Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Structural and biochemical studies of the *S. cerevisiae* DNA/RNA helicase Sen1

Bronislava Leonaitė-Pittelkov

aus

Utena, Litauen

2018

Erklärung

Diese Dissertation wurde im Sinne von §7 der Promotionsordnung vom 1. September 2014 von Frau Prof. Dr. Elena Conti betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 26.02.2018

……

Bronislava Leonaitė-Pittelkov

Dissertation eingereicht am 27.02.2018

1. Gutachterin	Prof. Dr. Elena Conti
2. Gutachter	Prof. Dr. Karl-Peter Hopfner

Mündliche Prüfung am 10.04.2018

PREFACE

Parts of this thesis have been published:

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

I presented parts of this thesis in an international conference:

EMBO Conference "Gene transcription in yeast: from chromatin to RNA and back", Sant Feliu de Guixols, Spain, 2016

"Structural and biochemical studies of the RNA helicase Sen1"

The atomic coordinates and structure factors for Sen1 were deposited at the Protein Data Bank (PDB) with the accession code 5MZN

Summary

The RNA polymerase II (Pol II) is known to play a central role in transcribing all protein coding genes and non-coding RNAs (ncRNAs) in eukaryotic cells. Intriguingly, the majority of short ncRNAs are immediately degraded in the nucleus and therefore referred to as cryptic unstable transcripts (CUTs). Studies in *S. cerevisiae* have revealed that the Nab3-Nrd1-Sen1 (NNS) complex couples the short ncRNA transcription termination and RNA degradation by the nuclear exosome. Sen1 (252 kDa) is a well-conserved 5'→3' RNA helicase and a key player in transcription termination.

In order to understand better the mechanism of termination, the helicase core domain of Sen1 (94 kDa) was expressed, purified and crystallized, and the crystal structure was solved. As shown in this work, Sen1 helicase domain has a very similar overall structure to that of Upf1-like helicases. Surprisingly, the structure reveals a unique feature, the "brace", which fastens the accessory subdomains to RecA1 and frames the helicase in a favorable conformation for RNA binding. Moreover, structure based biochemical studies reveal that the "prong" is an essential element for 5'→3' unwinding and releasing the transcription complex from the template. Finally, I discuss the mechanism of RNA helicase translocation in the 5'→3' direction and propose a structure based model for Pol II elongation complex dissociation.

Zusammenfassung

Die RNA Polymerase II (Pol II) spielt bekanntermaßen eine zentrale Rolle in der Transkription aller proteinkodierenden Gene und nicht-kodierender RNAs (ncRNAs, von English: *non-coding RNAs*) in Eukaryoten. Interessanterweise wird ein Großteil der kleinen ncRNAs noch im Zellkern umgehend abgebaut und deshalb als kryptische instabile Tranksripte (CUTs, von English: *cryptic unstable transcripts*) bezeichnet. Studien in *S. cerevisiae* haben gezeigt, dass der Nab3-Nrd1-Sen1 (NNS) Komplex die Transkriptionstermination kleiner ncRNAs mit dem exosomalen RNA Abbau im Zellkern verbindet. Sen1 (252 kDa) ist eine hochkonservierte 5'→3' RNA-Helikase und nimmt eine Schlüsselrolle in der Transkriptionstermination ein.

Um den Terminationsmechanismus besser zu verstehen wurde die zentrale Helikasedomäne von Sen1 (94 kDa) exprimiert, aufgereinigt und kristallisiert, sowie deren Kristallstruktur gelöst. Wie in dieser Arbeit beschrieben, besitzt Sen1 eine Helikasedomäne mit einer ähnlichen Gesamtstruktur wie Upf1-artige Helikasen. Überraschenderweise weist die Struktur als einzigartiges Merkmal den sogenannten „*brace*“ (Deutsch: Klammer) auf, der die zusätzlichen Unterdomänen an RecA1 fixiert und die Helikase in einer günstigen Konformation für die RNA-Bindung hält. Desweiteren zeigen strukturbasierte biochemische Analysen, dass der sogenannte „prong“ (Deutsch: Zinke) eine wichtige Komponente für die 5'→3' Entwindung und die Freisetzung des Transkriptionskomplexes vom Matrizenstrang ist. Abschließend wird der Mechanismus der RNA-Helikasetranslokation in 5'→3' Richtung diskutiert und ein Modell für die Dissoziation des Pol II Elongationskomplexes vorgeschlagen, basierend auf den Proteinstrukturen.

Acknowledgements

This journey has been empowered by many people that I have met along the way to where I am today.

I am whole-heartedly grateful to my supervisor, Elena Conti, for her guidance, support, kindness and enthusiasm. Thank you, Elena, for your trust and for giving me the full responsibility for my projects. The way of how you have been leading your team over these years taught me a lot about great leadership, it was my pleasure working with you.

The work that I have presented in this thesis has been accomplished working shoulder to shoulder with our collaborators. Odil, thank you for the close communication and coordination, we really made it straightforward! Domenico, it was a great pleasure to discuss with you every time we met. Thank you for coming to my TAC meeting and for your enthusiasm and encouragement that you have shared. And thank you, Zhong, I really appreciate all the work that you have done and wish you all the best in the future.

I also owe mountains of gratitude to all the people in our department; I could not have ended up in a better place to develop as a scientist and a person. Judith, thank you for sharing your lab experience and, more importantly, your patience and wisdom of life: at times when it got too hectic you helped me to ground. Thank you, Jérôme, for sharing your expertise in crystallography, exciting trips to SLS or in other times collecting data for me, and the squash sessions with Steffi, of course. And, Steffi, thank you for always being on my side and listening to my complaints about "luxury problems". My days were definitely brighter in your company!

Also, I would like to say many thanks to Peter and Walter for making sure that all machines in the lab run smoothly, and the cleaning kitchen ladies, Monika, Christel and Sylvia, for making sure that we have all lab's glassware ready to use. Thomas, thank you for preparing plates, media and buffers for us. Fabien, it was so nice to have been working with you, thank you for teaching me safe and precise work in the hotlab and also for cheering me up (at the end I did not need to do quantitative studies of my crystals). My special thanks go to the MPIB crystallization facility. Karina and Sabine, thank you for the set-up of lots and lots of crystallization trials. Ariane, I additionally thank you for test-expressions and factorial buffer screens. The same holds for the MPIB core facility.

Many thanks extend to Claire for the biophysics measurements and for being a great lunch companion. Thank you, Rajan, for the scripts that you have written for us. Elfriede and Marc, thank you for taking care of our insect cell cultures. Petra and Tatjana, thank you for your help in cloning and for being nice desk neighbors. Lisa, thank you for your optimism and organizing our Christmas parties! Steffen, I appreciate our small talks a lot. Petra and Ulrike, thank you for your help with administrative issues. Jörg, thank you for taking care of our finances and safe work.

For the critical reading of my thesis I want to thank Christian, Piotr and Yair. I thank Sebastian and Ingmar for their suggestions and sharing extensive knowledge throughout these years. Also, I would like to say thank you to Mahesh, Michaela, Alex, Achim, Jana, Ksenia and Jan. Moreover, I appreciate a lot to have met the people from Lorentzen, Mizuno and Biertümpfel labs. And the people that have already left the lab: Felix, Shoots, Masami, Debora, Varun, Humayun, Ajla, Ben, Katharina, and Sevim. I hope one day our paths will cross again. Eva, thank you for supervising me during my Master's thesis, you did a great job warming me up in the lab.

My special thanks go to QBM Graduate Program, especially to Markus, Filiz, Maren, Michael and Julia. My special thanks go to Ulrike for the childcare support, which was a huge help. And of course to all you, Qubies, it was amazing to connect with you!

Last but not least, I deeply appreciate my family support. Ačiū jums, mielieji, kad mane išleidot taip toli vieną, kad jaudinotės, palaikėt ir tikėjot. Didžiausias ačiū ir jums, Justina ir Emili, jūs esat pats nuostabiausias turtas ir laimėjimas turėti jus šiandien.

Abbreviations

Å	Ångstrom (=10^{-10} m)
ADP	adenosine diphosphates
ALS4	amyotrophic lateral sclerosis type 4
Amp	ampicillin
AMPPNP	5'-adenylyl-imido-triphosphate
AOA2	ataxia ocular apraxia type 2
ATP	adenosine triphosphates
bp	base pair
BSA	bovine serum albumin
CF	cleavage factor
CPD	cysteine protease domain
CPF	cleavage polyadenylation factor
CTD	C-terminal domain of Rpb1 (Pol II)
CUT	cryptic unstable transcripts
dd	double distilled
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
ds	double stranded
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
GST	glutathione *S*-transferase
HEPES	4-(2-hydroxylethyl)-1-piperazineethanesulfonic acid
H. sapiens	*Homo sapiens*
IP_6	inositol hexakisphosphate
IPTG	isopropyl β-D-1thiogalactopyranoside
MPI	Max Planck Institute
K_D	dissociation constant
kDa	kilodalton
LB	Luria-Bertani
M	molar
mRNA	messenger RNA
MW	molecular weight
MWCO	molecular weight cutoff
Nab3	nuclear polyadenylated RNA-binding protein 3

NEXT	nuclear exosome targeting complex
ncRNA	non-coding RNA
NLS	nuclear localization sequence
NNS	Nab3-Nrd1-Sen1
Nrd1	nuclear pre-mRNA down-regulation protein 1
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PEI	polyethylenimine
PSI	Paul Scherrer Institute
Pol II	RNA polymerase II
PROMPT	promoter upstream transcript
PVDF	polyvinylidene fluoride
r.m.s.d	root mean square deviation
RNA	ribonucleic acid
S. cerevisiae	*Saccharomyces cerevisiae*
S. pombe	*Schizosaccharamyces pombe*
S200	Superdex 200
SAD	single-wavelength anomalous diffraction
SDS	sodium dodecyl sulphate
Sen1	splicing endonuclease gene 1
SETX	senataxin
SF	superfamily
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleo proteins
snoRNA	small nucleolar RNA
ss	single stranded
TBE	Tris base, boric acid and EDTA containing buffer
TFIIH	transcription factor II H
Tm	melting temperature
TRAMP	Trf4-Air2-Mtr4 polyadenylation
tRNA	transfer RNA
Trx	thioredoxin tag
UTR	untranslated region
UV	ultraviolet

TABLE OF CONTENT

FIGURES AND TABLES

1 Introduction

Transcription is the first decoding step towards functional expression of the genomic information in the cell. About 85% of the *Saccharomyces cerevisiae* genome is transcribed (David *et al*, 2006) and the RNA polymerase II (Pol II) plays a central role in transcribing all protein coding genes as well as many non-coding RNAs (ncRNAs), such as cryptic unstable transcripts (CUTs), small nuclear RNAs (snRNAs), and small nucleolar RNAs (snoRNAs). In fact, the eukaryotic genome is pervasively transcribed to a large extent because Pol II binds DNA within the nucleosome-free regions and forms two adjacent preinitiation complexes, resulting in bidirectional transcription of coding and antisense strands of DNA (reviewed in Berretta & Morillon, 2009). Transcription is essential for a flow of life, however it is equally essential to terminate transcription in order to maintain a balance of RNA concentrations within a cell. Therefore, eukaryotic cells have developed several mechanisms which restrict the extent of pervasive transcription, e.g. by limiting transcription initiation of ncRNAs via gene looping or chromatin remodeling (Terzi *et al*, 2011; Tan-Wong *et al*, 2012; Whitehouse *et al*, 2007). But the main transcriptome surveillance mechanism mostly relies on early termination of unwanted transcription and immediate degradation of transcripts (reviewed in Jensen *et al*, 2013).

1.1 Two main Pol II transcription termination pathways in *S.cerevisiae*

Eukaryotic cells maintain tight control over every transcription step through a large number of specific factors that bind to the C-terminal domain (CTD) of Rpb1, the largest subunit of the Pol II core complex. The CTD contains a conserved heptapetide Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (YSPTSPS, repeated 26 times in yeast and 52 times in humans) and is differentially phosphorylated at Tyr1, Ser2, Ser5, and Ser7 throughout the transcription cycle. In *S. cerevisiae*, the mechanism of initiation of all transcripts by Pol II appears to be similar, but, dependent on the phosphorylation pattern of the CTD and the sequence of new transcripts, two distinct pathways, the canonical and the NNS-dependent pathways, are utilized for transcription termination (Figure 1-1).

When Pol II binds to a promoter, the CTD is dephosphorylated. Once Pol II has escaped from the promoter, the CTD is phosphorylated on Ser5 by transcription factor TFIIH, resulting in a conformational change of the CTD (Komarnitsky *et al*, 2000; Buratowski,

2009; Kim *et al*, 2009; Heidemann *et al*, 2013). In *S. cerevisiae*, this modification is recognized by the Nab3-Nrd1-Sen1 (NNS) complex (Kubicek *et al*, 2012; Vasiljeva *et al*, 2008), which then binds to Pol II and scans the emerging nascent RNA for specific sequence motifs to terminate the transcription at early elongation (Wlotzka *et al*, 2011; Mischo & Proudfoot, 2013; Grzechnik *et al*, 2015). If these motifs are not detected, Pol II continues to transcribe RNA and the conformation of the CTD is altered by a decrease of Ser5 and increase of Tyr1 and Ser2 phosphorylation levels, which in turn leads to the dissociation of the NNS complex. Phosphorylation level of Tyr1 and Ser2 increases over the gene length, but Tyr1 phosphorylation level sharply decreases at the 3'-end of the gene enabling binding of the cleavage and polyadenylation CPF-CF complex to the CTD for canonical termination of Pol II transcription (Grzechnik *et al*, 2015; Heidemann *et al*, 2013; Mayer *et al*, 2012).

Notably, there is no clear, pronounced separation between the two termination pathways but rather a gradual decrease in efficiency with an overlap that might function as a termination fail-safe mechanism (Gudipati *et al*, 2008; Grzechnik *et al*, 2015; Rondón *et al*, 2009). The efficiency of early termination can be modulated in response to environmental changes (e.g. nutrient availability), suggesting that ncRNA termination could work in concert with regulatory mechanisms like the Ras pathway (Darby *et al*, 2012). Moreover, for certain genes transcription termination may occur by both pathways alternatively and the choice of the pathway is autoregulated by a stem-loop near the polyA site in the 3'-UTR, or by certain RNA-binding proteins that might cover premature termination sequence motifs (Gudipati *et al*, 2012a; Kim & Levin, 2011). Thus, the NNS complex can function in transcriptional attenuation (Kim & Levin, 2011; Chen *et al*, 2017a).

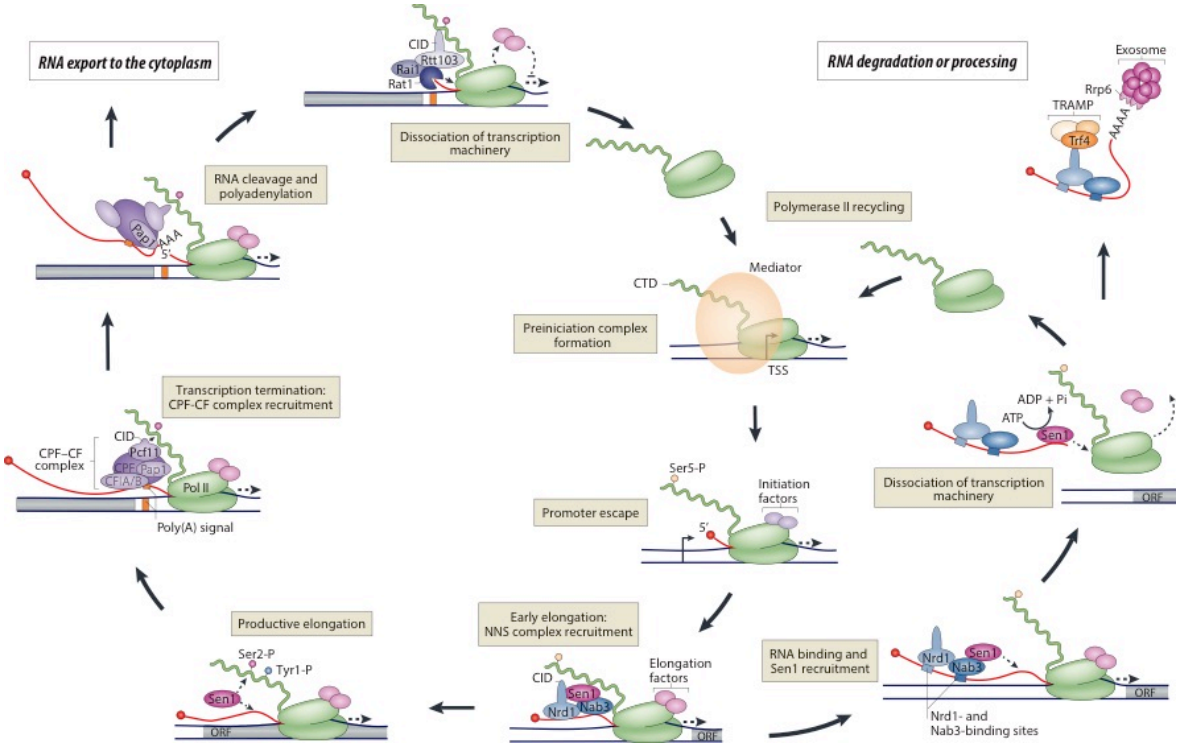


Figure 1-1. Two major pathways of RNA polymerase II transcription termination.
Phosphorylation levels of the CTD of Pol II throughout the transcription cycle coordinate the recruitment of transcription factors and termination complexes. The Nab3-Nrd1-Sen1 (NNS) complex terminates transcription of short ncRNA and promotes degradation of CUTs or processing of sn/snoRNA by the nuclear exosome, whereas the canonical transcription termination of Pol II by the cleavage and polyadenylation CPF-CF complex leads to mRNA export to the cytoplasm. Both complexes bind to the transcription machinery and scan the emerging nascent RNA for conserved termination motifs; however, the NNS complex is recruited much earlier and binds transcript within a few hundreds nucleotides after transcription start site whereas in the CPF-CF pathway the complex is recruited at 3'-UTR of the RNA. This timing advantage allows the NNS complex to detect and control pervasive transcription early during elongation in order to control pervasive transcription. Modified from Porrua & Libri, 2015b.

Termination factors involved in the two pathways bind to the CTD of all Pol II complexes regardless of the template DNA (Heo *et al*, 2013; Lenstra *et al*, 2013), indicating that the termination pathway is determined by the distance of the Pol II from the transcription start site (i.e., the phosphorylation ratio of Ser5 to Ser2), rather than the template sequence. Thus, short ncRNAs are terminated by the NNS pathway within 1kb downstream of the transcription start site, whereas mRNAs are terminated via the CPF-CF pathway (Richard & Manley, 2009; Kuehner *et al*, 2011; Marquardt *et al*, 2011).

Generally, the termination pathway determines the fate of transcripts. In the CPF-CF pathway, the nascent mRNAs are cleaved and polyadenylated at the 3'-end and the stable

mature transcripts are then exported to the cytoplasm (reviewed in Mischo & Proudfoot 2013). In the NNS pathway, the transcripts are linked to the nuclear RNA exosome for rapid degradation of CUTs or processing of sn/snoRNA (reviewed in Porrua & Libri, 2015b). Therefore, the NNS complex has a pivotal role in transcriptome surveillance by selective termination and degradation of cryptic transcripts (Jensen et al. 2013).

It is important to note that the NNS pathway is not evolutionary conserved. Nrd1 and Nab3 are yeast-specific proteins, however, even within various yeast species the function of Nrd1 and Nab3 orthologs differ. For example, in *S. pombe*, Nrd1 ortholog Seb1 is involved in canonical mRNA transcription termination rather than forming a typical NNS-like complex (Lemay *et al*, 2016). By contrast, human transcription termination-coupled decay of ncRNAs (PROMPTs) requires the nuclear cap binding complex and the nuclear exosome targeting complex NEXT (Andersen *et al*, 2013).

1.2 The NNS complex recruitment for transcription termination

The current transcription termination model postulates that the NNS complex is recruited first to CTD of Pol II and in the presence of termination signal – to the nascent RNA. Nrd1 and Nab3 contain RNA-recognition motif (RRM) domains (Figure 1-2) (Hobor *et al*, 2011; Franco-Echevarría *et al*, 2017) with sequence specificity for consensus binding sites GUAA/G and UCUU(G), respectively (Wlotzka *et al*, 2011; Mischo & Proudfoot, 2013; Carroll *et al*, 2004; 2007). Importantly, these sequence motifs occur at high frequencies in antisense direction but are depleted in sense direction of protein coding regions, thus distinguishing ncRNAs from protein coding mRNAs (Cakiroglu *et al*, 2016). *In vitro* data suggests that Nrd1 and Nab3 form a heterodimer to have a cooperative binding to RNA (Carroll *et al*, 2004; Creamer *et al*, 2011; Porrua *et al*, 2012).

Upon Nrd1-Nab3 binding, Sen1 is brought to the nascent RNA via direct interaction with Nab3 (Chinchilla *et al*, 2012; Porrua & Libri, 2015b), however the mechanism of Sen1 recruitment is not understood. *In vitro* studies show that Sen1 alone is sufficient for termination of a stalled polymerase (Porrua & Libri, 2013; Han *et al*, 2017). It has also been shown that Sen1 can bind to the CTD phosphorylated at Ser2 (i.e. during productive elongation) independently of Nrd1-Nab3 (Chinchilla *et al*, 2012). Thus, it is possible that

Nrd1-Nab3 is needed to increase transcription termination efficiency rather than Sen1 recruitment alone. Supporting this possibility is the finding that Nrd1-Nab3 dimer binds to both the extending RNA and Pol II and contributes to polymerase pausing (Schaughency *et al*, 2014).

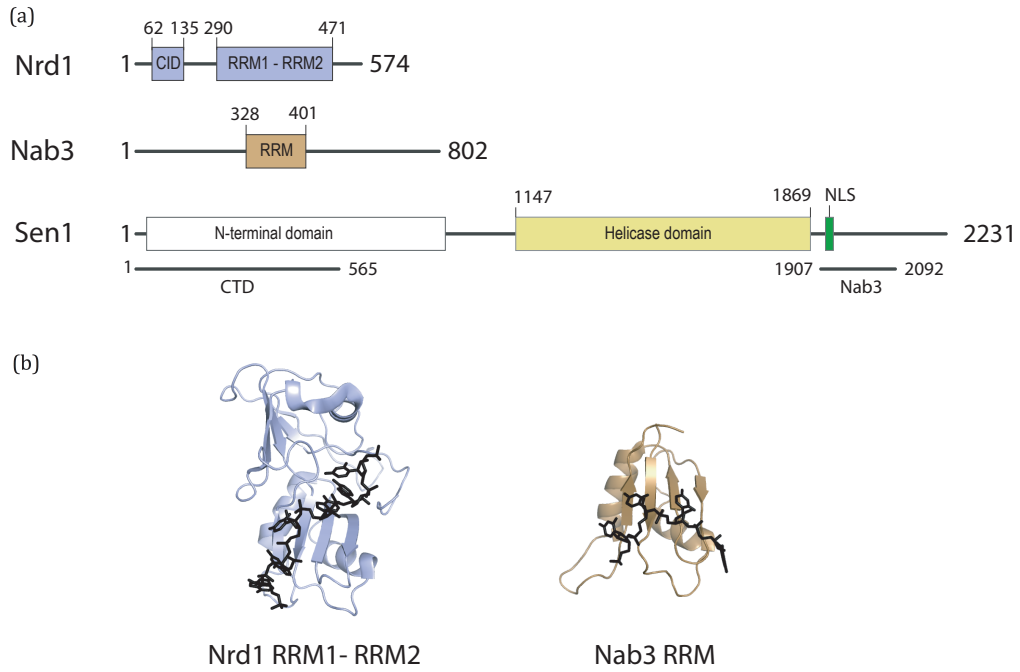


Figure 1-2. The Nrd1-Nab3-Sen1 (NNS) complex.
(a) Schematic presentation of the domain composition of *S. cerevisiae* proteins. Nrd1 contains a CTD-interacting domain (CID) that binds to CTD phosphorylated at Ser5. Nrd1 also contains two RNA-recognizing motif (RRM) domains that are fused together. Nab3 harbors only one RRM domain and forms a heterodimer with Nrd1 to bind RNA cooperatively. Sen1 binds RNA in a sequence independent manner; however, it can be recruited to RNA via interaction with Nab3 or CTD phosphorylated at Ser2. The bars indicate the regions of Sen1 interaction with other proteins. Sen1 contains a nuclear localization sequence (NLS). (b) RRM domain structures. Nrd1 is bound UUAGUAAUCC (PDB: 2LO6) and Nab3 is bound UCUU (PDB: 2L41).

Interestingly, Nrd1 and Nab3 have an unusual sequence stretch of 8 and 16 glutamines, respectively, at the C-termini and therefore Nrd1-Nab3 heterodimers can polymerize onto nascent RNA to assemble a large ribonucleoprotein complex (Carroll *et a*l, 2007; Loya *et al*, 2013a; 2013b). Moreover, Sen1 also has a stretch of polar residues at the N-terminus. All together, this hints to a possibility of all three proteins forming nuclear foci that may enhance the control of pervasive transcription at highly expressed Pol II genes (Loya *et al*, 2013b; Bacikova *et al*, 2014; O'Rourke *et al*, 2015).

1.3 The NNS complex links short ncRNA to TRAMP and the nuclear exosome

Contrary to the CPF-CF pathway, termination by the NNS-mediated termination is coupled to degradation of CUTs or trimming of the precursors of sn(o)RNAs by the nuclear exosome (Kubicek *et al*, 2012). Transcriptomic studies of yeast strains with a catalytically inactivated nuclear exosome have revealed up to 1600 CUTs that are otherwise immediately degraded and not detectable in wild type cells (Wyers *et al*, 2005; Xu *et al*, 2009; Gudipati *et al*, 2008; 2012b).

