (Chen *et al*, 2014). Moreover, a W1166S mutant also has been shown to be defective *in vivo* (Chen *et al*, 2014). Thus, the “brace” is an important element for Sen1 function both *in vitro* and *in vivo*.

The structural analysis of Sen1_{Hel} suggests that the “brace” firmly connects the “barrel” to the “stalk” helices. Although the conformation we observe might also be partly stabilized by lattice contacts, the extensive intramolecular interactions of the “brace” and the short connections described above appear to genuinely restrain the position of the “barrel” on top of the RecA1 domain. We compared the position of the

“barrel” of Sen1_{Hel} with that of other SF1B helicases (Figure 4A and Figure EV2B). In the apo structure of UPF1_{Hel}, the “barrel” is in close contact with RecA1 and interferes with the RNA-binding surface (Cheng *et al*, 2007). Upon RNA binding, the Upf1 “barrel” moves away from RecA1, effectively sandwiching part of the nucleic acid (Chakrabarti *et al*, 2011). In the apo structure of Sen1_{Hel}, however, the “barrel” is already displaced from the equivalent RNA-binding surface and adopts the position (albeit not the orientation) observed in the RNA-bound state of Upf1_{Hel}.

The RNA-binding properties of Sen1

Generally, the RNA-binding interactions of Sen1_{Hel} are expected to be similar to those observed in the structure of Upf1_{Hel} - U₆ - ADP:AlF₄⁻ (Chakrabarti *et al*, 2011). Sen1_{Hel} shares the conserved RecA2 residues that interact with ribonucleotides 1 and 2, at the 5' end of the RNA (Tyr1752_{Sen1} and Arg1813_{Sen1}, corresponding to yeast Tyr732_{Upf1} and Arg794_{Upf1}) (Figure EV2C). It also shares residues in the RecA1-RecA2 linker that approaches ribonucleotides 3 and 4, in the central portion of the RNA (Pro1622_{Sen1} and Thr1623_{Sen1}, corresponding to Pro604_{Upf1} and Val605_{Upf1}). Finally, it shares residues in the RecA1 domain and in the “stalk” that interact with ribonucleotides 5 and 6, at the 3' end of the RNA (Thr1289_{Sen1}, Arg1293_{Sen1} and Asn1413_{Sen1} corresponding to Thr356_{Upf1}, Arg360_{Upf1} and Asn462_{Upf1}) (Figure EV2B). In support of this structural analysis, a Sen1_{Hel} T1289A, R1293A double mutant was impaired in RNA binding, ATP hydrolysis and transcription termination *in vitro* (Figure EV2D-F).

A major difference between Sen1 and Upf1 is that the latter contains an additional CH domain, which regulates RNA binding in an allosteric manner. In Upf1_{CH-Hel}, binding of the CH domain onto RecA2 effectively pushes the “barrel” towards the “prong”, creating a binding site for two additional ribonucleotides (7 and 8) at the 3' end of the RNA (Chakrabarti *et al*, 2011) (Figure 4A). No CH domain is present in the sequence of Sen1. We assessed the RNA-binding properties of Sen1_{Hel} in RNase protection assays (Figure 4B). As previously reported, Upf1_{CH-Hel} protected longer RNA fragments (~10-11 ribonucleotides) than Upf1_{Hel} (~9 ribonucleotides) (Chamieh *et al*, 2008), consistent with the structural data (Chakrabarti *et al*, 2011). In similar RNase

protection assays, Sen1_{Hel} protected 11-ribonucleotide long fragments (Figure 4B), more similar to Upf1_{CH-Hel} than to Upf1_{Hel}. This finding raises the question of which intrinsic feature of Sen1_{Hel} mimics the effect of the separate CH domain of Upf1_{CH-Hel} on the RNA footprint. The structural analysis pointed to the “brace”, which appears to pull the “barrel” of Sen1_{Hel} towards the “prong” and pre-position it for RNA binding. If our hypothesis is correct, the extended footprint we observed should be directly dependent on the “prong”.

In order to test this prediction, we engineered two mutants in Sen1_{Hel} in which we truncated only the upper part of the “prong” that is missing in the present structure (Sen1_{Hel}Δ1471-1538 or Sen1_{Hel}ΔUP for upper “prong” deletion) or the entire solvent-exposed portion of the “prong” (Sen1_{Hel}Δ1461-1554 or Sen1_{Hel}ΔLP for lower “prong” deletion) (Figure 5A). In RNase protection assays, Sen1_{Hel}ΔUP had an RNA footprint similar to that of the wild-type protein, but Sen1_{Hel}ΔLP resulted in a smaller footprint, with the protection of fragments of ~9 ribonucleotides (Figure 5B). This pattern is consistent with our model that the “barrel” and the “prong” of Sen1_{Hel} come together to interact with additional nucleotides at the 3' end.

The 5'-3' RNA-unwinding features of Sen1

As mentioned above, both the SF1 and the SF2 helicases bind RNA with the same directionality across the two RecA domains. In the case of SF2 Ski2-like helicases (which unwind RNA duplexes processively in the 3'-5' direction), the RNA-unwinding element has been identified as a β-hairpin that protrudes from RecA2, where the 5' end of a bound RNA resides, and separates the strands of an incoming duplex as it enters the helicase core (Büttner *et al*, 2007; Ozgur *et al*, 2015). For SF1B Upf1-like RNA helicases, which unwind RNAs processively in the opposite direction (5'-3'), we reasoned that the unwinding element might reside on the opposite side of the helicase to act on an incoming duplex (i.e. near the RecA1 domain). We analyzed the structure of Sen1_{Hel} and compared it to those of Upf1_{Hel}, Upf1_{CH-Hel} and IGBMH2_{Hel} to identify a possible structural element on the RecA1 side of the molecule, where the 3' end of a bound RNA resides. We noticed conserved structural features in the lower part of the “prong” (including Arg537_{Upf1} and Lys331_{IGHMBP2},

corresponding to Arg1552_{Sen1}) (Figure EV2B). In the Upfl_{CH-Hel} structure, this positively-charged residue approaches the very 3' end of the RNA (Chakrabarti *et al*, 2011).

In order to test a possible role of the “prong” in duplex unwinding, we analyzed the activities of the “prong” mutants described above. Both the Sen1_{Hel}ΔUP and the Sen1_{Hel}ΔLP mutants not only retained the footprint in RNase protection assays (Figure 5B), but also retained RNA-dependent ATPase activity (Figure 5C). Deletion of the disordered part of the “prong” in Sen1_{Hel}ΔUP did not decrease the capacity to dissociate an RNA:DNA duplex, but rather enhanced it. In contrast, the full deletion in Sen1_{Hel}ΔLP abolished the unwinding activity (Figure 5D). This was not due to a decrease in the affinity for the RNA, since we observed similar RNA-binding by Sen1_{Hel}ΔLP compared to Sen1_{Hel} (Figure EV3). We then assessed the behavior of the same mutants in *in vitro* transcription termination assays. The Sen1_{Hel}ΔUP mutant exhibited a moderate decrease in termination efficiency. Importantly, the Sen1_{Hel}ΔLP mutant that was inactive for duplex unwinding was not capable to elicit termination (Figure 5E). Consistent with the major role of the “prong” in termination *in vitro*, in the context of the full-length protein the LP deletion lead to lethality and provoked major transcription termination defects *in vivo* (Figure EV4). These results indicate that the “prong” is a critical determinant of the 5' to 3' unwinding and of the transcription termination activity of Sen1. In the context of termination, the “prong” might have additional functions besides unwinding. When analyzing the structure in detail, we noticed the presence of a hydrophobic residue exposed on the surface of the “prong” and involved in crystal contacts (Leu1549). Reasoning that lattice contacts often occur at protein-protein interaction interfaces, we tested the effect of replacing this residue. We found that the L1549D mutation provoked a less than 2-fold decrease in Sen1_{Hel} unwinding activity and only a mild decrease in the affinity for the RNA (~2-fold), but substantially decreased transcription termination (Figure 5 and Figure EV3). Importantly, a Sen1_{Hel}ΔUP L1549D double mutant exhibited levels of unwinding activity similar to the wt protein but was strongly affected in transcription termination (Figure 5). These results suggest that the “prong” might mediate protein-protein interactions required for Sen1 transcription termination activity.

Molecular basis for disease-associated mutations of SETX, the human orthologue of yeast Sen1

Analogously to *S. cerevisiae* Sen1, the human orthologue SETX, has been proposed to play crucial roles in transcription termination and in the maintenance of genome integrity (Suraweera *et al*, 2009; Skourti-Stathaki *et al*, 2011; Zhao *et al*, 2016). In line with its biological importance, mutations in SETX have been linked to two neurological disorders: ALS4 and AOA2. Many of the AOA2-causing mutations are missense mutations at the N-terminal domain and at the helicase domain of SETX, which shares about 30% sequence identity with yeast Sen1 (Figure EV5). Importantly, SETX possesses the key residues of the “brace” that are absent in other related helicases (Figure 2). Given the conservation, we took advantage of our Sen1_{Hel} structural data and biochemical tools to get insights into the molecular effects of AOA2-associated mutations in SETX. We mapped twenty-five missense AOA2 mutations on the Sen1_{Hel} structure (Appendix Figure S1). Two thirds of the mutations target residues buried inside the helicase core and their substitution is expected to disrupt the fold of the protein. A third of the mutations maps near the surface and/or the regions that in other helicases are important for either RNA recognition or ATP hydrolysis, suggesting that these mutations would affect SETX catalytic activity.

In order to test this prediction, we used Sen1_{Hel} as a surrogate of SETX. We introduced a subset of AOA2-associated substitutions at the equivalent positions in the yeast protein (Figure 6A) and analyzed their effect on Sen1_{Hel} activities (Figure 6 B-E and Table 2). First, we constructed a disease mutant affecting a residue buried at the interface between the two RecA domains, the Sen1_{Hel} D1616V mutant (D2207V in SETX), which resulted in an insoluble protein (unpublished observations). Second, we analyzed disease mutants expected to affect residues in direct contact with the RNA. The N1413S and T1779P mutants (N2010S and T2373P in SETX, respectively) were impaired in RNA binding, and consequently, duplex unwinding activity and *in vitro* transcription termination activity (Figure 6, B-D). The effect was similar to the double T1289A, R1293A mutant described before, which affects adjacent residues (Figure EV2C). The P1622L mutant (P2213L in SETX), which substitutes a residue expected to be about 10 Å away from the RNA, resulted in a significant decrease in RNA-binding affinity (k_D of ~6.4 μM, as compared to the ~0.6

μM of the wild-type protein). Accordingly, this mutant was also deficient in unwinding and transcription termination (Figure 6, D-E). Last, we analyzed a disease mutant predicted to involve a residue in direct contact with the γ -phosphate of ATP and therefore expected to be specifically impaired in ATP hydrolysis, the R1820Q mutant (R2414Q in SETX). As predicted, this mutant did not exhibit any detectable ATPase activity. Because transcription termination strictly depends on ATP hydrolysis, the R1820Q mutant was dramatically affected in this activity (Figure 6B-E). Taken together, our structure-function analyses support the idea that Sen1 is a good model to study the properties of SETX and the molecular basis of the diseases provoked by its mutation.

Conclusions

RNA helicases of the Upf1-like family take part in a variety of biological functions. Common to all Upf1-like helicases is the ability to unwind nucleic acids in the 5'-3' direction. We propose that this property is based on a common molecular mechanism of unwinding that depends on the presence of similar structural elements. In all members of the Upf1-like family studied to date, subdomain 1B (the “stalk” and the “barrel”) and subdomain 1C (that we refer to as the “prong”) extend on top of the RecA1 domain, where the nucleic acid is expected to enter the helicase channel. The conserved part of the “prong” has a similar conformation in all structures of SF1B RNA helicases determined to date and thus appears to be a rather rigid element. In contrast, the “barrel” can generally adopt different conformations and respond to RNA binding as well as to direct or indirect regulatory elements. Although the details are still unclear and the regulation is likely to differ in different members of the SF1B family, we envisage a general mechanism whereby closure of the two subdomains around the incoming RNA allows the “prong” to insert into the duplex, melting it. This strand-separation mechanism is significantly different from that proposed for DNA helicases of the SF1A family (Velankar *et al*, 1999). However, it is reminiscent of the 3'-5' unwinding mechanism proposed for RNA helicases of the SF2 Ski2-like family, where a structural element on the RecA2 domain (the β -hairpin) inserts into the duplex and melts the incoming base pairs (Büttner *et al*, 2007). Melting elements on opposite sides of the helicase core together with the movements of the two RecA

domains in response to ATP hydrolysis could thus underpin the opposite unwinding polarities of Upf1-like and of Ski2-like RNA helicases.

Notwithstanding the similarities in unwinding properties, different RNA helicases of the SF1B family have different biological functions. Sen1 has a specific function in termination of non-coding transcription in yeast cells (Steinmetz *et al*, 2006) and the endogenous full-length protein can indeed recapitulate transcription termination in reconstituted *in vitro* assays (Porrua and Libri, 2013). The most characteristic feature of Sen1 is the presence of a large N-terminal domain, which is important for Sen1 function *in vivo* (Ursic *et al*, 2004). However, we have found that the helicase core of Sen1 retains all the properties that are necessary for transcription termination *in vitro* (Figure 1). Although in an *in vivo* situation the additional domains are likely involved in the recruitment of Sen1 near the nascent RNA and/or regulation, the specificity for the termination reaction *in vitro* is embedded in the helicase core alone. One possibility is that specific features on the outer surfaces of the Sen1 helicase core might contact Pol II. Another, not necessarily exclusive, possibility is that specificity determinants are in the unwinding elements of this helicase.

The region of Sen1 where unwinding is expected to occur and where Pol II is expected to come in close proximity has indeed unique features, with an additional structural element (that we dubbed the “brace”) that encircles the RecA1 domain, the “barrel” and the “stalk” into a rather rigid unit. Residues in the “brace” are important *in vivo* (Chen *et al*, 2014) and are highly conserved in Sen1 orthologues, including the human SETX, further supporting the importance of this region. From the structural and biochemical analysis, the “brace” appears to push the “barrel” in a position competent for RNA-binding, similar to the effect exerted by the presence of the CH domain in Upf1. However, the “brace” does not lock Sen1 on the RNA as the CH domain does: the helicase core of Sen1 is indeed active in unwinding, similar to Upf1 when the CH domain has been displaced. We propose that the particular conformation that the “brace” imposes on the accessory domains as well as the distinctive characteristics of the “prong”, which we have shown to be essential for termination both *in vitro* and *in vivo*, are key determinants of the specific function of Sen1 in transcription termination. The current model is that Sen1 bind the nascent RNA and translocate along it until it encounters Pol II. Given that the integrity of the “prong” is

essential for dissociation of the elongation complex, it is tempting to speculate that the final step of termination involves the insertion of the “prong” into the Pol II RNA exit channel, which would lead to profound conformational changes and destabilization of the elongation complex. This process might be facilitated by the flexible nature of the upper part of the “prong” of Sen1 and/or by protein-protein interactions between the “prong” and specific surfaces of Pol II. Finally, all disease mutations of human SETX we mapped on yeast Sen1 resulted in a transcription termination defect *in vitro*, supporting the idea that the development of AOA2 might be associated with transcription termination defects.

MATERIALS AND METHODS

Protein expression and purification

For expression, we used a fusion protein with a cleavable C-terminal His tag coupled to *Vibrio cholerae* MARTX toxin cysteine protease domain (His₈-CPD) (Shen *et al*, 2009). His₈-CPD tagged Sen1_{Hel} (1095-1904) and its mutant derivatives were purified from *Escherichia coli* BL21 (DE3) STAR pRARE (Stratagene) cells grown in TB medium. Overexpression was induced by adding IPTG (0.5 mM final concentration) at 18°C overnight. Cells were lysed in buffer containing 20 mM sodium phosphate pH 8.0, 500 mM NaCl, 2 mM MgCl₂, 30 mM imidazole, 10% (v/v) glycerol, 1 mM β-mercaptoethanol, benzamide hydrochloride and protease inhibitors. The proteins were bound to a Ni²⁺-affinity chromatography column (HisTrap FF from GE Healthcare) and eluted by on-column tag cleavage using the 3C protease (Youell *et al*, 2011). CPD tag cleavage with 3C allowed us to overcome an unspecific cleavage product we otherwise obtained when using inositol hexakisphosphate (InsP6). The eluates were then subjected to heparin affinity chromatography on a HiTrap Heparin HP column (GE Healthcare) using buffer A for binding (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM MgCl₂, 1 mM DTT) and buffer B for elution (20 mM Tris-HCl pH 7.5, 1 M NaCl, 2 mM MgCl₂, 1 mM DTT). Size-exclusion chromatography (SEC) was performed as a final step of purification using a Superdex 200 column (GE Healthcare) and elution buffer containing 20 mM HEPES pH 7.5, 300 mM NaCl, 2

mM MgCl₂ and 1 mM DTT. Proteins were stored at -80°C on SEC buffer containing 50% (v/v) glycerol.

TAP-tagged full-length Sen1 was overexpressed from the *GALI* promoter in the presence of galactose in *S. cerevisiae* (strain YDL2556) and purified using a previously described protocol (Porrua & Libri, 2015b) with the following modifications: the concentration of NaCl in elution buffers was increased to 500 mM to improve the elution yield and proteins bound to IgG-beads were treated with 20 µg/ml of RNase A during elution at 4°C overnight.

RNA pol II (12 subunits) was purified from *S. cerevisiae* strain BJ5464 (Kireeva *et al*, 2003) by Ni²⁺-affinity chromatography followed by anion exchange essentially as previously described (Porrua & Libri, 2015b). Recombinant His₆-tagged Rpb4/7 heterodimer was purified Ni²⁺-affinity chromatography and gel filtration as previously described (Porrua & Libri, 2015b).

Crystallization and structure determination

Sen1_{Hel} was concentrated to 3 mg/mL and mixed with a 10-fold molar excess of ADP. Crystals were grown at 4°C by hanging-drop vapor diffusion from drops formed by equal volumes of protein and of crystallization solution (6% (w/v) PEG 8000, 8% (v/v) ethylene glycol, 0.1 M HEPES pH 7.5). Prior to flash freezing in liquid nitrogen, the crystals were briefly soaked in mother liquor containing 28% (v/v) ethylene glycol. The best diffracting crystals were obtained by removing a disordered loop (1471-1538) with a (Gly-Ser)₂ linker.

A single-wavelength anomalous diffraction experiment from intrinsic sulfur atoms (S-SAD) was performed at the macromolecular crystallography super-bending magnet beamline X06DA (PXIII) at the Swiss Light Source (Villigen, Switzerland). On a single crystal, 4x 360° data sets were collected at 100 K at a wavelength of 2.075-Å with 0.1° oscillation 0.1 s exposure in four different orientations of a multi-axis goniometer (Waltersperger *et al*, 2015), as previously described (Weinert *et al*, 2015). The sample-to-detector distance was set to 120 mm. The data were processed using XDS and scaled and merged with XSCALE (Kabsch, 2010). The high-resolution data

cutoff was based on the statistical indicators CC1/2 and CC* (Karplus & Diederichs, 2012). Substructure determination and phasing were performed with SHELXC/D/E (Sheldrick, 2010) using the HKL2MAP interface (Pape & Schneider, 2004). The successful SHELXD substructure solution, in a search for 25 sulfur sites, had a CCall and a CCweak of 36.9 and 18.2, respectively. Density modification resulted in a clear separation of hands. Three cycles of chain tracing resulted in the automatic building of 275 amino acids with SHELXE. An initial model was built automatically with BUCCANEER (Cowtan, 2006) and extended manually in the experimental electron density in COOT (Emsley & Cowtan, 2004) and refined against the native data with phenix.refine (Adams *et al*, 2010). The final model includes residues 1096-1875, with the exception of missing or disordered loops in subdomain 1C (residues 1471-1543), in RecA1 (residues 1382-1395) and RecA2 (residues 1705-1713 and 1799-1801).

ATP hydrolysis assays

ATPase assays were performed as previously described (Porrua & Libri, 2015b). Briefly, 5 nM of purified Sen1 proteins were assayed at 28°C in 10- μ l reactions containing 10 mM Tris-HCl pH 7.5, 75 mM NaCl, 1 mM MgCl₂, 1 mM DTT, 25% glycerol and 50 ng/ μ l polyU. The reaction started with the addition of a 250 μ M ATP solution containing 0.25 μ M of 800 Ci/mmol α^{32} P-ATP (final concentrations). Aliquots were taken at various times, mixed with one volume of quench buffer (10 mM EDTA, 0.5% SDS) and subjected to thin layer chromatography on PEI cellulose plates (Merck) in 0.35 M potassium phosphate (pH 7.5). Hydrolysis products were analysed by phosphorimaging (GE Healthcare).

Duplex unwinding assays

The RNA:DNA substrates for the unwinding assays were formed by annealing a short 5'-end labeled DNA oligonucleotide to either the 5' or 3' end of a longer RNA oligonucleotide (Appendix Table S1). The 44-mer RNA oligonucleotide was purchased from Integrated DNA Technologies whereas the 75-mer RNA was produced by *in vitro* transcription with the appropriate templates (see Appendix Table S1) using the MEGAscript T7 kit (Ambion). Unwinding assays were performed in unwinding buffer (10 mM Tris-HCl pH 7.5, 50 mM NaCl, 7.5 μ M ZnCl₂, 0.5 mM

DTT, 10% glycerol, 0.1mg/ml BSA) in 20- μ l reactions at 28°C. Sen1 proteins were preincubated with the corresponding duplex substrate and the reaction was initiated by adding a mixture containing ATP and MgCl₂ (2 mM in the reaction) and an excess of unlabeled DNA oligonucleotide (0.1 μ M) to trap the unwound RNA. Aliquots were taken at the indicated time-points and mixed with 1 volume of stop/loading buffer containing 50 mM EDTA, 1% SDS and 20% glycerol. Samples were subjected to electrophoresis on a 15% native PAGE and gels were directly scanned using a Typhoon scanner (GE Healthcare).

***In vitro* transcription-termination assays**

Termination assays were performed basically as previously described (Porrua and Libri, 2015 MiMB). Briefly, ternary ECs were assembled in a promoter-independent manner by first annealing a fluorescently labeled RNA (oligo DL2492, see Appendix Table S1) with the template DNA (oligo DL3352, see Appendix Table S1) and subsequently incubating the RNA: DNA hybrid with purified RNA pol II. Next, the non-template strand (oligo DL3353, see Appendix Table S1) and recombinant Rpb4/7 heterodimer were sequentially added to the mixture. The ternary ECs were then immobilized on streptavidin beads (Dynabeads MyOne Streptavidin T1 from Invitrogen) and washed with transcription buffer (TB) containing 20 mM Tris-HCl pH 7.5, 100 mM NaCl, 8 mM MgCl₂, 10 μ M ZnCl₂, 10% glycerol and 2 mM DTT; then with TB/0.1% Triton, TB/ 0.5 M NaCl and finally TB. The termination reactions were performed at 28°C in TB in a final volume of 20 μ l in the absence or in the presence of 20 to 80 nM of Sen1 proteins. Transcription was initiated after addition of a mixture of ATP, UTP and CTP (1 mM each as the final concentration in the reaction) to allow transcription through the G-less cassette up to the first G of a G-stretch in the non-template strand. The reactions were allowed for 15 min and then stopped by the addition of 1 μ l of 0.5 M EDTA. After separation of beads and supernatant fractions, beads fractions were resuspended in 8 μ l of loading buffer (1x Tris-borate-EDTA, 8 M urea) and boiled for 5 min at 95°C while RNAs in the supernatant fractions were ethanol precipitated and resuspended in 8 μ l of loading buffer. Transcripts were subjected to 10% (w/v) denaturing PAGE (8M urea) and gels were scanned with a Typhoon scanner.

RNase protection assays

Proteins (10 pmol each) were mixed with 5 pmol ³²P body-labeled RNA to a final 20 μL reaction volume in 50 mM HEPES pH 6.5, 150 mM NaCl, 1 mM magnesium diacetate, 10% (v/v) glycerol, 0.1% (w/v) NP-40, and 1 mM DTT. After incubation for 1 hr at 4°C, the reaction mixtures were digested with 1 μg RNase A/T1 mix and 2.5 U RNase T1 (Fermentas) for 20 min at 20°C. Protected RNA fragments were then extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1 (v/v), Invitrogen), precipitated with ethanol, separated on a denaturing 22% (w/v) polyacrylamide gel, and visualized by phosphorimaging (Fuji).

Fluorescence anisotropy

Fluorescence anisotropy measurements were performed with a 5'-end fluorescein-labeled 15-mer RNA (oligo ARE, see Appendix Table S1) at 20°C in 50 μl reactions on Infinite M1000 Pro (Tecan). The RNA was dissolved to a concentration of 10 nM and incubated with Sen1_{Hel} variants at different concentrations in a buffer containing 20 mM HEPES pH 7.5, 300 mM NaCl, 2 mM MgCl₂ and 1 mM DTT. The excitation and emission wavelengths were 485 nm and 535 nm, respectively. Each titration point was measured three times using ten reads. The data were analyzed by nonlinear regression fitting using the BIOEQS software (Royer, 1993).

Additional methods (Electrophoretic mobility shift assays and *in vivo* RNA expression analyses) are included in the Appendix.

AUTHOR CONTRIBUTIONS

B.L. solved the structure with help from J.B. and purified all proteins; Z.H. carried out the enzymatic assays and *in vivo* analysis; C.B. and F.B. carried out the fluorescence anisotropy and RNase protection assays, respectively; E.C., O.P. and D.L. initiated the project; E.C., O.P. and B.L. wrote the manuscript.

ACKNOWLEDGEMENTS

We thank the Max Planck Institute of Biochemistry (MPIB) Crystallization and Core Facilities; Claire Basquin for FA measurements; Vincent Olieric at PXIII at SLS for assistance with data collection and members of our labs for useful discussions. We also thank David Brow (University of Wisconsin) for critical reading of the manuscript and suggestions and Christian Biertumpfel (MPI Biochemistry) for the His₈-CPD expression vector. This study was supported by the Graduate School of Quantitative Biosciences Munich to B.L., by the Max Planck Gesellschaft, the European Commission (ERC Advanced Investigator Grant 294371) and the Deutsche Forschungsgemeinschaft (DFG SFB646, SFB1035, GRK1721, and CIPSM) to E.C.; by the CNRS, the Agence National pour la Recherche (ANR-08-Blan-0038-01 and ANR-12-BSV8-0014-01 to D.L. and ANR-16-CE12-0001-01 to O.P) and the Fondation pour la Recherche Medicale (programme Equipes 2013 to D. L). Z.H. was supported by PhD fellowships from the China Scholarship Council and La Ligue contre le Cancer.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACCESSION NUMBERS

The coordinates and structure factors of Sen1_{Hel} have been deposited in the Protein Data Bank with the accession code 5MZN.

REFERENCES

- Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC & Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**: 213–221
- Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E & Tollervey D (1999) Functions of the exosome in rRNA, snoRNA and snRNA synthesis. *EMBO J.* **18**: 5399–5410
- Arndt KM & Reines D (2015) Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annu. Rev. Biochem.* **84**: 150306093657004–404
- Bennett CL & La Spada AR (2015) Unwinding the Role of Senataxin in Neurodegeneration. *Discov Med* **103**: 127–136
- Bhattacharya A, Czaplinski K, Trifillis P, He F, Jacobson A & Peltz SW (2000) Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA* **6**: 1226–1235
- Büttner K, Nehring S & Hopfner K-P (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**: 647–652
- Carroll KL, Ghirlando R, Ames JM & Corden JL (2007) Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* **13**: 361–373
- Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H & Conti E (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol. Cell* **41**: 693–703
- Chamieh H, Ballut L, Bonneau F & Le Hir H (2008) NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat. Struct. Mol. Biol.* **15**: 85–93
- Chen X, Mueller U, Sundling KE & Brow DA (2014) *Saccharomyces cerevisiae* Sen1 as a Model for the Study of Mutations in Human Senataxin That Elicit Cerebellar Ataxia. *Genetics* **198**: 577–U180
- Cheng Z, Muhrad D, Lim MK, Parker R & Song H (2007) Structural and functional insights into the human Upf1 helicase core. *EMBO J.* **26**: 253–264
- Clerici M, Mourão A, Gutsche I, Gehring NH, Hentze MW, Kulozik A, Kadlec J, Sattler M & Cusack S (2009) Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.* **28**: 2293–2306
- Cordin O, Tanner NK, Doère M, Linder P & Banroques J (2004) The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J.* **23**: 2478–2487

- Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**: 1002–1011
- DeMarini DJ, Winey M, Ursic D, Webb F & Culbertson MR (1992) SEN1, a positive effector of tRNA-splicing endonuclease in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**: 2154–2164
- Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2126–2132
- Fairman-Williams ME & Jankowsky E (2012) Unwinding initiation by the viral RNA helicase NPH-II. *J. Mol. Biol.* **415**: 819–832
- Fiorini F, Bagchi D, Le Hir H & Croquette V (2015) Human Upf1 is a highly processive RNA helicase and translocase with RNP remodelling activities. *Nat Commun* **6**: 7581
- Jackson RN, Klauer AA, Hintze BJ, Robinson H, van Hoof A & Johnson SJ (2010) The crystal structure of Mtr4 reveals a novel arch domain required for rRNA processing. *EMBO J.* **29**: 2205–2216
- Jensen TH, Jacquier A & Libri D (2013) Dealing with pervasive transcription. *Mol. Cell* **52**: 473–484
- Kabsch W (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**: 133–144
- Karplus PA & Diederichs K (2012) Linking crystallographic model and data quality. *Science* **336**: 1030–1033
- Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845–858
- Kim HD, Choe J & Seo YS (1999) The *sen1(+)* gene of *Schizosaccharomyces pombe*, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. *Biochemistry* **38**: 14697–14710
- Kireeva ML, Lubkowska L, Komissarova N & Kashlev M (2003) Assays and Affinity Purification of Biotinylated and Nonbiotinylated Forms of Double-Tagged Core RNA Polymerase II from *Saccharomyces cerevisiae*. In *RNA Polymerases and Associated Factors, Part C* pp 138–155. Elsevier
- LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A & Tollervey D (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**: 713–724
- Lim SC, Bowler MW, Lai TF & Song H (2012) The Ighmbp2 helicase structure reveals the molecular basis for disease-causing mutations in DMSA1. *Nucleic Acids Res.* **40**: 11009–11022
- Linder P & Lasko P (2006) Bent out of shape: RNA unwinding by the DEAD-box helicase Vasa. *Cell* **125**: 219–221

- Martin-Tumasz S & Brow DA (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
- Mischo HE & Proudfoot NJ (2013) Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochim. Biophys. Acta* **1829**: 174–185
- Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A & Proudfoot NJ (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell* **41**: 21–32
- Nishimura K, Fukagawa T, Takisawa H, Kakimoto T & Kanemaki M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**:917-922.
- Ozgur S, Buchwald G, Falk S, Chakrabarti S, Prabu JR & Conti E (2015) The conformational plasticity of eukaryotic RNA-dependent ATPases. *FEBS J.* **282**: 850–863
- Porrua O & 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 & Libri D (2015a) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16**: 190–202
- Porrua O & Libri D (2015b) Characterization of the mechanisms of transcription termination by the helicase Sen1. *Methods Mol. Biol.* **1259**: 313–331
- Porrua O, Hobor F, Boulay J, Kubicek K, D'Aubenton-Carafa Y, Gudipati RK, Stefl R & Libri D (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *EMBO J.* **31**: 3935–3948
- Pyle AM (2008) Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys* **37**: 317–336
- Royer CA (1993) Improvements in the numerical analysis of thermodynamic data from biomolecular complexes. *Anal. Biochem.* **210**: 91–97
- Schulz D, Schwalb B, Kiesel A, Baejen C, Torkler P, Gagneur J, Soeding J & Cramer P (2013) Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell* **155**: 1075–1087
- Sengoku T, Nureki O, Nakamura A, Kobayashi S & Yokoyama S (2006) Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* **125**: 287–300
- Sheldrick GM (2010) Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**: 479–485

- Shen A, Lupardus PJ, Morell M, Ponder EL, Sadaghiani AM, Garcia KC & Bogyo M (2009) Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PLoS ONE* **4**: e8119
- Skourti-Stathaki K, Proudfoot NJ & Gromak N (2011) Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Mol. Cell* **42**: 794–805
- Söding J, Biegert A & Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**: W244–8
- Steinmetz EJ & Brow DA (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Mol. Cell Biol.* **16**: 6993–7003
- Steinmetz EJ, Warren CL, Kuehner JN, Panbehi B, Ansari AZ & Brow DA (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Mol. Cell* **24**: 735–746
- Suraweera A, Lim Y, Woods R, Birrell GW, Nasim T, Becherel OJ & Lavin MF (2009) Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum. Mol. Genet.* **18**: 3384–3396
- Tudek A, Porrua O, Kabzinski T, Lidschreiber M, Kubicek K, Fortova A, Lacroute F, Vanacova S, Cramer P, Stefl R & Libri D (2014) Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Mol. Cell* **55**: 467–481
- Ursic D, Chinchilla K, Finkel JS & Culbertson MR (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 & Keller W (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS Biol.* **3**: 986–997
- Velankar SS, Soultanas P, Dillingham MS, Subramanya HS & Wigley DB (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**: 75–84
- Waltersperger S, Olieric V, Pradervand C, Gletting W, Salathe M, Fuchs MR, Curtin A, Wang X, Ebner S, Panepucci E, Weinert T, Schulze-Briese C & Wang M (2015) PRIGo: a new multi-axis goniometer for macromolecular crystallography. *J Synchrotron Radiat* **22**: 895–900
- Weinert T, Olieric V, Waltersperger S, Panepucci E, Chen L, Zhang H, Zhou D, Rose J, Ebihara A, Kuramitsu S, Li D, Howe N, Schnapp G, Pautsch A, Bargsten K, Prota AE, Surana P, Kottur J, Nair DT, Basilico F, et al (2015) Fast native-SAD phasing for routine macromolecular structure determination. *Nat. Methods* **12**: 131–133

- Weir JR, Bonneau F, Hentschel J & Conti E (2010) Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *Proc. Natl. Acad. Sci. U.S.A.* **107**: 12139–12144
- Wlotzka W, Kudla G, Granneman S & Tollervy 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 JC, Dufour ME, Boulay J, Regnault B, Devaux F, Namane A, Seraphin B, Libri D & Jacquier A (2005) Cryptic Pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121**: 725–737
- Youell J, Fordham D & Firman K (2011) Production and single-step purification of EGFP and a biotinylated version of the Human Rhinovirus 14 3C protease. *Protein Expr. Purif.* **79**: 258–262
- Zhao DY, Gish G, Braunschweig U, Li Y, Ni Z, Schmitges FW, Zhong G, Liu K, Li W, Moffat J, Vedadi M, Min J, Pawson TJ, Blencowe BJ & Greenblatt JF (2016) SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* **529**: 48–53

FIGURE LEGENDS

Figure 1 - A recombinant version of the *S.cerevisiae* Sen1 helicase core retains the main biochemical properties of the full-length protein.

A Analysis of RNA-dependent ATPase activity of Sen1 proteins. Top: Schematic diagram of Sen1 full-length purified from yeast (ySen1 FL), and recombinant Sen1_{Hel} and Sen1_{Hel} E1591Q mutant. Bottom, left: SDS-PAGE analysis of the purified proteins used in these assays (M: molecular weight marker). 2 pmol of ySen1 FL and 25 pmol of Sen1_{Hel} proteins were loaded. Bottom, right: graphical representation of the ATP hydrolysed by the different Sen1 proteins as a function of time. Values represent the average and standard deviation (SD) from three independent experiments.

B Time course analysis of the ATP-dependent 5'-3' duplex unwinding activity of Sen1 proteins. Reactions contained 5 nM of Sen1 and 2 nM of substrate. An RNA:DNA duplex composed of a 44-mer RNA annealed to a 19-mer DNA molecule to provide a 5'-end 25-nt single strand overhang was used as the substrate (see Appendix Table S1 for sequence details). The asterisk (*) denotes the presence of a FAM at the 5'end of the DNA. The first lanes correspond to heat-denatured (95°C) samples and the last lanes are control reactions incubated with Sen1 proteins in the absence of ATP. The graph on the right show the fraction of duplex unwound as a function of time. Data were fitted with Kaleidagraph to the Michaelis–Menten equation. Values represent the average and standard deviation (SD) from three independent experiments.

C *In vitro* transcription termination (IVTT) assays with 20 nM of ySen1 FL and 40 nM of Sen1_{Hel} proteins. Left: Scheme of an IVTT assay. Ternary ECs composed of Pol II, fluorescently labelled nascent RNA and DNA templates are assembled and attached to streptavidin beads via the 5' biotin of the non-template strand to allow subsequent separation of beads-associated (B) and supernatant (S) fractions. The transcription template contains a G-less cassette followed by a G-stretch in the non-template strand. After adding an ATP, UTP, CTP mix, Pol II transcribes until it encounters the G-rich sequence. Sen1 dissociates ECs paused at the G-stretch and releases Pol II and the associated transcripts to the supernatant. Right: PAGE analysis

of RNAs from a representative IVTT assay. The fraction of transcripts released from ECs stalled at the G-stretch is used as a measure of the termination efficiency. Representative gel of one out of two independent experiments (values of RNA released in both experiments can be found in the corresponding data source file).

Figure 2 - Common and unique structural features of the Sen1 helicase core.

Structures of yeast Sen1_{Hel}-ADP (left), Upf1_{Hel}-AMPPNP (middle, PDB: 2GJK, Cheng et al., 2007) and IGHMBP2_{Hel} (right, PDB: 4B3F, Lim *et al*, 2012) determined in the absence of RNA are shown in a similar orientation after optimal superposition of their respective RecA1 domains (on the right in this front-view orientation). Dotted lines indicate disordered loops not modeled in the present structure. On top there is a scheme with the domain organization of the full-length proteins, with predicted structured and unstructured regions shown as rectangles and lines, respectively. The fragments crystallized are highlighted in color. The RecA1 and RecA2 domains are in yellow, the “stalk” in gray, subdomain 1B (the “barrel”) in orange and subdomain 1C (the “prong”) in red. In the case of Sen1_{Hel}, the N-terminal “brace” is shown in blue.

Figure 3 - The evolutionary conserved interactions of the N-terminal “brace” of Sen1.

A-B Zoom-in views (with the corresponding overall views) showing the extensive hydrophobic interactions the “brace” makes with RecA1, the “stalk” helices and the “barrel”. Selected residues are shown in stick representation. Panels A and B show the molecule in a side-view and back-view orientations (90° and 180° clockwise rotation around a vertical axis with respect to the front-view in Figure 2).

C Structure-based sequence alignment of the “brace” showing the amino-acid conservation (highlighted in blue) and the interactions with RecA1 (yellow circles), the “stalk” helices (gray circles) and the “barrel” (orange circles).

Figure 4 - Analysis of RNA binding features of Sen1.

A Comparison of the structures of yeast Sen1_{Hel}-ADP, human UPF1_{Hel}-AMPPNPP (PDB: 2GJK, Cheng et al., 2007), UPF1_{Hel}-ADP:AlF₄⁻-RNA (PDB: 2XZO, Chakrabarti et al., 2010), and yeast Upf1_{Hel-CH}-ADP:AlF₄⁻-RNA (PDB: 2XZL, Chakrabarti et al., 2010). Colors are the same as in Figure 2. The nucleotides and RNA are shown in black. On the bottom, schematic representation of the subdomain organization of Sen1 and Upf1 illustrating the different location of the “barrel” (in orange) and its repositioning in Upf1 upon RNA binding. Note that the CH domain of Upf1 pushes the “barrel” and changes its orientation extending the RNA-interaction region. The molecules in a side-view orientation are shown in Figure EV2B.

B RNase protection assays with yeast Sen1_{Hel}, Upf1_{Hel} and Upf1_{CH-Hel} in presence of different nucleotides (ADP:BeF₃⁻ and ADP:AlF₄⁻, which mimic the ground state and the transition state of the nucleotide in the ATPase cycle). A coomassie-stained SDS-PAGE with the different proteins used is shown on the left (M: molecular weight marker). RNA fragments were obtained by digesting ³²P body-labeled (CU)₂₈C 57-mer RNA in the presence of the indicated proteins with RNase A and RNase T1. The left lane was loaded with 10-mer and 15-mer radioactively labeled transcripts as size markers. The asterisks (*) identify minor fragments likely due to the contiguous binding of more than one protein to the same RNA.

Figure 5 - A critical role for the “prong” in duplex unwinding and transcription termination.

A On top is a schematic presentation of the Sen1_{Hel} variants analyzed in the experiments below. At the bottom is a zoom-in view of the “prong”. The dotted lines indicate the approximate positions of the “prong” deletion mutants ΔUP for the removal of the upper part and ΔLP for the removal of also lower part. Selected residues are shown in stick representation.

B RNase protection assays with Sen1 proteins as in Figure 4B in the presence of ADP:AlF₄⁻.

C Analysis of the ATPase activity of the different Sen1_{Hel} variants. Values correspond to the average and SD of three independent experiments.

D Analysis of the effect of the different “prong” mutations on Sen1_{Hel} unwinding activity. Reaction conditions are the same as in Figure 1B. The graph shows the fraction of duplex unwound as a function of time. Data were fitted with Kaleidagraph to the Michaelis–Menten equation. The values reflect the average and standard deviations (SD) from three independent measurements.

E IVTT assays in the absence and in the presence of the different Sen1_{Hel} versions (80 nM in the reaction). Representative gel of one out of two to three independent experiments (values of RNA released in additional experiments are provided in the corresponding data source file).

Figure 6 - Functional characterization of Sen1_{Hel} mutants harboring AOA2-associated substitutions

A Mapping of selected AOA2-associated substitutions (shown in magenta) introduced at the equivalent positions in Sen1_{Hel} for their functional analysis. The mutations are reported in Chen *et al*, 2014 and in the UCLA Neurogenetics SETX Database.

B Quantitative measurements of RNA-binding affinities of the mutants by fluorescence anisotropy using fluorescently-labeled AU-rich RNA as the substrate. The data were fitted^{[[[SEP]]]} to a binding equation describing a single-site binding model to obtain the dissociation constants (K_D , indicated on the left of the curves). The best fit was plotted as a solid line. The K_D s and their corresponding errors are the mean and standard deviation (SD) of a minimum of three independent experiments.

C Analysis of the ATPase activity of the Sen1_{Hel} R1820Q mutant predicted to be affected in nucleotide binding. Values correspond to the average and SD of three independent experiments.^{[[[SEP]]]}

D Assessment of the effect of several AOA2-associated mutations on Sen1_{Hel} unwinding activity.^{[[[SEP]]]} RNA:DNA duplex unwinding reactions using a 75-mer RNA annealed to a 20-mer DNA oligonucleotide (see Appendix Table S1) leaving a 5'-end

55 nt single-strand overhang as the substrate. Reactions contained 30 nM of Sen1_{Hel} variants and 0.5 nM of substrate. The efficiency of unwinding indicated corresponds to the fraction of substrate unwound by the different proteins at 30 minutes. Values correspond to the average and SD of three independent experiments. ^[1]_[SEP]

E Analysis of the impact of AOA2-associated mutations on the efficiency of transcription termination. IVTT assays performed in the presence of 80 nM of the different Sen1 variants. The values of nascent RNA released correspond to one out of two independent experiments. Quantification of both experiments are included in the corresponding data source file.

EXPANDED VIEW FIGURE LEGENDS

Figure EV1 - Identification and characterization of a recombinant helicase core of *S. cerevisiae* Sen1 suitable for structural studies.

A SDS-PAGE analysis of Sen1₉₇₆₋₁₈₈₀ tagged with C-terminal CPD-His₈. Lane E shows the elution fraction after Ni²⁺-affinity purification step and lane InsP6 shows the tag cleavage after the protein was incubated with 400 μM inositol hexakisphosphate (InsP6) for 20 min at 4°C. The protein before and after tag-cleavage is smaller than expected: theoretical molecular weights of Sen1₉₇₆₋₁₈₈₀-CPD-His₈ and Sen1₉₇₆₋₁₈₈₀ are ~126 kDa and ~102 kDa, respectively. Left lane M shows a molecular weight marker.

B Time course analysis of the ATP-dependent 3'-5' duplex unwinding activity of Sen1 proteins. Reactions performed in the presence of 5 nM of Sen1 and 2 nM of substrate. An RNA:DNA duplex composed of a 44-mer RNA annealed to a 19-mer DNA molecule to provide a 3'-end 25-nt single strand overhang was used as the substrate (see Appendix Table S1 for sequence details).

C Snapshots of the electron density maps at important regions of the structure described in the text. The 2Fo-Fc maps are contoured at 1.7σ.

Figure EV2 – Biochemical and structural properties of Sen1, and comparison with Upf1.

A Zoom-in view of the nucleotide binding site in Sen1 (left) and Upf1 (right) (PDB: 2XZO, Chakrabarti et al., 2010). The adenine ring is sandwiched between an apolar surface of RecA1 and an aromatic residue protruding from the short linker that connects RecA1 to RecA2 (Tyr1655, corresponding to Tyr638_{Upf1} and Tyr442_{IMGMBP2}). In addition, the conserved side chain of Gln1339 (corresponding to Gln413_{Upf1} and Gln196_{IMGMBP2}) forms a bidentate hydrogen-bond interaction with the N6 and N7 moieties of the adenine ring.

B Comparison of the structures of yeast Sen1_{Hel}-ADP, human UPF1_{Hel}-AMPPNPP (PDB: 2GJK, Cheng et al., 2007), UPF1_{Hel}-ADP:AlF₄⁻-RNA (PDB: 2XZO, Chakrabarti et al., 2010), and yeast Upf1_{Hel-CH}-ADP:AlF₄⁻-RNA (PDB: 2XZL, Chakrabarti et al., 2010). The molecules in a side-view orientation (90° clockwise rotation around a vertical axis with respect to the front-view in Figure 4A).

C Comparison of the RNA-binding sites of Sen1 (left) and Upf1 (right) (PDB: 2XZO, Chakrabarti et al., 2010).

D-F Functional analysis of the Sen1_{Hel} T1289A, R1293A mutant harboring substitutions at conserved positions at the predicted RNA-binding surface. **D**) Fluorescence anisotropy assays. Curves represent 3 independent measurements **E**) ATP hydrolysis assays. Values correspond to the average and SD of three independent experiments. **F**) IVTT assays performed in the same conditions as in Figure 1C. The images correspond to different gels migrated and processed in parallel. The values of nascent RNA released correspond to one out of two independent experiments.

Figure EV3- Analysis of the impact of the “prong” mutations on the affinity of Sen1_{Hel} for the RNA.

Electrophoretic Mobility Shift Assay (EMSA) using a 5'-end fluorescently labeled 44-mer RNA as the substrate (DL3316, see Appendix Table S1) at 2 nM and Sen1_{Hel}

variants at 10, 20, 40, 80 and 160 nM at the final concentrations. Gels were migrated and processed in parallel. The values correspond to the mean of two independent experiments. At high protein concentrations Sen1_{Hel} forms high-order complexes with the RNA that are retained in the wells of the gel.

Figure EV4- Analysis of the phenotype of the Δ LP mutant *in vivo*.