It has been suggested that the NNS complex links transcription with nuclear RNA surveillance after Nrd1 was shown to associate with Rrp6, the catalytic subunit of the nuclear exosome (Vasiljeva & Buratowski, 2006). Follow-up studies have revealed that the CID of Nrd1, the domain that contributes to the NNS complex recruitment to the early elongating Pol II, is also required for direct interaction with Trf4 subunit of the TRAMP complex (Tudek *et al*, 2014) or another nuclear exosome cofactor Mpp6 (Kim *et al*, 2016). Structural studies have confirmed that CID interaction with CTD and Trf4 (Figure 1-3) is mutually exclusive (Kubicek *et al*, 2012). Thus, Nrd1 can interact with the nuclear exosome cofactors TRAMP and Mpp6 or directly with the nuclear exosome, however this direct interaction is possible only when Nrd1 is not bound to Pol II.

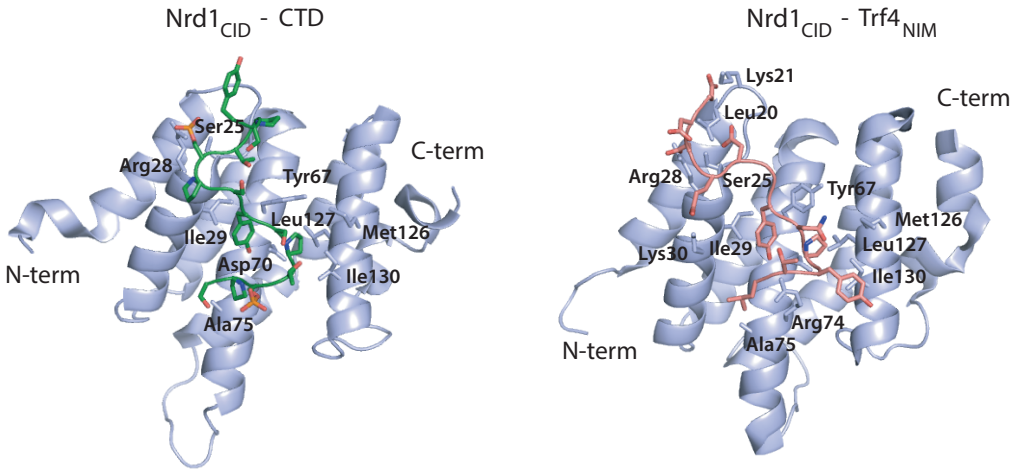


Figure 1-3. Nrd1 CID domain binding to phosphorylated CTD and Trf4.
Structure of $Nrd1_{CID}$ bound with CTD (green) phosphorylated at Ser5 (orange) is shown in the left panel (PDB: 2LO6). Structure of $Nrd1_{CID}$ bound with Nrd1-interacting motif (NIM) of Trf4, a subunit of TRAMP is shown in the right panel (PDB: 2MOW). Nrd1 interaction with CTD and Trf4 is mutually exclusive.

The nuclear exosome is a barrel-shape complex, which has two catalytic subunits, Rrp44 and Rrp6, associated at the bottom and on the top of the barrel, respectively. RNA can be either threaded through the central channel to reach Rrp44 or trimmed by Rrp6 on the top of the complex. Both enzymes are 3'→5' exonucleases, however, Rrp44 possesses an additional endonuclease activity (reviewed in Butler & Mitchell, 2011; Kilchert *et al*, 2016). According to the current model, the nuclear exosome can be in a "closed" (bound Rrp6 only) or "open" (bound Rrp6 and cofactor Mtr4/Mpp6) conformation (Makino *et al*, 2015; Schuch *et al*, 2014). The exosome is active in both conformations, however the conformational change determines whether the RNA will pass through the central channel of the exosome for degradation by Rrp44 or whether it will be trimmed at the surface of the exosome. Exo- and endonuclease activities of the exosome are coordinate by TRAMP and Mpp6 cofactors (Makino *et al*, 2015). In addition, fate of the RNA is largely dependent on its secondary structure (single-stranded vs. bulky RNA).

Interestingly, Nrd1 also can influence the choice between degradation and 3'-end trimming of ncRNA (Vasiljeva & Buratowski, 2006), most likely because of mutually exclusive interactions with Rrp6, TRAMP and Mpp6 and multiple ways by which the NNS complex binds to the exosome (Kim *et al*, 2016). Moreover, Nab3 can also interact with Rrp6 and enhance its catalytic activity independently of Nrd1 (Fasken *et al*, 2015). No direct interaction of Sen1 and exosome or its cofactors has been detected yet, thus the main function of Sen1 is most likely limited to transcription machinery dissociation.

Surprisingly, Nrd1-Nab3 is not limited to transcripts of Pol II but has also been shown to act in the nuclear surveillance of aberrant transcripts of RNA polymerase III, e.g. pre-tRNAs and pre-RPR1, by recognizing consensus binding motifs or structural abnormalities in the RNA and recruiting TRAMP complex (Wlotzka *et al*, 2011).

1.4 DNA/RNA helicase classification

DNA and RNA helicases are essential for every step of nucleic acid metabolism, from chromatin remodeling and DNA replication to mRNA transcription and protein translation. Depending on their structure and function, helicases are classified into six superfamilies (SFs) (reviewed in Singleton *et al*, 2007). The helicases that form a

hexameric ring-shape structure (SF3 to SF6) have been found only in viruses and certain bacteria, but not in eukaryotes. In contrast, all eukaryotic DNA/RNA helicases have a core composed of two structurally similar domains, RecA1 and RecA2, which resemble the fold of the recombination protein RecA. Depending on the characteristic sequence motifs within and the accessory subdomains on the surface of RecA domains, the helicases are separated to SF1 and SF2 (Pyle, 2008; Fairman-Williams *et al*, 2010) (Figure 1-4). These motifs are either involve in ATP binding and hydrolysis (motifs Q, I, II and VI) or mediate DNA/RNA binding (motifs Ia, Ib, Ic, IIIa, IV and IVa). Additionally, the accessory subdomains can interact with RecA domains and/or nucleic acid to stabilize the helicase binding and therefore influence its catalytic activity. A defined coordination of nucleic acid binding and ATP hydrolysis enables the helicase to move along the nucleic acid chain, which may result in removal of secondary structure or associated proteins. In yeast, there are only three families of processive helicases that can translocate along RNA.

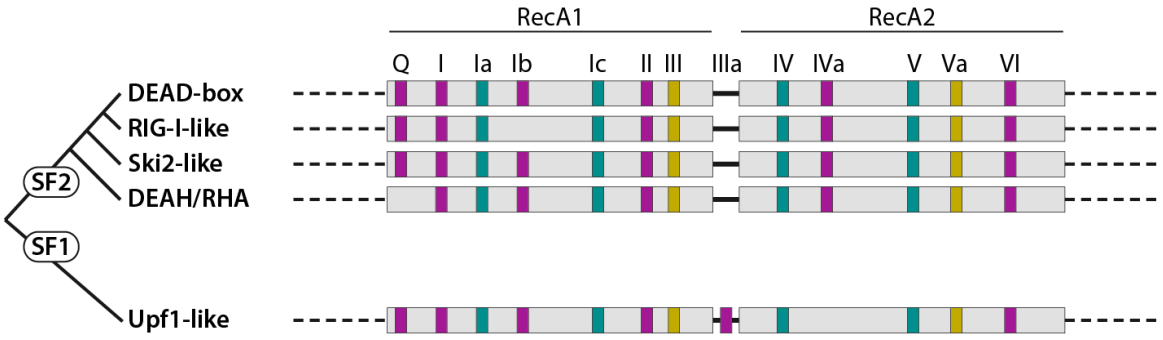


Figure 1-4. Conserved motifs of the RNA helicases in eukaryotes.
A schematic representation of conserved sequence motifs in the helicase core across eukaryotic RNA helicases of SF1 and SF2. Grey rectangles represent two RecA domains; a black line indicates a linker between the domains. Motifs Q, I, II, and VI (magenta rectangles) are involved to ATP binding and hydrolysis. Motif IIIa is found only in SF1, and together with Ib it contributes to adenine ring binding to the cleft, whereas this interaction in SF2 is made by motif IVa (magenta rectangles). Motifs Ia, Ic, IV, and V (turquoise rectangles) are involved in RNA binding. Motifs III and Va (dark yellow rectangles) coordinate the nucleotide and RNA binding. Modified from Jankowsky & Fairman, 2007.

DEAH/RHA family helicases (e.g. Prp43, Sub2, Brr2) and Ski2-like family helicases (e.g. Ski2, Mtr4, Suv3) belong to SF2 and can translocate only in the 3'→5' direction. These helicases are required for proper splicing, RNA degradation or processing by the exosome. Conversely, Upf1-like family helicases (e.g. Upf1 and Sen1) of SF1 translocate in an

opposite 5'→3' direction, although they bind the RNA in the same polarity as SF2 helicases. Upf1-like helicases have a function in nonsense-mediated decay (NMD) or transcription termination (Pyle, 2008; Fairman-Williams *et al*, 2010).

1.5 RNA helicase Sen1 in *S. cerevisiae*

Sen1 is the key enzyme in facilitating the termination of ncRNAs (Chinchilla *et al*, 2012); however, the helicase has a broader function beyond pervasive transcription control. Initially, Sen1 was identified as a splicing endonuclease gene 1 (*SEN1*) in a screen for mutations that inhibit pre-tRNA splicing (DeMarini *et al*, 1992). It was later discovered that mutations in *SEN1* result in a broad range of phenotypes, including disruption of nucleolar organization, defects in transcription, transcription-coupled DNA repair, RNA processing, and snRNP assembly (Ursic *et al*, 1995; Ursic, 1997; Steinmetz *et al*, 2006). Sen1 also binds to the transcription machinery to resolve DNA:RNA hybrid (R-loops) and protects the genome from R-loop mediated genome damage, particularly in higher eukaryotes (Mischo *et al*, 2011; Hamperl & Cimprich, 2014). Finally, it has been suggested that Sen1 binds to replicating forks as well to counteract the DNA:RNA hybrid formation at collision sites between transcription and replication machineries, thus, preventing DNA damage checkpoint activation (Alzu *et al*, 2012).

Interestingly, Sen1 is a low-abundance nuclear protein (125 copies/cell) (Ghaemmaghami *et al*, 2003; Ursic *et al*, 1995) in comparison to roughly 14,000 transcribing Pol II in the cell (Borggrefe *et al*, 2001), 19,000 and 5,800 molecules of Nrd1 and Nab3, respectively (Ghaemmaghami *et al*, 2003). Overexpression of *SEN1* does not lead to a significant increase of Sen1 concentration and the excess amount of the protein appears to be degraded by the ubiquitin-dependent 26S proteosome (DeMarini *et al*, 1995). Also, Sen1 is a low-processivity enzyme that disengages the RNA soon after its binding (Han *et al*, 2017). Together, this helps avoid spurious termination.

The broad spectrum of Sen1 functions is made possible by multiple protein-protein interactions (Ursic *et al*, 2004; Singh *et al*, 2015). For example, the N-terminal domain of Sen1 interacts with Pol II, the endonuclease Rad2 (required for nucleotide excision repair), and the RNase III endonuclease Rnt1 (involved in RNA processing) (Ursic *et al*, 2004;

Chinchilla *et al*, 2012; Li *et al*, 2016). Mutations that disrupt these interactions cause defects in transcription termination, transcription-coupled DNA repair and RNA processing. Furthermore, the C-terminal domain of Sen1 is required for interactions with Nab3 and phosphatase Glc7 (a subunit of CPF factor) (Jamonnak *et al*, 2011; Nedea *et al*, 2008; Ursic *et al*, 2004).

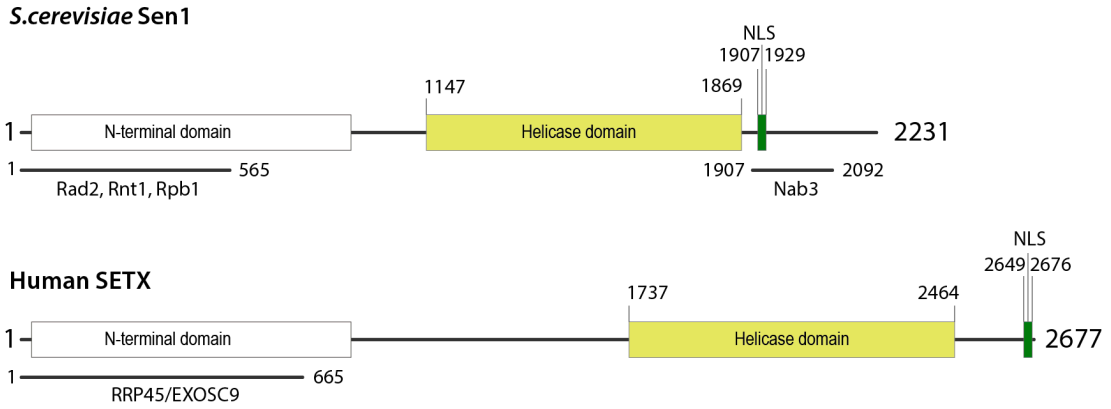


Figure 1-5. Comparison of domain organization in yeast Sen1 and human SETX.
Schematic presentation of a full-length *S. cerevisae* Sen1 and its human ortholog SETX. The numbers indicate the residue numbers. The bars indicate the regions of helicase interaction with other proteins. Both Sen1 and SETX contain a nuclear localization sequence (NLS) (shown in green boxes). The shortest known viable fragment of Sen1 comprises residues 1089-1929. SETX is not essential in human.

Surprisingly, the N- and C-terminal domains are dispensable for viability unless both are deleted at the same time (DeMarini *et al*, 1992; Chen *et al*, 2014; 2017b; Steinmetz *et al*, 2006). Moreover, Sen1 mutant strains that have abolished interaction with Pol II or Nab3 show an increase of certain yet different ncRNAs as a result of termination readthrough (Jenks *et al*, 2008; Chen *et al*, 2017b; Schaughency *et al*, 2014). This phenotype implies the presence of at least two alternative pathways for these transcripts. Since Sen1 can associated with Pol II, either by directly binding to Ser2-phosphorylated CTD (with the N-terminus) or through indirect interactions with the Ser5-phosphorylated CTD as part of the NNS complex (with the C-terminus) (Jamonnak *et al*, 2011; Chinchilla *et al*, 2012), it is possible that the deletion of one of the flanking domains is compensated by another domain. Nevertheless, Sen1 does not require the interaction neither with the CTD of Pol II nor Nab3 for termination reaction itself but rather for earlier steps of commitment to

termination, the helicase alone is sufficient to dissociate the paused polymerase *in vitro* (Porrua & Libri, 2013). However, Sen1 must be recruited to the RNA in close proximity to the transcription machinery and the polymerase should be relatively slow or stalled (Hazelbaker *et al*, 2013; Han *et al*, 2017).

Sen1 (252.5 kDa) is a DNA/RNA-dependent ATPase that translocates in the 5'→3' direction and is capable of unwinding RNA:DNA duplexes (Martin-Tumasz & Brow, 2015; Han *et al*, 2017; Kim *et al*, 1999; Hamperl & Cimprich, 2014). To date, the shortest known region (1089-1929) of Sen1 that is essential for yeast viability comprises a characteristic SF1 helicase core and a flanking region containing a nucleus localization signal (Chen *et al*, 2014)(Figure 1-5).

The human ortholog, Senataxin (SETX), has also been implicated to have similar functions to that of yeast Sen1. Additionally, SETX is required for the efficient transcription termination of protein-coding RNA, and is involved in regulation of the circadian rhythm and microRNA biogenesis (Bennett *et al*, 2013; Skourti-Stathaki *et al*, 2011). However, SETX seems to be non-essential in mammals; loss-of-function of the helicase leads to downregulation of mitochondrial biogenesis and oxidative stress (Bennett *et al*, 2013; Sariki *et al*, 2016). Over 40 missense mutations at the N-terminus and helicase core have been reported to cause progressive neurological diseases, e.g., amyotrophic lateral sclerosis (ALS4) or ataxia ocular apraxia type 2 (AOA2) (reviewed in Bennett & La Spada, 2015). The mechanisms underlying these diseases are not understood, but it is suggested that the mutations cause SETX dysfunction either directly through helicase inactivation or by disrupting protein–protein interactions.

2 Aims

The ultimate goal of this thesis work is to gain understanding of the mechanism of how the helicase Sen1 disrupts the elongation complex of Pol II. An approach combining biochemical and structural methods needs to be taken to define the molecular mechanism of Sen1 binding to RNA, translocation in 5'→3' direction and RNA:DNA duplex unwinding properties. A high-resolution crystal structure of the Sen1 helicase domain and its comparison to available structures of other SF1 RNA helicases would help to reveal why Sen1 is the only RNA helicase that can terminate the transcription of Pol II. The insights and the conclusions of this work should provide strategies for further studies of transcription termination.

3 Results

3.1 Characterization of the active helicase core

3.1.1 Identification of soluble Sen1 constructs

Prior to the initiation of this project there was no successful recombinant expression of Sen1 reported. Also here, expression of full-length Sen1 (1-2231) was not successful either in *E. coli* or in insect cells. Thus, a combination of structure prediction (PSIPRED, Phyre2) and sequence alignment (ClustalW2) was applied to design various constructs of Sen1 (Figure 3-1) (structure prediction, multiple alignment and a full list of constructs see in Appendix).

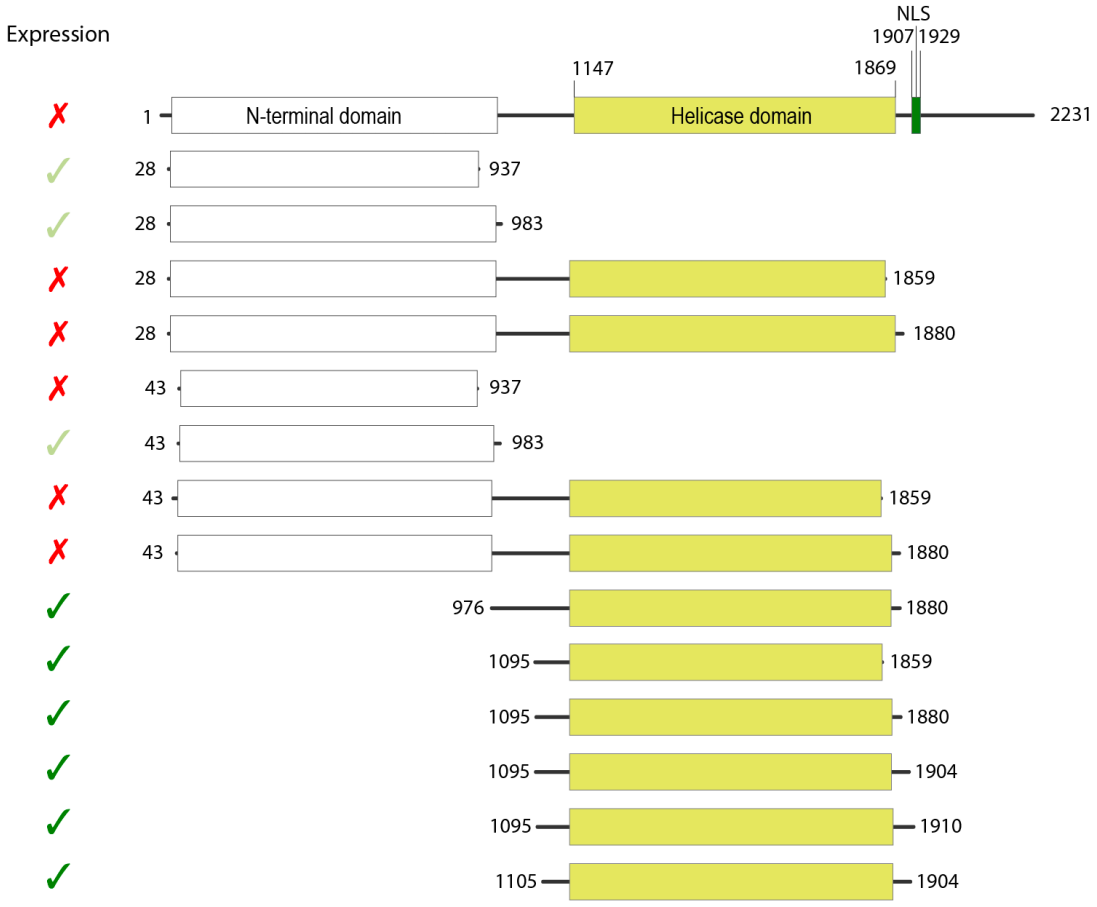


Figure 3-1. Tested constructs of Sen1.
Initially, a full-length Sen1 and the shorter constructs containing both the N-terminal and the helicase domains were tried to express. Also, the domains were expressed separately. The N-terminal domain expression levels were too low for the crystallization trials or not expressed at all and further only expression and purification of the helicase domain alone was continued.

The expression of the N-terminal domain was not suitable for crystallization, thus the main focus was on obtaining the helicase domain alone. At first, a construct 976-1880 that lacks both the N-terminus domain and the low-complexity C-terminus end was designed. The fragment was amplified from *S. cerevisiae* genomic DNA and cloned into a vector with a CPD-His_8 tag at the C-terminus. The construct was well expressed in *E. coli*, however, during purification it appeared that the protein was about 20 kDa smaller than it was expected (Figure 3-2, a). Mass-spectrometry and Edman sequencing confirmed the endogenous protein degradation of 130 amino acids at the N-terminus.

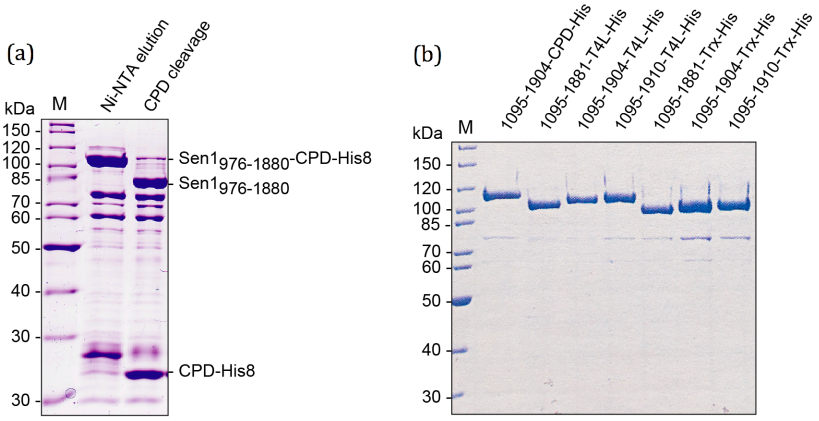


Figure 3-2. Test-purification of Sen1 helicase domain (SDS-PAGE).
(a) Sen1 helicase core [976-1880]. Expected molecular weight of the construct with the tag was 126 kDa but the protein ran just above 100 kDa. (b) Ni^{2+}-NTA elution samples of different Sen1 helicase core constructs.

Secondary structure prediction (see Appendix 1) suggested that residues 1095-1106 form an α-helix, therefore the following constructs starting at residue 1095 were designed. Moreover, for some of the constructs the C-terminal region was extended up to the nucleus localization signal (NLS) (residue 1910) with the idea to obtain longer fragments of the protein and do limited proteolysis. The constructs were designed for the expression in *E. coli* and tested for solubility. Additionally, T4 lysozyme (T4L) and thioredoxin (Trx) tags were tested to check whether a tag changes the expression levels of the proteins. It appeared that all the constructs were similarly well expressed (Figure 3-2, b). Some of the constructs were purified in a large scale and set-up for crystallization. A fragment 1095-1904 (hereafter referred to as $Sen1_{Hel}$) yielded to crystals and, therefore, the further studies were continued only on it.

3.1.2 Purification of the Sen1$_{Hel}$

Several rounds of expression tests and purification buffer screens were performed in order to determine the conditions for the highest yield of Sen1$_{Hel}$ (see Methods 6.2.2 and Methods 6.2.4). The best purity was achieved when Lysis buffer (20 mM sodium phosphate pH 8.0, 500 mM NaCl, 2 mM $MgCl_2$, 1 mM β-mercaptoethanol) was supplemented with 10 % (v/v) glycerol, 30 mM imidazole, benzonase, and protease inhibitors and the CPD-His_8 -tag was cleaved 'on column' by adding HRV 3C protease to Ni^{2+}-affinity beads and incubating at 4°C overnight (see Methods 6.2.4.2). A combination of Ni^{2+}-affinity and heparin ion exchange chromatography steps led to almost pure protein (Figure 3-3, a).

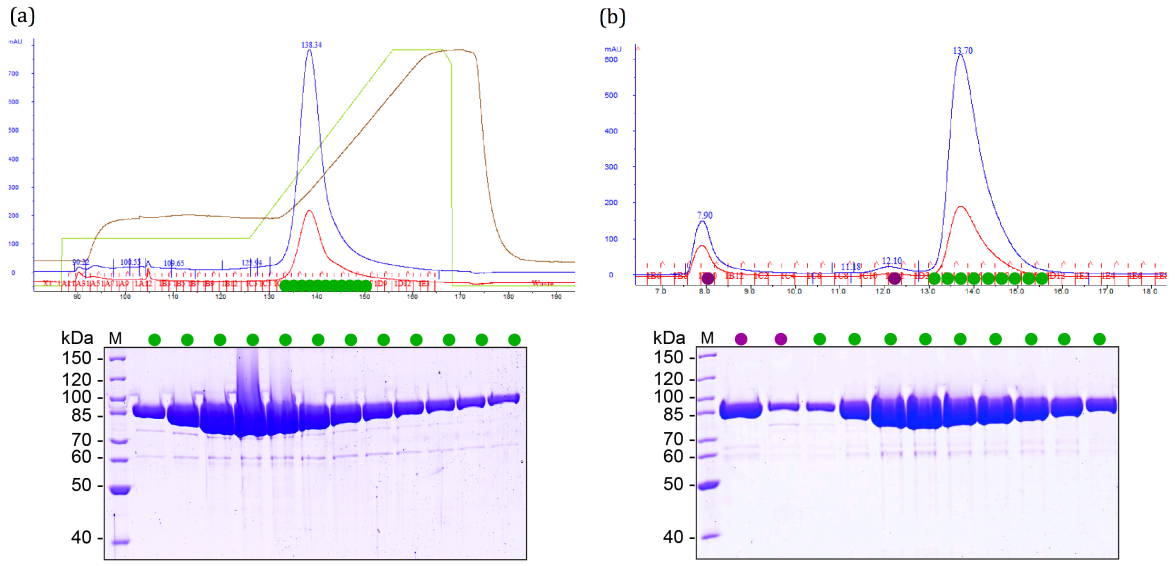


Figure 3-3. Purification of Sen1$_{Hel}$.
(a) A chromatogram of the elution from heparin sepharose column with a linear NaCl gradient. Red and blue lines correspond to the absorbance at 260 nm and 280 nm, respectively. Brown line is the conductivity and green line is percentage of Buffer B. A corresponding SDS-PAGE gel, stained with Coomassie blue, shown at the bottom. After elution from heparin sepharose column, Sen1$_{Hel}$ was concentrated and loaded onto a Superdex 200 [16/600] gel filtration column. (b) A chromatogram of SEC and a corresponding SDS-PAGE gel, stained with Coomassie blue, shown at the bottom. Red and blue lines correspond to the absorbance at 260 nm and 280 nm, respectively. In the main peak (green dots), the ratio of 260 nm and 280 nm is below 0.5, indicating that the sample is not contaminated with nucleic acids. A small fraction of the protein eluted in void-volume or as aggregate (first two peaks, violet dots).