A A Sen1 variant harboring the Δ LP cannot support cell viability. A Δ sen1 strain (YDL2767) covered by an *URA3*-containing plasmid (pFL38) expressing wt Sen1 was transformed with a *TRP1*-plasmid (pFL39) carrying either the wt or a Δ LP version of *SEN1*. After over-night growth in non-selective medium, cells initially harboring both plasmids were plated on minimal medium (CSM) containing 5-fluorootic acid (5-FOA) to select for cells that have lost the *URA3* plasmid (and can therefore survive thanks to the *TRP1* plasmid-borne *SEN1* copy). The absence of cells growing in 5-FOA and containing the *TRP1* plasmid expressing Sen1 Δ LP indicates that the Δ LP deletion is lethal.

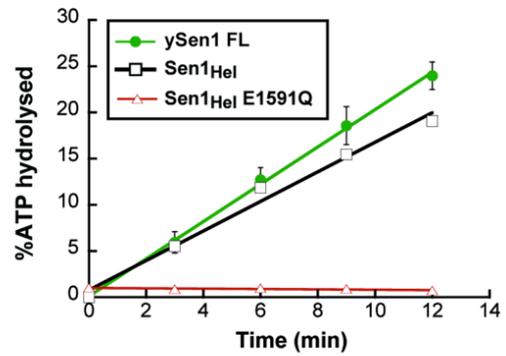
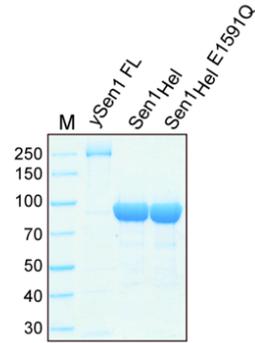
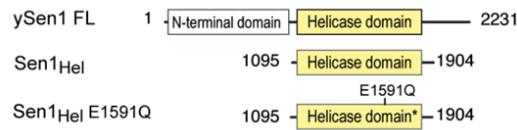
B The Sen1 Δ LP mutant is strongly defective in transcription termination *in vivo*. Northern blot analyses of two well-characterized NNS-targets, snR47 and snR13, in a Sen1-AID (auxin-induced degron, Nishimura et al, 2009) strain carrying a plasmid (pFL39) expressing either the wt or a Δ LP version of *SEN1*. A strain harboring an empty vector was included as a positive control for termination defects. To detect the primary products of NNS-dependent termination that are processed/degraded by the exosome, the strain was also deleted in the exonuclease *RRP6*. Sen1-AID was depleted for 1h by the addition of 100 μ M indole-3-acetic acid (a natural auxin) to monitor the capacity of the plasmid-borne versions of *SEN1* to induce transcription termination. The strong accumulation of longer RNA species in the *sen1* Δ LP mutant compared to the wt is indicative of major termination defects. Under non-depletion conditions, the strain harboring the mutant protein exhibits a dominant-negative phenotype (partial termination defects), indicating that Sen1 Δ LP has similar expression levels compared to the endogenous Sen1. The *ACT1* transcript is used as a loading control.

Figure EV5- Multiple sequence alignment of the helicase domain of Upf1-like helicases.

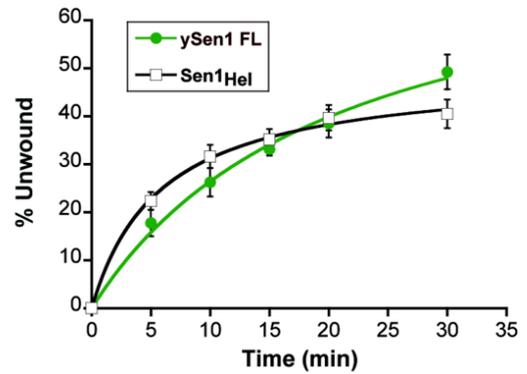
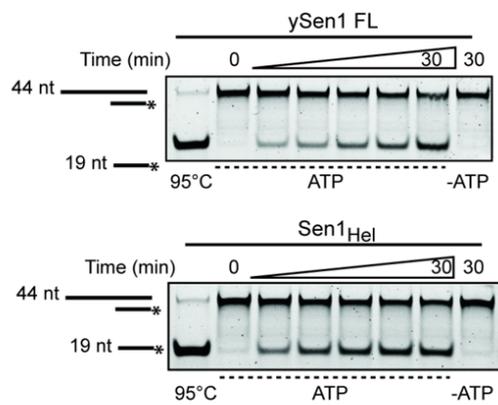
The multiple alignment was done using Clustal Omega and the conservation was calculated using BLOSUM62 and is shown in purple.

Figure 1

A



B



C

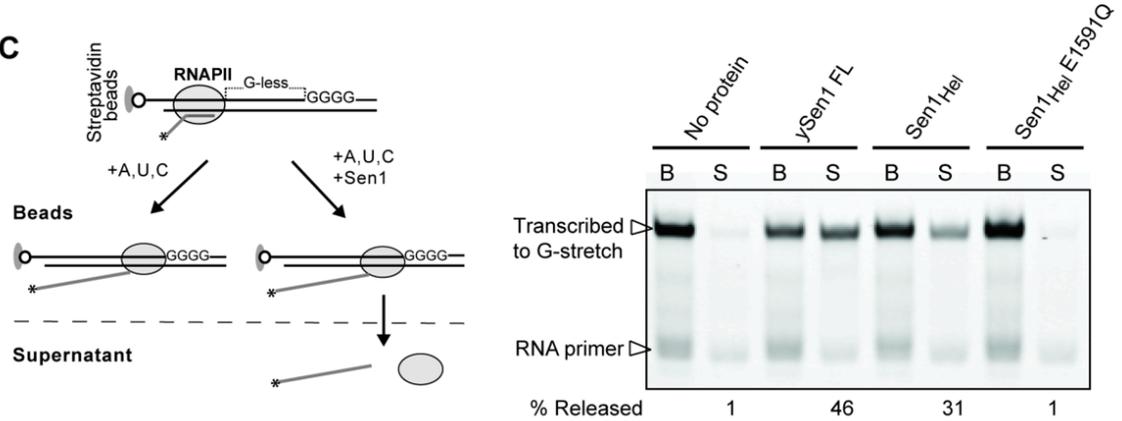


Figure 2

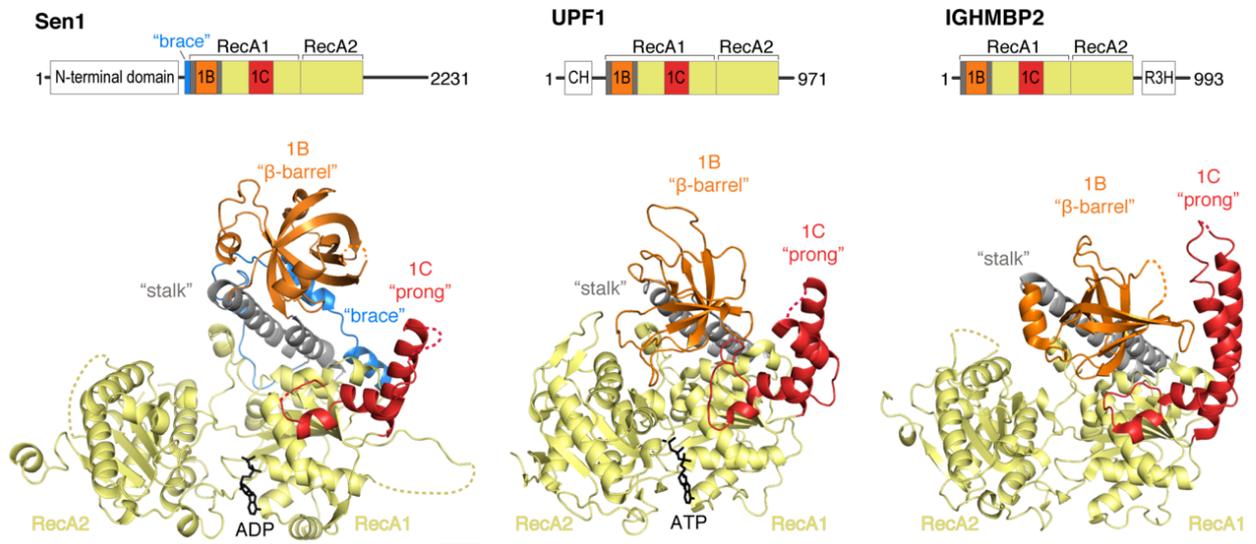


Figure 3

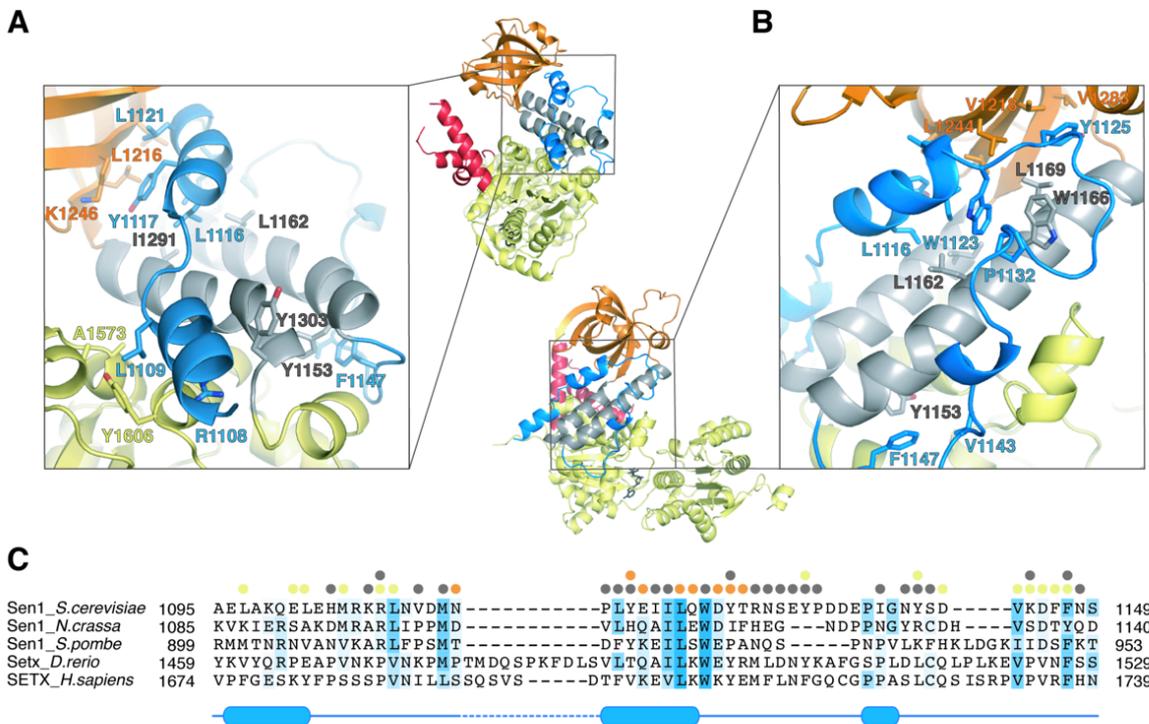
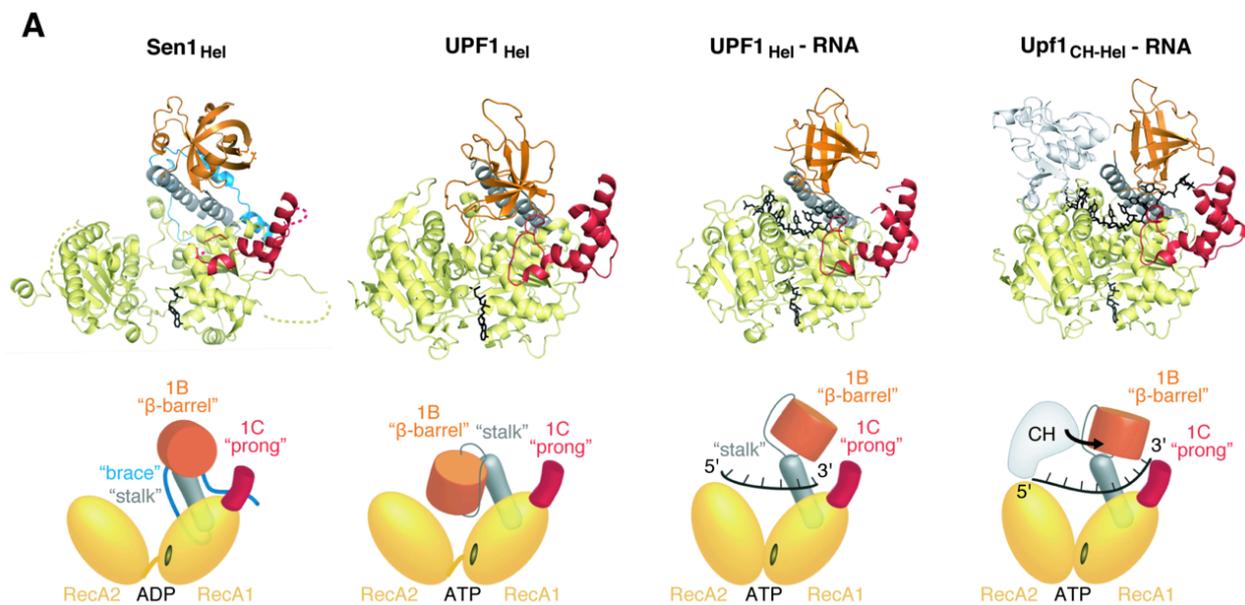


Figure 4



B

Upf1 _{Hel}	+	+	+	+	+	+	+
Upf1 _{CH-Hel}		+					
Sen1 _{Hel}			+				+
ADP	+	+	+	+	+	+	+
BeF ₃	+	+	+				
AIF ₄				+	+	+	
57-mer RNA	+	+	+	+	+	+	+
RNaseA/T1	+	+	+	+	+	+	+
		1	2	3	4	5	6
							7
							8

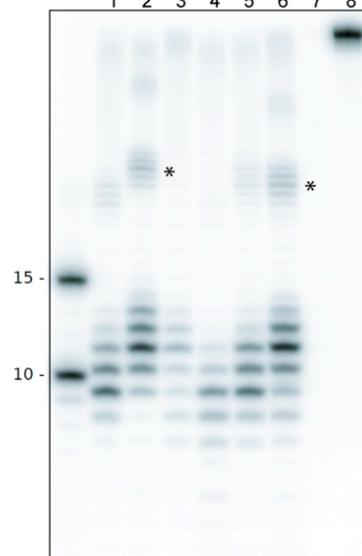
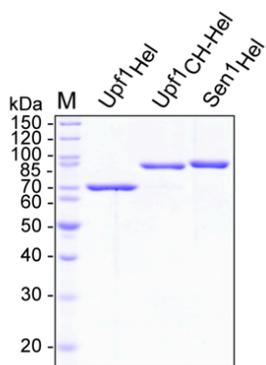


Figure 5

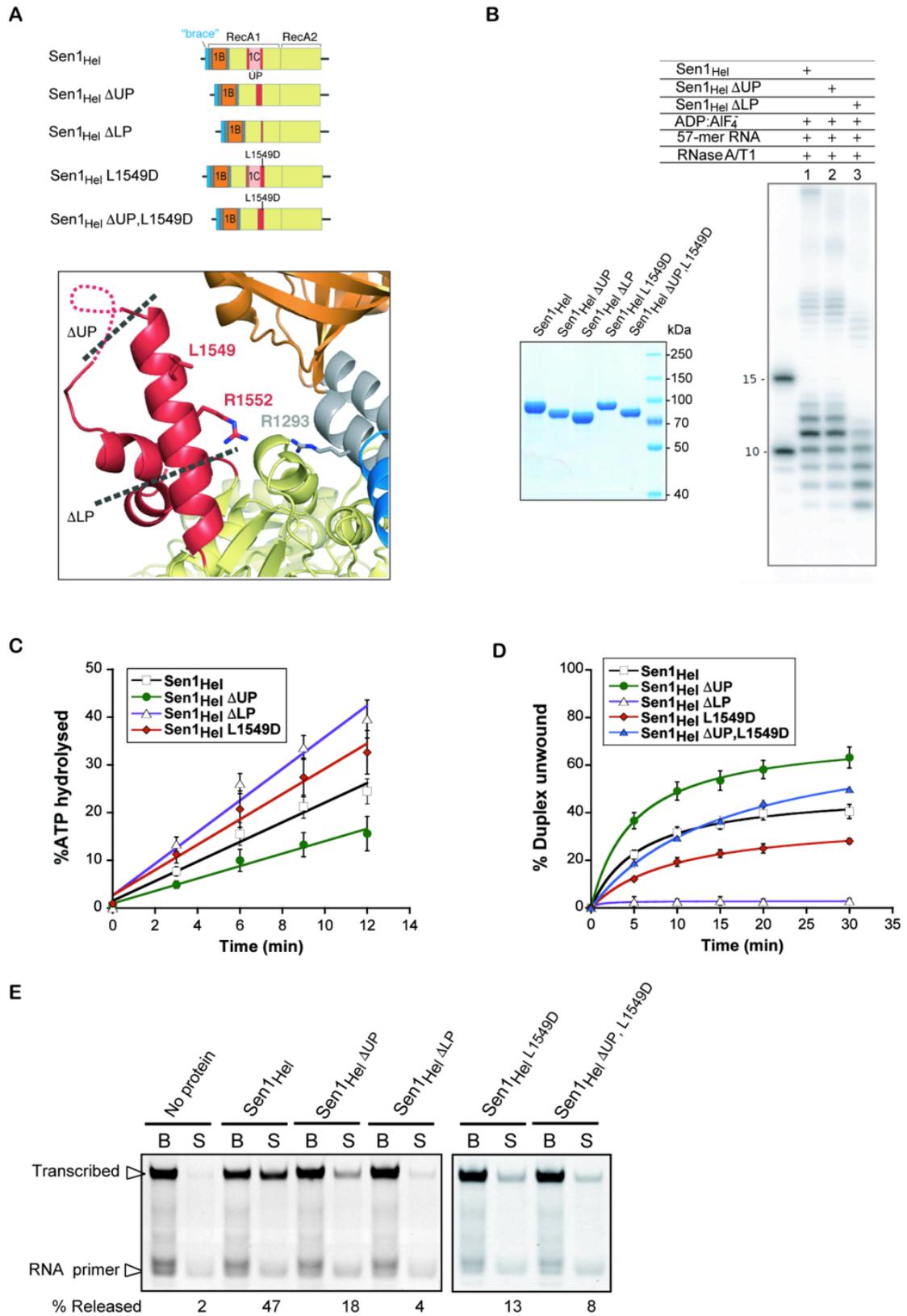


Figure 6

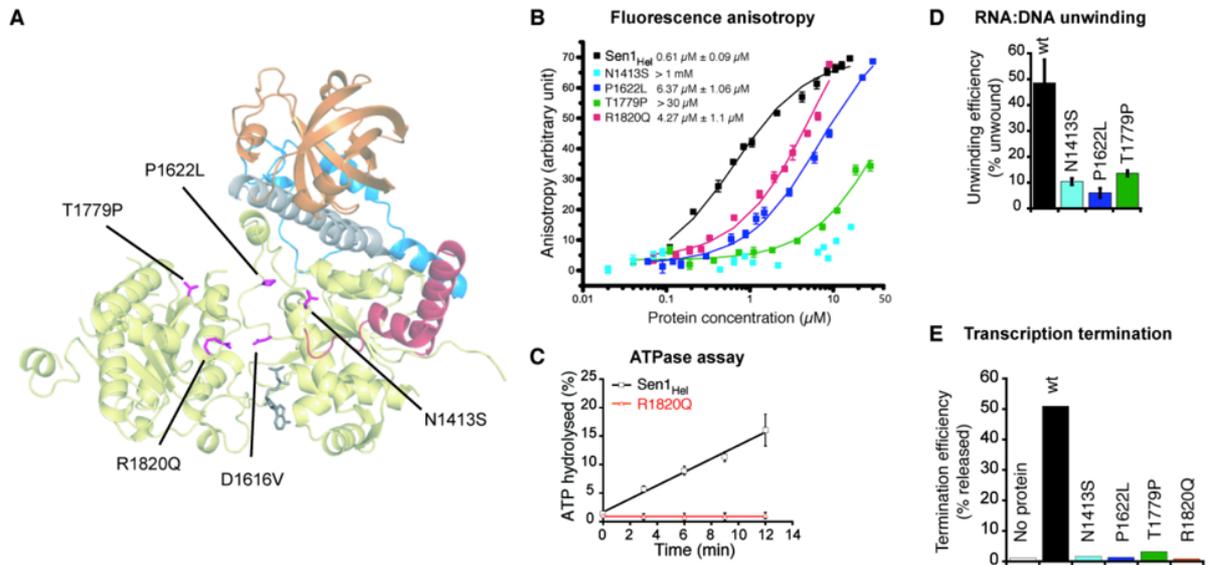


Table1: Crystallography statistics.

Data set	Sen1 _{Hel} native	Sen1 _{Hel} S-SAD
Data collection		
Space group	P 21 21 2	P 21 21 2
Unit cell (a, b, c in Å)	90.285, 171.944, 69.094	90.2, 171.66, 68.85
Wavelength (Å)	1.00	2.095
Resolution range (Å)	48.39 - 1.787 (1.851 - 1.787)	85.83 - 2.145 (2.221 - 2.144)
Total reflections	680302 (29401)	1643000
Unique reflections	100766 (2170)	114276 (8032)
Multiplicity	13.2 (13.5)	
Completeness (%)	98.27 (95.07)	93.9 (68.81)
Mean I/sigma(I)	18.8 (1.6)	22.47 (1.2)
Wilson B-factor	29.6	30.57
R-merge	0.085 (1.500)	N/D
R-meas	0.092	0.086
CC1/2	0.999 (0.610)	0.999 (0.69)
CC*	1 (1)	1 (1)
Refinement		
R-work (%)	15.28	
R-free (%)	18.36	
Number of non-hydrogen atoms	6970	
macromolecules	5543	
ligands	337	
water	1090	
Protein residues	682	
RMS(bonds)	0.011	
RMS(angles)	1.39	
Ramachandran favored (%)	98	
Ramachandran outliers (%)	0	
Clashscore	10.86	
Average B-factor	49.2	
macromolecules	42	
ligands	100.50	
solvent	66.4	

Statistics for the highest-resolution shell are shown in parentheses.

Table 2: Summary of the phenotypes of Sen1_{Hel} mutants

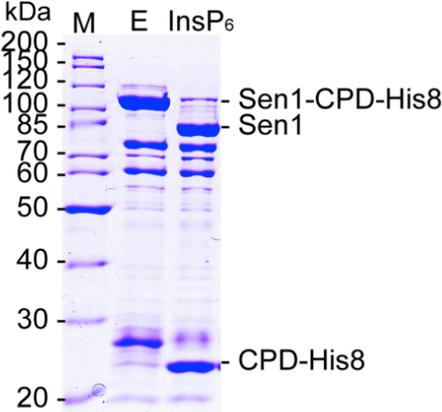
Sen1 _{Hel} version	ATP hydrolysis	RNA binding	Unwinding	Termination	Growth [°]
Wild-type	+++	+++	+++	+++	Normal
T1289A, R1293A	+/-	+/-	ND	-	ND
N1413S	ND	-	+/-	-	HS
Δ 1471-1538 (Δ UP)	++	+++	++++	++	ND
Δ 1461-1554 (Δ LP)	+++	+++	-	-	Lethal
L1549D	+++	ND	++	+/-	ND
Δ UP, L1549D	ND	ND	+++	+/-	ND
E1591Q	-	ND	ND	-	ND
P1622L	ND	+	-	-	Lethal
T1779P	ND	+/-	+/-	+/-	HS
R1820Q	-	+++	ND	-	Lethal

ND: Not done; HS: Heat sensitive.