The collected elution fractions from the heparin column were concentrated to ~12 mg/mL and then further purified by size exclusion chromatography (SEC)(Figure 3-3, b). In this step the aggregates were separated (void volume and peak 1) and the protein eluted in a peak at the expected elution volume for a globular protein of 90 kDa (14 ml on Superdex 200 [16/600] column). Every purification step was monitored by SDS-PAGE.

Usually, 3 L of bacterial culture yielded up to 20 mg of purified protein, which could be concentrated to 10 mg/mL or higher. In order to avoid aggregates forming upon freezing, 50 % (w/v) glycerol was added to the buffer. When kept on ice, $Sen1_{Hel}$ was generally very stable and could be stored for weeks.

3.1.3 Biochemical characterization of $Sen1_{Hel}$

To test whether $Sen1_{Hel}$ is a functionally relevant construct, Dr. Zhong Han and Dr. Odil Porrua from collaboration group of Dr. Domenico Libri (Institut Jacques Monod, Paris) tested the ATPase and helicase activities and did *in vitro* transcription termination assays (Porrua & Libri, 2013). It appeared that the helicase core domain alone has similar biochemical properties to those of endogenous full-length Sen1 (the results are discussed more in detail later). In parallel, the collaborators studied the endogenous Sen1 proteins. Their helicase domain contains 976-1880 fragment, the same that I could purify from *E. coli,* however, endogenously expressed protein did not degrade. The results showed that, indeed, the helicase domain alone is sufficient for transcription termination *in vitro* and that the N- and C-terminal domains are most relevant for processes *in vivo* (Han *et al*, 2017). Very similar conclusions were made by Brow's group, who studied $Sen1_{1095\text{-}1876}$ *in vitro* (Martin-Tumasz & Brow, 2015). Whether Sen1 is regulated through posttranscriptional modifications is still not clear, but the fact that the recombinant protein retains the capability of endogenous protein suggests that the modifications are not crucial, at least for the tested truncations and activities.

The optimal RNA length and sequence needed to obtain the $Sen1_{Hel}$-RNA crystal structure was determined by using an RNase protection assay and fluorescence anisotropy measurements (Figure 3-4). In the RNase protection assays, $Sen1_{Hel}$ was incubated with a ^{32}P body-labeled 57-mer RNA in the presence of different nucleotides for 1h on ice. After incubation, to the samples RNase A and RNase T1 were added to digest the RNA that was not bound by the helicase. The protected RNA fragments were then extracted and analyzed by denaturing PAGE, and the gels were visualized by phosphor-imaging (see Methods 6.2.14). $Sen1_{Hel}$ covers ~11-nucleotide fragments in the presence of ADP:AlF_x or ADP:BeF_x (Figure 3-4, a). Minor fragments of ~23 nucleotides were likely due the contiguous binding of more than one protein to the same RNA. No protection was observed when only ADP or AMPPNP was added.

In fluorescence anisotropy measurements, different $Sen1_{Hel}$ concentrations were incubated with a 5'-end fluorescein-labeled 15-mer RNA or ssDNA in a presence of ATP and binding affinities were measured as described in Methods 6.2.13. $Sen1_{Hel}$ showed strongest binding to ssDNA (K_D of 0.25 μM) and $polyU_{15}$ RNA (K_D of 1.11 μM).

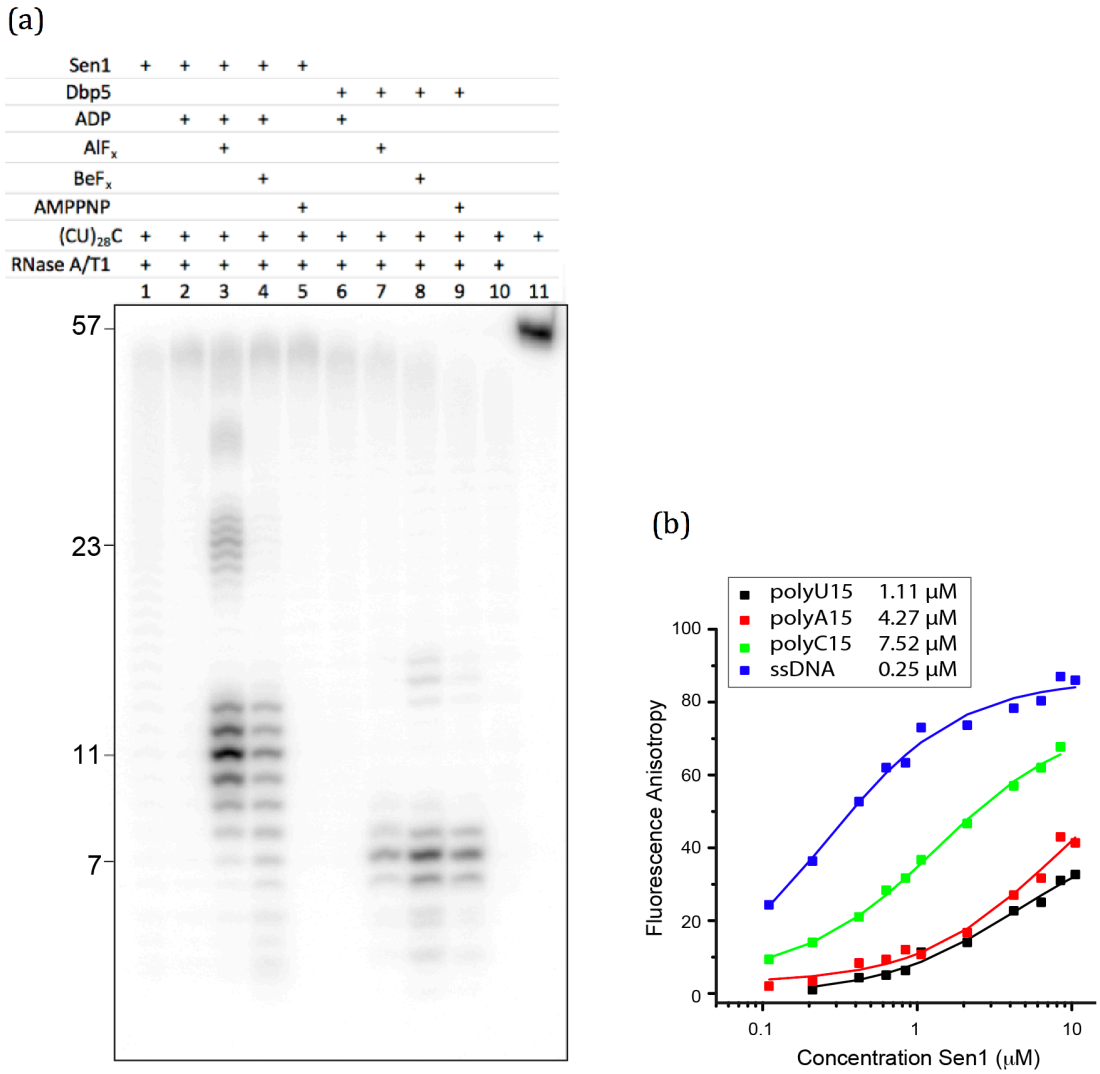


Figure 3-4. RNA binding by Sen1$_{Hel}$.
(a) RNase protection assay in the presence of different nucleotides. RNA fragments were obtained by digesting 57-mer RNA. Sen1$_{Hel}$ binds RNA only in a presence of ADP:BeF$_x$ or ADP:AlF$_x$ but not in a presence of ADP or AMPPNP. Sen1$_{Hel}$ protects 11-nucleotide fragments. For comparison, the footprints of helicase Dbp5 are shown on the right side of the gel. (b) Fluorescence anisotropy measurements. Sen1$_{Hel}$ can bind to both RNA and ssDNA, with the higher affinity for the latter.

3.2 Crystal structure determination of $Sen1_{Hel}$

3.2.1 $Sen1_{Hel}$ crystallization

Purified $Sen1_{Hel}$ was concentrated to 3-4 mg/mL (30-35 μM), mixed with a 1.2 molar excess of $polyU_{15}$ RNA and a 10-fold molar excess of freshly prepared nucleotides to set up for automated crystallization screening in the Crystallization facility of our department. The initial crystallization screening was done using several multicomponent screens by sitting-drop vapor diffusion method in 96-well plates. Most of crystallization hits were observed in crystallization conditions from the in-house Complex screen II solution (Table 3-1) at 4°C (Figure 3-5, a-d). The crystals were forming clusters of needles (a), plates (b-c) or three-dimensional rods (c-d) within 6-10 days. Adjusting precipitant pH and PEG concentrations as well as drop size helped to optimize the crystallization conditions further (Figure 3-5, e-h). Optimized crystals formed larger rod clusters (e), which could be broken apart, or single triangular prisms (f-h), and were mounted to nylon loops for data collection. Prior to flash freezing in liquid nitrogen, the crystals were briefly soaked in mother liquor supplemented with 25-28 % (w/v) ethylene glycol for cryo-protection. In some cases, crystals were soaked in mother liquor additionally supplemented with $polyU_{15}$ RNA.

3.2.2 X-ray data collection

Crystals were first pre-screened using in-house D8 VENTURE (Bruker) X-ray diffractometer to select well diffracting (up to 4 Å) and efficiently cryo-protected ones (no ice ring observed in diffraction pattern). X-ray data collection of data sets used for structure determination was performed at the super-bending magnet beamline X06DA (PXIII) at the Swiss Light Source (Villigen, Switzerland). A native data was collected at 1.0 Å wavelength and a single-wavelength anomalous diffraction (SAD) data set was collected at 2.095 Å.

Table 3-1. Complex screen II.

	1	2	3	4	5	6	7	8	9	10	11	12
A	20% PEG 400 0,1M Na acetate pH 4,5	20% PEG 400 0,1M tri-Na citrate pH 5,5	20% PEG 400 0,1M Na cacodylate pH 6,5	20% PEG 400 0,1M HEPES pH 7,5	20% PEG 400 0,1M Tris pH 8,5	18% PEG 2000 0,1M Na acetate pH 5,0	18% PEG 2000 0,1M tri-Na citrate pH 5,6	18% PEG 2000 0,1M Bis - Tris pH 6,2	18% PEG 2000 0,1M Na K phosphate pH 6,5	18% PEG 2000 0,1M Na cacodylate pH 6,0	18% PEG 2000 0,1M MES pH 6,0	18% PEG 2000 0,1M MES pH 6,5
B	18% PEG 2000 0,1M ADA pH 6,6	18% PEG 2000 0,1M HEPES pH 7,0	18% PEG 2000 0,1M HEPES pH 7,5	18% PEG 2000 0,1M MOPS pH 7,2	18% PEG 2000 0,1M Tris pH 8,0	18% PEG 2000 0,1M Tris pH 8,5	18% PEG 4000 0,1M Na acetate pH 5,0	18% PEG 4000 0,1M tri-Na citrate pH 5,6	18% PEG 4000 0,1M Bis-Tris pH 6,2	18% PEG 4000 0,1M Na K phosphate pH 6,5	18% PEG 4000 0,1M Na cacodylate pH 6,0	18% PEG 4000 0,1M MES pH 6,0
C	18% PEG 4000 0,1M MES pH 6,5	18% PEG 4000 0,1M ADA pH 6,6	18% PEG 4000 0,1M HEPES pH 7,0	18% PEG 4000 0,1M HEPES pH 7,5	18% PEG 4000 0,1M MOPS pH 7,2	18% PEG 4000 0,1M Tris pH 8,0	18% PEG 4000 0,1M Tris pH 8,5	18% MPEG 5000 0,1M Na acetate pH 5,0	18% MPEG 5000 0,1M tri-Na citrate pH 5,6	18% MPEG 5000 0,1M Bis - Tris pH 6,2	18% MPEG 5000 0,1M Na K phosphate pH 6,5	18% MPEG 5000 0,1M Na cacodylate pH 6,0
D	18% MPEG 5000 0,1M MES pH 6,0	18% MPEG 5000 0,1M MES pH 6,5	18% MPEG 5000 0,1M ADA pH 6,6	18% MPEG 5000 0,1M HEPES pH 7,0	18% MPEG 5000 0,1M HEPES pH 7,5	18% MPEG 5000 0,1M MOPS pH 7,2	18% MPEG 5000 0,1M Tris pH 8,0	18% MPEG 5000 0,1M Tris pH 8,5	15% PEG 6000 0,1M Na acetate pH 5,0	15% PEG 6000 0,1M tri-Na citrate pH 5,6	15% PEG 6000 0,1M Bis - Tris pH 6,2	15% PEG 6000 0,1M Na K phosphate pH 6,5
E	15% PEG 6000 0,1M Na cacodylate pH 6,0	15% PEG 6000 0,1M MES pH 6,0	15% PEG 6000 0,1M MES pH 6,5	15% PEG 6000 0,1M ADA pH 6,6	15% PEG 6000 0,1M HEPES pH 7,0	15% PEG 6000 0,1M HEPES pH 7,5	15% PEG 6000 0,1M MOPS pH 7,2	15% PEG 6000 0,1M Tris pH 8,0	15% PEG 6000 0,1M Tris pH 8,5	15% PEG 8000 0,1M Na acetate pH 5,0	15% PEG 8000 0,1M tri-Na citrate pH 5,6	15% PEG 8000 0,1M Bis - Tris pH 6,2
F	15% PEG 8000 0,1M Na K phosphate pH 6,5	15% PEG 8000 0,1M Na cacodylate pH 6,0	15% PEG 8000 0,1M MES pH 6,0	15% PEG 8000 0,1M MES pH 6,5	15% PEG 8000 0,1M ADA pH 6,6	15% PEG 8000 0,1M HEPES pH 7,0	15% PEG 8000 0,1M HEPES pH 7,5	15% PEG 8000 0,1M MOPS pH 7,2	15% PEG 8000 0,1M Tris pH 8,0	15% PEG 8000 0,1M Tris pH 8,5	12% PEG 20 000 0,1M Na acetate pH 5,0	12% PEG 20 000 0,1M tri-Na citrate pH 5,6
G	12% PEG 20 000 0,1M Bis - Tris pH 6,2	12% PEG 20 000 0,1M Na K phosphate pH 6,5	12% PEG 20 000 0,1M Na cacodylate pH 6,0	12% PEG 20 000 0,1M MES pH 6,0	12% PEG 20 000 0,1M MES pH 6,5	12% PEG 20 000 0,1M ADA pH 6,6	12% PEG 20 000 0,1M HEPES pH 7,0	12% PEG 20 000 0,1M HEPES pH 7,5	12% PEG 20 000 0,1M MOPS pH 7,2	12% PEG 20 000 0,1M Tris pH 8,0	12% PEG 20 000 0,1M Tris pH 8,5	1,8M Amm. sulfate 0,1M Na acetate pH 5,0
H	1,8M Amm. sulfate 0,1M tri-Na citrate pH 5,6	1,8M Amm. sulfate 0,1M Bis - Tris pH 6,2	1,8M Amm. sulfate 0,1M Na K phosphate pH 6,5	1,8M Amm. sulfate 0,1M Na cacodylate pH 6,0	1,8M Amm. sulfate 0,1M MES pH 6,0	1,8M Amm. sulfate 0,1M MES pH 6,5	1,8M Amm. sulfate 0,1M ADA pH 6,6	1,8M Amm. sulfate 0,1M HEPES pH 7,0	1,8M Amm. sulfate 0,1M HEPES pH 7,5	1,8M Amm. sulfate 0,1M MOPS pH 7,2	1,8M Amm. sulfate 0,1M Tris pH 8,0	1,8M Amm. sulfate 0,1M Tris pH 8,5

In brown are highlighted precipitant solutions in which crystals of $Sen1_{Hel}$ were formed.

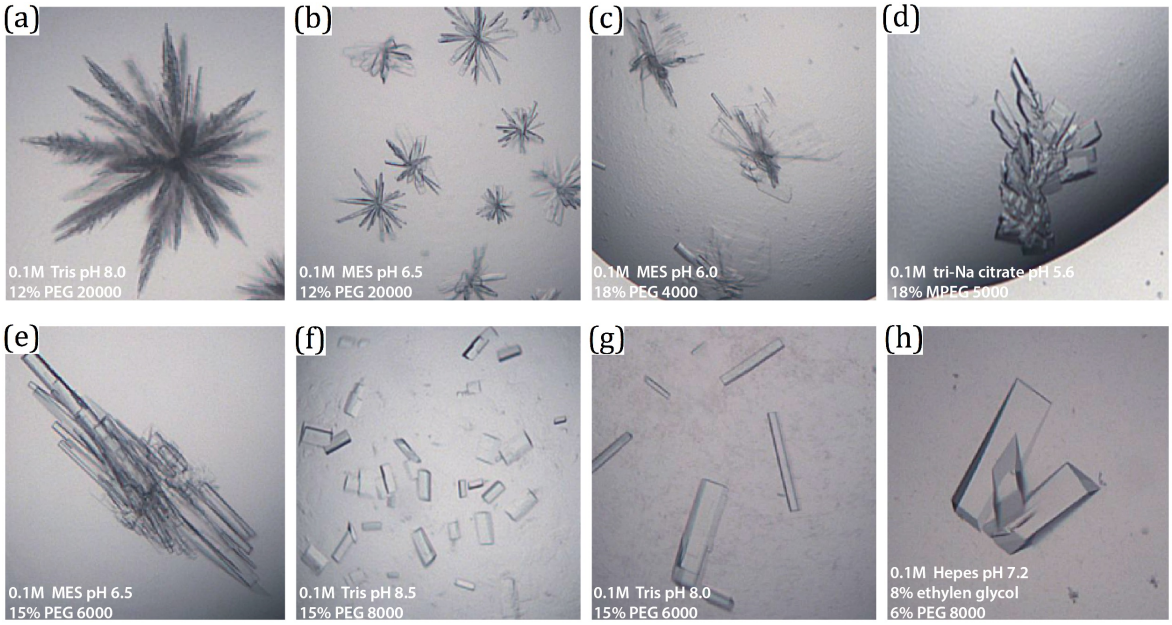


Figure 3-5. Crystal shape changes upon optimization of crystallization conditions.
(a-d) Initial crystals in multicomponent precipitant solution screens. (e-h) Crystals after optimization of precipitant solution. The best diffracting crystals (h) were grown at 4 °C by hanging-drop vapor diffusion from 2 µL drops formed by equal volumes of protein and of crystallization solutions (6 % (w/v) PEG 8000, 8 % (v/v) ethylene glycol, 0.1 M HEPES pH 7.2).

The best crystals diffracted to 1.8-2.5 Å and could bear up to 4 MGy dosage (<0.5 MGy per 360°) until the first signs of diffraction decay were visible, thus multiple 360° data sets could be collected in different orientations on a single crystal. This enabled doing a single-wavelength anomalous diffraction experiment from intrinsic sulfur atoms of cysteine and methionine residues (S-SAD) (see Methods 6.3.2). The weak anomalous signal of intrinsic sulfur could be assessed by measuring the individual reflections at different angles, i.e., multiple trajectories of reciprocal lattice points. Merging of the data sets increased the proportion of anomalous signal while the number of unique reflections remains nearly the same. Eventually, 4 x 360° data sets from one crystal were sufficient to increase the anomalous signal correlation by 16 % at 3 Å resolution (Table 3-2). The statistics of data collection are listed in the Table 3-3.

Table 3-2. Statistics of data processing.

Native data set *(collected at 1.00Å wavelength)*

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
5.64	47980	6230	6262	99.5%	3.3%	3.4%	47961	51.57	3.5%	99.9*	16*	0.964	2718
3.99	73711	11436	11490	99.5%	3.4%	3.5%	73706	44.08	3.7%	99.9*	8	0.885	5315
3.26	105113	14677	14792	99.2%	4.3%	4.2%	105112	37.67	4.6%	99.9*	9	0.911	6943
2.82	128373	17411	17569	99.1%	6.7%	6.5%	128373	25.49	7.2%	99.8*	3	0.867	8303
2.52	131434	19645	19911	98.7%	10.7%	10.6%	131434	15.73	11.7%	99.3*	2	0.831	9427
2.30	145455	21740	22014	98.8%	17.0%	16.8%	145455	10.58	18.4%	98.4*	3	0.826	10467
2.13	167785	23451	23936	98.0%	27.8%	27.4%	167784	7.03	30.0%	96.6*	2	0.821	11329
1.99	183811	25139	25704	97.8%	48.9%	48.4%	183810	4.17	52.6%	90.7*	0	0.789	12177
1.88	191101	26726	27392	97.6%	83.4%	82.8%	191100	2.37	89.8%	77.3*	0	0.757	12964
1.78	184024	28134	28946	97.2%	146.1%	145.3%	183973	1.24	158.8%	47.6*	1	0.708	13640
1.70	201704	29231	30345	96.3%	249.7%	249.5%	201568	0.72	269.9%	26.2*	2	0.668	14120
1.63	211329	30212	31947	94.6%	397.0%	397.9%	210866	0.42	428.0%	13.3*	1	0.628	14423
1.56	212820	30692	33183	92.5%	626.2%	627.9%	212341	0.23	675.4%	5.7	0	0.576	14719
1.51	199997	30492	34445	88.5%	898.9%	903.6%	199370	0.13	974.7%	2.5	0	0.564	14365
1.46	149145	26657	35762	74.5%	1466.5%	1476.0%	147678	0.04	1611.8%	0.3	1	0.533	12125
1.41	105055	19557	36908	53.0%	3365.7%	3409.5%	104162	-0.03	3717.6%	-1.0	2	0.514	9047
1.37	74162	14897	38219	39.0%	-99.9%	-99.9%	72966	-99.00	-99.9%	-4.3	2	0.497	6495
1.33	47441	10923	39230	27.8%	-99.9%	-99.9%	46226	-99.00	-99.9%	-7.3	1	0.487	4597
1.29	23260	8085	40458	20.0%	-99.9%	-99.9%	22213	-99.00	-99.9%	-4.8	4	0.498	3212
1.26	5721	3707	41361	9.0%	-99.9%	-99.9%	3542	-99.00	-99.9%	-15.5	4	0.428	528
total	2589421	399042	559874	71.3%	13.8%	13.8%	2579640	6.94	15.0%	99.9*	2	0.691	186914

Merged data set for S-SAD *(collected at 2.095Å wavelength)*

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
9.59	28953	1260	1269	99.3%	3.2%	3.6%	28953	92.60	3.3%	100.0*	79*	2.452	497
6.78	49713	2324	2324	100.0%	3.9%	4.1%	49713	70.94	4.0%	100.0*	65*	1.997	1020
5.54	59485	3016	3016	100.0%	5.4%	5.4%	59485	52.52	5.5%	99.9*	61*	1.798	1368
4.79	76317	3566	3566	100.0%	5.0%	5.0%	76317	58.01	5.1%	100.0*	50*	1.560	1642
4.29	86063	4001	4001	100.0%	4.9%	4.8%	86063	62.29	5.0%	100.0*	42*	1.389	1859
3.91	93163	4461	4461	100.0%	5.6%	5.3%	93163	54.48	5.7%	99.9*	31*	1.233	2089
3.62	91689	4860	4860	100.0%	6.4%	6.0%	91689	45.60	6.5%	99.9*	18*	1.080	2288
3.39	104472	5238	5238	100.0%	7.5%	7.3%	104472	38.93	7.7%	99.9*	16*	1.014	2475
3.20	112083	5533	5533	100.0%	9.3%	9.2%	112083	31.41	9.6%	99.9*	13*	0.956	2626
3.03	118184	5871	5871	100.0%	11.8%	12.0%	118184	24.33	12.1%	99.8*	16*	0.947	2795
2.89	116986	6128	6129	100.0%	14.5%	15.0%	116986	19.07	14.9%	99.7*	10	0.861	2923
2.77	112337	6451	6451	100.0%	18.4%	19.2%	112337	14.51	19.0%	99.4*	7	0.825	3085
2.66	118556	6731	6731	100.0%	24.0%	25.3%	118556	11.25	24.7%	98.9*	4	0.800	3224
2.56	120752	6982	6982	100.0%	28.0%	29.5%	120752	9.56	28.9%	98.7*	2	0.790	3346
2.48	120815	7234	7234	100.0%	34.9%	37.0%	120815	7.62	36.0%	97.7*	-1	0.777	3473
2.40	96621	7547	7547	100.0%	40.4%	43.2%	96618	5.61	42.1%	95.8*	4	0.779	3629
2.33	63297	7566	7703	98.2%	45.6%	49.4%	63136	3.81	48.6%	92.0*	4	0.756	3535
2.26	42820	7396	7943	93.1%	55.0%	59.5%	42516	2.56	60.3%	83.1*	1	0.720	3376
2.20	24191	6824	8186	83.4%	62.2%	66.0%	23327	1.69	72.2%	69.3*	0	0.729	2456
2.14	6503	3560	8378	42.5%	70.1%	75.9%	4854	0.96	90.6%	49.0*	0	0.708	352
total	1643000	106549	113423	93.9%	8.3%	8.4%	1640019	22.47	8.6%	99.9*	13*	0.968	48058

Red line indicates where the data was cut-off.

3.2.3 Crystal structure determination and evaluation

The data could be indexed and processed in orthorhombic space group $P2_12_12$ with one molecule per symmetric unit. The data was processed using XDS and scaled and merged with XSCALE (Kabsch, 2010). The high-resolution data cutoff was based on the statistical indicators $CC_{1/2}$ and CC* (Karplus & Diederichs, 2012). Substructure determination and phasing were performed with SHELXC/D/E (Sheldrick, 2010) using HKL2MAP (Pape & Schneider, 2004). The successful SHELXD substructure solution, in a search for 25 sulfur sites, had a CC_{all} and a CC_{weak} 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.

The initial model was built automatically using BUCCANEER (Cowtan, 2006) and corrected and completed manually using COOT (Emsley & Cowtan, 2004) and the experimental electron density map in. Further model refinements were done against native data with PHENIX.refine (Adams *et al*, 2010). When building in ADP:AlF_x into the model, it appeared that only ADP molecule was bound to $Sen1_{Hel}$ and no electron density for the AlF_x moiety could be obtained (Figure 3-6). Finally, also ligands (glycerol and ethylene glycol) and water molecules were built in into the model.

The final model was refined at 1.8-Å resolution with an R_{free} of 18.4 % and R_{factor} of 15.3 % and good stereochemistry (Table 3-3). In particular, 98 % of residues are in the most favored regions of the Ramachandran plot and the model has no outliers (Figure 3-7, b). Based on the wwPDB evaluation tool, which compares all available structures of the same resolution, $Sen1_{Hel}$ model has relatively good percentile ranks also for key global quality indicators like Clashscore and outliers of Sidechain or Real-space R-value Z-score (RSRZ) (Figure 3-7, a).

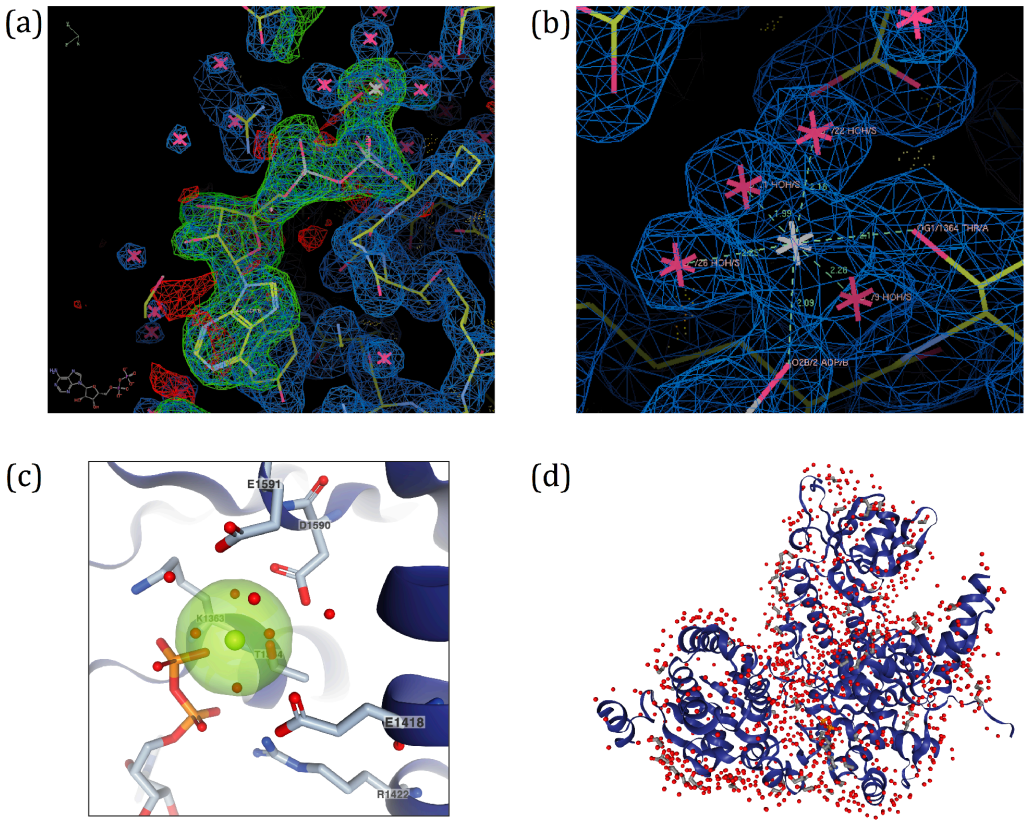


Figure 3-6. Sen1$_{Hel}$ model building
(a) ADP molecule and magnesium ion are built into the F_o-F_c electron density map (green). (b) A hexahydrated octahedral architecture of magnesium dication and six oxygen atoms, four from water molecules and two from Thr$_{1364}$ residue and β-phosphate of ADP. The $2F_o$-F_c electron density map is contoured at 1.7σ. (c) Magnesium (green sphere) is coordinated at the active site of Sen1 by amino acid residues (Thr1364, Asp1590, Glu1591, Glu1418, Arg1422 and Lys1363), ADP and is surrounded with six waters. (d) The overall view of final Sen1$_{Hel}$ model (blue). Water molecules and ligands are displayed in red and grey, respectively.

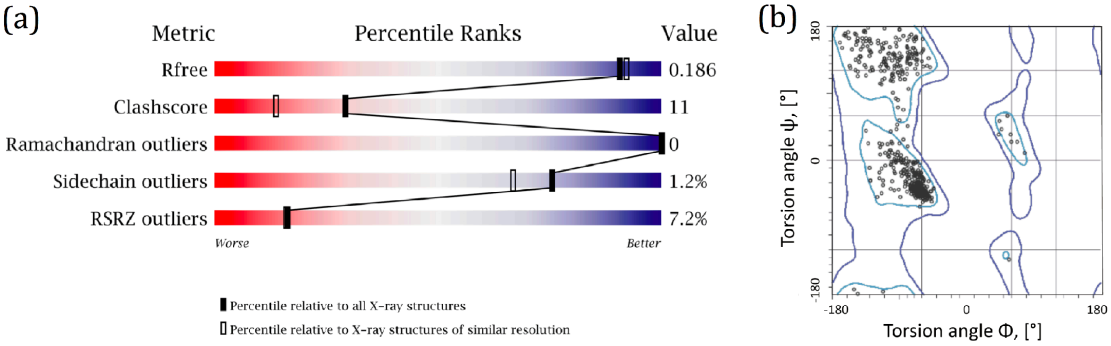


Figure 3-7. Sen1$_{Hel}$ model validation.
(a) Overall X-ray structure quality validation with the respect to all structures in the Protein Data Bank (http://rcsb.org). (b) Ramachandran plot of Φ and Ψ torsion angles of the α-chain of the protein.

Table 3-3. Crystallographic data collection and refinement 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	680,302 (29,401)	1,643,000
Unique reflections	100,766 (2,170)	114,276 (8,032)
Multiplicity	13.2 (13.5)	
Completeness (%)	98.27 (95.07)	93.9 (68.81)
Mean I/sigma(I)	28.21 (19.71)	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	6,970	
macromolecules	5,543	
ligands	337	
water	1,090	
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.

3.2.4 Trials to change crystal-packing in order to obtain protein-RNA structure

One of the project goals was to achieve the structure of $Sen1_{Hel}$ in complex with RNA, however in the first attempts no RNA density could be observed. $Sen1_{Hel}$ was again set up with 11-nucleotide polyU RNA or polyT ssDNA and ADP:AlF_x or ADP:BeF_x for

crystallization trials. Additionally, catalytically inactive $Sen1_{Hel}$ (E1591Q) was set up with ATP and nucleic acid. Despite extensive trials, including UV-cross linking and crystal soaking with RNA, no density for nucleic acid chain in the structure could be obtained. Most likely reason for that was the tight crystal packing that was unfavorable for nucleic acid binding, as the helicases have different RNA-bound and RNA-free conformations. Moreover, one of the crystal-packing contacts (residue Leu1549) seemed to be at the site where 3'-end of RNA would reside.

Therefore, several mutants that might disrupt the packing were cloned, expressed and purified as described for $Sen1_{Hel}$ (Methods 6.2.4). Firstly, the disordered region within 1C subdomain (residues 1471–1543) was deleted. Secondly, three sites of protein-protein interactions in the crystal packing (see Figure 3-8) were mutated: the N-terminus end was truncated to residue 1105 (ΔN-term) and two point mutations, L1549D and G1769R, were introduced. Thirdly, crystallization trials with uncleaved tags (CPD, Trx or T4 lysozyme) were set up. Nevertheless, the unit cell parameters in all cases were the same and the RNA/DNA density could never be observed.

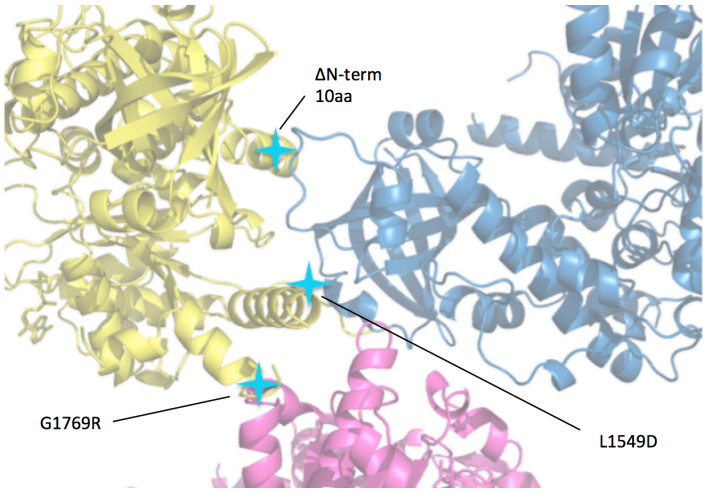


Figure 3-8. Attempts to disrupt protein-protein crystal-packing contacts.
$Sen1_{Hel}$ has three extensive sites of protein-protein interactions in crystal packing. The topside of β-barrel (blue molecule) interacts with the N-terminal end as well as the side of RecA1 domain (yellow molecule). The bottom of RecA2 domain of the third molecule (magenta) has a contact to the backside of β-barrel (blue) and the side of RecA1 domain (yellow molecule). All together the three molecules "wrap" around the 1C subdomain (yellow molecule). In order to disrupt the contact sites, a mutation in the 1C subdomain (L1549D and G1769R) as well as a N-terminus truncation of 10 amino acids was introduced into new constructs.

3.3 The structure of Sen1$_{Hel}$

The model of Sen1$_{Hel}$ has only one molecule in the asymmetric unit. The majority of the cloned construct could be modeled with exception of a disordered region in the subdomain 1C (residues 1471–1543), the C-terminus (residues 1876–1904) and a loop region in the RecA1 domain (residues 1382–1395). Also, the regions encompassing residues 1705–1713 and 1799–1801 have very poor electron density and could not be modeled. The model includes one ADP and one magnesium ion (Figure 3-9). Overall, Sen1$_{Hel}$ has a domain organization similar to that of the related helicases Upf1 (Upf1$_{Hel}$, also known as Upf1-ΔCH) (Cheng *et al*, 2007; Clerici *et al*, 2009; Chakrabarti *et al*, 2011) and IGHMBP2 (IGHMBP2$_{Hel}$) (Lim *et al*, 2012).

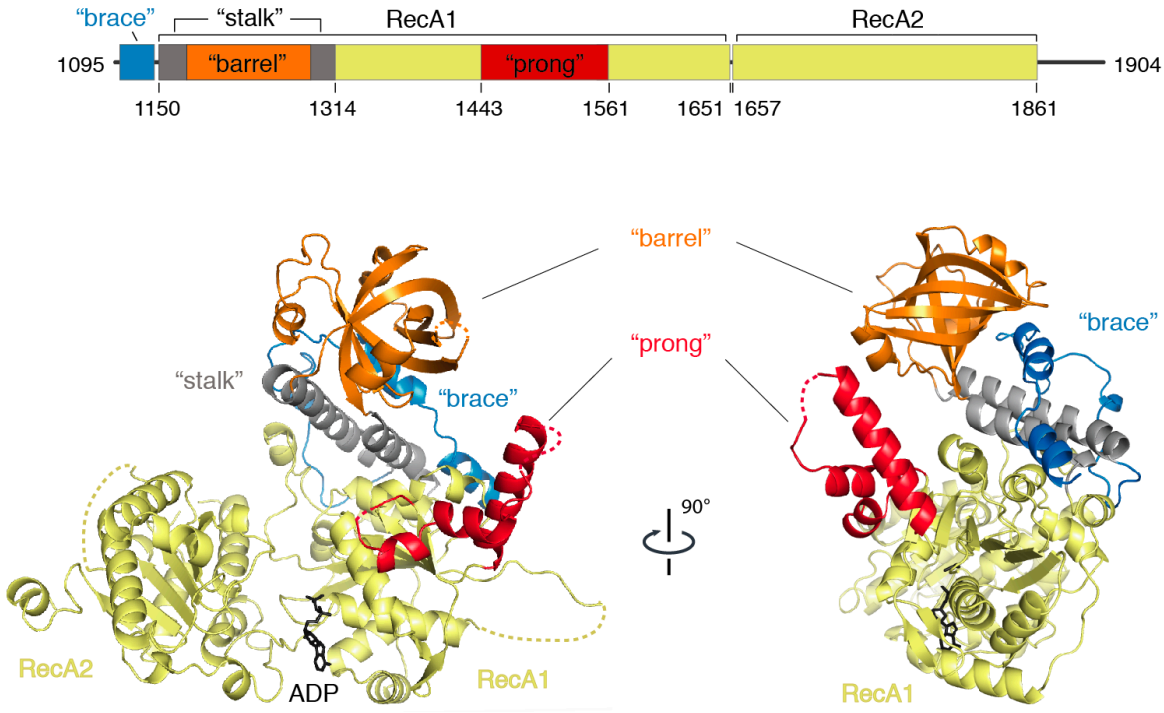


Figure 3-9. Crystal structure of yeast Sen1$_{Hel}$.
Top: schematic representation of the domain organization of the Sen$_{Hel}$. Bottom: the front and the side views of the Sen$_{Hel}$ structure. The core domain is formed of two RecA domains (yellow). The "prong" (red) and the "stalk" with the "barrel" on top of it are protruding from RecA1. The "brace" is wrapping around the "stalk" and is sandwiched between the "stalk" and the "barrel".

3.3.1 Architecture of Sen1$_{Hel}$

3.3.1.1 Two central RecA domains

The helicase core of Sen1 is composed of two small globular RecA domains (residues 1314-1651 and 1657-1861) that pack against each other forming a 10 Å-wide wedge-like cleft, and are connected with a short linker of five residues. RecA2 is rotated of about 30° from the position that is typical of helicases in an active RNA-ATP-bound conformation (Figure 3-14) (Pyle, 2008). Both domains have a similar topology of a central parallel β-sheet surrounded by eight α-helices. The loops between β-strands contain characteristic motifs for DNA/RNA and ATP binding (Figure 3-10). Tight coordination of the motif binding to nucleotides and RecA domain rotation upon ATP hydrolysis is a key for translocation on nucleic acid (Saikrishnan *et al*, 2009).

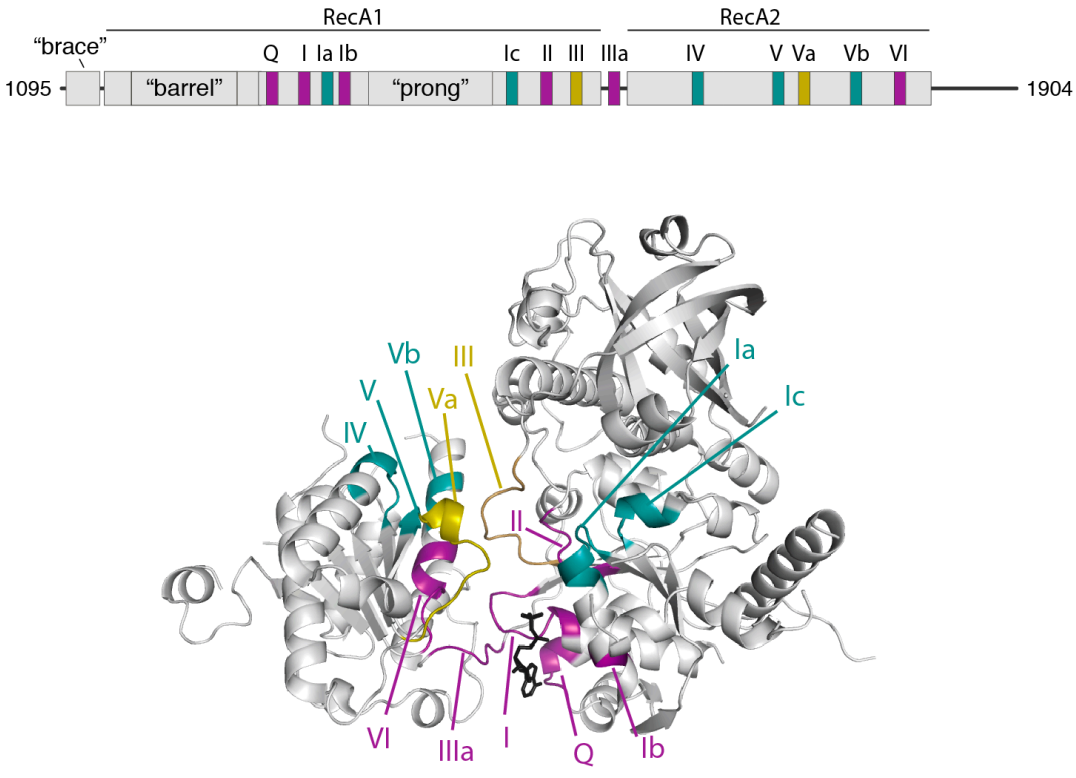


Figure 3-10. Conserved motifs for nucleic acid and ATP binding mapped on Sen1$_{Hel}$ structure. The characteristic sequence motifs are involved in ATP binding and hydrolysis (Q, I, Ib, II, IIIa, and VI, shown in magenta), in DNA/RNA binding (Ia, Ic, IV, V, and Vb, shown in turquoise) and coordination between nucleic acid and ATP binding (III and Va, shown in dark yellow). Importantly, Motif IIIa is a part of the linker between two RecA domains and is found only in SF1 helicases.

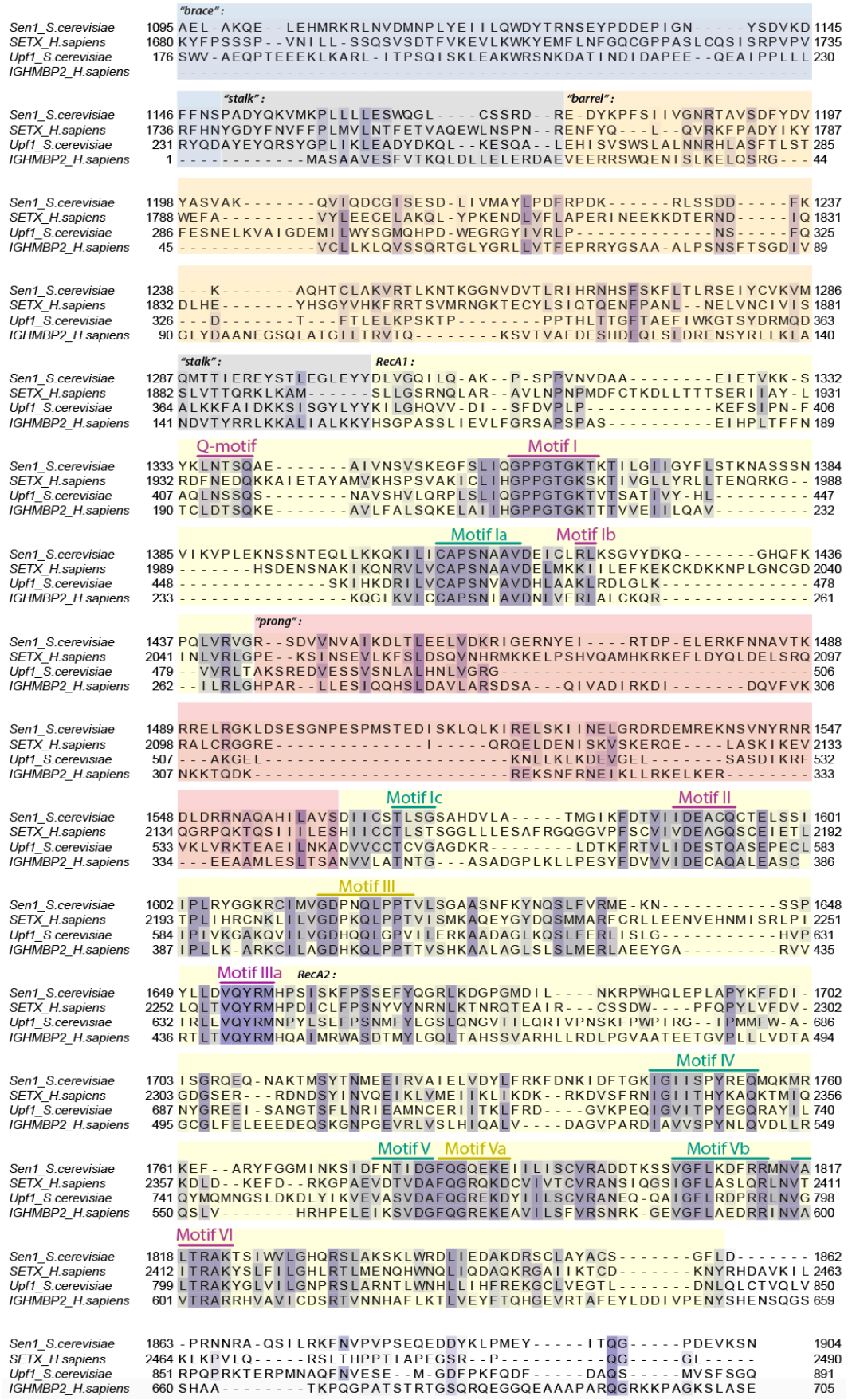


Figure 3-11. Multiple sequence alignment of Upf1-like helicases

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

3.3.1.2 ADP binding

ADP is bound to the RecA1 domain inside the cleft between RecA1 and RecA2 (Figure 3-9 and Figure 3-10). The adenine ring is sandwiched between an apolar surface of Lys1365 of RecA1 and Tyr1655 of the linker. This tyrosine is conserved among Upf1-like helicases, it corresponds to $Tyr638_{Upf1}$ and $Tyr422_{IGHMBP2}$ (see alignment Figure 3-11), and is a part of the Motif IIIa (Gln-Tyr-Arg-Met), which is present only in SF1 helicases but not in SF2 (Fairman-Williams *et al*, 2010). In SF2 helicases, e.g. in yeast Ski2 (Halbach *et al*, 2012), the adenine ring is also sandwiched between the two RecA domains by Phe328 and Arg767 (part of Motif IVa that is found only in SF2 helicases), however the linker to the nucleotide binding is not involved (Figure 3-13). In addition, the adenine amino group is recognized by conserved Gln1339 side chain (corresponding to $Gln413_{Upf1}$, $Gln196_{IMGMBP2}$ and $Gln334_{Ski2}$) of the Q-motif. This Gln-based specificity is common to all SF1 and SF2 helicases that have substrate preference to ATP over other nucleotides (Cordin *et al*, 2004). Gly1362, Lys1363, Thr1364 and Arg1422 coordinate α- and β-phosphates of the ADP (Figure 3-13).

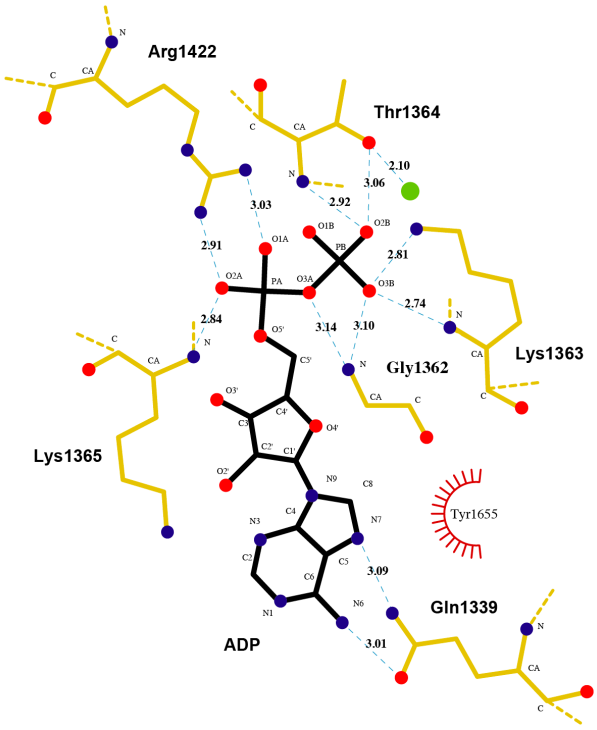
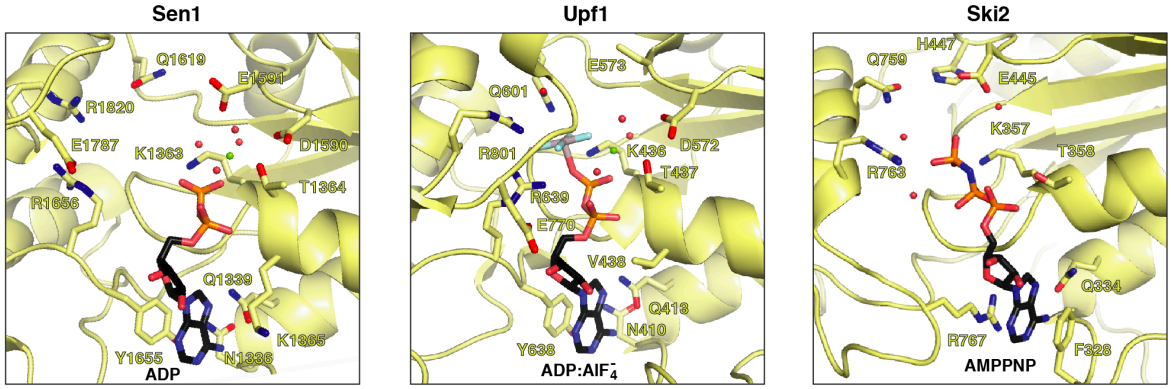


Figure 3-12. $Sen1_{Hel}$ ADP binding plot.
A two-dimensional representation of the residues involved in ADP binding by $Sen1_{Hel}$. ADP is shown in black, the residues are shown in yellow. The spheres in red represent oxygen, in blue – nitrogen, in green – magnesium ion. Blue dashed lines indicate hydrogen bonds, and numbers indicate bond length in Å. The adenine ring is sandwiched between an apolar surface of Lys1356 and an aromatic ring of Tyr1655. The ADP binding plot was calculated using LigPlot.