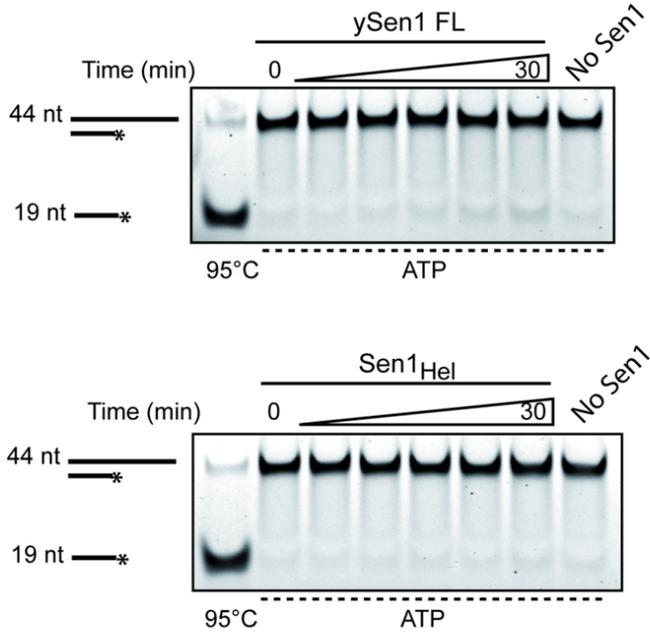
[°]Growth of yeast expressing the indicated version of Sen1 according to Chen et al, 2014, except for Δ LP (see figure EV4).

Figure EV1

A



B



C

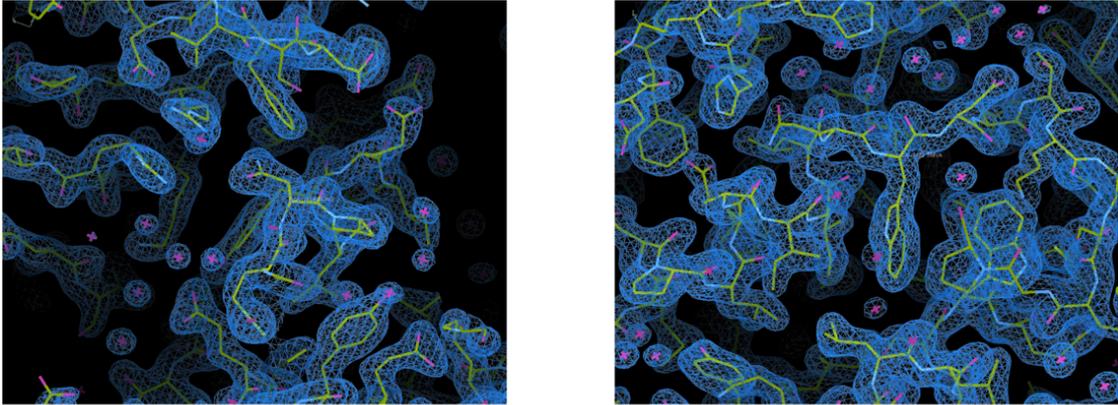
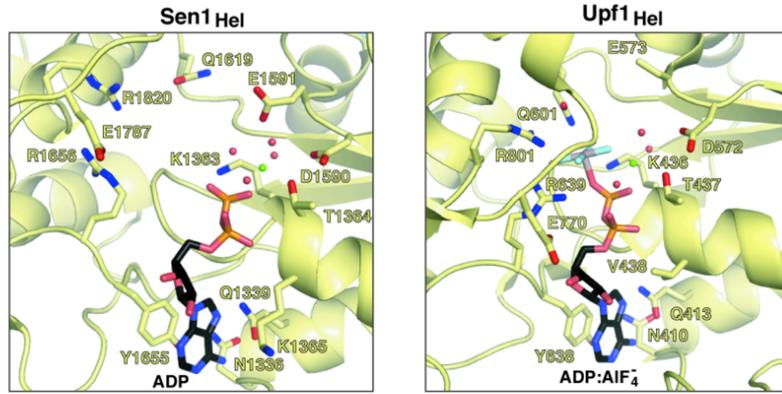
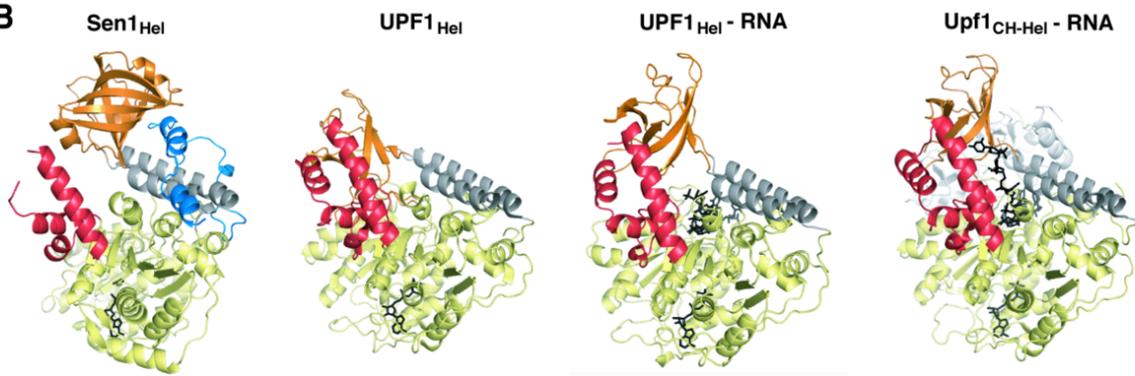


Figure EV2

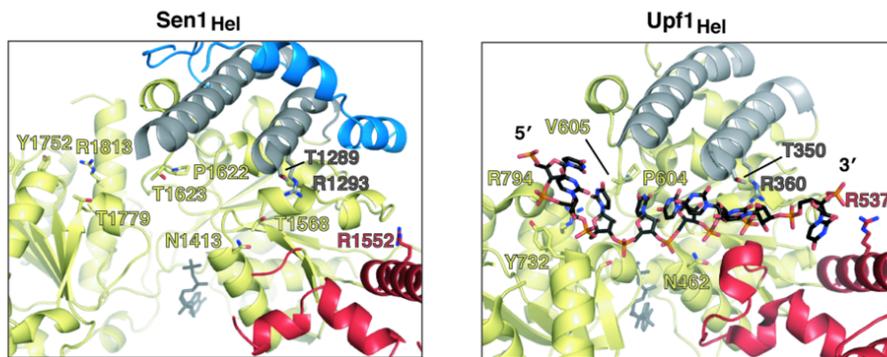
A



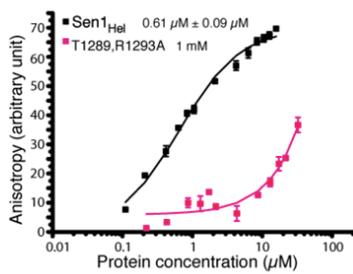
B



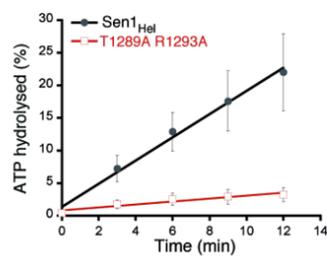
C



D Fluorescence anisotropy



E ATPase assay



F Transcription termination

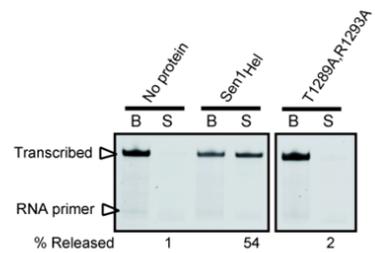


Figure EV3

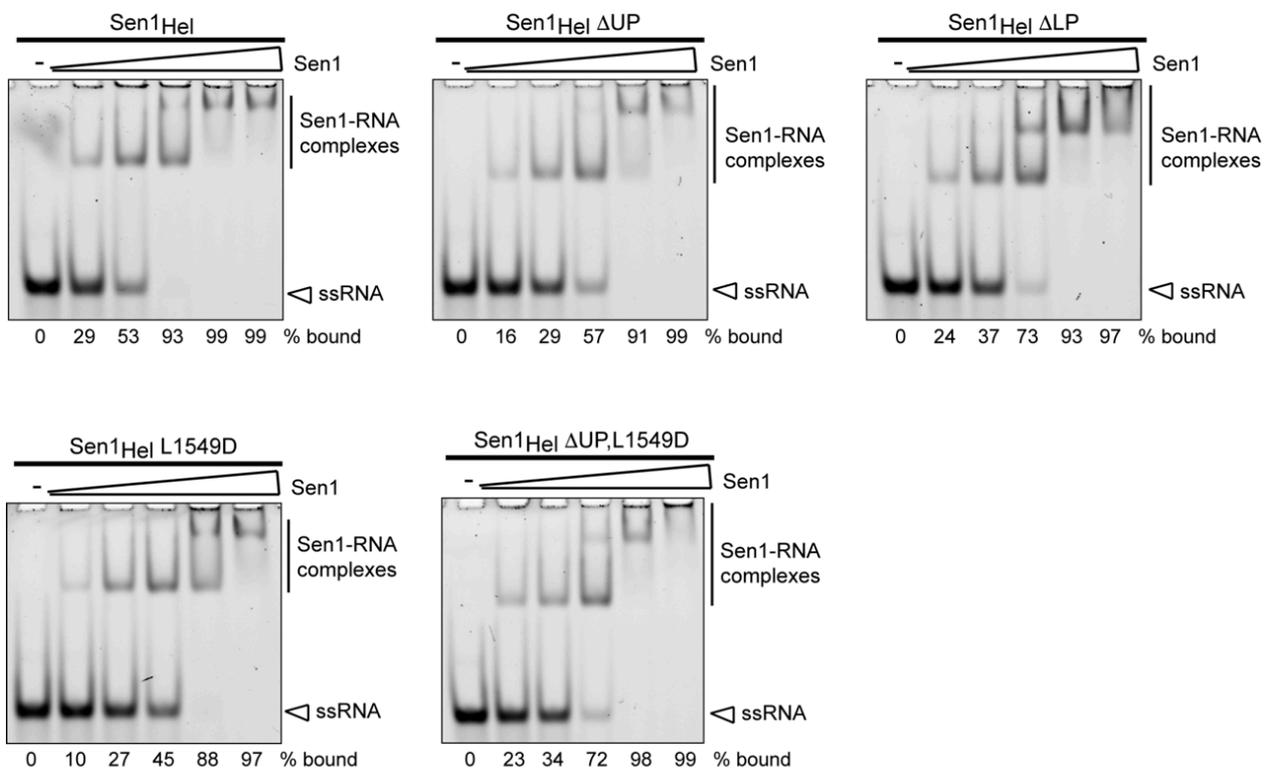
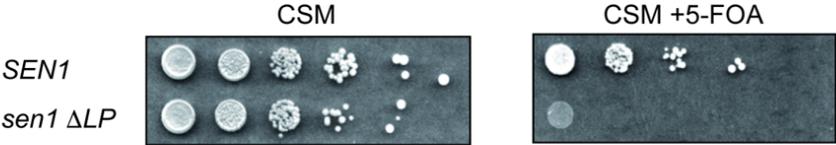
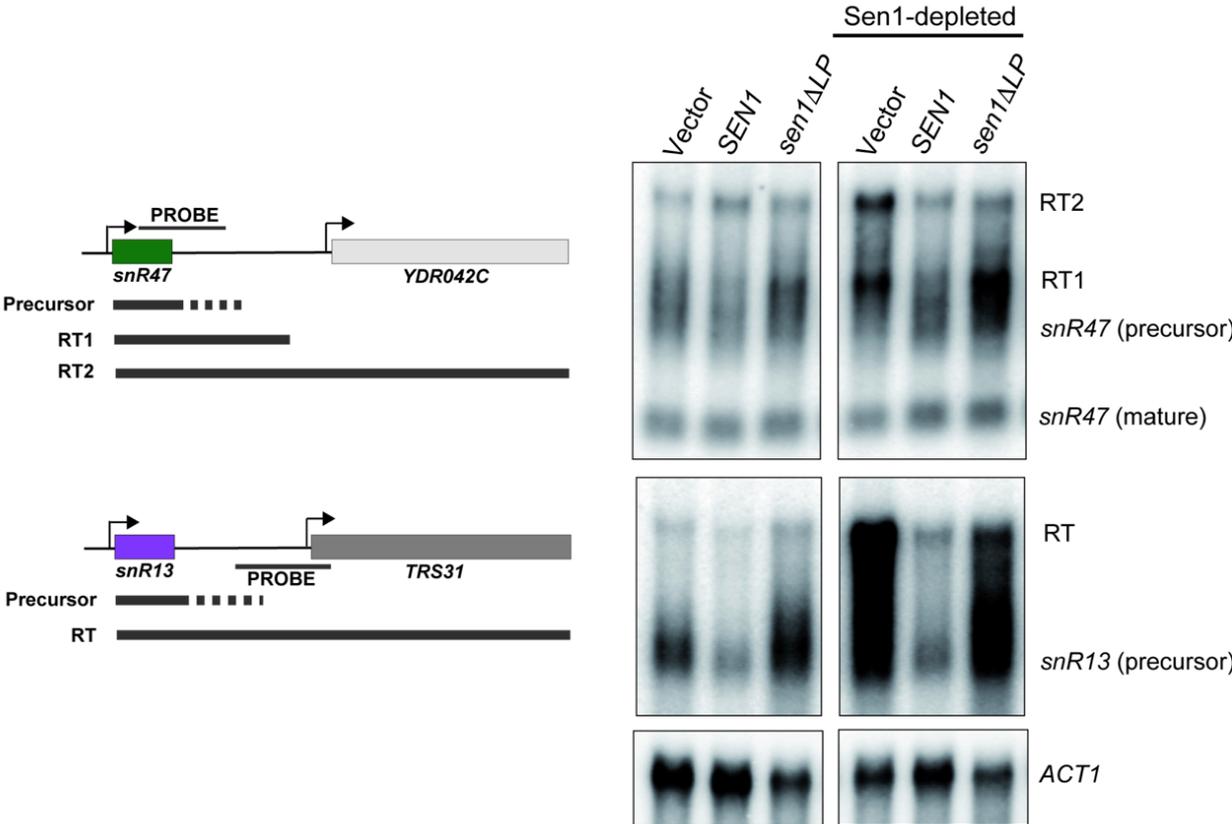


Figure EV4

A



B



Appendix:

Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family

Supplementary information:

Figure S1

Table S1

Supplementary Methods

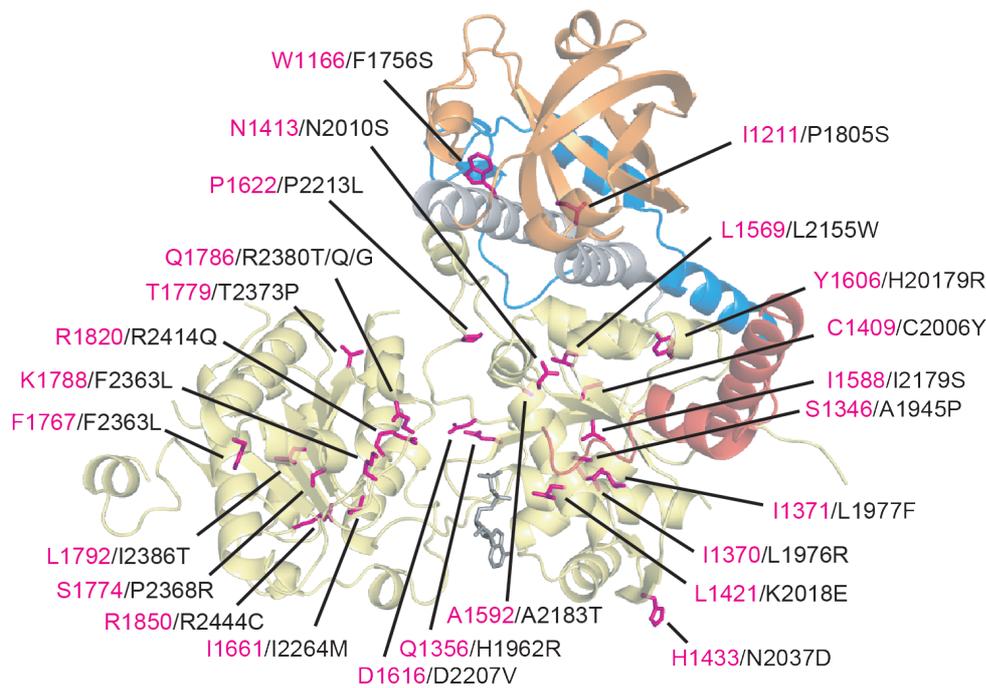


Figure S1 - AOA2-associated missense mutations.

Conserved AOA2-associated mutations mapped on the Sen1_{Hel} structure. The residues that represent SETX missense mutations (black labels) are shown as sticks in magenta.

Table S1: List of oligonucleotides used in this work.

Name	Sequence (5'-3')	Information/use
DL1119	AAGTGACGAAGTTCATGCTA	Forward oligo to generate by PCR a probe to detect snR13 read-through region.
DL1154	CCTATAACAACAACAACATG	Forward oligo to generate by PCR a probe to detect snR47 RNA.
DL1157	ATAGCCATTAGTAAGTACGC	Reverse oligo to generate by PCR a probe to detect snR47 RNA.
DL1367	GGCCCAACAGTATATTCATATCC	Reverse oligo to generate by PCR a probe to detect snR13 read-through region.
DL2492	UGCAUUUCGACCAGGC	16-mer RNA oligonucleotide used for promoter-independent assembly of ECs. 5'-end FAM labeled.
DL2661	TAATACGACTCACTATAGGGGACCG GATGAGCGGTATTGAGTTTGAATTT ATCGATGGTATCAGATCTGGATCCT CGAGAAGCTGCGGGTACC	Non-template strand to produce by <i>in vitro</i> transcription the 75-mer RNA component of RNA:DNA duplexes used in unwinding assays in figure 6.
DL2662	GGTACCCGCAGCTTCTCGAGGATCC AGATCTGATAACCATCGATAAATTC AACTCAATACCGCTCATCCGGTCCC CTATAGTGAGTCGTATTA	Template strand to produce by <i>in vitro</i> transcription the 75-mer RNA component of RNA:DNA duplexes used in unwinding assays in figure 6.
DL3316	UUCAUUUCAGACCAGCACCCACUC ACUACAACUCACGACCAGGC	44-mer RNA oligonucleotide used to form the RNA:DNA duplexes used in unwinding assays in figures 1, 5 and S1.
DL3352	CTAGAGGAAACAAACTATAGGAAA CGACCAGGCCCTCAACATCTCTCAC CCATCTCCACACGGGGGTACCCGG CCTGCA	Non-template strand used in IVTT assays. Labeled with biotin at the 5'-end.
DL3353	GGCCGGGTAACCCCGTGTGGAGAT GGGTGAGAGATGTTGAGGGCCTGGT CGTTTCTATAGTTTGTTTCT	Template strand used in IVTT assays. Labeled with biotin at the 5'-end.
DL3522	GGTACCCGCAGCTTCTCGAG	Short oligonucleotide for forming the RNA:DNA duplex containing a 5' single-strand overhang used in unwinding assays in figure 6.
DL3791	GCCTGGTCGTGAGTTGTAG	Short oligonucleotide that annealed to DL3316 forms the RNA:DNA duplex used in unwinding assays in figures 1 and 6. 5'-end FAM labeled.
DL3810	GCCTGGTCGTGAGTTGTAG	Same sequence as DL3791 but without the FAM label. Used as a competitor to prevent reannealing of the unwound oligo.
DL3841	GGTGCTGGTCTGAAATGAA	Short oligonucleotide that annealed to DL3316 forms an RNA:DNA duplex containing a 3' overhang used in unwinding assays in figure S1. 5'-end FAM labeled.
DL3842	GGTGCTGGTCTGAAATGAA	Same sequence as DL3841 but without the FAM label. Used as a competitor to prevent reannealing of the unwound oligo.
ARE	UUUCUAUUUAUUUUG	15-mer RNA oligonucleotide used for fluorescence anisotropy measurements. 5'-end FL labeled.

Supplementary Methods:

Electrophoretic mobility shift assay

Assays were performed in 10 μ l-reactions containing 10 mM Tris-HCl pH 7.5, 70 mM NaCl, 2 mM MgCl₂, 7.5 μ M ZnCl₂, 10 μ g/ml BSA, 10% glycerol and 1 mM DTT. The RNA substrate (2 nM final concentration) was incubated with increasing concentrations of Sen1_{Hel} or Sen1_{Hel}DLP at 28°C for 15 min. Reactions were loaded on 6% native polyacrylamide gels and subjected to electrophoresis in 0.5XTBE at 100V for 80 min. Gels were scanned directly using a Typhoon scanner (GE Healthcare).

RNA expression analyses

RNAs were prepared using standard methods. Samples were separated by electrophoresis on 1.2% agarose gels, and then transferred to nitrocellulose membranes and 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 -³²P dCTP (3000 Ci/mmol). Oligonucleotides used to generate the PCR probes are listed in Appendix Table S1. Hybridizations were performed using a commercial buffer (Ultrasch, Ambion) and after washes, membranes were analysed by phosphorimaging.

DISCUSSION

1 The mechanism of Sen1-dependent transcription termination

1.1 Sen1 helicase domain is sufficient for transcription termination *in vitro*

In a previous work, it has been demonstrated that full-length Sen1 can elicit RNAP II transcription termination *in vitro* (95). The integrity of helicase domain is required for Sen1-mediated transcription termination *in vivo* (164, 268) and the N-terminal and C-terminal domains flanking the helicase core are involved in mediating the interaction of Sen1 with RNAP II and Nab3, respectively (71, 94, 267). However, whether the Sen1 helicase core alone can elicit transcription termination *in vitro* and whether N-terminal and C-terminal domains of Sen1 participate in the step of elongation complex (EC) dissociation have remained open questions.

In my thesis, I have performed a functional dissection of Sen1 to explore these aspects. I have found that the helicase domain can bind single-stranded nucleic acids and it is active for ATPase and helicase functions. Importantly, it is sufficient for the step of dissociation of the EC, indicating that this region contains all the properties that are essential for the final step of termination. The N-terminal and C-terminal domains of Sen1 do not play a direct role in the step of EC dissociation. However, they might be important for the regulation of Sen1 activity *in vivo* (See discussion later).

In the previous report, it has been shown that the capacity to dismantle an EC is not an unspecific property of any RNA helicase (95). Thus, it seems interesting to undertake structural analyses of Sen1 helicase domain to identify the determinants of Sen1 termination activity.

1.2 Distinctive structural features of Sen1 helicase domain involved in transcription termination

Our collaborators from Elena Conti's Lab have managed to determine the crystal structure of Sen1 helicase domain at high resolution. In addition to the catalytic and auxiliary subdomains characteristic of SF1B helicases, Sen1 has a distinct and evolutionary conserved structural feature at the N-terminus of the helicase domain that we have dubbed the "brace". This subdomain appears to fasten and stabilize the overall fold of the protein *since in vitro*, deletions of "brace" residues result in insoluble proteins and *in vivo*, a deleted variant of Sen1 lacking part of the "brace" is inviable (268).

We also find that accessory subdomain 1C (the "prong") is an essential element for 5'-3' unwinding activity and for Sen1-mediated transcription termination *in vitro*. A Sen1 variant lacking most of the "prong" can still bind to single-stranded nucleic acids and hydrolyze ATP. However, it can neither unwind double-stranded nucleic acids nor release the elongation complex. Preliminary single-molecule analyses performed with several variants of Sen1 helicase domain (Shuang Wang and Terence Strick, unpublished) indicate that the "prong"-deleted mutant can still translocate on single-stranded nucleic acids, which strongly suggests that the "prong" specifically helps melting both strands of a duplex region. Because the "prong" is critical for transcription termination both *in vitro* and *in vivo*, it is tempting to speculate that upon encountering that RNAPII, further translocation of Sen1 might provoke the invasion of the RNA exit channel by the "prong". The consequent opening of the channel would lead to profound conformation changes and destabilization of the elongation complex (Figure 27). This process might be facilitated by the partially unstructured nature of the "prong" and by specific interactions between the "prong" and RNAPII. Indeed, combining a deletion of the disordered portion of the "prong" with a mutation in a residue of the "prong" (L1549D) putatively involved in protein-protein

interactions resulted in a substantial impairment of the termination efficiency, whereas the unwinding activity remained almost unaffected.