Front view:



Bottom view:

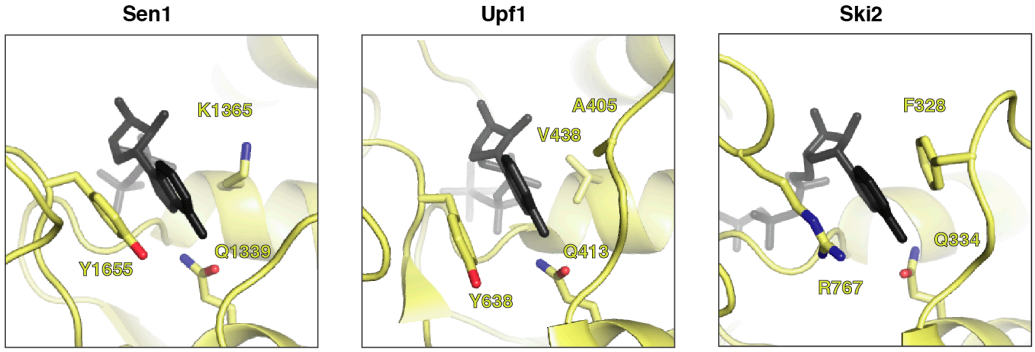


Figure 3-13. Comparison of nucleotide binding by Sen1, Upf1 (SF1) and Ski2 (SF2).
Residues involved in nucleotide binding are shown as sticks. The RNA helicases belong to different superfamilies and yet have many conserved residues. The adenine amino group is recognized by $Gln1339_{Sen1}$/$Gln413_{Upf1}$/$Gln334_{Ski2}$ (Q-motif). The α-phosphate of the nucleotide is coordinated by $Gly1362_{Sen1}$/$Gly435_{Upf1}$/$Gly356_{Ski2}$ (Motif I) and β-phosphate by $Lys1363_{Sen1}$/$Lys436_{Upf1}$/$Lys357_{Ski2}$ and $Thr1634_{Sen1}$/$Thr437_{Upf1}$/$Thr358_{Ski2}$. At this site in Sen1 and Upf1 structures is also to see bound magnesium (green sphere). Waters are shown in red spheres. The γ-phosphate is coordinated by $Arg1820_{Sen1}$/$Arg801_{Upf1}$/$Arg763_{Ski2}$ (Motif VI). When a helicase forms a tertiary complex with ATP and RNA, the side chain of $Gln1619_{Sen1}$/$Gln601_{Upf1}$ (Motif III) also binds to the γ-phosphate.

3.3.1.3 Accessory subdomains: the "stalk" and the "barrel" (1B)

Additionally to two RecA domains, Sen1 also has two subdomains, 1B and 1C, characteristic to Upf1-like helicases protruding from the RecA1. Subdomain 1B is the 160-residue long region of two antiparallel helices that form a so-called "stalk" and a 6-stranded β-barrel (referred to as the "barrel"). The helices of the "stalk" pack against each other and against the side of RecA1 with extensive hydrophobic interactions and are rather in a fixed position alike to the "barrel", which is connected to the "stalk" with short linkers, and hovers over RecA1. In comparison to $UPF1_{Hel}$ and $IGHMBP2_{Hel}$ structures, the "barrel" of $Sen1_{Hel}$ has more elaborated topology and shorter loop regions. But perhaps

more importantly, the "barrel" in apo-structure of Sen1$_{Hel}$ is fixed to the top of the "stalk" whereas in the other helicases the "barrel" is positioned on the top of RecA domains where the RNA would reside (Figure 3-14).

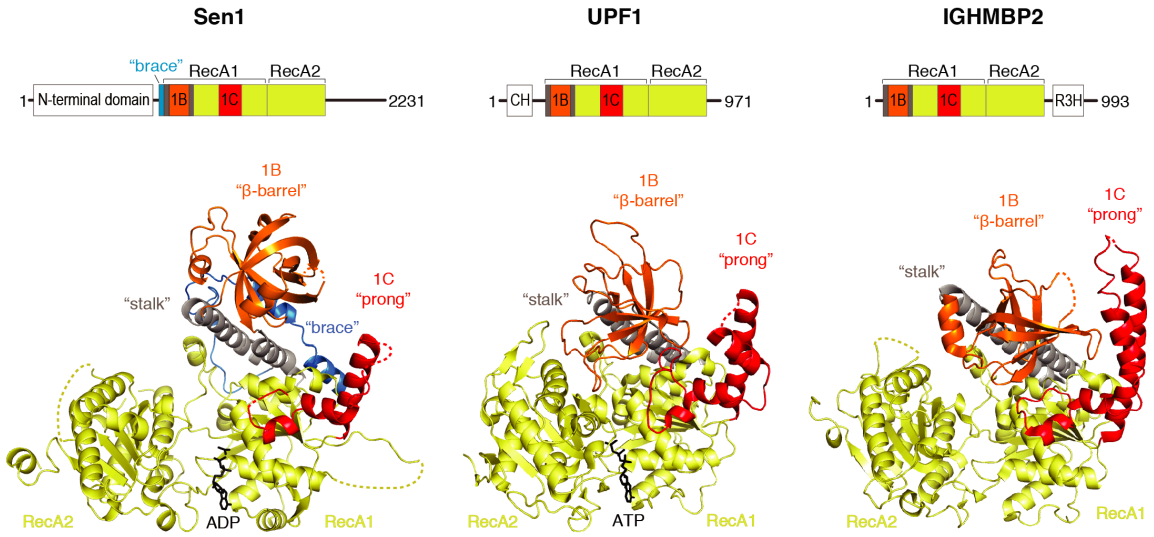


Figure 3-14. Structural comparison of Sen1$_{Hel}$ with related helicases.
The apo-structures of the helicase core of yeast Sen1, human Upf1 (PDB: 2GJK) and IGHMBP2 (PDB: 4B3F). The structures are shown in the orientation of superposed RecA1 domains. Dashed lines indicate missing parts of the structure. The schematic representations of full-length helicase domain organization are shown on top, with structured domains shown as rectangles and unstructured regions as a black line. Domains in scheme and the structure are highlighted in the same colors. Among Upf1-like helicases only Sen1 has an additional fragment to the helicase core, which braces the "barrel" and the "stalk" together, and therefore is referred to as the "brace".

3.3.1.4 Accessory subdomains: the "brace" is unique to Sen1

Structural comparison of Sen1$_{Hel}$ with Upf1$_{Hel}$ and IGHMBP2$_{Hel}$ revealed key differences in subdomain 1B. First of all, the "barrel" has a short 3-residue linker to the ascending helix of the "stalk" in contrast to a 10-residue linker observed in Upf1$_{Hel}$ or to a helix in IGHMBP2$_{Hel}$. Similarly, the "barrel" is directly connected to the descending helix of the "stalk" in Sen1$_{Hel}$, whereas it has a short linker in case of Upf1$_{Hel}$ and IGHMBP2$_{Hel}$. Thus, in contrast to other Upf1-like helicases, the "barrel" of Sen1 seems to be spatially restricted. In addition, Sen1 has a unique N-terminal segment (residues 1096-1149) that we refer to as the "brace" (Figure 3-14). The "brace" fastens together the RecA1, the "stalk" and the "barrel" with extensive hydrophobic interactions.

The "brace" starts with a short helix (α1) that inserts via aliphatic side chains (Leu1109 and Arg1108) into a hydrophobic surface groove formed between the RecA1 domain (Ala1573, Ala1578 and Tyr1606) and the descending helix of the "stalk" (Tyr1303) (Figure 3-15). The polypeptide chain then continues with a second helix (α2) sandwiched between the "barrel" and both helices of the "stalk". Here, hydrophobic residues on one side (Tyr1117 and Leu1121) are engaged in van der Waals interactions with aliphatic side chains of the "barrel" (with Leu1216, Leu1244 and Lys1246). Hydrophobic residues on the other side (Leu1116, Ile1120 and Trp1123) make apolar interactions with residues on the ascending and descending helices of the "stalk" (Ile1291 and Leu1162, Trp1166, and Leu1169, respectively). 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 3-15).

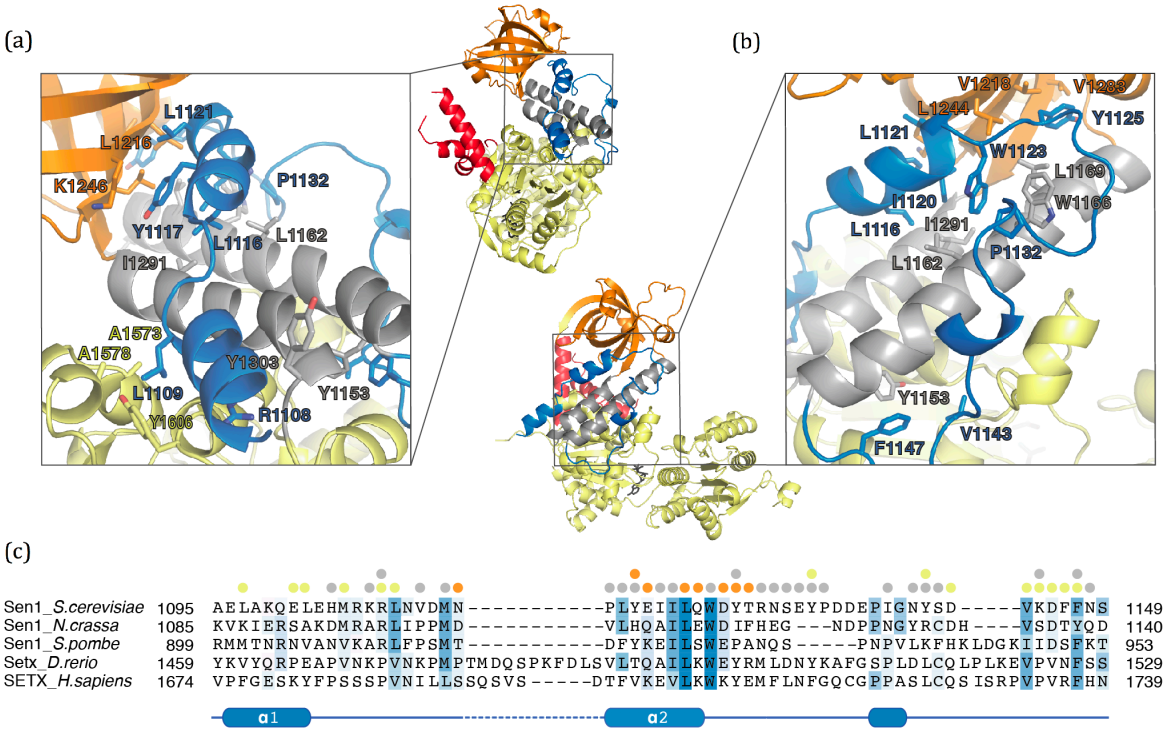


Figure 3-15. Evolutionary conserved interactions of the "brace".
(a) and (b) are the zoom-in views of the $Sen1_{Hel}$ model , shown in a 90° and 180° clockwise rotation around vertical axis of the view in Figure 3-9. Selected residues are highlighted as sticks and show the extensive hydrophobic interactions between the "brace" and RecA1, the "stalk" or the "barrel". (c) Structure-based sequence alignment of the "brace" of Sen1 and its orthologs. The conserved amino acids are highlighted in blue. The circles (yellow for RecA1, grey for the "stalk"", and orange for the "barrel") show with which domain a certain residue is interacting.

The deletions of the "brace" (constructs 1128-1904, 1136-1094, 1149-1904, and 1178-1880) resulted in insoluble protein, most likely because upon removal of the "brace" the above-mentioned hydrophobic residues become exposed to the solvent. Consistently, also the proteins of point mutants where interaction between the "brace" and the "barrel" were disrupted (L1121R, W1123R, and Y1117R) were insoluble. Thus, it seems that the "brace" is a distinct structural feature that stabilizes the overall fold of $Sen1_{Hel}$ and enables to promote transcription termination.

3.3.1.5 Accessory subdomains: the "prong" (1C)

The subdomain 1C is the second structural feature characteristic to Upf1-like helicases (Figure 3-14). In $Sen1_{Hel}$ this 120-residue prong-like protrusion from the RecA1 domain (residues 1443-1563) is composed of an α-helix almost parallel to the "stalk", two short α-helices at the base and a 75-residue region in between that could not be modeled (Figure 3-9), although this region is predicted (PSIPRED) to form α-helices as well. In contrast, the "prong" subdomains in Upf1 and IGHMBP2 are 65 and 80-residue long, respectively. Although the overall structures of the three "prong" subdomains are very similar, their sequences are poorly conserved with the exception of a few residues that are important for RNA binding (see multiple sequence alignment Figure 3-11).

3.3.2 RNA-binding

Attempts to get a structure of $Sen1_{Hel}$ in complex with RNA failed and the following studies were based on the comparison of the available structures of $Sen1_{Hel}$-ADP with $Upf1_{CH\text{-}Hel}$-ADP:AlF_4^--RNA (Chakrabarti *et al*, 2011) and $IGHMBP2_{Hel}$-ADP-RNA (Lim *et al*, 2012). All three proteins share conserved RNA binding motifs that enable the helicase to bind or translocate along the RNA. Particularly interesting is the comparison of RNA binding by Motif Ia and Motif III in "open" and "close" conformations of the helicases (Figure 3-16).

Motif III ($GDPNQLPPT_{Sen1}$, $GDHQQLGPV_{Upf1}$, $GDHKQLPPT_{IGHMBP2}$) together with Motif Vb coordinates RNA binding and ATPase activity (Jankowsky & Fairman, 2007). In Sen1, the GDPNQ fragment (residues 1615-1619) stacks on top of Motif I with

Gln1619 (corresponding to $Gln597_{Upf1}$ and $Gln399_{IGHMBP2}$) at the position where γ-phosphate of ATP would reside. At the other end of Motif III, the hydrophobic Leu1620 and non-polar Thr1624 side chains stack with the interaction site between Motif Va, Motif V and the RNA backbone.

Motif Ia ($CAPSNAAV_{Sen1}$, $CAPSNVAV_{Upf1}$, $CAPSNIAV_{IGHMBP2}$) is particularly characteristic to both DNA and RNA helicases of SF1 that translocate in 3'→5' direction. As there are more structures of DNA helicases available, the mechanism of how the helicases translocate on DNA is better understood. Here, structural studies show, that during translocation the base at position 5 flips in and out of Motif Ia pocket and stacks to Val/Thr of Motif III (Saikrishnan *et al*, 2009) (Figure 3-16, bottom right panel). In the available structures of RNA helicases such base flipping has not been observed although the same conserved residues are involve in RNA binding. However, in an "open" conformation there is enough space between Motif Ia and Motif III to fit two nucleotides whereas in a "close" conformation - only one nucleotide could fit. Probably, the main structural difference between the structures of DNA and RNA helicase of SF1 is the presence of accessory subdomains on RecA1. In DNA helicases, the N-terminal domain is pushing ssDNA to topside of RecA domains, whereas in RNA helicases the "stalk" and the "barrel" are pushing the RNA towards the "prong".

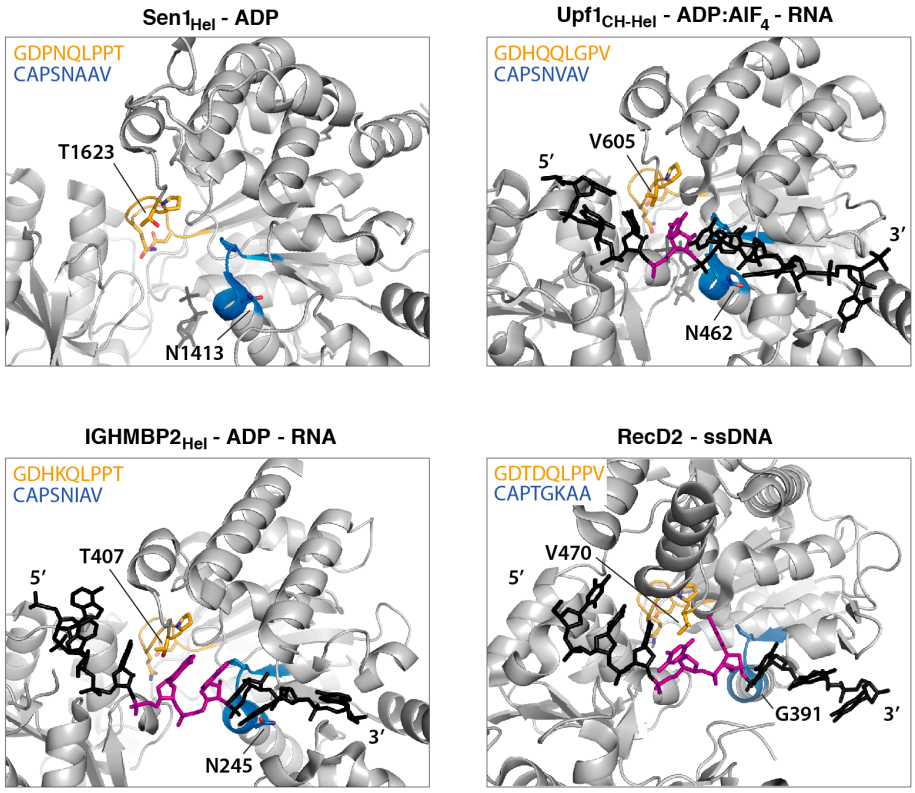


Figure 3-16. RNA binding by Motifs Ia and III.
Comparison of top views of nucleic acid binding sites of Sen1, IGHMBP2 (PDB: 4B3G) and Upf1 (PDB: 2XZL) and DNA helicase RecD2 (PDB: 3GP8). Motif Ia is shown in blue and Motif III is shown in orange. The nucleotides bound between the two motifs are highlighted in magenta. The tertiary helicase structure changes upon DNA/RNA-dependent ATP hydrolysis, the helicase is in an "open" conformation when bound ADP or a "close" conformation when bound to an ADP:AlF_4 (right top panel).

$Sen1_{Hel}$ shares the conserved residues that interact with RNA (Figure 3-17). RecA2 residues ($Tyr1752_{Sen1}$ and $Arg1813_{Sen1}$ corresponding to yeast $Tyr732_{Upf1}$ and $Arg794_{Upf1}$) interact with nucleotides 1 and 2 at the 5'-end of the RNA, whereas RecA1 residues ($Pro1622_{Sen1}$, $Thr1623_{Sen1}$, $Asn1413_{Sen1}$, and $Thr1289_{Sen1}$ corresponding to $Pro604_{Upf1}$, $Val605_{Upf1}$, $Asn462_{Upf1}$, and $Thr356_{Upf1}$) approach the central portion of the RNA. Finally, $Sen1_{Hel}$ also shares conserved residues in the "stalk" ($Thr1289_{Sen1}$, $Arg1293_{Sen1}$ corresponding to $Thr356_{Upf1}$, $Arg360_{Upf1}$) and in the "prong" ($Arg1552_{Sen1}$ corresponding to $Arg537_{Upf1}$) that interact with ribonucleotides 6 and 7, respectively, at the 3'-end of the RNA. Thus unsurprisingly, a double mutant T1289A, R1293A of $Sen1_{Hel}$ was impaired in RNA binding, ATPase activity and transcription termination *in vitro* (Figure 3-18).

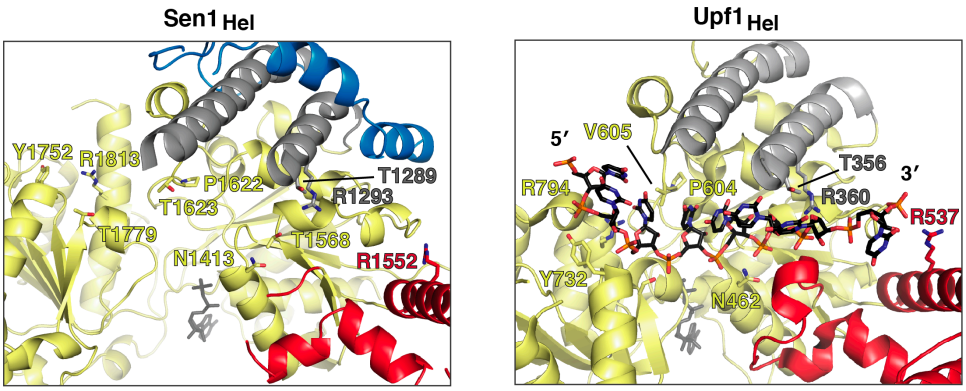


Figure 3-17. Comparison of the RNA-binding sites of Sen1 and Upf1.
Most of the residues of the RNA-binding site of Sen1 and Upf1 (PDB: 2XZO) are conserved.

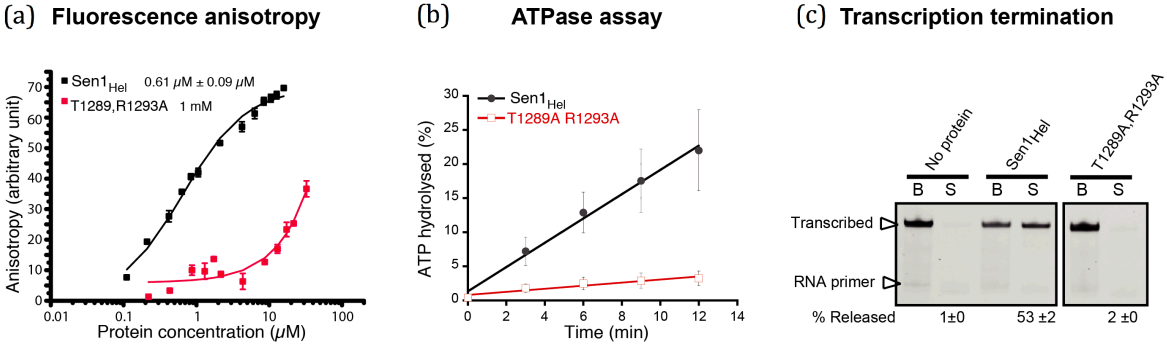


Figure 3-18. Sen1$_{Hel}$ double mutant T1289A, R1293A has lost its ability to bind RNA.
(a) RNA binding measurements by florescence anisotropy with fluorescein labeled AU-rich RNA. Curves represent the average of three independent experiments. The double mutation T1289A, R1293A of the residues of the "stalk" drastically decreases protein affinity to RNA. The measurements were done by Dr. Claire Basquin. (b) ATP hydrolysis assay. Values correspond to the average of three independent measurements. Sen1 is an RNA-dependent ATPase and impaired RNA binding leads to loss of ATPase activity. The measurements were done by Dr. Zhong Han. (c) *In vitro* transcription termination assay. The double mutant T1289A, R1293A is inactive for transcription termination. The assay was done by Dr. Zhong Han.

3.3.3 The "brace" pre-positions the "barrel" for RNA binding

One of most the prominent differences between structures of Sen1 and Upf1 helicases is that the latter has an additional CH domain outside the helicase core (referred to as Upf1$_{CH\text{-}Hel}$). The CH domain regulates helicase activity in an allosteric manner. It binds onto RecA2 and pushes the "barrel" towards the "prong" (Figure 3-19), this extends RNA binding by two nucleotides at the 3'-end of the RNA (Chakrabarti *et al*, 2011; Chamieh *et al*, 2008). Similarly, IGHMBP2 contains an additional R3H domain, which also has been shown to modulate RNA binding and ATPase activity (Lim *et al*, 2012). No analogous

allosteric regulation domain is present in Sen1 but, surprisingly, in RNase protection assays (see Methods 6.2.14) $Sen1_{Hel}$ protected the 11-ribonucleotide long fragments, more similar to $Upf1_{CH\text{-}Hel}$ than to $Upf1_{Hel}$ (Figure 3-19). In fact, $Sen1_{Hel}$ has an intrinsic feature, the "brace", which might mimic the effect of the CH domain of $Upf1_{CH\text{-}Hel}$ on the RNA-binding footprint by pre-positioning the "barrel" of $Sen1_{Hel}$ for RNA binding.