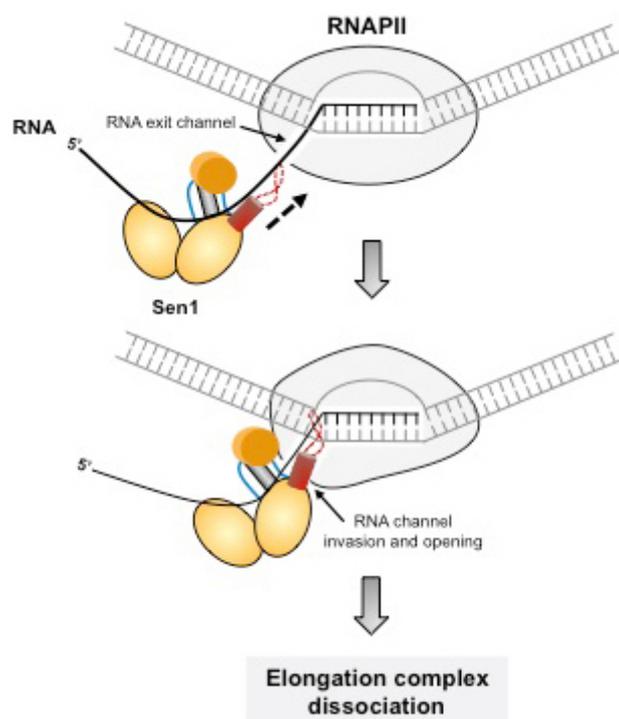


Figure 27: An allosteric model of transcription termination by Sen1. See the main text for details.

Understanding whether Sen1 interacts with specific regions of RNAP II to elicit termination remains a major challenge. One possible strategy to tackle this question is to perform cross-linking coupled with mass spectrometry (XL-MS) experiments. By freezing the transient interactions through the formation of covalent bonds, XL-MS might provide vital insights into both the structure and organization of Sen1 and RNAP II during the Sen1-mediated EC dissociation. In addition, since the “prong” plays a crucial role in RNAP II release, it would be interesting to try swapping the auxiliary subdomain of Sen1 with that of its closest homologue, the helicase Upf1, which is inactive for transcription termination *in vitro*. If the “prong” subdomain is indeed the key determinant of Sen1 function in releasing EC, it is possible that Upf1 helicase core harboring Sen1 “prong” would gain the capacity to induce transcription

termination while a Sen1 variant carrying the Upf1 auxiliary subdomain would lose the ability to release the elongation complex. However, it remains possible that the specific activity of Sen1 in termination is a consequence of not only the distinctive features of the “prong”, but also the presence of the “brace” that imposes a particular conformation on other subdomains like the “barrel”, which would also theoretically be located close to or in contact with RNAPII during the termination process. In that case, constructing a termination-proficient chimeric Upf1 protein would be extremely challenging. The XL-MS experiments proposed above followed by mutational analyses could help understanding the role of subdomains of Sen1 other than the “prong” in EC dissociation.

1.3 Features of the elongation complex with a potential impact on Sen1-mediated termination

The elongation complex adopts different translocation states during transcription. According to the Brownian ratchet model, the EC is in an equilibrium between the pre- and post-translocation states during active transcription (135, 139) and it has also been shown that RNAP II can undergo backtracking at some sequence-specific pausing sites (151, 152). In addition, the efficiency of termination by Rho helicase, with which Sen1 shares important mechanistic traits, is closely related to the translocation state of RNAP since the rate of the EC dissociation is reduced if the EC is backtracked (227). It is unclear whether part or the totality of the ECs paused in our *in vitro* transcription termination assays undergo backtracking. The existence of backtracking could be unveiled by RNase footprinting analyses of the EC in our conditions. It would be interesting to test whether similar to Rho, Sen1 works less efficiently on backtracked RNAPIIs. If that would be the case, we could perform complementary *in vitro* termination assays in the presence of TFIIS, an elongation factor that, as detailed in a previous section (see section 1.2.2.2), promotes the rescue of backtracked ECs. In addition, it would be important to test whether Sen1 exhibits a

“preference” for RNAPIIs in the pre-translocated or in the post-translocated state. There are protocols available that allow characterizing the translocation state of RNAPs (293).

In addition, a structural element within the active site of bacterial RNAP, the trigger loop, has been found to be important for Rho-dependent EC dissociation (6). The antibiotic tagetitoxin (Tgt) can bind RNAP trigger loop in the vicinity of the catalytic site and restrict the movement of the trigger loop (232). Interestingly, ECs composed of Tgt-bound RNAP are refractory to Rho-mediated transcription termination (227, 231). It was later found that Tgt can increase the stability of the pre-translocated state of the EC by stabilizing a folded conformation of the trigger loop, which inhibits forward and backward translocation of the complex (294). Similarly, the mushroom toxin α -amanitin can also contact the bridge helix and trigger loop of RNAP II, trapping the EC in a new conformation that is intermediate between pre- and post-translocation structures (136, 144). It would be interesting to test whether α -amanitin can affect Sen1-dependent transcription termination and explore the possible underlying mechanisms. For instance, if α -amanitin can indeed prevent termination it will be important to understand whether the inhibition is due to the fact that the toxin blocks RNAPII conformational changes that are critical for termination or is a consequence of the action of this molecule on the translocation state of RNAPII (or both).

Elucidating these aspects is crucial not only to attain a full comprehension of the mechanisms of Sen1-mediated termination, but also eventually to spot critical steps of the process of termination that could be a target of regulation.

2 Regulation of Sen1 function

Helicases are enzymes that often have little substrate specificity and that require additional factors to specify their biological function and regulate their activity (158, 295). During my thesis work, I have found that *in vitro*, the presence of the N-terminal domain decreases Sen1 ATPase activity and the efficiency of Sen1-mediated termination. This behavior is reminiscent of that of Upf1. In the case of Upf1, intra-molecular interactions mediated by N-terminal and C-terminal extensions of the helicase domain induce autorepression. The autoinhibition exerted by the N-terminal domain, is relieved upon interaction with the Upf1 partner Upf2 (175, 181). It remains to be tested whether the N-terminal domain of Sen1 also mediates intra-molecular interactions. We have an ongoing collaboration with the group of Vlad Pena (Max Planck, Gottingen), which has recently succeeded in producing recombinant full-length Sen1 and are currently attempting structural analyses by cryo-electron microscopy. We expect to obtain a more complete picture of the architecture and the network of intra-molecular interactions of Sen1 within the next few months.

I have also observed that the presence of the C-terminal domain improves Sen1 helicase activity, especially on RNA:DNA duplexes. Helicases often possess additional nucleic-acid binding domains at the extensions of their core helicase domains (174). It is therefore possible that the C-terminal domain contains an additional RNA-binding domain. Such possible RNA-binding activity could be particularly relevant for termination *in vivo*, where the presence of a multitude of competing RNA-binding proteins (e.g. RNP factors) might limit the access to the nascent RNA. Indeed, we have recently purified a recombinant form of the C-terminal domain of Sen1 (aa 1930-2231) that is soluble and preliminary EMSA assays indicate that this domain can bind nucleic acids *in vitro*. Our collaborators from R. Stefl lab (CEITEC, Brno) are working to solve the structure in solution of this domain and characterize in more detail its nucleic acid-binding activity.

Sen1 is the catalytic subunit of the Nrd1-Nab3-Sen1 complex (NNS) and the C-terminal domain of Sen1 plays the role in mediating interaction with Nrd1 and Nab3 (71). We considered the possibility that Nrd1 and Nab3 play a role in regulating directly Sen1 catalytic activity. For that reason, I performed a set of *in vitro* experiments with recombinant Nrd1 and Nab3 together with full-length Sen1 purified from yeast to test whether the biochemical activities of Sen1 are affected by the interaction with Nrd1 and Nab3. Although we could reproduce the interaction of Sen1 with recombinant Nrd1 and Nab3 *in vitro*, we did not detect any clear effect of Nrd1-Nab3 on any of the measurable activities of Sen1 (data not shown). These results suggest either that Nrd1 and Nab3 cannot to directly modulate Sen1 catalytic activity or that this modulation requires protein modifications that are absent in the recombinant proteins.

Unlike bacterial Rho helicase, which can bind Rho utilization sites (*rut*) on the nascent RNA and recruit itself to the termination substrates, Sen1 cannot recognize specific sequences on the nascent RNA. Instead, it has been proposed that the recruitment of Sen1 is mediated by the other two components of the NNS complex, Nrd1 and Nab3 (62). In addition, the N-terminal domain of Sen1 has been proposed to interact with Ser2-phosphorylated CTD of RNAP II (94, 267), which might also be involved in the recruitment of Sen1. Further investigations are required for better understanding of the mechanism of Sen1 recruitment to termination sites.

The C-terminal domain of Sen1 has also been shown to interact with Glc7, a yeast type1 phosphatase that dephosphorylates the CTD of RNAPII phosphorylated at Tyr1, and the removal of Glc7 cause termination defects at snoRNA genes (71). However, the mechanism by which Glc7 is involved in the transcription termination of snoRNA targets and the role of the interaction between Sen1 and Glc7 are still unclear. It is possible that Glc7 regulate the function of the NNS complex via modulating the phosphorylation state of the NNS components. Other members of our laboratory are currently exploring this possibility.

In addition, it would be interesting to identify new facilitators or inhibitors of Sen1-dependent termination. Quite a few antitermination factors for Rho have been identified in *E. coli* (296). For example, RfaH associates with RNAP and reduces EC pausing, thereby decreasing Rho-dependent termination efficiency (297, 298). YaeO binds directly to Rho and block the interaction between Rho and *rut* sites (296, 299). On one hand, an *in vivo* genetic screening could be performed in *sen1* hypomorphic mutants using an appropriate reporter system to try to identify factors that can suppress the termination defects when overexpressed (i.e. facilitators of Sen1 action) or when mutated (i.e. antiterminators). On the other hand, partially purified yeast ECs can be used in our *in vitro* transcription termination assays to try to identify RNAPII-associated proteins that can affect the efficiency of termination by Sen1. Finally, there is an ongoing project in our lab that aims at characterizing in detail the network of Sen1-interacting factors to try to detect proteins that can modulate the efficiency of termination *in vivo*.

3 Implications for the function of senataxin

Although many studies have linked mutations in SETX to amyotrophic lateral sclerosis type 4 (ALS4) and ataxia-ocular apraxia type 2 (AOA2) (162-164, 284), for the moment it is unclear how SETX mutations can lead to these neurodegenerative diseases.

SETX mutants present a plethora of phenotypes, none of which can univocally be spotted as the main cause of ALS4 and/or AOA2. For instance, SETX depletion or mutation leads to defective transcription termination and/or decreased expression of several protein-coding genes (285, 291, 292), accumulation of R-loops (285, 287-289), an increased basal level of oxidative DNA damage and an increased sensitivity to agents that cause oxidative stress (282, 300).

Recently, using yeast Sen1 as a surrogate for SETX the group of D. Brow investigated if the AOA2 mutations observed in patients could be associated with defects in transcription termination. Interestingly, out of the 13 missense AOA2 mutations studied that are located in the SETX helicase domain, 10 (77%) caused lethality or growth defects and decreased termination efficiency *in vivo* (268), which lead the authors of that study to suggest that defects in transcription termination may be a primary cause of the AOA2 disease.

We extended the above *in vivo* results with our structural and biochemical analyses performed in collaboration with E. Conti's lab. Comparing the primary sequence of Sen1 and SETX, we found that the unique "brace" subdomain, which stabilizes the overall fold of the helicase domain, also exists in SETX, and one AOA2-associated mutation (F1756S, W1166S in Sen1) affects indeed one of the key residues on the 1B subdomain that interacts with the "brace". Given the high conservation between Sen1 and SETX we took advantage of our structural data on Sen1 to get insight into the molecular effect of SETX mutations. Mapping of 30 disease-associated mutations on

Sen1 structure allowed us to predict that 2/3 of them would disrupt the overall folding of the protein, whereas 1/3 of them would affect the catalytic activity. We produced several Sen1 variants harboring some of these AOA2 disease-associated mutations predicted to impact directly Sen1 catalytic activity and we perform a complete biochemical characterization of the different mutants *in vitro*. We found that these mutations either affected ATP hydrolysis or decrease the affinity of Sen1 for the RNA, thereby impairing the unwinding activity as well as the termination efficiency. These results strongly suggest that AOA2 is a consequence of the dysfunction of the helicase domain of SETX. However, because the determinants of the duplex unwinding activity more often coincide with those of the transcription termination activity, the *in vivo* study mentioned above (268) and our present results cannot establish a clear link between the termination defects and AOA2. This neurodegenerative disease could still be due to other functions of Sen1 that more directly depend on its capacity to unwind RNA:DNA duplexes (i.e. the resolution of R-loops).

Nevertheless, in the absence of a recombinant version of SETX, Sen1 remains a powerful model to investigate mutations associated with AOA2 and ALS4. Using the biochemical assays we have set up, additional disease mutants could be tested and might possibly lead to the identification of residues that are important for transcription termination but not for duplex unwinding. For instance, the L1549D mutant, which locates in Sen1 “prong” subdomain, cause a substantial decrease in termination efficiency with a very minor effect on Sen1 unwinding activity. According to the alignment of Sen1 and SETX sequences, one ALS4-associated mutation (R2136H/C in SETX) is expected to locate in the same helix as L1549 in Sen1, within the “prong” subdomain. It would be interesting to test whether the R2136H/C mutations specifically impair transcription termination.

Concerning the mechanism of SETX-dependent transcription termination it has been proposed that SETX could resolve R-loops that naturally form at termination sites so that the nascent RNA becomes accessible to the 5'-3' “torpedo” exonuclease XRN2

(285, 291). Subsequently, XRN2 degrades the nascent transcript and induces RNAP II release. However, it has recently been shown that XRN2 is not required for SETX-mediated early transcription termination at virus-induced genes (292). The precise mechanism by which SETX mediates RNAP II transcription termination requires further investigation. To address this aspect, it will be crucial to analyse the behavior of SETX *in vitro* using a highly purified transcription system to test whether it can elicit the dissociation of elongation complex as its yeast counterpart. Our collaborators from V. Pena lab are currently attempting the purification of a recombinant version of SETX, if they succeed we will be able to perform an *in vitro* characterization of SETX and address this important question.

BIBLIOGRAPHY

1. Yusupov MM, *et al.* (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292(5518):883-896.
2. Sharp SJ, Schaack J, Cooley L, Burke DJ, & Soll D (1985) Structure and transcription of eukaryotic tRNA genes. *CRC critical reviews in biochemistry* 19(2):107-144.
3. Matera AG, Terns RM, & Terns MP (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nature Reviews Molecular Cell Biology* 8(3):209-220.
4. Ambros V (2004) The functions of animal microRNAs. *Nature* 431(7006):350-355.
5. Jensen TH, Jacquier A, & Libri D (2013) Dealing with pervasive transcription. *Molecular cell* 52(4):473-484.
6. Porrua O, Boudvillain M, & Libri D (2016) Transcription Termination: Variations on Common Themes. *Review Trends Genet* 32(8):508-522.
7. Martinez-Rucobo FW & Cramer P (2013) Structural basis of transcription elongation. *Biochimica et biophysica acta* 1829(1):9-19.
8. Sainsbury S, Bernecky C, & Cramer P (2015) Structural basis of transcription initiation by RNA polymerase II. *Nature reviews. Molecular cell biology* 16(3):129-143.
9. Basehoar AD, Zanton SJ, & Pugh BF (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116(5):699-709.
10. Huisinga KL & Pugh BF (2004) A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Molecular cell* 13(4):573-585.
11. Nikolov DB, *et al.* (1992) Crystal structure of TFIID TATA-box binding protein. *Nature* 360(6399):40-46.
12. Kim JL, Nikolov DB, & Burley SK (1993) Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 365(6446):520-527.
13. Yang C, Bolotin E, Jiang T, Sladek FM, & Martinez E (2007) Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* 389(1):52-65.
14. Rhee HS & Pugh BF (2012) Genome-wide structure and organization of eukaryotic

- pre-initiation complexes. *Nature* 483(7389):295-301.
15. Chalkley GE & Verrijzer CP (1999) DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *The EMBO journal* 18(17):4835-4845.
 16. Nikolov DB, *et al.* (1995) Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* 377(6545):119-128.
 17. Bushnell DA, Westover KD, Davis RE, & Kornberg RD (2004) Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Angstroms. *Science* 303(5660):983-988.
 18. Rani PG, Ranish JA, & Hahn S (2004) RNA polymerase II (Pol II)-TFIIF and Pol II-mediator complexes: the major stable Pol II complexes and their activity in transcription initiation and reinitiation. *Molecular and cellular biology* 24(4):1709-1720.
 19. He Y, Fang J, Taatjes DJ, & Nogales E (2013) Structural visualization of key steps in human transcription initiation. *Nature* 495(7442):481-486.
 20. Flores O, Maldonado E, & Reinberg D (1989) Factors involved in specific transcription by mammalian RNA polymerase II. Factors IIE and IIF independently interact with RNA polymerase II. *The Journal of biological chemistry* 264(15):8913-8921.
 21. Ohkuma Y, Hashimoto S, Wang CK, Horikoshi M, & Roeder RG (1995) Analysis of the role of TFIIE in basal transcription and TFIIF-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE- α . *Molecular and cellular biology* 15(9):4856-4866.
 22. Grunberg S, Warfield L, & Hahn S (2012) Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature structural & molecular biology* 19(8):788-796.
 23. Sainsbury S, Niesser J, & Cramer P (2013) Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 493(7432):437-440.
 24. Kostrewa D, *et al.* (2009) RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature* 462(7271):323-330.
 25. Pal M & Luse DS (2003) The initiation-elongation transition: lateral mobility of RNA in RNA polymerase II complexes is greatly reduced at +8/+9 and absent by +23. *Proceedings of the National Academy of Sciences of the United States of America*

100(10):5700-5705.

26. Saunders A, Core LJ, & Lis JT (2006) Breaking barriers to transcription elongation. *Nature reviews. Molecular cell biology* 7(8):557-567.
27. Sims RJ, 3rd, Belotserkovskaya R, & Reinberg D (2004) Elongation by RNA polymerase II: the short and long of it. *Genes & development* 18(20):2437-2468.
28. Harlen KM & Churchman LS (2017) The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature reviews. Molecular cell biology*.
29. Hsin JP & Manley JL (2012) The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes & development* 26(19):2119-2137.
30. Lu H, Flores O, Weinmann R, & Reinberg D (1991) The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proceedings of the National Academy of Sciences of the United States of America* 88(22):10004-10008.
31. Wong KH, Jin Y, & Struhl K (2014) TFIIF phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Molecular cell* 54(4):601-612.
32. Bharati AP, *et al.* (2016) The mRNA capping enzyme of *Saccharomyces cerevisiae* has dual specificity to interact with CTD of RNA Polymerase II. *Scientific reports* 6:31294.
33. Peterlin BM & Price DH (2006) Controlling the elongation phase of transcription with P-TEFb. *Molecular cell* 23(3):297-305.
34. Qiu H, Hu C, Gaur NA, & Hinnebusch AG (2012) Pol II CTD kinases Bur1 and Kin28 promote Spt5 CTR-independent recruitment of Paf1 complex. *The EMBO journal* 31(16):3494-3505.
35. Sun M, Lariviere L, Dengl S, Mayer A, & Cramer P (2010) A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II C-terminal repeat domain (CTD). *The Journal of biological chemistry* 285(53):41597-41603.
36. Kim H, *et al.* (2010) Gene-specific RNA polymerase II phosphorylation and the CTD code. *Nature structural & molecular biology* 17(10):1279-1286.
37. Komarnitsky P, Cho EJ, & Buratowski S (2000) Different phosphorylated forms of

- RNA polymerase II and associated mRNA processing factors during transcription. *Genes & development* 14(19):2452-2460.
38. Licatalosi DD, *et al.* (2002) Functional interaction of yeast pre-mRNA 3' end processing factors with RNA polymerase II. *Molecular cell* 9(5):1101-1111.
 39. Jasnovidova O, Krejcikova M, Kubicek K, & Stefl R (2017) Structural insight into recognition of phosphorylated threonine-4 of RNA polymerase II C-terminal domain by Rtt103p. *EMBO reports*.
 40. Zaborowska J, Egloff S, & Murphy S (2016) The pol II CTD: new twists in the tail. *Nature structural & molecular biology* 23(9):771-777.
 41. Harlen KM, *et al.* (2016) Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell reports* 15(10):2147-2158.
 42. Hsin JP, Li W, Hoque M, Tian B, & Manley JL (2014) RNAP II CTD tyrosine 1 performs diverse functions in vertebrate cells. *eLife* 3:e02112.
 43. Descostes N, *et al.* (2014) Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. *eLife* 3:e02105.
 44. Mayer A, *et al.* (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* 336(6089):1723-1725.
 45. Kwak H & Lis JT (2013) Control of transcriptional elongation. *Annual review of genetics* 47:483-508.
 46. Mason PB & Struhl K (2005) Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Molecular cell* 17(6):831-840.
 47. Li W, Giles C, & Li S (2014) Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic acids research* 42(11):7069-7083.
 48. Martinez-Rucobo FW, Sainsbury S, Cheung AC, & Cramer P (2011) Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *The EMBO journal* 30(7):1302-1310.
 49. Cheung AC & Cramer P (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* 471(7337):249-253.

50. Galburt EA, *et al.* (2007) Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature* 446(7137):820-823.
51. Kettenberger H, Armache KJ, & Cramer P (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* 114(3):347-357.
52. Hsieh FK, *et al.* (2013) Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* 110(19):7654-7659.
53. Xiang K, Tong L, & Manley JL (2014) Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Molecular and cellular biology* 34(11):1894-1910.
54. Zhao J, Kessler MM, & Moore CL (1997) Cleavage factor II of *Saccharomyces cerevisiae* contains homologues to subunits of the mammalian Cleavage/polyadenylation specificity factor and exhibits sequence-specific, ATP-dependent interaction with precursor RNA. *The Journal of biological chemistry* 272(16):10831-10838.
55. Zenklusen D, Larson DR, & Singer RH (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology* 15(12):1263-1271.
56. Creamer TJ, *et al.* (2011) Transcriptome-wide binding sites for components of the *Saccharomyces cerevisiae* non-poly(A) termination pathway: Nrd1, Nab3, and Sen1. *PLoS genetics* 7(10):e1002329.
57. Richard P & Manley JL (2009) Transcription termination by nuclear RNA polymerases. *Genes & development* 23(11):1247-1269.
58. Gilmour DS (2007) Transcription termination factor Pcf11 limits the processivity of Pol II on an HIV provirus to repress gene expression. *Genes & development* 21(13):1609-1614.
59. Gilmour DS (2006) Pcf11 is a termination factor in *Drosophila* that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Molecular cell* 21(1):65-74.
60. Gilmour DS (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes & development* 19(13):1572-1580.