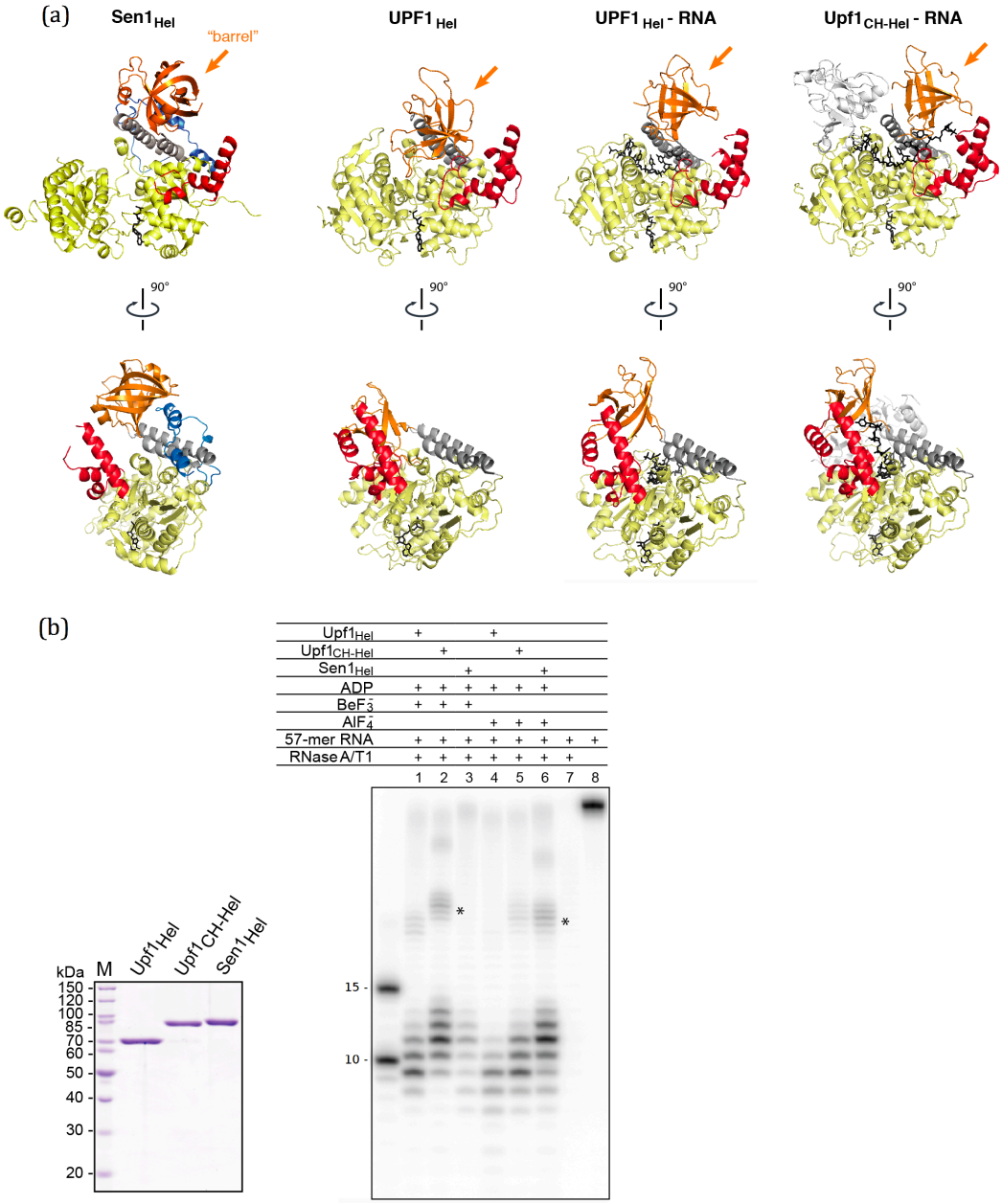


Figure 3-19. Comparison of the "barrel" positioning in Sen1 and Upf1.
(a) Comparison of the structures of yeast $Sen1_{Hel}$, human $UPF1_{Hel}$ (PDB: 2GJK) and $UPF1_{Hel}$-RNA (PDB: 2XZO), and yeast $Upf1_{CH\text{-}Hel}$-RNA (PDB: 2XZL). The molecules are shown in a front view (top) and their 90° clockwise rotations around a vertical axis (bottom). (b) RNase protection assay with yeast $Sen1_{Hel}$, $Upf1_{Hel}$, and $Upf1_{CH\text{-}Hel}$ in the presence of ADP:BeF_x or ADP:AlF_x. The assay was done by Fabien Bonneau. The asterisk (*) marks fragments protected by helicases continuously. On the left panel, a Coomassie stained SDS-PAGE gel of the proteins used in the assay.

In order to test this hypothesis, two "prong" mutants in $Sen1_{Hel}$ were engineered, cloned, expressed and purified (see Figure 3-20). In the first mutant, the part of the sequence in the "prong" that could not be modeled (residues 1471–1538) was replaced with a linker Gly-Ser-Gly-Ser ($Sen1_{Hel}$ΔUP for *u*pper "*p*rong" deletion). In the second mutant, the entire solvent-exposed portion (residues 1461-1554) was replaced with a linker Ser-Gly-Gly ($Sen1_{Hel}$ΔLP for *l*ower "*p*rong" deletion). 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 shorter protection of ~8-9-ribonucleotide fragments (Figure 3-20).

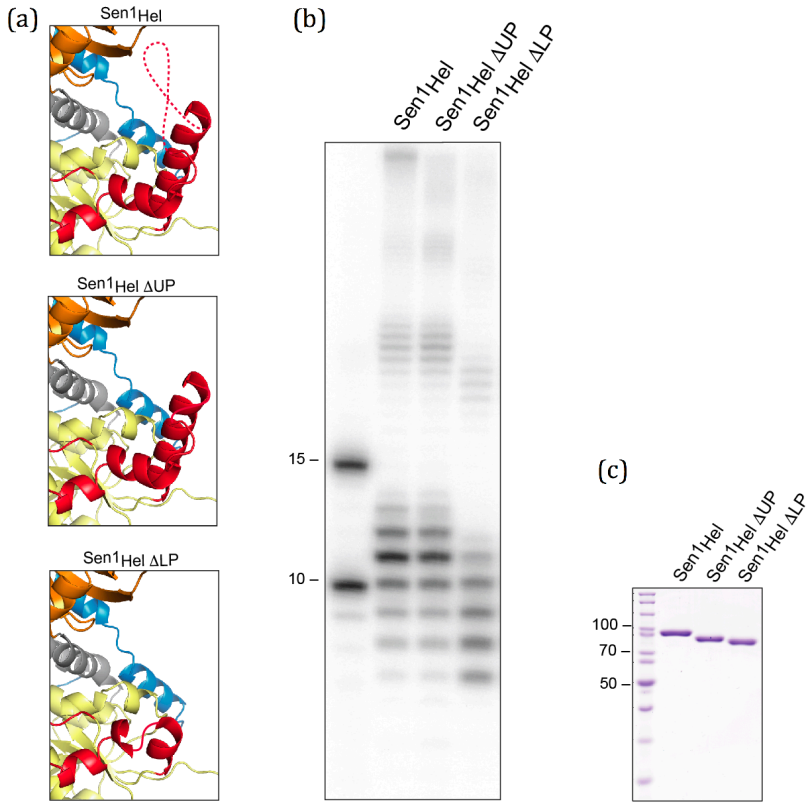


Figure 3-20. Deletions reveal contribution of the "prong" to RNA binding.
(a) The structures show the deletions of the "prong". The dashed line represents a region of the upper part of the prong that was disordered in the crystal structure. The two structures depicted below, represent models of the two deletion constructs, $Sen1_{Hel}$ΔUP and $Sen1_{Hel}$ΔUP, used in the assay. (b) RNase protection assay with yeast $Sen1_{Hel}$, $Sen1_{Hel}$ΔUP, and $Sen1_{Hel}$ΔLP in the presence of ADP:AlF_x. RNA fragments were obtained by digesting ^{32}P body-labeled 57-mer $(CU)_{28}C$ RNA with RNases A and T1. The assay was done by Fabien Bonneau. (c) A Coomassie stained SDS-PAGE gel of the proteins used in the assay.

3.3.4 The "prong" is critical for RNA unwinding and transcription termination

One function of Sen1 in the cell is processive RNA unwinding. The helicase translocates on ssRNA in 3'→5' direction and removes secondary structure of RNA, DNA from RNA:DNA duplex, or RNA-bound proteins (Jankowsky & Fairman, 2007). It is reasonable to expect the unwinding element to be located at the RNA entrance to the helicase, i.e., RecA1 domain. A close inspection of the "prong" of Sen1 and related helicases points out two residues.

Firstly, $Sen1_{Hel}$ shares a conserved residue $Arg1552_{Sen1}$ (corresponding to $Arg537_{Upf1}$ and $Lys331_{IGHMBP2}$), which in the $Upf1_{CH\text{-}Hel}$ structure is located near the last nucleotide of the RNA (Chakrabarti *et al*, 2011). Secondly, $Sen1_{Hel}$ has a hydrophobic residue, Leu1549, which is involved in protein-protein interaction in crystal packing as mentioned above (see Chapter 3.2.4). The residue is not only exposed to the surface, but also is in a close proximity to where the RNA is entering and the helicase might have a protein-protein interaction with encountered RNA-bound protein (i.e., RNA polymerase). Therefore, in collaboration with Dr. Zhong Han and Dr. Odil Porrua, we checked the effect of the "prong" deletions, $Sen1_{Hel}\Delta UP$ and $Sen1_{Hel}\Delta LP$, as well as L1549D mutation for the helicase activity.

All three mutants retained RNA-dependent ATPase activity (Figure 3-21, a), however, the full "prong" deletion in $Sen1_{Hel}\Delta LP$ abolished the unwinding activity (Figure 3-21, b). This was not due to a decrease in the affinity for the RNA, since we observed similar RNA binding by $Sen1_{Hel}\Delta LP$ compared to $Sen1_{Hel}$ (Figure 3-20). In contrast, the deletion of the disordered part of the "prong" in $Sen1_{Hel}\Delta UP$ had the lowest ATPase activity but showed the highest capacity to dissociate the duplex. Interestingly, the higher the ATPase activity that was measured, the less active was the mutant in RNA:DNA unwinding, suggesting that the processivity of the helicase is not directly dependent on ATP hydrolysis rate.

More intriguingly, $Sen1_{Hel}\Delta UP$ mutant, which was the most active in unwinding assays, exhibited a moderate decrease in termination efficiency *in vitro*. (Figure 3-21, c). In addition, the double $Sen1_{Hel}\Delta UP$,L1549D mutant exhibited levels of duplex unwinding activity similar to the wild-type protein but was strongly affected in transcription termination. In contrast, $Sen1_{Hel}\Delta LP$ mutant was inactive for both duplex unwinding and

transcription termination. Consistently, the full "prong" deletion leads to lethality and transcription termination defects *in vivo* (Leonaitė *et al*, 2017). Thus, for the first time we could show that both translocation and RNA unwinding are required for Sen1 to terminate Pol II transcription.

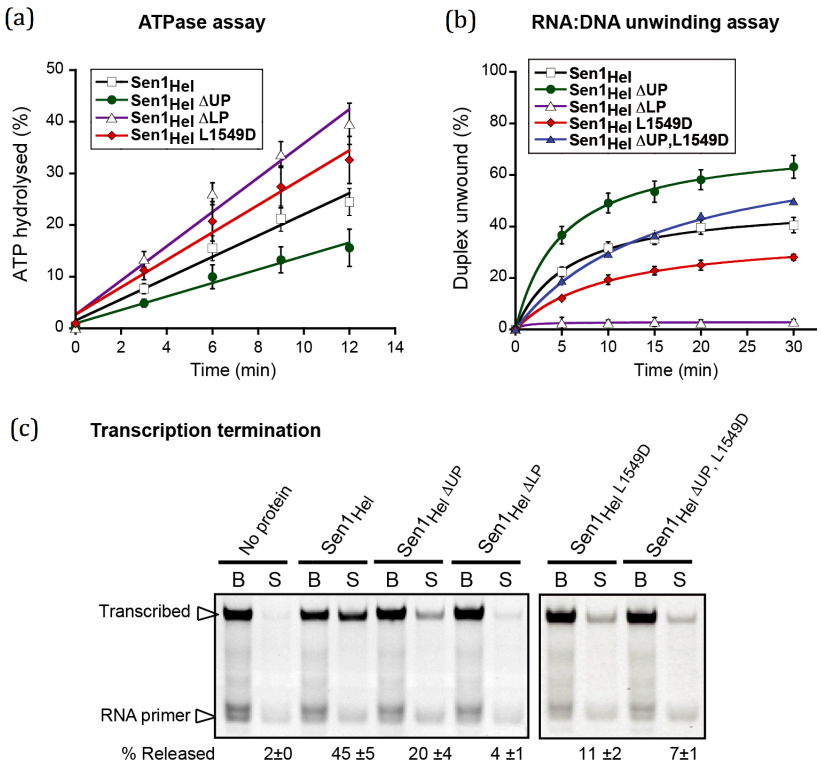


Figure 3-21. "Prong" in transcription termination.
(a) ATP hydrolysis assay. Values correspond to the average of three independent measurements. (b) RNA:DNA duplex unwinding assay. (c) *In vitro* transcription termination assay. The assays were done by Dr. Zhong Han.

3.4 Sen1$_{Hel}$ as a model for AOA2-associate mutations in SETX

Based on secondary structure prediction and sequence alignment, several constructs of the helicase domain (fragments 1671-2478, 1671-2488, 1694-2478, and 1694-2488) of human SETX were cloned for expression in *E. coli*, as described in Methods 6.2.1. The designed constructs were similar length to crystallized Sen1$_{Hel}$, however the proteins were not expressed.

The helicase domain of SETX shares about 30 % sequence identity with $Sen1_{Hel}$, including the key residues of the "brace" that are absent in other SF1 helicases. Thus, the disease-associated mutations in *SETX* gene were tested using $Sen1_{Hel}$ as a surrogate. In total, 26 missense AOA2-associated mutations (reported in LOVD Database) could be mapped onto $Sen1_{Hel}$ structure (Figure 3-22). The majority of the mutated residues are buried inside the helicase core, meaning that the substitution disrupt the protein fold and hinders the helicase activity. Other mutations affect SETX catalytic activity directly, as the mutated side chains are in places important for RNA recognition or ATP hydrolysis. Only a few residues were mapped at the surface of the helicase, however, those residues were not conserved.

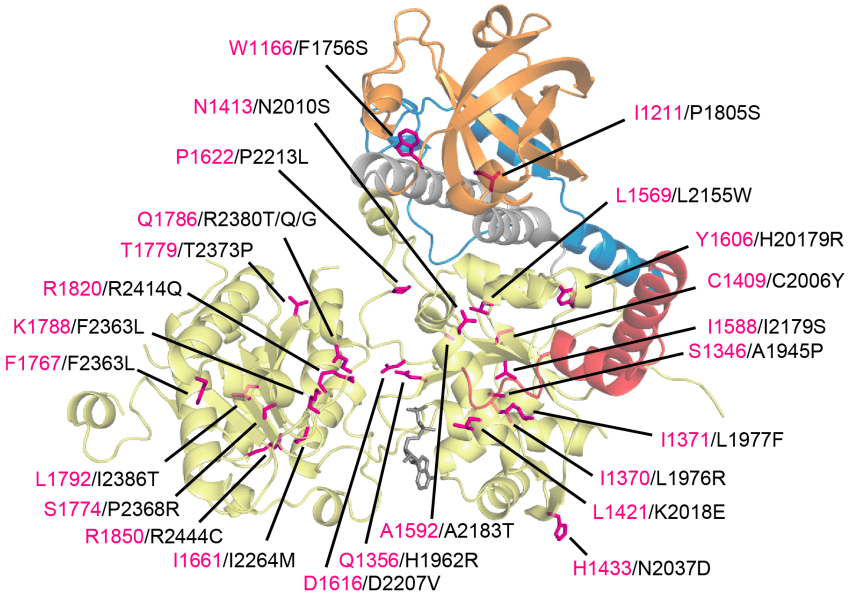


Figure 3-22. AOA2-associated missense mutations.
Conserved AOA2-associated mutations mapped onto the $Sen1_{Hel}$ structure. The residues that represent SETX missense mutations (black labels) are shown as sticks in magenta. All known SETX mutations are reported in LOVD Database.

A subset of AOA2-associated substitutions was cloned into $Sen1_{Hel}$, the helicase was purified and tested for its enzymatic activity (Figure 3-23). A D1616V mutation of the residue buried at the interface between the two RecA domains resulted in an insoluble protein. The N1413S, P1622L and T1779P mutations targeted the residues that were expected to have a direct contact with RNA. These mutants were impaired in RNA

binding (FA measurements, Figure 3-23), and, consequently, did not elicit duplex unwinding and *in vitro* transcription termination. Finally, the R1820Q mutant that supposedly has a direct contact to γ-phosphate of the ATP and the substitution of which would impair ATP hydrolysis, was inactive in all assays.

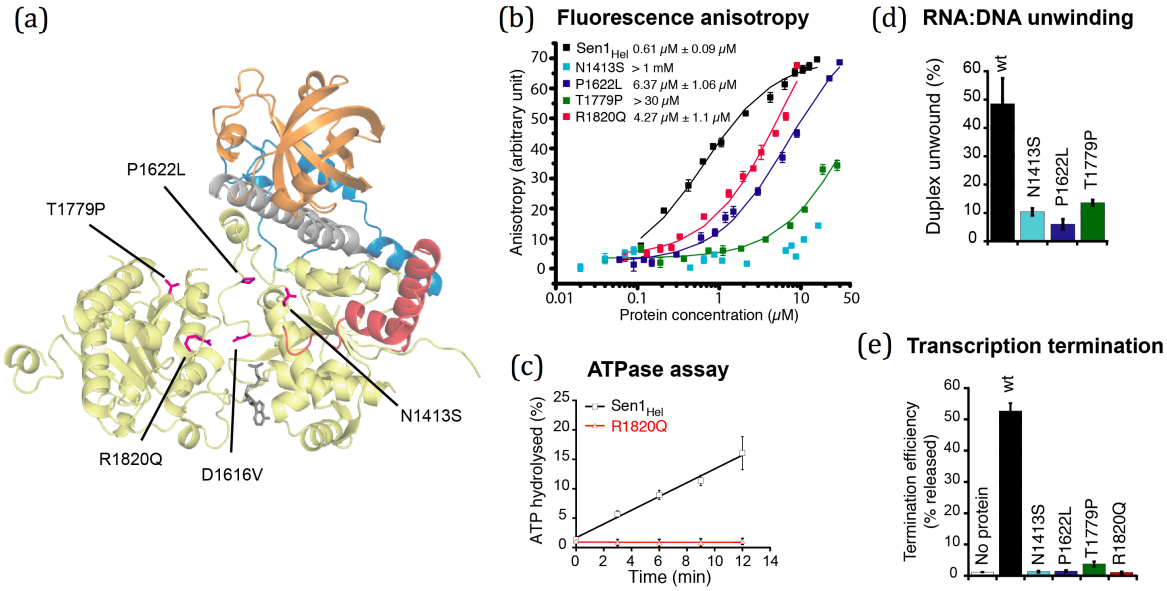


Figure 3-23. Characterization of the Sen1$_{Hel}$ harboring AOA2-associated mutations.
(a) Selected Sen1$_{Hel}$ point mutations that mimic the disease mutations of SETX mapped in Sen1$_{Hel}$ structure. (b) Comparison of RNA binding affinity of the mutants measured by fluorescence anisotropy using fluorescently labeled AU-rich RNA as the substrate. (c) ATPase activity analysis of the R1820Q mutant. Arg1820 is highly conserved and coordinates γ-phosphate of the ATP. The point mutation was predicted to affect nucleotide binding. (d) The mutant activity in RNA:DNA duplex unwinding assays. The efficiency of the unwinding measured at 30 min. (e) Analysis of mutation impact on efficiency of transcription termination. All values correspond to the average and SD of three independent measurements. The assays were done by Dr. Zhong Han.

4 Discussion and conclusions

4.1 $Sen1_{Hel}$ structure is silimilar to other Upf1-like RNA helicases of SF1

The RNA binding and unwinding activities of Sen1 require the RecA1-RecA2 core domain as well as two distinct accessory subdomains that protrude from RecA1: the "stalk" with the "barrel" and the "prong". Overall, the domain organization of $Sen1_{Hel}$ is very similar to the helicase core of other RNA helicases of the SF1. Consistently, $Sen1_{Hel}$-ADP is in an "open" conformation, in which the RecA2 domain is rotated 30° outwards from the position observed in a "close" conformation of $Upf1_{Hel}$-ATP bound structure. This domain movement upon ATP binding and hydrolysis is important for helicases translocation on a nucleic acid chain.

In the $Sen1_{Hel}$ apo structure, ADP is bound to RecA1 inside the cleft. The adenine ring is sandwiched between RecA1 and Motif IIIa (see Figure 3-13). Motif IIIa is characteristic for Upf1-like helicases and resides in the linker that connects the two RecA domains. In SF2 helicases the adenine ring is sandwiched directly between RecA1 and RecA2 (Motif IVa) domains. In addition, the adenine amino group is recognized by a conserved Q-motif, common to all SF1 and SF2 helicases that have substrate preference for ATP over other nucleotides (Cordin *et al*, 2004). $Sen1_{Hel}$ also harbors Motif I and Motif II (also known as Walker A and Walker B) that are crucial components for ATP hydrolysis.

Importantly, $Sen1_{Hel}$ binds RNA in an ATP-dependent manner. Although in RNase assays ADP:AlF_x and ADP:BeF_x were bound to $Sen1_{Hel}$, no electron density for AlF_x or BeF_x moiety but for ADP was obtained. Moreover, the protein has crystallized in an "open" conformation that most likely leads to the reduced affinity for RNA binding. The structure of $Sen1_{Hel}$ bound to RNA was not obtained.

However, the RNA binding motifs are well conserved and superposition with the structure of $Upf1_{Hel}$-RNA suggests the key residues for RNA binding (see Figure 3-17). Most of the RNA contacts are made with sugar phosphate backbone at the top of RecA domains; therefore the helicases do not discriminate between specific nucleotides or "read" the sequence of RNA and can also bind ssDNA. Additionally, Upf1-like helicases have an extended RNA binding site on the RecA1 domain. The conserved residues of Upf1-like helicases at nearly identical positions on surfaces of the "stalk" and the "prong" bind

additional three to four nucleotides. The "barrel" positioning is very important for RNA binding at this site and.

4.2 The "brace" is unique to Sen1

Specialized domains gained during evolution serve helicases by regulating RNA binding. Additional to the accessory domains ("stalk", "barrel", and "prong"), Upf1 and IGHMBP2 have domains outside the helicase core that coordinate the "barrel" positioning to tune RNA-binding affinity and ATPase activity. When this project was started there was no indication that Sen1 had a similar allosteric regulation domain to modulate the positioning of the "barrel". However, the crystal structure revealed that Sen1 has a unique feature, the "brace", that has not been predicted as a part of the helicase domain before. It is a ~50-residue fragment (residues 1096-1149) located upstream of the helicase domain and is well conserved across species. From the structure prediction or sequence alignment it was not possible to predict the importance of the "brace". Although some residues, e.g. Leu1121 and W1123, are conserved in all organisms, homology-based structural modelling has not predicted these two residues to interact with the barrel. However, studies *in vivo* had demonstrated that the region 1089-1135 is important for yeast viability (Chen *et al.*, 2014), but the authors suggested that this region serves as a secondary nuclear localization sequence (NLS). The sequence motif MRKRL could indeed be assigned to Class 3 of NLS for the helicase import into nucleus (Kosugi *et al*, 2009).

In $Sen1_{Hel}$ structure, the "brace" fastens the RecA1 domain, the "stalk" and the "barrel" together, thus the "barrel" is in a rather fixed position. Also, in comparison to other helicases, the linkers between the "barrel" and the "stalk" of Sen1 are the shortest and the "barrel" movement would be limited even if there were no "brace" present. Moreover, results of the RNase protection assays confirm that in $Sen1_{Hel}$ the "barrel" is in a position similar to that of the $Upf1_{CH\text{-}Hel}$ structure (see Figure 3-19). All together, this suggests that the N-terminal domain is not regulating the helicase in a similar manner to the CH or R3H domains of Upf1 and IGHMBP2, respectively.

It is possible that the "barrel" conformation in the crystal structure of $Sen1_{Hel}$ is partially stabilized by crystal lattice contacts. However, the "brace" buries 2400 $Å^2$ of the $Sen1_{Hel}$

surface and is absolutely required. It has been shown previously that a single point mutant within the "brace", W1166S, is defective *in vivo* (Chen *et al*, 2014). Moreover, the deletion of N-terminal residues 1-1134 in Sen1 is lethal, whereas the deletion of residues 1-1088 does not have such a drastic effect (Chen *et al*, 2014).

4.3 Structural basis of 5'-3' translocation of Sen1 and related RNA helicases

Translocation and unwinding are one of the fundamental properties that define helicases. Structural studies of DNA helicases of SF1 revealed that the 5'→3' translocation directionality is achieved via alteration of the grip to DNA between RecA1 and RecA2 domains (Saikrishnan *et al*, 2009). During ATP binding and closing of the cleft, the helicase grips the ssDNA most tightly with RecA1 domain. During ATP hydrolysis and opening of the cleft, the tightest grip is with RecA2 domain and as the consequence the "relaxed" RecA1 domain is sliding towards 3'-end of the ssDNA.

The mechanism of how RNA helicases of SF1 translocate in 5'→3' direction is not known. However, Sen1 and the related helicases share conserved residues on the accessory subdomains that contribute to RNA binding. As it has previously been reported for Upf1 (Chakrabarti *et al*, 2011), Sen1 also achieves an extended RNA-binding through conserved residues on the accessory domains, particularly on the "prong". Another interesting aspect is that in transition between the "open" and "closed" conformations, the RecA2 domain rotates about 30°. In IGHMBP2$_{Hel}$-ADP-RNA (PDB: 4B3G), the cleft is "open" and the RNA is clearly twisted but relaxed on RecA2. In the case of Upf1$_{CH\text{-}Hel}$-ADP:AlF_4-RNA (PDB: 2XZL), the cleft is "closed" and the RNA is "squeezed" (Figure 3-16). Here, similarly to DNA helicases (Saikrishnan *et al*, 2009), the fifth nucleotide of the RNA chain is switching between Motif III and Ia. Unfortunately, the structure of both "open" and "close" conformations are not available, neither for Sen1, Upf1, nor IGHMBP2, and only a comparison between the helicases is possible. Moreover, in RNase protection assays, the footprint of ground and transition states of Sen1 and Upf1 is different (Figure 3-17). Nevertheless, Sen1$_{Hel}$ shares all the conserved residues of Motifs Ia, III and V, thus the mechanism of how Sen1 translocates in 5'→3' direction most plausibly is similar to that of DNA helicases of SF1.

4.4 Proposed mechanism of transcription termination by Sen1

In the NNS transcription termination pathway, the recognition of Nrd1 and Nab3 consensus motifs is a critical first step but Sen1 is a key enzyme in facilitating the elongation complex dissociation. It has continuously been demonstrated that Sen1 alone is sufficient for termination of the elongation complex. $Sen1_{Hel}$ is the smallest enzyme variant that has been studied in transcription termination *in vitro*. The N- and C-terminal domains of the helicase are limiting Sen1 functions only *in vivo* (Han *et al*, 2017), as they are needed for recruitment to termination sites. As mentioned, the CTD of Pol II directly or through Nab3-Nrd1 mediates timely recruitment of Sen1. However, it has not been shown whether a direct contact of Sen1 with surface-exposed regions of Pol II could help to disassemble the elongation complex. There are though examples of other transcription factors, e.g. CF IA, that have been shown to bind to the evolutionary conserved flap loop above the RNA exit channel and to induce an allosteric signal to terminate (Pearson & Moore, 2014). The current model suggests that, similarly to bacterial Rho-factor, Sen1 translocates on the nascent RNA up to the transcribing Pol II and displaces it. However, Rho-factor is incapable to disassemble yeast elongation complex (Porrua & Libri, 2013). Consistently, it has been shown that the termination efficiency increases if the polymerase decelerates, thus the kinetic competition between Sen1 and the Pol II plays a role.

The results of this thesis reveal that not only translocation but also unwinding properties are required for Pol II termination. $Sen1_{Hel}$ has residues on the surface of the "prong" that do not have a direct effect to RNA binding or ATP hydrolysis, however are important for the elongation complex dissociation. For example, $Sen1_{Hel}\Delta UP$ mutant, which was the most active in unwinding assays, exhibited a moderate decrease in termination efficiency *in vitro* (Figure 3-21). But probably more importantly, point mutation L1549D acerbated the effect of deletion of the upper part of the "prong". The double mutant retained unwinding activity similar to wild-type protein, however termination efficiency dropped by 5-fold.

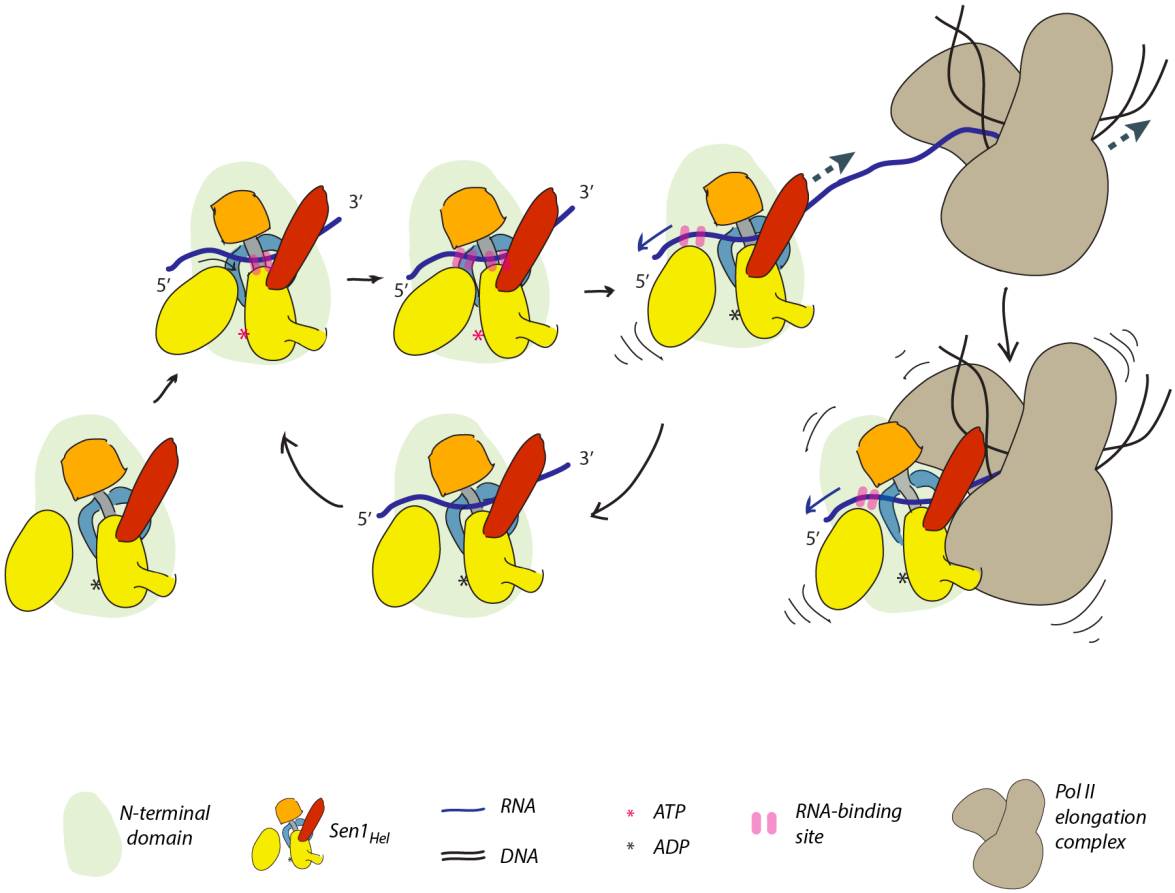


Figure 4-1. Sen1 in transcription termination.
A schematic step-wise representation of Sen1 translocation along newly transcribed RNA in 5'→3' direction. Sen1 is a strong RNA-dependent ATPase, however it has a transient state after ATP hydrolysis in which the helicase is weakly bound to RNA, thus Sen1 is a low processivity enzyme. For efficient transcription termination it is crucial that Sen1 is either recruited in a close proximity to Pol II and/or that Pol II is stalled. *In vivo*, Sen1 is recruited to Ser2-phosphorylated Pol II via interaction with Nrd1-Nab3 heterodimer but also Sen1 can bind to Ser5-phospborylated Pol II directly. It is still not clear whether the N-terminal domain of Sen1 is binding to the RNA and whether the domain is in front or behind of the helicase core during the translocation. However, Sen1 translocates on the transcript RNA until collides with Pol II, leading to disruption of the elongation complex. Most likely the close contact causes allosteric changes in Pol II, it also could be that Sen1 destabilizes the transcription bubble by pulling out the RNA from active site.

4.5 SETX

The majority of ataxia ocular apraxia type 2 (AOA2)-associated mutations are recessive missense, nonsense and deletion mutations that result in loss of helicase function. We have mapped the missense mutations onto the structure of $Sen1_{Hel}$ and tested 6 disease mimicking $Sen1_{Hel}$ variants in ATPase, unwinding and transcription termination assays. $D1616V_{Sen1}$ was insoluble, and all other proteins have partially or completely lost their enzymatic activity. Other AO2-mutations, for example E65K, W305C and P413L abolish binding of Rrp45 (Richard & Manley, 2014), which indicates that the N-terminal domain

of SETX functions as the exosome cofactor. Interestingly, none of ALS mutations cause direct inactivation of the helicase. For example, H1956R ($Q1356_{Sen1}$) is at the back site of RecA2 and K2018E ($L1421_{Sen1}$) is at the front site of RecA1 domain, and R2136C/H ($K1539_{Sen1}$) is in the disordered upper part of the sidewall. The exact role of ALS4-related mutations remains unclear, but these mutations are dominant missense mutations that most likely have a gain-of-function phenotype. It could be that ALS4-related mutations impair SETX interaction with other proteins and ALS4 phenotype is not directly related to transcription termination. For instance, the most common ALS4 mutation, L389S, results in increased binding of SUMOylation proteins or to a peptide encoded by BCYRN1-reverse complement (Bennet *et al*, 2013).

4.6 Conclusions

In this thesis the identified ~94 kDa helicase core of *S. cerevisiae* Sen1 has biochemical properties similar to those of the endogenous full-length protein in *in vitro* experiments. $Sen1_{Hel}$ has a very similar overall structure to that of Upf1-like helicases. Although the helicases share the same conserved sequence motifs for RNA binding and ATP hydrolysis, Upf1 is not capable to terminate stalled Pol II *in vitro*. The most prominent difference in structures of Sen1 and Upf1 is the "brace", which fastens the RecA1, the "stalk" and the "barrel" into a rigid domain. Sen1 has a relatively high ATPase activity and is capable to disrupt a tiny fraction of Pol II even when the "prong" is fully deleted. Interestingly, the processivity of the helicase seems not to be directly dependent on ATP hydrolysis but rather on affinity to RNA (Han *et al*, 2017). It could be, that relatively low processivity and a low copy number of Sen1 molecules per cell prevent spurious termination events.

All together the results of this thesis support the view that Sen1 uses the nascent RNA as a guide to catch up with the elongation complex of Pol II moving in the same direction. Most likely, at direct contact, Sen1 inserts the "prong" into the RNA exit channel of Pol II to open the cleft of the polymerase. Additionally, surface interaction between the two proteins might provoke a conformational change in the elongation complex. It is also plausible that the elongation complex is destabilized using a "ratchet" mechanism for pulling the nascent RNA out of transcription burble.

5 Outlook

To date it is not clear at which side of Sen1 structure the N-terminal domain resides. It also is not known whether the N-terminal domain can bind to RNA directly. Unfortunately, we could not test it in RNase assays because of the limited amount of the protein. And although the presence of the N-terminal domain decreases the termination efficiency *in vitro* (Han *et al*, 2017), it is not clear whether the domain has any regulatory role or the decrease is due to the large size. It is possible that the N-terminal domain partially interferes in interaction between the helicase domain and Pol II. These should be the main questions trying to solve the structures of the pre-termination complex of Sen1 and Pol II in the future.

Nevertheless, in *in vitro* experiments the helicase domain alone is capable to disassemble the elongation complex of Pol II. A closer inspection of the structures suggested at which surface of $Sen1_{Hel}$ and the elongation complex (PDB: 3HOV (Sydow *et al*, 2009)) could make a contact (Figure 5-1). At the collision of $Sen1_{Hel}$ and Pol II, the "brace", the "barrel", the "prong" together with the part of the RecA1 domain would bed in to the side of Pol II where the transcript is exiting.

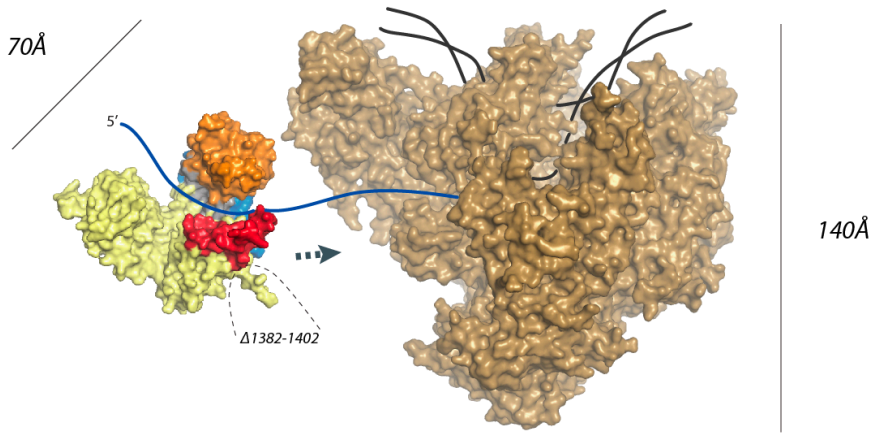


Figure 5-1. Interaction site on the surface of $Sen1_{Hel}$ and Pol II.
At the collision of $Sen1_{Hel}$ with the Pol II elongation complex (PDB: 3HOV), $Sen1_{Hel}$ would bed in closely to the RNA exit channel.

At the side of the RecA1 domain, $Sen1_{Hel}$ has a 20-residue loop (residues 1382-1402) that protrudes from the domain (Figure 5-1). PSIPRED predicted this region to be disordered

and only a part of the loop could be built into the model. Deletion of the loop had no effect on RNA binding or ATPase activity but, surprisingly, the mutant has increased transcription termination activity (60 % termination compared to 47 % of the wild-type $Sen1_{Hel}$). We have not studied the mutant further, however this loop together with other conserved residues on the Sen1 surface should be investigated in the future. For example, we have also found that a double mutation on the top of the "barrel" (R1188D, Y1195F) has a decreased activity in transcription termination assays (40 % termination).

On the other side, it is not known which regions of Pol II are mediating the interaction with Sen1 at the collision. Pearson and Moore have demonstrated that the evolutionary conserved flap loop of Rpb2 interacts with Pcf11 (essential for canonical transcription termination after polyA site), and the deletion of that loop leads to termination defects even at NNS-dependent noncoding genes (Pearson & Moore, 2014). Furthermore, a few point mutations of Rpb3 and Rpb11 at the bottom at the Pol II provoke significant transcription termination defects *in vivo* (Steinmetz *et al*, 2006). These residues seem to be too far for the helicase domain of Sen1 to reach but it could be that this is the site where the N-terminal domain would make the contact.

Finally, it is not known at which distance from Pol II the helicase should be loaded on the transcript RNA. Currently it is suggested that Sen1 processivity is ~20-40 nt (Han *et al.*, 2017). However, the RNA length during Sen1 recruitment via Nrd1-Nab3 or direct interaction with CTD is not known and should be addressed in the future.

6 Materials and methods

6.1 Materials

6.1.1 Chemicals and reagents

All common chemicals and reagents were ordered from Sigma-Aldrich or Fluka unless otherwise stated. Chemicals fot crystallization were additionally order from Hampton.

6.1.2 Enzymes

For DNA amplification by PCR, Phusion polymerase from Thermo Fisher Scientific was used. For RNase protection assays RNase A and RNase T1 (Fermentas) were used. Rhinovirus 3C protease was expressed and purified in-house.

6.1.3 DNA and RNA oligonucleotides

DNA and RNA oligonucleotides were ordered from Sigma-Aldrich and Purimex, respectively. All oligonucleotides were ordered desalted and lyophilized, RNA was additionally purified by PAGE. DNA was dissolved in ddH_2O and RNA – in 100 mM HEPES pH 7.5 and stored at -20°C.

Table 6-1. DNA oligonucleotides used as primers for cloning.

Name	Sequence (5' to 3')
1_F	cacccgggatctcgagcATGAATTCCAACAATCCTG
OneStrep_R	cccccatctcccggtaccgCTATTTCTCGAACTGCGGGTGGC
2231_R	cccccatctcccggtaccgTCATGATCTAGGCTTTC
SM_1105_F	aactttaagaaggagatatacatATGAGAAAGAGATTAAATGTTG
SM_1105_R	CAACATTTAATCTCTTTCTCATatgtatatctccttcttaaagtt
ins3C_F	gtggttctgttccaggggcccCTCGCGGGCGGTAAAATACTCCATAATC
ins3C_1881_R	ctggaacagaaccaccagTGATGGTACGGGTACATTGAACTTTC
ins3C_1904_R	ctggaacagaaccaccagATTGGATTTTACCTCATCGGGGCCTTGCG
ins3C_1910_R	ctggaacagaaccaccagTCTCTTCTTTGTGTCCTTATTGGATTTTACC
T4L_1_R	GGTGGTGGTGGTGGTGGTGGTGtccCAGGTTCTTGTAAGCGTCCC
T4L_2_F	GGGACGCTTACAAGAACCTGggaCACCACCACCACCACCACCACC
Trx_1_R	GGTGGTGGTGGTGGTGGTGGTGtccCAGGTTAGCGTCGAGGAACTC
Trx_2_F	GAGTTCCTCGACGCTAACCTGggaCACCACCACCACCACCACCACC
43_1_F	GAAAGTTCAATGTACCCGTACCATCACTGGAAGTTCTGTTCCAGGG GCCCATGAACATCTTCGAGATGCTGCG
43_2_R	CGCAGCATCTCGAAGATGTTCATGGGCCCCTGGAACAGAACTTCCA GTGATGGTACGGGTACATTGAACTTTC
44_1_F	GGCCCCGATGAGGTAAAATCCAATCTGGAAGTTCTGTTCCAGGGGC CCATGAACATCTTCGAGATGCTGCG

44_2_R	CGCAGCATCTCGAAGATGTTCATGGGCCCCTGGAACAGAACTTCCAGATTGGATTTTACCTCATCGGGGCC
45_1_F	ATCCAATAAGGACACAAAGAAGAGACTGGAAGTTCTGTTCCAGGGGCCCATGAACATCTTCGAGATGCTGCG
45_2_R	CGCAGCATCTCGAAGATGTTCATGGGCCCCTGGAACAGAACTTCCAGTCTCTTCTTTGTGTCCTTATTGGAT
46_1_F	AAGTTCAATGTACCCGTACCATCACTGGAAGTTCTGTTCCAGGGGCCCATGAGCGATAAAATTATTCACCTG
46_2_R	CAGGTGAATAATTTTATCGCTCATGGGCCCCTGGAACAGAACTTCCAGTGATGGTACGGGTACATTGAACTT
47_1_F	GGCCCCGATGAGGTAAAATCCAATCTGGAAGTTCTGTTCCAGGGGCCCATGAGCGATAAAATTATTCACCTG
47_2_R	CAGGTGAATAATTTTATCGCTCATGGGCCCCTGGAACAGAACTTCCAGATTGGATTTTACCTCATCGGGGCC
48_1_F	GTAAAATCCAATAAGGACACAAAGAAGAGACTGGAAGTTCTGTTCCAGGGGCCCATGAGCGATAAAATTATTCACCTG
48_2_R	CAGGTGAATAATTTTATCGCTCATGGGCCCCTGGAACAGAACTTCCAGTCTCTTCTTTGTGTCCTTATTGGATTTTAC
IA2_41_R2	CAGTCTGAGGCAGATTTCGTCCACAGCGGCATTACTGGGGGC
IA4_41_R2	ccttcttaaagttaaacaaaattatttctagagggg
IA6_41_R2	TACACGGACCAATTGAGGTTTAAATTGATGCCCC
IA7_41_R2	CATTACAATTAAGTCAGATTCAGAAATACCAC
IA1_41_F	GGTTACTTTTTGTCTACAAAGAATGCCTCAggtggtAAGCAAAAAATTCTAATCTGTGCCCCCAG
IA1_41_R	AATAATACCCAGTATAGTCTTTG
IA2_41_F	GAAATCTGCCTCAGACTGAAGAGTGGAggtggtggtTTGGTCCGTGTAGGTAGGTC
IA2_41_R	GTCCACAGCGGCATTACTGG
IA3_41_F	GGACCTTACTTTAGAAGAACTTagcggtggtCAGGCCCATATCTTGGCGGTTAGTG
IA3_41_R	TTAATTGCAACGTTTACAACGTC
IA4_41_F	gtttaactttaagaaggagatatacatATGAATAGTGAATATCCTGATGATGAGCCTATTGG
IA4_41_R	aaaattatttctagagggg
IA5_4_IA4_F	CATGAAGCCTTTACTACTTCTAGAATCTTcGCAAGGTaacTGTTCTTCACGTGATAGAGAGGAC
IA5_4_IA4_R	ACTTTTTGGTAGTCAGCAGGAG
IA6_41_F	CCTCAATTGGTCCGTGTAGGTggcggcggcATTAAGGACCTTACTTTAG
IA6_41_R	TTTAAATTGATGCCCCTG
IA7_41_F	GACTTAATTGTAATGGCTTATggcggcCATACATGTTTAGCAAAGGTG
IA7_41_R	AGATTCAGAAATACCACAATCC
IA8_41_F	CAACATACATGTTTAGCAgAGGTGAGAACATTGAAAAATACC
IA8_41_R	TGCTTTCTTGAAATCATCACTGGAC
IA9_41_F	CCTCTATATGAAATCATAagGCAATGGGATTACACCAG
IA9_41_R	GTTCATATCAACATTTAATCTCTTTCTCATATGTTC
IA10_41_F	CCTCTATATGAAATCATAagGCAAtgGGATTACACCAGAAATAGTG

IA11_41_F	GTTGATATGAACCCTCTAaggGAAATCATATTGCAATGGG
IA11_41_R	ATTTAATCTCTTTCTCATATGTTC
IA12_41_F	GTTGATATGAACCCTCTAaggGAAATCATAagGCAATGGGATTACAC CAG
IA13_41_F	CACAACATACATGTTTAGCAAAGGTGgaAACATTGAAAAATACCAA AGGTGG
IA13_41_R	CTTTCTTGAAATCATCACTGGAC
IA14_41_F	GGTAATGTTGATGTCACATTAgaaATTCATAGAAATCATTCTTTCAG TAAATTTTTGAC
IA14_41_R	ACCTTTGGTATTTTTCAATGTTCTCACC
IA15_41_F	gtttaactttaagaaggagatatacatATGTCTCCTGCTGACTACCAAAAAGTCATGA AGCC
IA16_41_F	gtttaactttaagaaggagatatacatATGCCTATTGGTAATTATTCTGACGTAAAG GATTTCTTC
IA10_41_F	CCTCTATATGAAATCATAagGCAAagGGATTACACCAGAAATAGTG
IA19_41_F	GAGATTAAATGTTGATATGAACCCTCTAgctGAAATCATAgcgCAAgcg GATTACACCAGAAATAGTGAATATCCTG

Small letters show overhangs of vector sequence or introduced mutations. Capital letters correspond to the sequence of the gene of interest.

Table 6-2. ssDNA/RNA used in assays.

Name	Sequence
16-mer RNA oligonucleotide used for assembly of ECs. 5'-end FAM labeled	UGCAUUUCGACCAGGC
Template DNA strand used for assembly of ECs. Labeled with biotin at the 5'-end	GGCCGGGTAACCCCCGTGTGGAGATGGGTGAGAGATGTTG AGGGCCTGGTCGTTTCCTATAGTTTGTTTCCT
Non-template strand usedfor assembly of Ecs. Labeled with biotin at the 5'-end	CTAGAGGAAACAAACTATAGGAAACGACCAGGCCCTCAAC ATCTCTCACCCATCTCCACACGGGGGTTACCCGGCCTGCA
ssDNA/RNA for ATPase assay 5'-end FL labeled	UUUUUUUUUUUUUUU
	CCCCCCCCCCCCCCC
	AAAAAAAAAAAAAAA
	CCCTAACCCTAAC
RNA for FA measurements	UUUCUAUUUAUUUUG

6.1.4 Constructs

All *S. cerevisiae* variants were cloned into pCB-A-bax10 plasmid (kindly shared by Dr. Christian Biertümpfel) for expression in *E. coli* and into pFL$^{\Delta SpeI}$ donor plasmid (Invitrogen) for expression in insect cells. See list of all cloned mutants in Appendix 3.

6.1.5 Cloning kits

Wizard Gel and PCR Clean-up (Promega) and Qiaquick Spin Miniprep (Qiagen) kits were used.

6.1.6 *E. coli* strains and insect cell lines

Table 6-3. *E. coli* strains used for cloning and expression.

E. coli strains	Genotype
One Shot™ OmniMAX™ 2 T1 (Thermo Fisher Scientific)	F' {proAB+ lacIq lacZΔM15 Tn10(TetR) Δ(ccdAB)} mcrA Δ(mrr-hsdRMS-mcrBC)ΔM15φ80(lacZ) Δ(lacZYA-argF) U169endA1 recA1 supE44 thi-1 gyrA96 relA1 tonA panD
DH10MultiBacY (Invitrogen)	F^- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697galUgalKλ$^-$ rpsLnupG/bMON14272/pMON7124
BL21 (DE3) STAR pRARE (Stratagene)	F^- ompT $hsdS_B$ ($r_B^- m_B^-$) gal dcm rne131 (DE3) (pRARE araW argU glyT ileX leuW proL metT thrT tyrU thrU Cam^R)

For protein expression in insect cells the *Autographa californica* cell lines Sf21 and HighFive (Invitrogen) were used.

6.1.7 Media and buffers

Table 6-4. Media used for cloning and protein expression.

Media	Components
S.O.C. Medium	2 % (w/v) bacto tryptone 0.5 % (w/v) bacto yeast extract 10 mM NaCl 1 mM $MgCl_2$ 2.5 mM KCl 10 mM $MgSO_4$ 0.4 % (w/v) glucose adjust pH to 7.2
Luria-Bertani (LB) (Miller, 1972)	1 % (w/v) bacto tryptone 0. 5% (w/v) bacto yeast extract 170 mM NaCl adjust pH to 7.6
LB agarose plates	1.5 % (w/v) bacto agar in LB 100 µg/mL ampicillin
Terrific Broth (TB) (Sambrook & Russell, 2001)	1.2 % (w/v) bacto tryptone 2.4 % (w/v) bacto yeast extract 0.4 % (v/v) glycerol ddH_2O to 900 ml 0.017 M KH_2PO_4 0.072 M K_2HPO_4
Sf-900™ SFM (insect cell culture medium)	Purchased from Invitrogen
ESF 921 (insect cell culture medium)	Purchased from Expression Systems

Table 6-5. Protein purification buffers.