61. Gilmour DS (2004) Analysis of polymerase II elongation complexes by native gel electrophoresis. Evidence for a novel carboxyl-terminal domain-mediated termination mechanism. *The Journal of biological chemistry* 279(22):23223-23228.
62. Porrua O & Libri D (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature reviews. Molecular cell biology* 16(3):190-202.
63. West S, Gromak N, & Proudfoot NJ (2004) Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432(7016):522-525.
64. Kim M, *et al.* (2004) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432(7016):517-522.
65. Park J, Kang M, & Kim M (2015) Unraveling the mechanistic features of RNA polymerase II termination by the 5'-3' exoribonuclease Rat1. *Nucleic acids research* 43(5):2625-2637.
66. Pearson EL & Moore CL (2013) Dismantling promoter-driven RNA polymerase II transcription complexes in vitro by the termination factor Rat1. *The Journal of biological chemistry* 288(27):19750-19759.
67. Baejen C, *et al.* (2017) Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Molecular cell* 66(1):38-49 e36.
68. Fong N, *et al.* (2015) Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Molecular cell* 60(2):256-267.
69. Zhang H, Rigo F, & Martinson Harold G (2015) Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. *Molecular cell* 59(3):437-448.
70. Luo W, Johnson AW, & Bentley DL (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes & development* 20(8):954-965.
71. Nedeia E, *et al.* (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Molecular cell* 29(5):577-587.
72. Vasiljeva L & Buratowski S (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Molecular cell* 21(2):239-248.

73. Schulz D, *et al.* (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155(5):1075-1087.
74. Wlotzka W, Kudla G, Granneman S, & Tollervey D (2011) The nuclear RNA polymerase II surveillance system targets polymerase III transcripts. *The EMBO journal* 30(9):1790-1803.
75. Lunde BM, Horner M, & Meinhart A (2011) Structural insights into cis element recognition of non-polyadenylated RNAs by the Nab3-RRM. *Nucleic acids research* 39(1):337-346.
76. Hobor F, *et al.* (2011) Recognition of transcription termination signal by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *The Journal of biological chemistry* 286(5):3645-3657.
77. Vasiljeva L, Kim M, Mutschler H, Buratowski S, & Meinhart A (2008) The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain. *Nature structural & molecular biology* 15(8):795-804.
78. Porrua O, *et al.* (2012) In vivo SELEX reveals novel sequence and structural determinants of Nrd1-Nab3-Sen1-dependent transcription termination. *The EMBO journal* 31(19):3935-3948.
79. Carroll KL, Ghirlando R, Ames JM, & Corden JL (2007) Interaction of yeast RNA-binding proteins Nrd1 and Nab3 with RNA polymerase II terminator elements. *RNA* 13(3):361-373.
80. Kubicek K, *et al.* (2012) Serine phosphorylation and proline isomerization in RNAP II CTD control recruitment of Nrd1. *Genes & development* 26(17):1891-1896.
81. Steinmetz EJ & Brow DA (1996) Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Molecular and cellular biology* 16(12):6993-7003.
82. Steinmetz EJ & Brow DA (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. *Proceedings of the National Academy of Sciences of the United States of America* 95(12):6699-6704.
83. Carroll KL, Pradhan DA, Granek JA, Clarke ND, & Corden JL (2004) Identification of cis elements directing termination of yeast nonpolyadenylated snoRNA transcripts. *Molecular and cellular biology* 24(14):6241-6252.

84. Bacikova V, Pasulka J, Kubicek K, & Stefl R (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic acids research*.
85. Nicholas K. Conrad SMW, †,1 Eric J. Steinmetz,‡ Meera Patturajan,* ,2 David A. Brow,‡ Maurice S. Swanson† and Jeffrey L. Corden* (2000) A Yeast Heterogeneous Nuclear Ribonucleoprotein Complex Associated With RNA Polymerase II.
86. Loya TJ, O'Rourke TW, & Reines D (2013) Yeast Nab3 protein contains a self-assembly domain found in human heterogeneous nuclear ribonucleoprotein-C (hnRNP-C) that is necessary for transcription termination. *The Journal of biological chemistry* 288(4):2111-2117.
87. Loya TJ, O'Rourke TW, Degtyareva N, & Reines D (2013) A network of interdependent molecular interactions describes a higher order Nrd1-Nab3 complex involved in yeast transcription termination. *The Journal of biological chemistry* 288(47):34158-34167.
88. Loya TJ, O'Rourke TW, & Reines D (2012) A genetic screen for terminator function in yeast identifies a role for a new functional domain in termination factor Nab3. *Nucleic acids research* 40(15):7476-7491.
89. Bacikova V, Pasulka J, Kubicek K, & Stefl R (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic acids research* 42(12):8024-8038.
90. Alberti S, Halfmann R, King O, Kapila A, & Lindquist S (2009) A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137(1):146-158.
91. Martin-Tumasz S & Brow DA (2015) *S. cerevisiae* Sen1 Helicase Domain Exhibits 5' to 3' Helicase Activity With a Preference for Translocation on DNA Rather Than RNA. *The Journal of biological chemistry*.
92. DeMarini DJ, Winey M, Ursic D, Webb F, & Culbertson MR (1992) SEN1, a positive effector of tRNA-splicing endonuclease in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 12(5):2154-2164.
93. Finkel JS, Chinchilla K, Ursic D, & Culbertson MR (2010) Sen1p performs two genetically separable functions in transcription and processing of U5 small nuclear RNA in *Saccharomyces cerevisiae*. *Genetics* 184(1):107-118.
94. Ursic D, Chinchilla K, Finkel JS, & Culbertson MR (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*

research 32(8):2441-2452.

95. Porrua O & Libri D (2013) A bacterial-like mechanism for transcription termination by the Sen1p helicase in budding yeast. *Nature structural & molecular biology* 20(7):884-891.
96. Hazelbaker DZ, Marquardt S, Wlotzka W, & Buratowski S (2013) Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Molecular cell* 49(1):55-66.
97. Wyers F, *et al.* (2005) Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121(5):725-737.
98. Thiebaut M, Kisseleva-Romanova E, Rougemaille M, Boulay J, & Libri D (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Molecular cell* 23(6):853-864.
99. Arigo JT, Eyler DE, Carroll KL, & Corden JL (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular cell* 23(6):841-851.
100. Gudipati RK, *et al.* (2012) Extensive degradation of RNA precursors by the exosome in wild-type cells. *Molecular cell* 48(3):409-421.
101. Neil H, *et al.* (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457(7232):1038-1042.
102. Xu Z, *et al.* (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457(7232):1033-1037.
103. van Dijk EL, *et al.* (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475(7354):114-117.
104. Wu DD, Irwin DM, & Zhang YP (2011) De novo origin of human protein-coding genes. *PLoS genetics* 7(11):e1002379.
105. Colin J, Libri D, & Porrua O (2011) Cryptic transcription and early termination in the control of gene expression. *Genet Res Int* 2011:653494.
106. Castelnuovo M & Stutz F (2015) Role of chromatin, environmental changes and single cell heterogeneity in non-coding transcription and gene regulation. *Current opinion in cell biology* 34:16-22.

107. Malabat C, Feuerbach F, Ma L, Saveanu C, & Jacquier A (2015) Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *eLife* 4.
108. Marquardt S, *et al.* (2014) A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* 157(7):1712-1723.
109. Kilchert C, Wittmann S, & Vasiljeva L (2016) The regulation and functions of the nuclear RNA exosome complex. *review Nat Rev Mol Cell Biol.*
110. Thoms M, *et al.* (2015) The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* 162(5):1029-1038.
111. Makino DL, *et al.* (2015) RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* 524(7563):54-58.
112. E.Conti (2010) Structural analysis reveals the characteristic features of Mtr4, a DExH helicase involved in nuclear RNA processing and surveillance. *PNAS*.
113. LaCava J, *et al.* (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* 121(5):713-724.
114. Jia H, Wang X, Anderson JT, & Jankowsky E (2012) RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proceedings of the National Academy of Sciences of the United States of America* 109(19):7292-7297.
115. Vanacova S, *et al.* (2005) A new yeast poly(A) polymerase complex involved in RNA quality control. *PLoS biology* 3(6):e189.
116. Tudek A, *et al.* (2014) Molecular Basis for Coordinating Transcription Termination with Noncoding RNA Degradation. *Molecular cell* 55(3):467-481.
117. Tudek A, *et al.* (2014) Molecular basis for coordinating transcription termination with noncoding RNA degradation. *Molecular cell* 55(3):467-481.
118. Costello JL, Stead JA, Feigenbutz M, Jones RM, & Mitchell P (2011) The C-terminal region of the exosome-associated protein Rrp47 is specifically required for box C/D small nucleolar RNA 3'-maturation. *The Journal of biological chemistry* 286(6):4535-4543.
119. Butler JS & Mitchell P (2011) Rrp6, rrp47 and cofactors of the nuclear exosome. *Advances in experimental medicine and biology* 702:91-104.
120. Schuch B, *et al.* (2014) The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *The EMBO journal*

33(23):2829-2846.

121. Milligan L, *et al.* (2008) A yeast exosome cofactor, Mpp6, functions in RNA surveillance and in the degradation of noncoding RNA transcripts. *Molecular and cellular biology* 28(17):5446-5457.
122. Kim K, Heo DH, Kim I, Suh JY, & Kim M (2016) Exosome Cofactors Connect Transcription Termination to RNA Processing by Guiding Terminated Transcripts to the Appropriate Exonuclease within the Nuclear Exosome. *The Journal of biological chemistry*.
123. Fasken MB, Laribee RN, & Corbett AH (2015) Nab3 facilitates the function of the TRAMP complex in RNA processing via recruitment of Rrp6 independent of Nrd1. *PLoS genetics* 11(3):e1005044.
124. Cramer P, Bushnell DA, & Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292(5523):1863-1876.
125. Bushnell DA & Kornberg RD (2003) Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proceedings of the National Academy of Sciences of the United States of America* 100(12):6969-6973.
126. Armache KJ, Kettenberger H, & Cramer P (2003) Architecture of initiation-competent 12-subunit RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* 100(12):6964-6968.
127. Cramer P (2004) Structure and function of RNA polymerase II. *Advances in protein chemistry* 67:1-42.
128. Nudler E (2009) RNA polymerase active center: the molecular engine of transcription. *Annual review of biochemistry* 78:335-361.
129. Gnatt AL, Cramer P, Fu J, Bushnell DA, & Kornberg RD (2001) Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* 292(5523):1876-1882.
130. Bernecky C, Herzog F, Baumeister W, Plitzko JM, & Cramer P (2016) Structure of transcribing mammalian RNA polymerase II. *Nature* 529(7587):551-554.
131. Kireeva ML, Komissarova N, Waugh DS, & Kashlev M (2000) The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *The Journal of biological chemistry* 275(9):6530-6536.

132. Barnes CO, *et al.* (2015) Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble. *Molecular cell* 59(2):258-269.
133. Naji S, Bertero MG, Spitalny P, Cramer P, & Thomm M (2008) Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic acids research* 36(2):676-687.
134. Westover KD, Bushnell DA, & Kornberg RD (2004) Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303(5660):1014-1016.
135. Kireeva ML, *et al.* (2008) Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Molecular cell* 30(5):557-566.
136. Kaplan CD, Larsson KM, & Kornberg RD (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Molecular cell* 30(5):547-556.
137. Jensen GJ, Meredith G, Bushnell DA, & Kornberg RD (1998) Structure of wild-type yeast RNA polymerase II and location of Rpb4 and Rpb7. *The EMBO journal* 17(8):2353-2358.
138. Svetlov V, Vassylyev DG, & Artsimovitch I (2004) Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *The Journal of biological chemistry* 279(37):38087-38090.
139. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, & Block SM (2005) Direct observation of base-pair stepping by RNA polymerase. *Nature* 438(7067):460-465.
140. Feig M & Burton ZF (2010) RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation. *Biophysical journal* 99(8):2577-2586.
141. Imashimizu M, Shimamoto N, Oshima T, & Kashlev M (2014) Transcription elongation: Heterogeneous tracking of RNA polymerase and its biological implications. *Transcription* 5(1):e28285.
142. Malinen AM, *et al.* (2012) Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic acids research* 40(15):7442-7451.
143. Bar-Nahum G, *et al.* (2005) A ratchet mechanism of transcription elongation and its control. *Cell* 120(2):183-193.
144. Brueckner F & Cramer P (2008) Structural basis of transcription inhibition by

- alpha-amanitin and implications for RNA polymerase II translocation. *Nature structural & molecular biology* 15(8):811-818.
145. Landick R (2006) The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society transactions* 34(Pt 6):1062-1066.
 146. Touloukhonov I, Zhang J, Palangat M, & Landick R (2007) A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Molecular cell* 27(3):406-419.
 147. Wang D, *et al.* (2009) Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science* 324(5931):1203-1206.
 148. Zenkin N, Yuzenkova Y, & Severinov K (2006) Transcript-assisted transcriptional proofreading. *Science* 313(5786):518-520.
 149. Sydow JF, *et al.* (2009) Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Molecular cell* 34(6):710-721.
 150. Larson MH, *et al.* (2014) A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* 344(6187):1042-1047.
 151. Sigurdsson S, Dirac-Svejstrup AB, & Svejstrup JQ (2010) Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Molecular cell* 38(2):202-210.
 152. Palangat M & Landick R (2001) Roles of RNA:DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II. *Journal of molecular biology* 311(2):265-282.
 153. Artsimovitch I & Landick R (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):7090-7095.
 154. Kireeva ML, *et al.* (2005) Nature of the nucleosomal barrier to RNA polymerase II. *Molecular cell* 18(1):97-108.
 155. Nudler E, Mustaev A, Lukhtanov E, & Goldfarb A (1997) The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell* 89(1):33-41.
 156. Komissarova N & Kashlev M (1997) Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded.

Proceedings of the National Academy of Sciences of the United States of America 94(5):1755-1760.

157. Kireeva M, Kashlev M, & Burton ZF (2010) Translocation by multi-subunit RNA polymerases. *Biochimica et biophysica acta* 1799(5-6):389-401.
158. Fairman-Williams ME, Guenther UP, & Jankowsky E (2010) SF1 and SF2 helicases: family matters. *Current opinion in structural biology* 20(3):313-324.
159. Kevin D. Raney AKB, and Suja Aarattuthodiyil (2013) Structure and Mechanisms of SF1 DNA Helicases. *Advances in experimental medicine and biology*.
160. Sharma S, Doherty KM, & Brosh RM, Jr. (2006) Mechanisms of RecQ helicases in pathways of DNA metabolism and maintenance of genomic stability. *The Biochemical journal* 398(3):319-337.
161. Shen JC, *et al.* (1998) Werner syndrome protein. I. DNA helicase and dna exonuclease reside on the same polypeptide. *The Journal of biological chemistry* 273(51):34139-34144.
162. Chen YZ, *et al.* (2004) DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *American journal of human genetics* 74(6):1128-1135.
163. Moreira MC, *et al.* (2004) Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature genetics* 36(3):225-227.
164. Groh M, Albulescu LO, Cristini A, & Gromak N (2016) Senataxin: Genome Guardian at the Interface of Transcription and Neurodegeneration. *Journal of molecular biology*.
165. Bennett CL & La Spada AR (2015) Unwinding the role of senataxin in neurodegeneration. *Discovery medicine* 19(103):127-136.
166. Grohmann K, *et al.* (2001) Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. *Nature genetics* 29(1):75-77.
167. Singleton MR, Dillingham MS, & Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annual review of biochemistry* 76:23-50.
168. Berger JM (2008) SnapShot: nucleic acid helicases and translocases. *Cell* 134(5):888-888 e881.

169. Medagli B & Onesti S (2013) Structure and mechanism of hexameric helicases. *Advances in experimental medicine and biology* 767:75-95.
170. O'Shea VL & Berger JM (2014) Loading strategies of ring-shaped nucleic acid translocases and helicases. *Current opinion in structural biology* 25:16-24.
171. Leipe DD, Wolf YI, Koonin EV, & Aravind L (2002) Classification and evolution of P-loop GTPases and related ATPases. *Journal of molecular biology* 317(1):41-72.
172. Seidel R, Bloom JG, Dekker C, & Szczelkun MD (2008) Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *The EMBO journal* 27(9):1388-1398.
173. Bizebard T, Ferlenghi I, Iost I, & Dreyfus M (2004) Studies on three E. coli DEAD-box helicases point to an unwinding mechanism different from that of model DNA helicases. *Biochemistry* 43(24):7857-7866.
174. Gilhooly NS, Gwynn EJ, & Dillingham MS (2013) Superfamily 1 helicases. *Front Biosci (Schol Ed)* 5:206-216.
175. Fiorini F, Boudvillain M, & Le Hir H (2013) Tight intramolecular regulation of the human Upf1 helicase by its N- and C-terminal domains. *Nucleic acids research* 41(4):2404-2415.
176. Tanner NK (2003) The newly identified Q motif of DEAD box helicases is involved in adenine recognition. *Cell Cycle* 2(1):18-19.
177. Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, & Wigley DB (1999) Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* 97(1):75-84.
178. Wigley DB (2009) Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell* 137(5):849-859.
179. Soultanas P, Dillingham MS, Velankar SS, & Wigley DB (1999) DNA binding mediates conformational changes and metal ion coordination in the active site of PcrA helicase. *Journal of molecular biology* 290(1):137-148.
180. Saikrishnan K, Powell B, Cook NJ, Webb MR, & Wigley DB (2009) Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell* 137(5):849-859.
181. Chakrabarti S, *et al.* (2011) Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Molecular cell* 41(6):693-703.

182. Brendza KM, *et al.* (2005) Autoinhibition of Escherichia coli Rep monomer helicase activity by its 2B subdomain. *Proceedings of the National Academy of Sciences of the United States of America* 102(29):10076-10081.
183. Lohman TM, Tomko EJ, & Wu CG (2008) Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nature reviews. Molecular cell biology* 9(5):391-401.
184. Matson SW & Robertson AB (2006) The UvrD helicase and its modulation by the mismatch repair protein MutL. *Nucleic acids research* 34(15):4089-4097.
185. Manelyte L, *et al.* (2009) The unstructured C-terminal extension of UvrD interacts with UvrB, but is dispensable for nucleotide excision repair. *DNA repair* 8(11):1300-1310.
186. Boudvillain M, Figueroa-Bossi N, & Bossi L (2013) Terminator still moving forward: expanding roles for Rho factor. *Curr Opin Microbiol* 16(2):118-124.
187. Ray-Soni A, Bellecourt MJ, & Landick R (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annual review of biochemistry*.
188. Kriner MA, Sevostyanova A, & Groisman EA (2016) Learning from the Leaders: Gene Regulation by the Transcription Termination Factor Rho. *Trends Biochem Sci* 41(8):690-699.
189. Grylak-Mielnicka A, Bidnenko V, Bardowski J, & Bidnenko E (2016) Transcription termination factor Rho: a hub linking diverse physiological processes in bacteria. *Microbiology* 162(3):433-447.
190. D'Heygere F, Rabhi M, & Boudvillain M (2013) Phyletic distribution and conservation of the bacterial transcription termination factor Rho. *Microbiology* 159(Pt 7):1423-1436.
191. Peters JM, *et al.* (2009) Rho directs widespread termination of intragenic and stable RNA transcription. *Proceedings of the National Academy of Sciences of the United States of America* 106(36):15406-15411.
192. Peters JM, *et al.* (2012) Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes & development* 26(23):2621-2633.
193. de Hoon M, Makita Y, Nakai K, & Miyano S (2005) Prediction of Transcriptional Terminators in Bacillus subtilis and Related Species. *PLoS Computational Biology* preprint(2005):e25.

194. Matsumoto Y, Shigesada K, Hirano M, & Imai M (1986) Autogenous regulation of the gene for transcription termination factor rho in Escherichia coli: localization and function of its attenuators. *Journal of bacteriology* 166(3):945-958.
195. Nicolas P, *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science* 335(6072):1103-1106.
196. Proshkin S, Rahmouni AR, Mironov A, & Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328(5977):504-508.
197. Leela JK, Syeda AH, Anupama K, & Gowrishankar J (2013) Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 110(1):258-263.
198. Gowrishankar J, Leela JK, & Anupama K (2013) R-loops in bacterial transcription: their causes and consequences. *Transcription* 4(4):153-157.
199. Hamperl S & Cimprich KA (2014) The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA repair* 19:84-94.
200. Bogden CE, Fass D, Bergman N, Nichols MD, & Berger JM (1999) The structural basis for terminator recognition by the Rho transcription termination factor. *Molecular cell* 3(4):487-493.
201. Bear DG, *et al.* (1988) Interactions of Escherichia coli transcription termination factor rho with RNA. II. Electron microscopy and nuclease protection experiments. *Journal of molecular biology* 199(4):623-635.
202. McSwiggen JA, Bear DG, & von Hippel PH (1988) Interactions of Escherichia coli transcription termination factor rho with RNA. I. Binding stoichiometries and free energies. *Journal of molecular biology* 199(4):609-622.
203. Morgan WD, Bear DG, & von Hippel PH (1983) Rho-dependent termination of transcription. II. Kinetics of mRNA elongation during transcription from the bacteriophage lambda PR promoter. *The Journal of biological chemistry* 258(15):9565-9574.
204. Alifano P, Rivellini F, Limauro D, Bruni CB, & Carlomagno MS (1991) A consensus motif common to all Rho-dependent prokaryotic transcription terminators. *Cell* 64(3):553-563.
205. Chen CY, Galluppi GR, & Richardson JP (1986) Transcription termination at lambda

- tR1 is mediated by interaction of rho with specific single-stranded domains near the 3' end of cro mRNA. *Cell* 46(7):1023-1028.
206. Richardson JP (1982) Activation of rho protein ATPase requires simultaneous interaction at two kinds of nucleic acid-binding sites. *The Journal of biological chemistry* 257(10):5760-5766.
 207. Canals A, Uson I, & Coll M (2010) The structure of RNA-free Rho termination factor indicates a dynamic mechanism of transcript capture. *Journal of molecular biology* 400(1):16-23.
 208. Skordalakes E & Berger JM (2003) Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. *Cell* 114(1):135-146.
 209. Thomsen ND & Berger JM (2009) Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* 139(3):523-534.
 210. Lowery-Goldhammer C & Richardson JP (1974) An RNA-dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. *Proceedings of the National Academy of Sciences of the United States of America* 71(5):2003-2007.
 211. Skordalakes E & Berger JM (2006) Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. *Cell* 127(3):553-564.
 212. Kim DE & Patel SS (2001) The kinetic pathway of RNA binding to the Escherichia coli transcription termination factor Rho. *The Journal of biological chemistry* 276(17):13902-13910.
 213. Thomsen ND, Lawson MR, Witkowsky LB, Qu S, & Berger JM (2016) Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor. *Proceedings of the National Academy of Sciences of the United States of America* 113(48):E7691-E7700.
 214. Brennan CA, Dombroski AJ, & Platt T (1987) Transcription termination factor rho is an RNA-DNA helicase. *Cell* 48(6):945-952.
 215. Gogol EP, Seifried SE, & von Hippel PH (1991) Structure and assembly of the Escherichia coli transcription termination factor rho and its interaction with RNA. I. Cryoelectron microscopic studies. *Journal of molecular biology* 221(4):1127-1138.
 216. Koslover DJ, Fazal FM, Mooney RA, Landick R, & Block SM (2012) Binding and translocation of termination factor rho studied at the single-molecule level. *Journal of molecular biology* 423(5):664-676.

217. Washburn RS & Gottesman ME (2015) Regulation of transcription elongation and termination. *Biomolecules* 5(2):1063-1078.
218. Steinmetz EJ & Platt T (1994) Evidence supporting a tethered tracking model for helicase activity of Escherichia coli Rho factor. *Proceedings of the National Academy of Sciences of the United States of America* 91(4):1401-1405.
219. Gocheva V, Le Gall A, Boudvillain M, Margeat E, & Nollmann M (2015) Direct observation of the translocation mechanism of transcription termination factor Rho. *Nucleic acids research* 43(4):2367-2377.
220. Peters JM, Vangeloff AD, & Landick R (2011) Bacterial transcription terminators: the RNA 3'-end chronicles. *Journal of molecular biology* 412(5):793-813.
221. Walmacq C, Rahmouni AR, & Boudvillain M (2004) Influence of substrate composition on the helicase activity of transcription termination factor Rho: reduced processivity of Rho hexamers during unwinding of RNA-DNA hybrid regions. *Journal of molecular biology* 342(2):403-420.
222. Walmacq C, Rahmouni AR, & Boudvillain M (2006) Testing the steric exclusion model for hexameric helicases: substrate features that alter RNA-DNA unwinding by the transcription termination factor Rho. *Biochemistry* 45(18):5885-5895.
223. Kassavetis GA & Chamberlin MJ (1981) Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. *The Journal of biological chemistry* 256(6):2777-2786.
224. Lau LF & Roberts JW (1985) Rho-dependent transcription termination at lambda R1 requires upstream sequences. *The Journal of biological chemistry* 260(1):574-584.
225. Stewart V, Landick R, & Yanofsky C (1986) Rho-dependent transcription termination in the tryptophanase operon leader region of Escherichia coli K-12. *Journal of bacteriology* 166(1):217-223.
226. Galloway JL & Platt T (1988) Signals sufficient for rho-dependent transcription termination at trp t' span a region centered 60 base pairs upstream of the earliest 3' end point. *The Journal of biological chemistry* 263(4):1761-1767.
227. Dutta D, Chalissery J, & Sen R (2008) Transcription termination factor rho prefers catalytically active elongation complexes for releasing RNA. *The Journal of biological chemistry* 283(29):20243-20251.
228. Richardson JP (2002) Rho-dependent termination and ATPases in transcript termination. *Biochimica et biophysica acta* 1577(2):251-260.

229. Walstrom KM, Dozono JM, & von Hippel PH (1997) Kinetics of the RNA-DNA helicase activity of Escherichia coli transcription termination factor rho. 2. Processivity, ATP consumption, and RNA binding. *Biochemistry* 36(26):7993-8004.
230. Park JS & Roberts JW (2006) Role of DNA bubble rewinding in enzymatic transcription termination. *Proceedings of the National Academy of Sciences of the United States of America* 103(13):4870-4875.
231. Epshtein V, Dutta D, Wade J, & Nudler E (2010) An allosteric mechanism of Rho-dependent transcription termination. *Nature* 463(7278):245-249.
232. Vassylyev DG, *et al.* (2005) Structural basis for transcription inhibition by tagetitoxin. *Nature structural & molecular biology* 12(12):1086-1093.
233. Artsimovitch I, *et al.* (2011) Tagetitoxin inhibits RNA polymerase through trapping of the trigger loop. *The Journal of biological chemistry* 286(46):40395-40400.
234. Lang WH, Platt T, & Reeder RH (1998) Escherichia coli rho factor induces release of yeast RNA polymerase II but not polymerase I or III. *Proceedings of the National Academy of Sciences of the United States of America* 95(9):4900-4905.
235. Wu SY & Platt T (1993) Transcriptional arrest of yeast RNA polymerase II by Escherichia coli rho protein in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 90(14):6606-6610.
236. Mooney RA, *et al.* (2009) Regulator trafficking on bacterial transcription units in vivo. *Molecular cell* 33(1):97-108.
237. Kalyani BS, Muteeb G, Qayyum MZ, & Sen R (2011) Interaction with the nascent RNA is a prerequisite for the recruitment of Rho to the transcription elongation complex in vitro. *Journal of molecular biology* 413(3):548-560.
238. Chambers AL, Smith AJ, & Savery NJ (2003) A DNA translocation motif in the bacterial transcription--repair coupling factor, Mfd. *Nucleic acids research* 31(22):6409-6418.
239. Park JS, Marr MT, & Roberts JW (2002) E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. *Cell* 109(6):757-767.
240. Selby CP (2017) Mfd Protein and Transcription-Repair Coupling in Escherichia coli. *Photochemistry and photobiology* 93(1):280-295.
241. Roberts J & Park JS (2004) Mfd, the bacterial transcription repair coupling factor:

- translocation, repair and termination. *Curr Opin Microbiol* 7(2):120-125.
242. Bohr VA, Smith CA, Okumoto DS, & Hanawalt PC (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* 40(2):359-369.
 243. Mellon I, Spivak G, & Hanawalt PC (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 51(2):241-249.
 244. Fan J, Leroux-Coyau M, Savery NJ, & Strick TR (2016) Reconstruction of bacterial transcription-coupled repair at single-molecule resolution. *Nature* 536(7615):234-237.
 245. Spivak G (2016) Transcription-coupled repair: an update. *Archives of toxicology* 90(11):2583-2594.
 246. Deaconescu AM, *et al.* (2006) Structural basis for bacterial transcription-coupled DNA repair. *Cell* 124(3):507-520.
 247. Selby CP & Sancar A (1995) Structure and function of transcription-repair coupling factor. II. Catalytic properties. *The Journal of biological chemistry* 270(9):4890-4895.
 248. Selby CP & Sancar A (1995) Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. *The Journal of biological chemistry* 270(9):4882-4889.
 249. Westblade LF, *et al.* (2010) Structural basis for the bacterial transcription-repair coupling factor/RNA polymerase interaction. *Nucleic acids research* 38(22):8357-8369.
 250. Smith AJ & Savery NJ (2005) RNA polymerase mutants defective in the initiation of transcription-coupled DNA repair. *Nucleic acids research* 33(2):755-764.
 251. Pakotiprapha D, Samuels M, Shen K, Hu JH, & Jeruzalmi D (2012) Structure and mechanism of the UvrA-UvrB DNA damage sensor. *Nature structural & molecular biology* 19(3):291-298.
 252. Deaconescu AM, Sevostyanova A, Artsimovitch I, & Grigorieff N (2012) Nucleotide excision repair (NER) machinery recruitment by the transcription-repair coupling factor involves unmasking of a conserved intramolecular interface. *Proceedings of the National Academy of Sciences of the United States of America* 109(9):3353-3358.
 253. Selby CP & Sancar A (1993) Molecular mechanism of transcription-repair coupling.

Science 260(5104):53-58.

254. Manelyte L, Kim YI, Smith AJ, Smith RM, & Savery NJ (2010) Regulation and rate enhancement during transcription-coupled DNA repair. *Molecular cell* 40(5):714-724.
255. Smith AJ, Pernstich C, & Savery NJ (2012) Multipartite control of the DNA translocase, Mfd. *Nucleic acids research* 40(20):10408-10416.
256. Smith AJ, Szczelkun MD, & Savery NJ (2007) Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. *Nucleic acids research* 35(6):1802-1811.
257. Murphy MN, *et al.* (2009) An N-terminal clamp restrains the motor domains of the bacterial transcription-repair coupling factor Mfd. *Nucleic acids research* 37(18):6042-6053.
258. Kad NM, Wang H, Kennedy GG, Warshaw DM, & Van Houten B (2010) Collaborative dynamic DNA scanning by nucleotide excision repair proteins investigated by single- molecule imaging of quantum-dot-labeled proteins. *Molecular cell* 37(5):702-713.
259. Damsma GE, Alt A, Brueckner F, Carell T, & Cramer P (2007) Mechanism of transcriptional stalling at cisplatin-damaged DNA. *Nature structural & molecular biology* 14(12):1127-1133.
260. Brueckner F, Hennecke U, Carell T, & Cramer P (2007) CPD damage recognition by transcribing RNA polymerase II. *Science* 315(5813):859-862.
261. Howan K, *et al.* (2012) Initiation of transcription-coupled repair characterized at single-molecule resolution. *Nature* 490(7420):431-434.
262. Graves ET, *et al.* (2015) A dynamic DNA-repair complex observed by correlative single-molecule nanomanipulation and fluorescence. *Nature structural & molecular biology* 22(6):452-457.
263. Mellon I & Hanawalt PC (1989) Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* 342(6245):95-98.
264. Arndt KM & Reines D (2015) Termination of Transcription of Short Noncoding RNAs by RNA Polymerase II. *Annual review of biochemistry* 84:381-404.
265. Steinmetz EJ, *et al.* (2006) Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular cell* 24(5):735-746.

266. Steinmetz EJ, Conrad NK, Brow DA, & Corden JL (2001) RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413(6853):327-331.
267. Chinchilla K, *et al.* (2012) Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in *Saccharomyces cerevisiae*. *Eukaryot Cell* 11(4):417-429.
268. Chen X, Muller U, Sundling KE, & Brow DA (2014) *Saccharomyces cerevisiae* Sen1 as a Model for the Study of Mutations in Human Senataxin That Elicit Cerebellar Ataxia. *Genetics* 198(2):577-590.
269. Cheng Z, Muhlrad D, Lim MK, Parker R, & Song H (2007) Structural and functional insights into the human Upf1 helicase core. *The EMBO journal* 26(1):253-264.
270. Steinmetz EJ, Ng SB, Cloute JP, & Brow DA (2006) cis- and trans-Acting determinants of transcription termination by yeast RNA polymerase II. *Molecular and cellular biology* 26(7):2688-2696.
271. Gudipati RK, Villa T, Boulay J, & Libri D (2008) Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nature structural & molecular biology* 15(8):786-794.
272. Loya TJ & Reines D (2016) Recent advances in understanding transcription termination by RNA polymerase II. *F1000Research* 5.
273. Rondon AG, Mischo HE, Kawauchi J, & Proudfoot NJ (2009) Fail-safe transcriptional termination for protein-coding genes in *S. cerevisiae*. *Molecular cell* 36(1):88-98.
274. Kawauchi J, Mischo H, Braglia P, Rondon A, & Proudfoot NJ (2008) Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes & development* 22(8):1082-1092.
275. Schaugency P, Merran J, & Corden JL (2014) Genome-wide mapping of yeast RNA polymerase II termination. *PLoS genetics* 10(10):e1004632.
276. Lepore N & Lafontaine DL (2011) A functional interface at the rDNA connects rRNA synthesis, pre-rRNA processing and nucleolar surveillance in budding yeast. *PloS one* 6(9):e24962.
277. Mischo HE, *et al.* (2011) Yeast Sen1 helicase protects the genome from transcription-associated instability. *Molecular cell* 41(1):21-32.

278. Alzu A, *et al.* (2012) Senataxin Associates with Replication Forks to Protect Fork Integrity across RNA-Polymerase-II-Transcribed Genes. *Cell* 151(4):835-846.
279. Bermejo R, Lai MS, & Foiani M (2012) Preventing replication stress to maintain genome stability: resolving conflicts between replication and transcription. *Molecular cell* 45(6):710-718.
280. Li W, Selvam K, Rahman SA, & Li S (2016) Sen1, the yeast homolog of human senataxin, plays a more direct role than Rad26 in transcription coupled DNA repair. *Nucleic acids research*:gkw428.
281. Winey M & Culbertson MR (1988) Mutations affecting the tRNA-splicing endonuclease activity of *Saccharomyces cerevisiae*. *Genetics* 118(4):609-617.
282. Suraweera A, *et al.* (2007) Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. *The Journal of cell biology* 177(6):969-979.
283. Chen Y-Z, *et al.* (2006) Senataxin, the yeast Sen1p orthologue: Characterization of a unique protein in which recessive mutations cause ataxia and dominant mutations cause motor neuron disease. *Neurobiology of Disease* 23(1):97-108.
284. James PA & Talbot K (2006) The molecular genetics of non-ALS motor neuron diseases. *Biochimica et biophysica acta* 1762(11-12):986-1000.
285. Skourti-Stathaki K, Proudfoot NJ, & Gromak N (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Molecular cell* 42(6):794-805.
286. Padmanabhan K, Robles MS, Westerling T, & Weitz CJ (2012) Feedback regulation of transcriptional termination by the mammalian circadian clock PERIOD complex. *Science* 337(6094):599-602.
287. Zhao DY, *et al.* (2016) SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* 529(7584):48-53.
288. Hatchi E, *et al.* (2015) BRCA1 recruitment to transcriptional pause sites is required for R-loop-driven DNA damage repair. *Molecular cell* 57(4):636-647.
289. Sollier J, *et al.* (2014) Transcription-coupled nucleotide excision repair factors promote R-loop-induced genome instability. *Molecular cell* 56(6):777-785.
290. Richard P, Feng S, & Manley JL (2013) A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2,

- targets the exosome to sites of transcription-induced DNA damage. *Genes & development* 27(20):2227-2232.
291. Wagschal A, *et al.* (2012) Microprocessor, Setx, Xrn2, and Rrp6 co-operate to induce premature termination of transcription by RNAPII. *Cell* 150(6):1147-1157.
 292. Miller MS, *et al.* (2015) Senataxin suppresses the antiviral transcriptional response and controls viral biogenesis. *Nature immunology* 16(5):485-494.
 293. Kireeva ML & Kashlev M (2009) Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* 106(22):8900-8905.
 294. Yuzenkova Y, Roghanian M, Bochkareva A, & Zenkin N (2013) Tagetitoxin inhibits transcription by stabilizing pre-translocated state of the elongation complex. *Nucleic acids research* 41(20):9257-9265.
 295. Jankowsky E (2011) RNA helicases at work: binding and rearranging. *Trends Biochem Sci* 36(1):19-29.
 296. Santangelo TJ & Artsimovitch I (2011) Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol* 9(5):319-329.
 297. Belogurov GA, Mooney RA, Svetlov V, Landick R, & Artsimovitch I (2009) Functional specialization of transcription elongation factors. *The EMBO journal* 28(2):112-122.
 298. Muteeb G, Dey D, Mishra S, & Sen R (2012) A multipronged strategy of an anti-terminator protein to overcome Rho-dependent transcription termination. *Nucleic acids research* 40(22):11213-11228.
 299. Gutierrez P, *et al.* (2007) Solution structure of YaeO, a Rho-specific inhibitor of transcription termination. *The Journal of biological chemistry* 282(32):23348-23353.
 300. Becherel OJ, *et al.* (2015) A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Hum Mol Genet* 24(20):5759-5774.

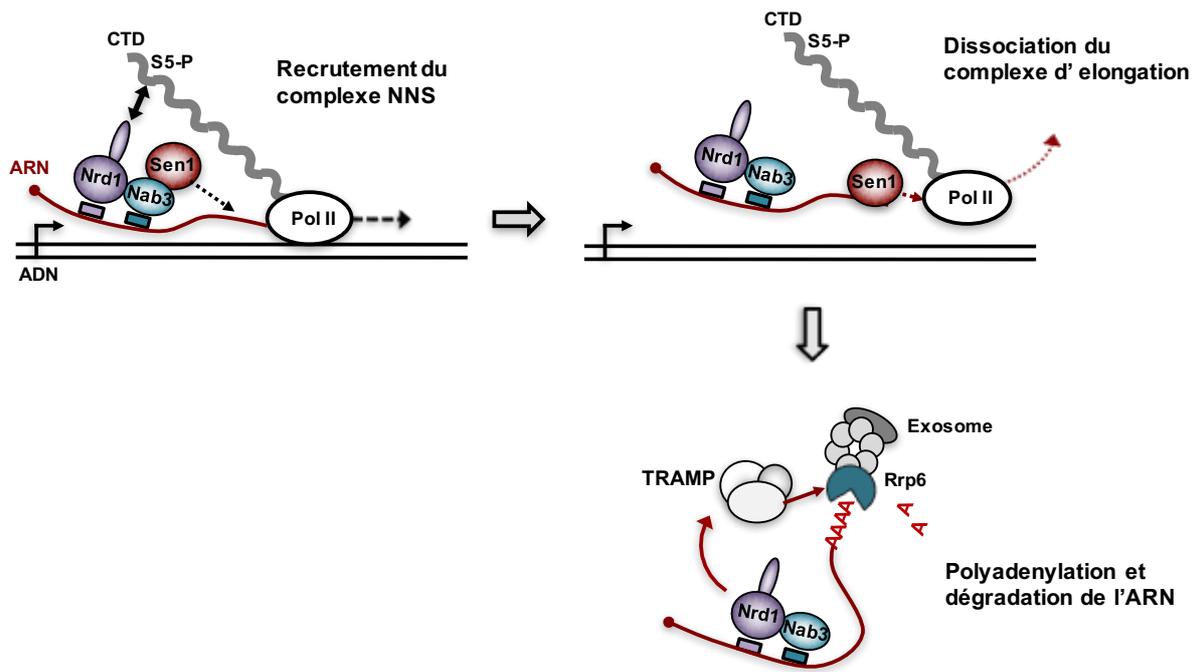
RESUMÉ EN FRANÇAIS

Mes travaux de thèse ont eu pour but de caractériser les mécanismes d'action de l'hélicase Sen1, principalement dans la terminaison de la transcription non-codante. Mes résultats ont généré deux publications en premier auteur qui sont résumées ici :

1) Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. (*Han et al, NAR 2017*)

La transcription cachée est un phénomène répandu aussi bien chez les eucaryotes que chez les procaryotes. Elle se caractérise par une production massive d'ARNs non-codants au niveau de régions non-annotées du génome. Ce phénomène est potentiellement dangereux pour la cellule car il peut interférer avec l'expression normale des gènes. Chez *Saccharomyces cerevisiae*, un des acteurs majeurs dans le contrôle de la transcription cachée est le complexe Nrd1-Nab3-Sen1 (NNS) qui induit la terminaison précoce de la transcription non-codante et favorise la dégradation des ARNs générés par l'exosome nucléaire. Le complexe NNS se compose de trois protéines essentielles: les facteurs de liaison à l'ARN Nrd1 et Nab3, qui reconnaissent des motifs spécifiques dans les ARNs cibles, et l'hélicase Sen1, qui dissocie le complexe d'élongation de façon dépendante de l'hydrolyse de l'ATP.

En effet, des études réalisées à l'échelle du génome ont montré que la déplétion des composants du complexe NNS conduit à la dérégulation de centaines de gènes, ce qui met en évidence l'importance de la terminaison précoce de la transcription non-codante pour maintenir l'expression correcte des gènes.

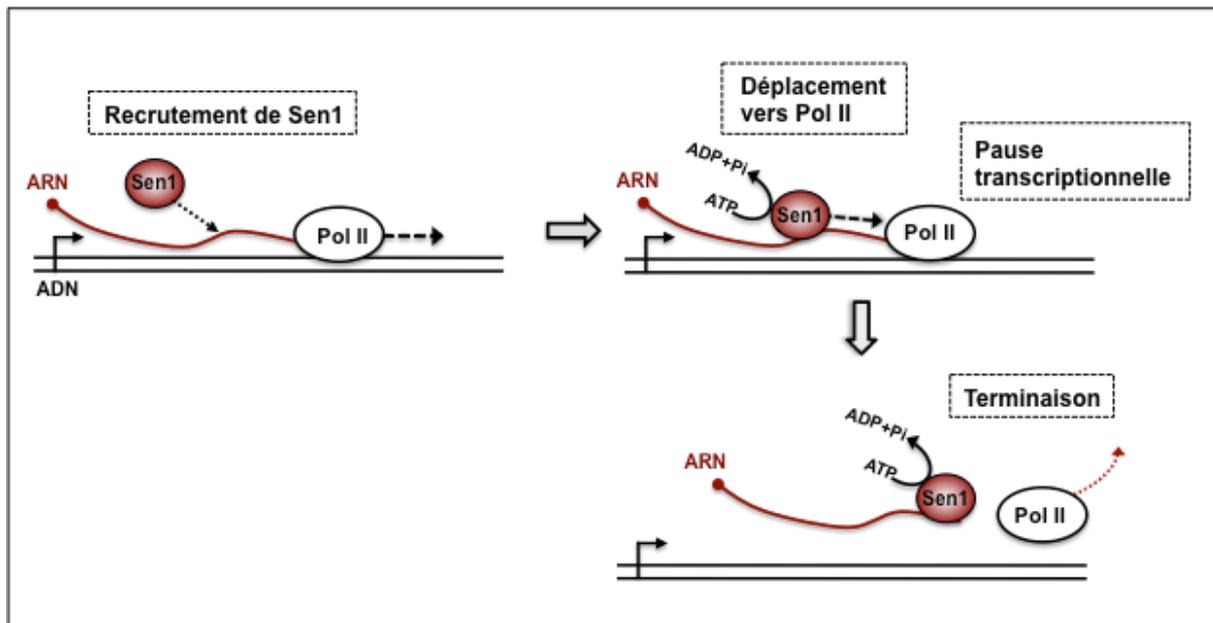


Résumé des étapes principales de la voie de terminaison de la transcription non-codante dépendante du complexe NNS. Le complexe NNS est recruté aux cibles de terminaison d'une part par la reconnaissance de séquences spécifiques qui sont enrichies dans les ARN non-codants par Nrd1 et Nab3 et d'autre part par l'interaction de Nrd1 avec le domaine C-terminal (CTD) de l'ARN polymérase II. Notamment, Nrd1 s'associe au CTD qui est phosphorylé dans la serine 5 (S5-P). Ensuite, Sen1 est relâché sur l'ARN et utilise l'hydrolyse de l'ATP pour induire la dissociation du complexe d'elongation. Finalement, Nrd1 et Nab3 stimulent la polyadenylation et la dégradation des ARNs par l'exosome nucléaire, qui contient l'exonuclease Rrp6, et son cofacteur, le complexe TRAMP.

Les hélicases sont des enzymes qui utilisent l'énergie de l'hydrolyse de l'ATP pour modifier des acides nucléiques ou des complexes protéine-acide nucléiques. Sen1 est très conservée et des mutations dans son homologue humain, senataxin, ont été associées à des maladies neurodégénératives. Malgré de nombreuses recherches menées sur ces protéines, leurs propriétés biochimiques ainsi que leurs mécanismes d'action restent peu connus. Dans cette étude, nous avons caractérisé biochimiquement les activités de Sen1 et nous avons étudié les mécanismes par lesquels elle induit la terminaison de la transcription.

Dans ce but, nous avons utilisé un ensemble de techniques *in vitro*, notamment un système de transcription-terminaison minimale qui contient uniquement des composants purifiés: Sen1, l'ARN polymérase II et les ADN matrices. Ce système nous permet de modifier les différents éléments de façon contrôlée afin de comprendre leur rôle précis dans la terminaison de la transcription. Nous avons tout d'abord analysé la fonction des différents domaines de Sen1 dans la terminaison. Sen1 est une protéine de taille importante (252 kDa) qui possède un domaine central catalytique (acides aminés 1095-1880) flanqué par deux domaines qui jouent un rôle dans l'interaction avec d'autres facteurs, notamment les partenaires Nrd1 et Nab3. Nous avons montré que le domaine central hélicase est suffisant pour déclencher la terminaison de la transcription *in vitro*, ce qui suggère que les autres domaines sont importants pour d'autres processus (régulation de l'activité, localisation dans le noyau, etc). Ensuite, nous avons montré que Sen1 utilise l'énergie de l'hydrolyse de l'ATP pour se déplacer sur des acides nucléiques simple bras, aussi bien sur l'ARN que sur l'ADN, dans le sens 5' vers 3'. Nous avons alors étudié le rôle des différents acides nucléiques du système (ARN naissant et matrice d'ADN) dans la terminaison par Sen1. Nos résultats indiquent que l'interaction de Sen1 avec l'ADN n'est pas nécessaire pour la terminaison ; en revanche Sen1 doit s'associer à l'ARN naissant et se déplacer vers la polymérase. Nous avons également montré qu'une fois que Sen1 entre en collision avec l'ARN polymérase II, elle y exerce une

action mécanique qui conduit à la terminaison uniquement quand la polymérase marque une pause. Cela indique que la terminaison est fortement dépendante de la pause transcriptionnelle. En conclusion, nos résultats constituent une avancée dans la compréhension des mécanismes de terminaison de la transcription par Sen1 et apportent des nouvelles pistes sur l'origine des maladies causées par des mutations dans senataxin.



2) Sen1 has unique structural features grafted on the architecture of the Upf1-like helicase family. (Leonaité*, Han* et al, EMBO J 2017)

*Co-premier auteurs.

Les hélicases sont des enzymes qui utilisent l'énergie de l'hydrolyse de l'ATP pour modifier des acides nucléiques ou des complexes protéine-acide nucléiques. Les hélicases existent dans tous les organismes et sont impliquées dans pratiquement tous les processus liés au métabolisme des acides nucléiques. Parmi les multiples fonctions possibles des hélicases, Sen1 joue un rôle majeur dans la terminaison de la transcription non-codante ainsi que dans le contrôle des R-loops. Ceux-ci sont des structures potentiellement dangereuses pour l'intégrité du génome qui se forment pendant la transcription quand l'ARN naissant envahit l'ADN et forme un hybride ARN-ADN avec le brin complémentaire. Aussi bien la terminaison de la transcription que l'élimination des R-loops par Sen1 dépendent de son activité catalytique. Les déterminants structuraux de ces fonctions restent cependant inconnus.

Dans cette étude nous avons d'abord identifié la région minimale de Sen1 qui est capable d'induire la terminaison de la transcription et de dissocier des hybrides ARN-ADN et nous avons déterminé sa structure cristalline avec une résolution de 1.8 Å. Sen1 présente une organisation similaire à celle d'autres hélicases de la même famille qui consiste en un core composé de deux domaines de type RecA desquels plusieurs domaines auxiliaires émergent : le "stalk", le "barrel" et le "prong". En général, le core est très conservé au sein des hélicases proches, alors que les domaines accessoires exhibent des caractéristiques distinctes qui confèrent des propriétés spécifiques aux différentes hélicases. En effet, nous avons identifié un sous-domaine spécifique à Sen1 mais conservé au cours de l'évolution que nous avons appelé le "brace". Ce sous-domaine établit un grand nombre d'interactions intramoléculaires avec d'autres sous-domaines (RecA2, le "stalk" et le "barrel"). Ces interactions sont

nécessaires pour un repliement correct de la protéine car la délétion ou la mutation des résidus clés du “brace” génèrent une protéine insoluble *in vitro* et sont fortement délétères *in vivo*. Nous avons également détecté des différences notables au niveau de deux autres domaines accessoires, quand nous comparons la structure du domaine hélicase de Sen1 avec celui d’hélicases proches. Concernant le “barrel”, nous avons remarqué une position différente, très rapprochée de RecA2 et du “prong” qui est probablement une conséquence des interactions avec le “brace”. Nous avons aussi observé que le “prong” a une structure beaucoup moins élaborée que chez d’autres hélicases de la même famille. Notamment, une bonne partie du “prong” est non-structuré. Nous avons pu montrer que ce domaine est essentiel aussi bien pour la dissociation des hybrides ARN-ADN que pour la terminaison de la transcription par Sen1 *in vitro*. Nous avons généré un mutant de délétion du “prong” pour analyser le rôle de ce sous-domaine *in vivo* et nous avons pu constater que délétion du “prong” est létale, ce qui confirme l’importance majeure de ce sous-domaine pour la fonction de Sen1.

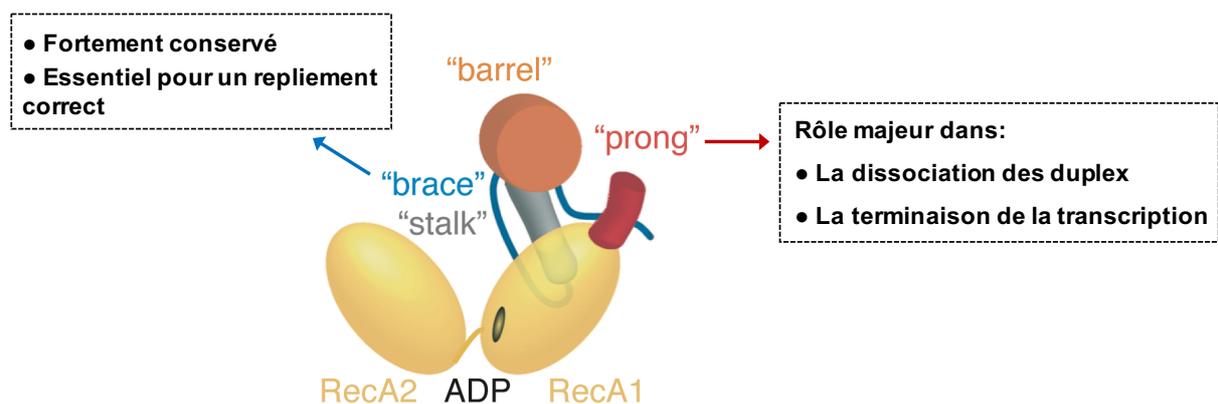
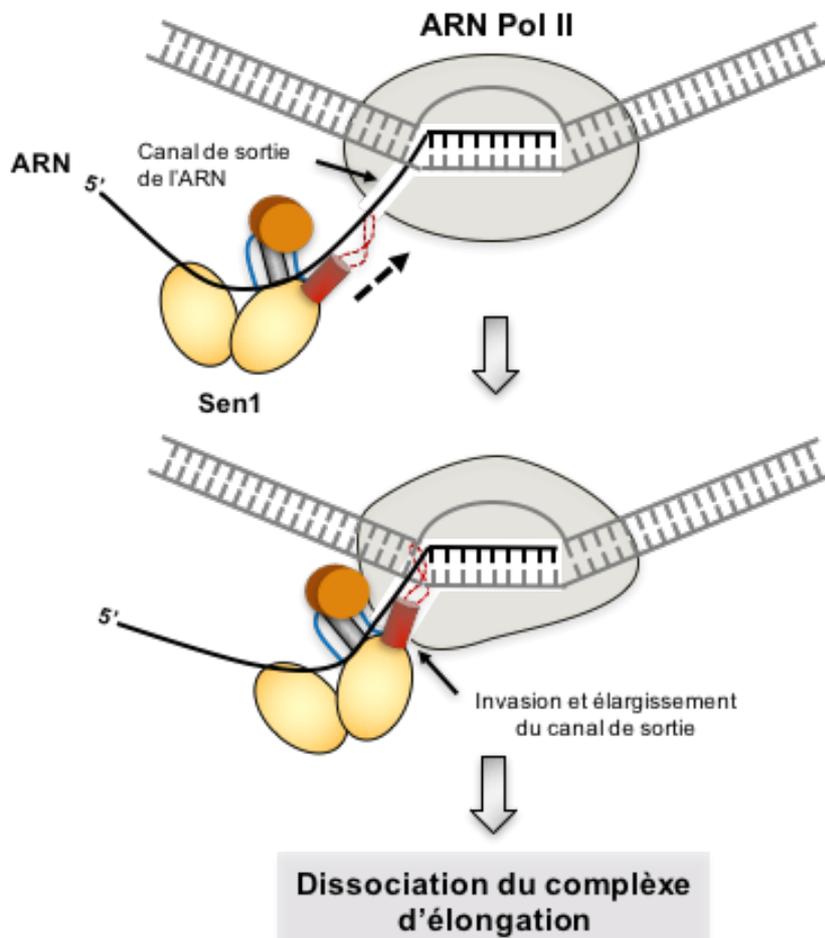


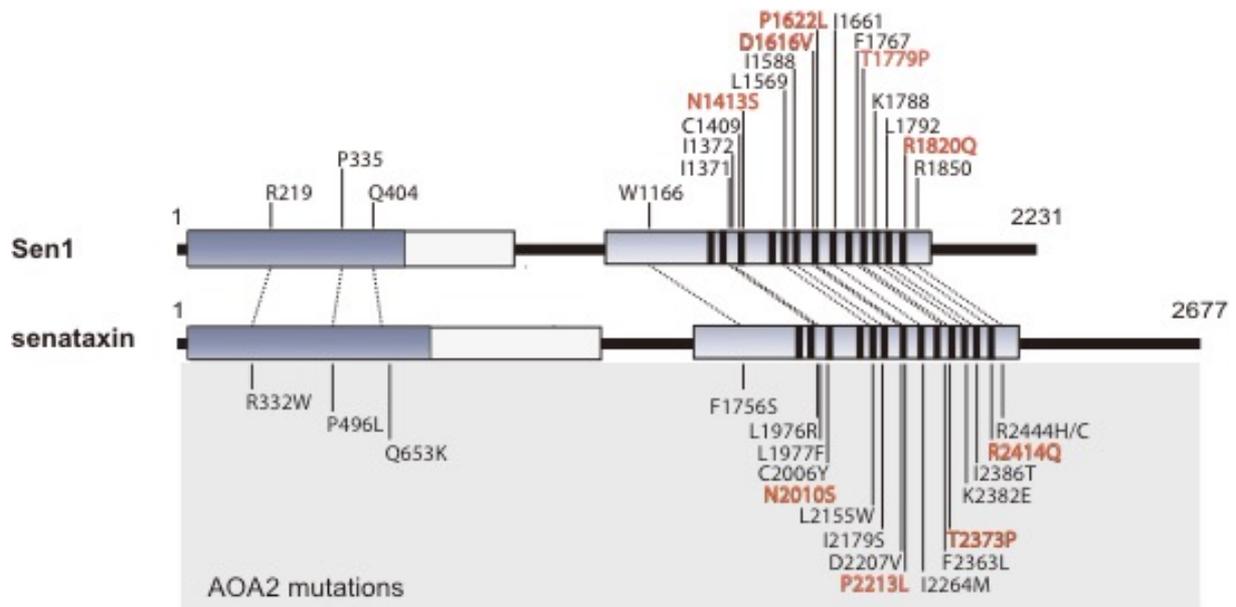
Schéma-résumé des principales caractéristiques structurales révélés par cette étude.

La capacité de dissocier des duplex d’acide nucléique est une propriété générale des hélicases,

tandis que la faculté d'induire la terminaison de la transcription est particulière à Sen1. Nos données suggèrent que les caractéristiques structurales spécifiques de Sen1 que nous avons révélées dans cette étude sont des déterminants majeurs de son activité dans la terminaison de la transcription. Particulièrement, le rôle crucial du “prong” dans la terminaison de la transcription nous a mené à proposer un modèle de terminaison de type “allosterique”, d’après lequel, une fois Sen1 entrera en collision avec l’ARN polymérase II, Sen1 pourrait continuer à transloquer par l’ARN, ce qui permettrait dans un deuxième temps l’invasion du canal de sortie de l’ARN nascent par le “prong”. Cet événement pourrait provoquer l’élargissement du canal et une série des changements conformationnels au niveau de l’ARN polymérase II conduisant à la dissociation du complexe d’élongation.



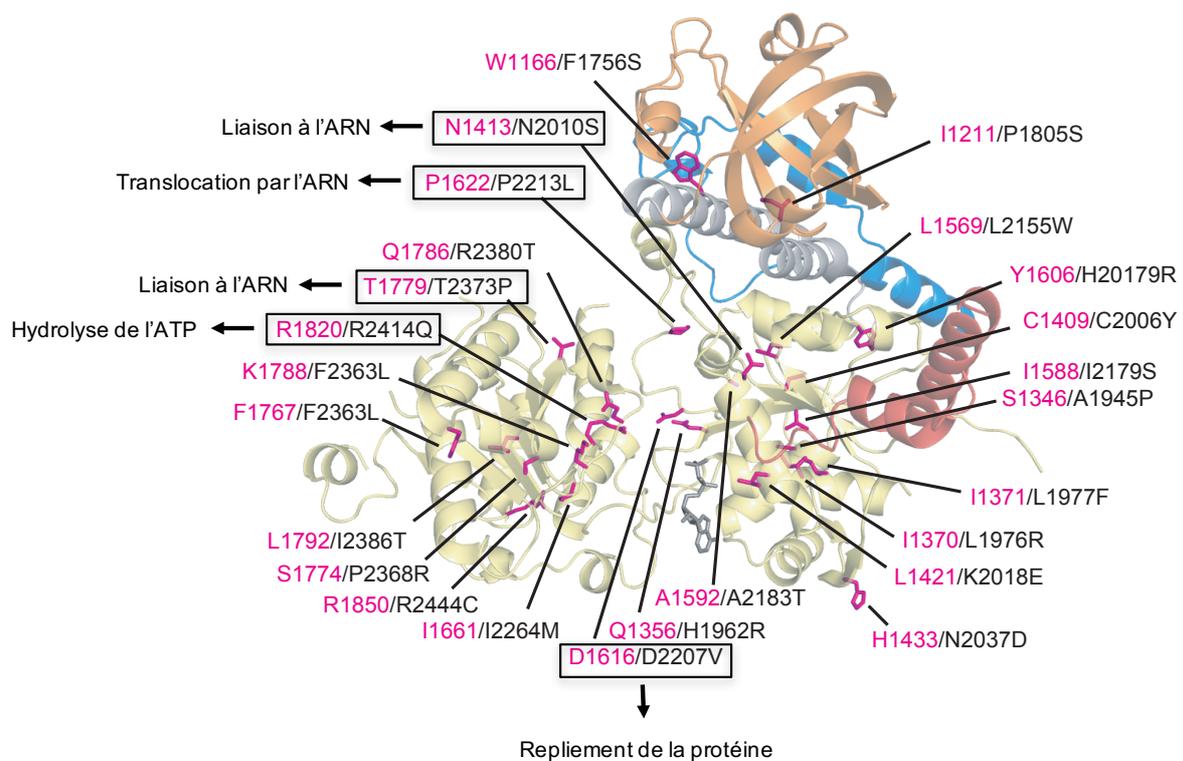
Dans une deuxième partie de ce travail nous avons utilisé Sen1 comme modèle pour étudier la fonction de son homologue chez l'homme, senataxin. Des nombreuses études ont proposé pour senataxin des fonctions dans la terminaison de la transcription et le contrôle de la formation des R-loops. Des mutations dans senataxin sont à l'origine de deux maladies neurodégénératives : sclérose amyotrophique latérale de type 4 (ALS4) et ataxie avec apraxie oculomotrice de type 2 (AOA2). Sen1 et senataxin partagent une organisation similaire avec un domaine central catalytique, un domaine N-terminale de taille importante qui est impliqué dans des interactions avec d'autres protéines et une région C-terminale qui n'a pas de structure secondaire prédite. Le domaine catalytique ou hélicase de senataxin est 50% identique à celui de Sen1 et notamment les résidus clés du "brace" sont aussi conservés chez senataxin, ce qui suggère que senataxin pourrait effectivement posséder les mêmes propriétés que Sen1. En plus, une des mutations qui sont à l'origine de AOA2 (F1756S, W1166S dans Sen1) empêche précisément une des interactions intramoléculaires médiées par le "brace". Puisque jusqu'à présent il n'existe aucune donnée structurale sur senataxin, nous avons décidé de profiter de la forte conservation entre Sen1 et senataxin pour étudier au niveau moléculaire l'effet des mutations associées à des maladies.



Comparaison des architectures de Sen1 et senataxin. Les régions qui n'ont pas de structure secondaire prédite sont indiquées par une ligne. Les régions qui sont les plus conservées sont montrées en gris. Une série de mutations associées à AOA2 qui affectaient des positions conservées sont indiquées, celles qui ont été analysées dans cet étude apparaissent en rouge. Une bonne partie des mutations localisées dans le domaine hélicase correspondent à des positions dans des motifs conservés dans la plupart des hélicases de la même famille (motifs représentés par des lignes noirs).

La plupart des mutations associées à ALS4 et AOA2 se localisent au niveau du domaine hélicase. Nous avons mappé 30 mutations liées à des maladies sur la structure du domaine hélicase de Sen1, ce qui nous a permis de prédire que 2/3 d'entre elles perturberaient considérablement le repliement de la protéine, car elles se situent dans des résidus qui sont noyés dans la structure. En revanche, 1/3 des mutations se localisent plutôt dans la surface du domaine hélicase et souvent correspondent à des acides aminés qui sont conservés et impliqués dans des fonctions clés des hélicases de la même famille. Donc, ces mutations affecteraient plus directement l'activité catalytique de senataxin. Dans le but de valider nos prédictions, nous avons introduit chez Sen1 une partie de ce dernier groupe de mutations et

nous avons réalisé une caractérisation biochimique complète de chaque mutant *in vitro*. Nos résultats nous ont permis d'identifier précisément l'activité affectée par chacune des mutations (hydrolyse de l'ATP, liaison de l'ARN, etc). Nous avons également montré que toutes les mutations sont fortement délétères pour la dissociation d'hybrides ARN-ADN et la terminaison de la transcription. En conclusion, nos résultats ont montré que Sen1 est un modèle très puissant qui peut être utilisé pour améliorer la compréhension de l'origine d'ALS4 and AOA2.



Mappage de 30 mutations liées à AOA2 sur la structure du domaine hélicase de Sen1.

Les mutations identifiées dans senataxin sont indiquées en noir alors que les positions correspondantes dans Sen1 apparaissent en magenta. Les mutations analysées dans ce travail ainsi que les activités que nous avons identifiées comme étant affectés par chaque mutation sont aussi indiquées.

Titre : Caractérisation des mécanismes de terminaison de la transcription par l'hélicase Sen1

Mots clés : Sen1, terminaison de la transcription, ARN

Résumé : La transcription cachée est un phénomène répandu aussi bien chez les eucaryotes que chez les procaryotes. Elle se caractérise par une production massive d'ARNs non-codants au niveau de régions non-annotées du génome et est potentiellement dangereuse pour la cellule car elle peut interférer avec l'expression normale des gènes. Chez *S. cerevisiae*, l'hélicase Sen1 induit la terminaison précoce de la transcription non-codante et joue ainsi un rôle clé dans le contrôle de la transcription cachée. Sen1 est très conservée et des mutations dans son homologue humain, senataxin (SETX), ont été associées à des maladies neurodégénératives. Malgré de nombreuses recherches menées sur ces protéines, leurs propriétés biochimiques ainsi que leurs mécanismes d'action restent peu connus. Durant ma thèse, j'ai étudié le mécanisme de terminaison par Sen1.

Premièrement, j'ai caractérisé les activités biochimiques de Sen1 et analysé comment elles permettent d'induire la terminaison. Pour cela, j'ai utilisé un ensemble de techniques *in vitro*, notamment un système de transcription-terminaison qui contient uniquement des composants purifiés : Sen1, l'ARN polymérase II (Pol II) et les ADN matrices. Ce système permet de modifier les différents éléments de façon contrôlée afin de comprendre leur rôle précis dans la terminaison. J'ai tout d'abord analysé la fonction des différents domaines de Sen1 dans la terminaison. Sen1 est une protéine de taille importante qui possède un domaine central catalytique flanqué par deux domaines impliqués dans l'interaction avec d'autres facteurs. J'ai montré que le domaine hélicase est suffisant pour déclencher la terminaison de la transcription *in vitro*. Ensuite, j'ai montré que Sen1 utilise l'énergie de l'hydrolyse de l'ATP pour se déplacer sur des acides nucléiques simple bras (ARN et ADN) dans le sens 5' vers 3'. J'ai alors étudié le rôle des différents acides nucléiques du système dans la terminaison par Sen1 et j'ai montré que l'interaction de Sen1 avec l'ADN n'est pas nécessaire; en revanche

Sen1 doit s'associer à l'ARN naissant et se déplacer vers la polymérase. J'ai aussi montré qu'une fois que Sen1 entre en collision avec la Pol II, elle y exerce une action mécanique qui conduit à la terminaison uniquement quand la Pol II marque une pause. Cela indique que la terminaison est fortement dépendante de la pause transcriptionnelle.

Deuxièmement, en collaboration avec le groupe d'E. Conti, nous avons réalisé une analyse structure-fonction du domaine hélicase de Sen1. Nous avons observé que Sen1 présente une organisation similaire à celle d'autres hélicases proches avec un core composé de deux domaines de type RecA avec plusieurs domaines auxiliaires. En général, le core est très conservé au sein des hélicases proches, alors que les domaines accessoires ont des caractéristiques distinctes qui confèrent des propriétés spécifiques aux différentes hélicases. En effet, nous avons identifié un sous-domaine spécifique à Sen1 mais conservé au cours de l'évolution que nous avons appelé le "brace". Nous avons également détecté des différences notables au niveau d'un autre domaine accessoire que nous avons nommé le "prong". Nous avons pu montrer que le "prong" est essentiel pour la terminaison par Sen1. Nos données suggèrent que les caractéristiques structurales spécifiques de Sen1 que nous avons révélées sont des déterminants majeurs de son activité dans la terminaison de la transcription.

Finalement, nous avons utilisé Sen1 comme modèle pour étudier des mutations dans SETX qui sont associées à des maladies neurodégénératives. Nous avons introduit chez Sen1 une partie des mutations liées à des maladies et nous avons réalisé une caractérisation biochimique complète de chaque mutant. Nous avons ainsi montré que toutes les mutations sont fortement délétères pour la terminaison de la transcription. En conclusion, nos résultats ont permis d'améliorer la compréhension de l'origine des maladies provoquées par des mutations dans SETX.

Title : Characterization of the mechanisms of transcription termination by the helicase Sen1

Keywords : Sen1, transcription termination, RNA

Abstract : Pervasive transcription is a common phenomenon both in eukaryotes and prokaryotes that consists in the massive production of non-coding RNAs from non-annotated regions of the genome. Pervasive transcription poses a risk that needs to be controlled since it can interfere with normal transcription of canonical genes. In *S.cerevisiae*, the helicase Sen1 plays a key role in restricting pervasive transcription by eliciting early termination of non-coding transcription. Sen1 is highly conserved across species and mutations in the human Sen1 orthologue, senataxin (SETX), are associated with two neurological disorders. Despite the major biological relevance of Sen1 proteins, little is known about their biochemical properties and precise mechanisms of action. During my PhD I have studied in detail the mechanisms of termination by Sen1.

In a first project, I have characterized the biochemical activities of Sen1 and investigated how these activities partake in termination. To this end I have employed a variety of *in vitro* approaches, including a minimal transcription-termination system containing only purified Sen1, RNA polymerase II (RNAPII) and DNA transcription templates that allows modifying the different elements of the system in a controlled manner to understand their role in termination. First, I have analysed the function of the different domains of Sen1 in termination. Sen1 is a large protein composed of a central catalytic domain flanked by additional domains with proposed roles in protein-protein interactions. I have demonstrated that the central helicase domain is sufficient to elicit transcription termination *in vitro*. Next, I have shown that Sen1 can translocate along single-stranded nucleic acids (both RNA and DNA) from 5' to 3'. Then, I have analysed the role of the different nucleic acid components of the elongation complex (i.e. nascent RNA and DNA transcription templates) in termination. My results indicate that termination does not involve the interaction of Sen1 with the DNA but requires Sen1 translocation on the nascent

RNA towards the RNAPII. Importantly, I have shown that upon encountering RNAPII, Sen1 can apply a mechanical force on the polymerase that results in transcription termination when RNAPII is paused under certain conditions. This indicates that RNAPII pausing is a strict requirement for Sen1-mediated termination.

In a second project, in collaboration with the group of E. Conti we have performed a structure-function analysis of the helicase domain of Sen1. Comparison of Sen1 structure with that of other related helicases has revealed an overall similar organization consisting in two tandem RecA-like domains from which additional accessory subdomains protrude. In general, the core RecA-like domains are very well conserved among related helicases and most variation is found in the accessory subdomains, which often confer specific characteristics to different helicases. Indeed, we have found that Sen1 contains a unique but evolutionary conserved structural feature that we have dubbed the "brace". In addition, Sen1 is different from other helicases in an auxiliary subdomain that we have named the "prong". Importantly, we have shown that the integrity of this subdomain is critical transcription termination by Sen1. We propose that the specific features identified in our structural analyses are important determinants of the transcription termination activity of Sen1.

Finally, we have used Sen1 as a model to investigate the molecular effect of SETX mutations linked to neurodegenerative diseases. We have introduced a set of disease-associated mutations in Sen1 and performed a complete biochemical characterization of the different mutants *in vitro*. Importantly, we found that all mutants were severely affected in transcription termination. Taken together, our results elucidate the key structural determinants of the function of Sen1 and shed light on the molecular origin of the diseases associated with SETX mutations.