Buffer	Components
Lysis buffer	20 mM sodium phosphate pH 8.0 500 mM Nacl 2 mM $MgCl_2$ 30 mM imidazole 10 % (v/v) glycerol 1 mM β-mercaptoethanol
Salt wash buffer	20 mM sodium phosphate pH 8.0 1000 mM Nacl 2 mM $MgCl_2$ 30 mM imidazole 1 mM β-mercaptoethanol
ATP wash buffer	20 mM sodium phosphate pH 8.0 250 mM Nacl 2 mM $MgCl_2$ 50 mM KCl 10 mM $MgSO_4$ 2 mM ATP 1 mM β-mercaptoethanol
Tag cleavage buffer	20 mM Tris-HCl pH 7.5 250 mM Nacl 2 mM $MgCl_2$ 10 % (v/v) glycerol 1 mM β-mercaptoethanol
Buffer A	20 mM Tris-HCl pH 7.5 250 mM Nacl 2 mM $MgCl_2$ 1 mM DTT
Buffer B	20 mM Tris-HCl pH 7.5 1000 mM Nacl 2 mM $MgCl_2$ 1 mM DTT
SEC buffer	20 mM HEPES pH 7.5 250 mM NaCl 2 mM $MgCl_2$ 1 mM DTT

Table 6-6. Sample freezing buffer.

Buffer	Components
Sample freezing buffer (2x)	40 mM HEPES pH 7.5 250 mM NaCl 2 mM $MgCl_2$ 99 % (v/v) glycerol 1 mM DTT

Table 6-7. Buffers used in assays.

Buffer	Components
TBE (20 x)	1 M Tris-HCl 0.89 M boric acid 20 mM EDTA pH 8.0
SDS-PAGE running	0.25 M Trizma Base 1.92 M Glycine 1% (v/v) SDS
SDS loading buffer (2x)	100 mM Tris pH 6.8 4% (v/v) SDS 0.2% (v/v) bromphenol bue 20% (v/v) glycerol 10% β-mercaptoethanol
RNase protection (2x)	100 mM HEPES pH 7.5 100 mM NaCl 2 mM $MgAc_2$ 0.2 % (v/v) NP-40 20 % (v/v) glycerol
Splicing dilution buffer	100 mM Tris pH 7.5 150 mM NaCl 300 mM Na acetate pH 5.2 10 mM EDTA 1% (v/v) SDS
100 % BFE	0.1 % bromphenol blue 0.1 % xylene cyanole FF 10 mM EDTA pH 8.0 formamide

Table 6-8. PAGE gels.

Gel	Components
Stacking SDS-PAGE gel	2.63 mL ddH_2O 2 mL 1.5 M Tris pH 8.7 800 µL 30 % (w/v) N,N'-methylene-bisacrylamide (37.5:1) 40 µL 10 % (w/v) SDS 24 µL 10 % (w/v) ammonium persulfate (APS) 8 µL 100 % N, N, N'N'-tetramethylethylene diamine (TEMED)
Separating SDS-PAGE (10 %)	3.4 mL ddH_2O 2 mL 1.5 M Tris pH 8.7 2.6 mL 30 % (w/v) N,N'-methylene-bisacrylamide (37.5:1) 80 µL 10 % (w/v) SDS 36 µL 10 % (w/v) ammonium persulfate (APS) 12 µL 100 % N, N, N'N'-tetramethylethylene diamine (TEMED)
Denaturing PAGE gel (22 %)	7 M urea 22 mL 30 % (w/v) N,N'-methylene-bisacrylamide (19:1) 180 µL 10 % (w/v) ammonium persulfate (APS) 30 µL 100 % N, N, N'N'-tetramethylethylene diamine (TEMED) to 30 mL 1x TBE

6.1.8 Equipment

Table 6-9. List of equipment.

Equipment	Model	Manufacturer
Pipettes	Eppendorf Research	Eppendorf
Scales	LA1200S & TE1502S	Sartorius
pH meter	Lab860	Schott
PCR machine	Mastercycle	Eppendorf
UV transilluminator	Safe Imager	Invitrogen
Vortex mixer	Vortex-Genie	Scientific Industries
Electroporator	Gene Pulser/Micro Pulser	Bio-Rad
Electro-cuvette	Gene pulser 0.1 cm electrode gap	Bio-Rad
Bacteria shaker	KS-15/Climo-shaker ISF1X	Kühner
Thermo shaker	Thermomixer Comfort	Eppendorf
Cell lysis sonicator	Sonifyer VS70T	Bandelin Electronics
Insect cell shaker	Kühner shaker	Kühner
Laminal flow hood	Holten LaminAir	Thermo electron corporation
Cell counter	Vi-Cell-XR	Beckman Coulter
Dounce Homogenizer	Borosilicate Glass	Thermo Fisher Scientific
Centrifuge	Avanti J-20 XP	Beckman Coulter
	Micro centrifuge 5417C & 5810	Eppendorf
Columns	Ni^{2+}-NTA beads & HisTrap FF HiTrap Heparin HP Superdex 200 [16/600]	GE Healthcare
Peristaltic pump	Ismatec 78001-40	Cole-Parmer
Chromatography FPLC	ÄKTA Prime ÄKTA Purifier	GE Healthcare
Protein concentrator	Amicon Ultra Centrifugal filters	Ultracel
Spectrophotometer	Nanodrop	PeqLab
Gel imaging	Gel visualization	Roth
Crystallization plates	96-well MRC96T	SwissCI
	24-well vdx	Hampton Research
Crystallization pipetting robot	Phoenix	Art Robbins Instruments
Crystallization visualization system	Xtal-Focus	ExploraNova La Rochelle
Cryo loops	CryoLoop 20 micro	Hampton Research
Caps and vials	Magnetic caps and vials	Molecular Dimensions
X-ray diffractometer	PX scanner	Rigaku
	D8 venture	Bruker
Fluorescence spectrometer	Infinite M1000 Pro	Tecan
Phosphorimager	Typhoon FLA 9500	Fuji

6.1.9 X-ray sources and synchrotron facility

Crystals were tested with in-house X-ray diffractometers PX-scanner (Rigaku) and D8 venture (Bruker). Some of the diffraction data sets were also collected on in-house D8 venture, however the higher resolution data sets were obtained at Swiss Light Source (SLS) synchrotron at the Paul Scherrer Institute (PSI), Villingen, Switzerland.

6.1.10 Software and web servers

The following software was used to write this thesis, analyze data and generate figures:

Adobe Illustrator (www.adobe.com/products/illustrator)

ApE (http://biologylabs.utah.edu/jorgensen/wayned/ape/)

BIOEQS (http://abcis.cbs.cnrs.fr/BIOEQS/)

Buccaneer (http://www.ccp4.ac.uk/html/cbuccaneer.html)

Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/)

CodonCode Aligner (http://www.codoncode.com/aligner/)

Coot (http://www2.mrc-lmb.cam.ac.uk/Personal/pemsley/coot/)

EndNote X7 (http://endnote.com)

Hampton Research Make-tray tool (http://hamptonresearch.com/make_tray.aspx)

Jalview (http://www.jalview.org/)

Microsoft Office (www.microsoft.com)

LigPlot (https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/)

LOVD Database (https://databases.lovd.nl/shared/variants/SETX/unique)

Phenix (http://www.phenix-online.org)

Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2)

ProtParam (http://web.expasy.org/cgi-bin/protparam)

PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/)

PyMOL (http://pymol.org)

SHELX (http://shelx.uni-ac.gwdg.de/SHELX/)

UNIPROT (http://www.uniprot.org)

wwPDB (https://www.wwpdb.org)

XDS (http://xds.mpimf-heidelberg.mpg.de)

6.2 Methods

6.2.1 Cloning for expression in *E. coli*

For cloning a one-step ligation-independent method was used. The method relies on DNA repair machinery of *E. coli* cells. For each transformation a set of a linearized vector and an insert with the >25 bp extensions were designed (Figure 6-1, a). In case of cloning a point mutation, a vector was linearized using primers that harbored the mutation (Figure 6-1, b). After PCR, purified unprocessed DNA was directly transformed into cells. All the constructs were cloned into a pCB-A-bax10 plasmid (kindly shared by Dr. Christian Biertümpfel) that contained a cleavable C-terminal His tag coupled to *Vibrio cholerae* MARTX toxin cysteine protease domain (CPD-His_8) (Shen *et al*, 2009).

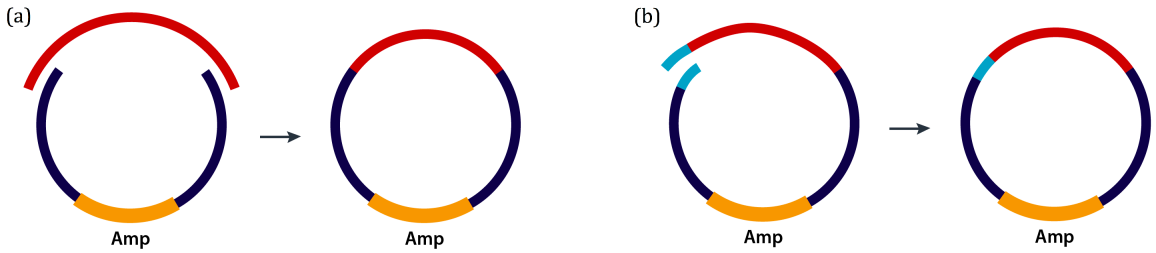


Figure 6-1. Schematic presentation one-step ligation-independent cloning.
(a) Cloning with a linearized vector and an insert. The insert (red) encodes Sen1 sequence and the >25 bp extensions at both ends. These extensions are identical to the sequence of the vector (navy). After transformation, under selective ampicillin pressure, the cell is forced to repair the plasmid and recombines the insert with the DNA containing the Ampicillin-resistance cassette (yellow). (b) Cloning of a point mutation. The vector is linearized using primers that encode a point mutation (light blue). Both ends if linearized vector are identical. After transformation the cells repairs the plasmid to gain ampicillin resistance.

6.2.1.1 PCR

The PCR reactions for vector linearization and the inserts were set up in a total volume of 50 µL:

5 µL forward primer (10 µM)
5 µL reverse primer (10 µM)
20 ng template DNA
2.5 µL DMSO
25 µM 2x Master Mix buffer (kit)
0.5 µM Phusion Polymerase (1 U/µL)

The following thermocycling conditions were used:

1-3 min - 98°C initial denaturation
10 s - 98°C denaturation*
10 s - Tm-5°C annealing*
30 s per kb - 72°C extension*
10 min - 72°C final extension

** Steps were repeated in 30-35 cycles. The annealing temperature varied depending on the melting temperature (Tm) of the primers. In some cases touch-down PCR was applied with the decrease of the annealing temperature of 1 °C per cycle.*

6.2.1.2 Agarose gel electrophoresis

DNA fragments of PCR products were separated on 0.75 % (w/v) agarose gels prepared in 1x TBE, supplemented with SYBR Safe (Invitrogen) stock 1:10000 dilution as a dye for visualization. Prior to loading samples were mixed with 6x loading buffer (Orange Loading Dye, Fermentas). 0.5 μg of the 1Kb Plus DNA ladder (Fermentas) was used as a marker in a separate lane. Gel electrophoresis was performed at 6 V/cm in 1x TBE buffer. DNA fragments were visualized and excited from the gel using a blue light transilluminator (Invitrogen) emitting light at a wavelength of ~470 nm.

6.2.1.3 DNA fragment purification

The separated DNA bands from an agarose gel were purified with the Wizard SV Gel and PCR Clean-up system (Promega) according the manufacturer's protocol. All DNA fragments were eluted in 30 μL ddH_2O.

6.2.1.4 Transformation

The purified linearized vector and the insert (1.5 μL and 3.5 μL each) were added to 50 μL aliquots of OmniMAX chemically supercompetent *E. coli* cells. Cells were then gently mixed and incubated on ice for 5 min. A heat shock was performed in order to transfer the vectors into bacteria cells. Therefore, the mixture was incubated 45 s at 42°C and then chilled on ice for 2 min followed by addition of 600 μL of S.O.C. medium and incubation at 37°C, 600 rpm for 30 min. Next, 100-300 μL of the cells were plated on LB agar plates, supplemented with 100 μg/mL ampicillin, and incubated over night at 37°C.

6.2.1.5 Plasmid amplification and isolation

For plasmid amplification a single-cell-derived colony from the *E.coli* agar plate was picked with a pipette tip and added to 5 mL LB medium supplemented with 100 µg/mL ampicillin. The cell culture was incubated over night 160 rpm at 37°C. The cells were harvested by centrifugation (4000 rpm / 10 min / 4°C). Plasmid DNA was purified with Wizard Gel and PCR Clean-up System kit (Promega) and eluted in 50 µL ddH_2O. The concentration of plasmid was measured with the Nanodrop spectrophotometer.

6.2.1.6 DNA Sequencing

All plasmid DNA was sequenced by the core facility of MPI Biochemistry or Eurofins. Custom sequence primers were ordered with the aim for a GC content of at least 45% and Tm ~65°C. The received nucleotide sequence was analyzed and compared to the nucleotide sequence of the original gene using CodonCode Aligner and ApE programs.

6.2.2 Recombinant protein expression in *E. coli*

For the recombinant protein expression in *E. coli* a plasmid was transformed to BL21 (DE3) STAR pRARE cells as described before. The successful transformation was selected by incubating 100 µL of the cells on LB agar plates containing 100 µg/mL chloramphenicol and 100 µg/mL ampicillin. Next day a small amount of cells was scraped from a single colony into 70 mL LB media supplemented with respective antibiotics and incubated at 37°C and 200 rpm for 2-3 hours. 10 mL of pre-culture was added to pre-warmed 500 mL TB media supplemented with respective antibiotics, and incubated further at 37°C and 220 rpm until OD_{600} ~1.5 (3-4 hours). Then the temperature was decreased to 18°C and 0.5 mM IPTG was added to induce protein expression. After over night expression (~16 hours), cells were harvested by centrifugation (9 000 x g, 10 min). The cell pellet was either used immediately or flash frozen in liquid nitrogen and stored at -80°C.

6.2.3 Cloning and expression in insect cells

For recombinant protein expression in cultured insect cells a recombinant baculovirus has to be generated in two steps (Figure 5-1). First, a donor plasmid is created and then the plasmid is transformed into competent *E. coli* cells that contain the bacmid (baculovirus shuttle vector) for site-specific transposition. Afterwards, the recombinant bacmid DNA is

used for insect cell transfection (virus generation), followed by virus amplification and recombinant protein expression.

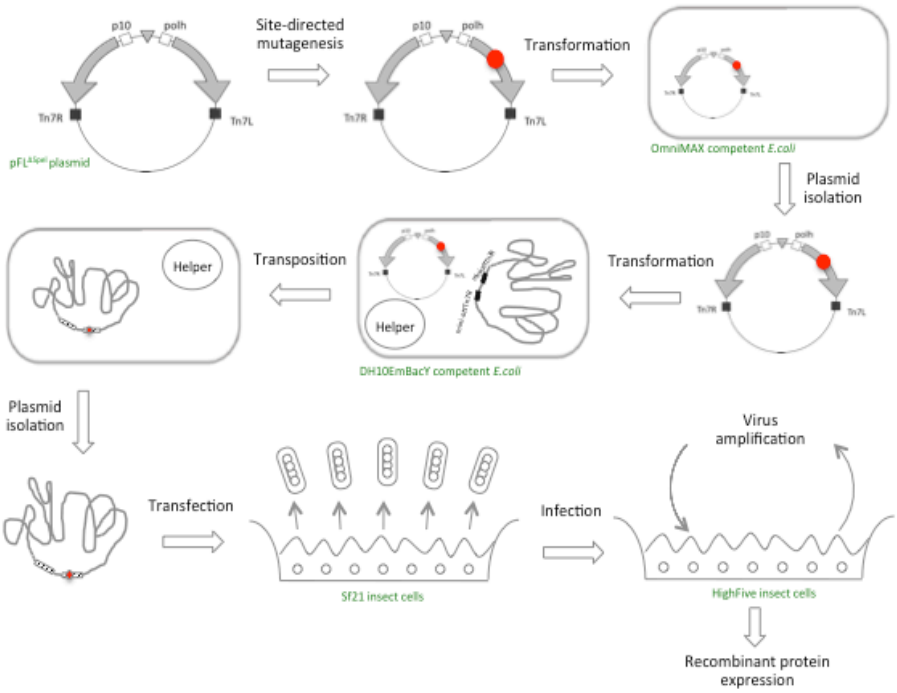


Figure 6-2. Generation of recombinant baculovirus and protein expression.
After a recombinant plasmid is generated, it has to be transformed into bacmid. The recombinant bacmid DNA then is used for insect cell transfection to generate the virus. During virus cycle the infected cells are killed which leads to virus release to media and infection of new insect cells. After the virus has been amplified (P1 → P2 → P3), the recombinant protein expression can be tested.

6.2.3.1 Donor plasmid cloning and bacmid generation

All the cloning procedures for expression in insect cells were done as described in Methods 6.2.1 with the following exceptions. DNA encoding full-length *S. cerevisiae* Sen1 was cloned into multi cloning site 1 (MCS1) of pFL$^{\Delta SpeI}$ donor plasmid (Figure 6-3). Sen1 was placed downstream *p10* promoter that is required for the expression in insect cells when using the *Bac-to-Bac Baculovirus Expression System* (Invitrogen). In this system, the transcriptional control of the *Autographa californica* nuclear polyhedrosis virus results in abundant heterologous gene expression at the late stage of infection, while host proteins synthesis is diminished.

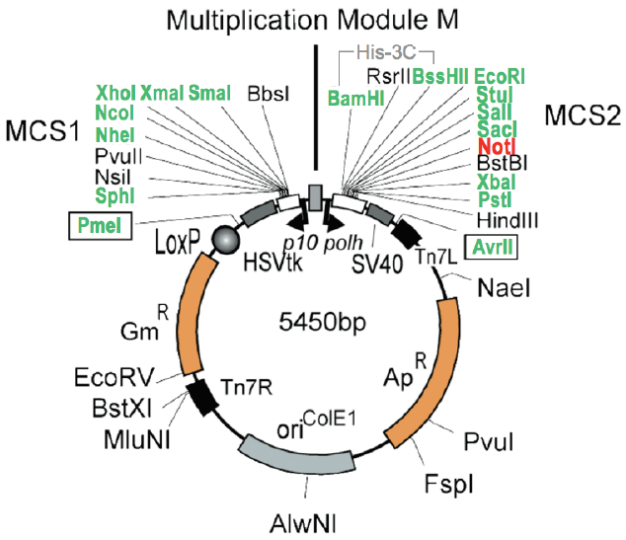


Figure 6-3. pFL$^{\Delta SpeI}$ plasmid map.
pFL$^{\Delta SpeI}$ is a donor plasmid used to generate recombinant viruses. The plasmid contains one origin of replication and can be propagated in *E. coli*. The multiple cloning sites (MCS1 and MCS2) contain several restriction sites for gene insertion and are flanked by Tn7 elements (Tn7L and Tn7R) for transposition to the *att*Tn7 attachment sites on the bacmid. Additionally, these elements contain an expression cassette consisting of a Gm^r and Ap^r genes and an SV40 poly(A) signals.

The newly cloned donor plasmid has to be then transposed into a bacmid. The transposition is based on site-specific transposition of an expression cassette into a bacmid DNA propagated in *E. coli*. The donor plasmid contains mini-Tn7 element that attaches to mini-*att*Tn7 on the bacmid. The required Tn7 transposase is provided by helper plasmid that is hosted by competent cell. The *lac* operator controls the successful transposition as the bacmid contains a segment of DNA encoding the *lacZα* peptide. Insertion of the mini-Tn7 into the mini-*att*Tn7 attachment site on the bacmid disrupts peptide's expression and the colonies containing the recombinant bacmid are white in background of blue non-transposed colonies. Additionally, the bacmid and the helper plasmid confer resistance to kanamycin and tetracycline, respectively.

Here, for the transposition the verified pFL$^{\Delta SpeI}$ plasmid was transformed into DH10EmBacY *E. coli* strain by electroporation. Therefore, 2 µL of plasmid DNA were added to 50 µL competent cells and incubated for 5 min on ice. Then the cells were transferred into electrophoresis cuvette. Immediately after the electroporation at 2.5 kV 900 µL S.O.C. medium was added and the culture was incubated at 37°C and 600 rpm for over night. The successful transformation was selected by blue-white screen. Therefore,

100-300 µL of the cells were plated on LB containing 50 µg/mL kanamycin, 7 µg/mL gentamicin, 10 µg/mL tetracyclin, 100 µg/mL Bluo-gal (dissolved in DMSO) and 80 µg/mL IPTG. The bacmid was amplified by growing 2 mL bacteria culture in LB from an isolated white colony as described previously. The selective pressure was maintained by adding 50 µg/mL kanamycin, 7 µg/mL gentamicin, and 10 µg/mL tetracyclin. Finally, the bacmid was isolated by alkaline lysis, using the Qiaprep Spin Miniprep kit, and DNA precipitation in isopropanol and dissolved in 30 µL ddH_2O.

6.2.3.2 Transfection of Sf21 insect cells and virus generation

During the transfection, Sf21 cells are infected with the recombinant bacmid DNA. To significantly increase the efficiency, the DNA has to be incubated with transfection agent such as polyethylenimine (PEI) that acts as an attachment factor for cell adherence. Both PEI are polycationic polymers and condense DNA into positively charged particles, which bind to anionic cell surface residues and are brought into the cell via endocytosis. Once the vesicle enters the cell, it undergoes osmotic swelling and bursts releasing the polymer-DNA complex into the cytoplasm. Then the polymer is replaced by other cations leading to unpacked bacmid DNA. The bacmid then can enter the nucleus where viral transcription and replication occurs. The baculovirus particles get budded outwards the cell by 24-72 h post-infection to spread the infection. In this time the recombinant protein is also expressed.

The initial virus titer and volume is too low for sufficient recombinant protein expression and thus several cycles (P1 → P2 → P3) of virus harvesting and new infection are required (Figure 6-4). Therefore, the infected insect cells were centrifuged and the supernatant with virus P1 was used to infect the larger amount of Sf21 insect cells. This procedure was repeated two more times and the final P3 virus was used further for protein expression in HighFive insect cell line. The cell growth and lysis was monitored throughout the virus generation using cell counter Vi-Cell-XR (Beckman Coulter).

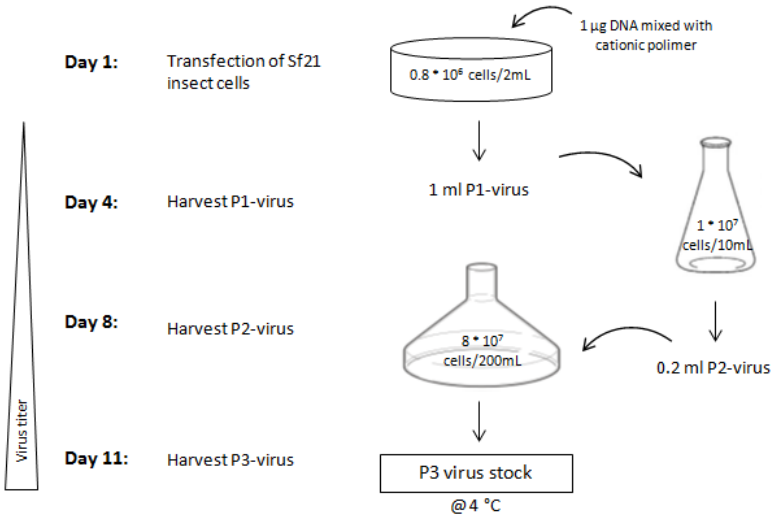


Figure 6-4. Virus generation in Sf21 insect cells.
Schematic representation of the virus generation. Three rounds of new cell culture infection were required to increase volume and virulence of virus stocks.

6.2.3.3 Recombinant protein test-expression in HighFive insect cells

In order to test recombinant protein expression in HighFive insect cell line, the same amounts of the cells (10^6 cells/mL in 3 mL) were infected with 0.15-300 µL of P3 virus to find out the optimal amount of virus. After 3 days of the expression, 400 µL of the infected cell cultures were centrifuged at 27°C and 800 x g for 15 min. The supernatant was removed and the pellet was dissolved in 4x SDS sample buffer for Western blot analysis.

6.2.4 Protein purification

6.2.4.1 Cell lysis

The cell pellet was thawed and resuspended in 100-300 mL lysis buffer supplemented with benzonase and protease inhibitor cocktail. All purification steps were done at 4°C. The cells were lysed using a sonificator at 40 % intensity for 0.5 s with 0.5 s break intervals until the cells were completely disrupted (8-15 min). After sonification cell debris and insoluble materials were removed by centrifugation (75 000 x g, 45 min) and the supernatant was filtered through 5-µm membrane (Millex-VG) for a further purification of the target protein.

6.2.4.2 Ni^{2+}-NTA affinity chromatography

The cleared lysate was passed over 5 mL Ni^{2+}-NTA beads (GE Healthcare), pre-equilibrated in the lysis buffer. The column was then washed with 250 mL high salt wash

buffer to remove contamination. An additional wash with 250 mL ATP wash buffer was performed to remove persistent chaperon contamination. Finally, 100 mL Ni Buffer A was passed through the column. The sample was eluted by 'on-column' tag cleavage with recombinant HRV 3C protease over night.

6.2.4.3 Ion exchange chromatography

To increase the purity of the protein, the elution fraction of Ni^{2+}-affinity was loaded to pre-equilibrated HiTrap Heparin HP column (GE Healthcare) using Buffer A for binding. A stepwise 20 % wash with Buffer B removed protein still containing nucleic acid contamination. Finally, Sen1 was eluted with a gradient over 4-6 CV to 100 % Buffer B. The purest fractions were pooled and used for further purification by size exclusion chromatography. Alternatively, the sample was centrifuged at 4000 g using 15 mL Amicon Ultra Centrifugal filters (Ultracel) concentrators with a molecular weight cut off (MWCO) of 50 kDa and flash frozen in liquid nitrogen and stored at -80°C.

6.2.4.4 Size exclusion chromatography

Shortly before injecting onto Superdex 200 [16/600] (GE Healthcare) the sample was again centrifuged in an eppendorf tube at 13.000 x g for 5 min to pellet insoluble components. The chromatography flow rate was set up at 0.5 mL/min, fractions of 0.5-1 mL were pooled and concentrated to 10-15 mg/mL.

6.2.5 SDS-PAGE

Every step of protein purification was analyzed on Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels with 10% (w/v) acrylamide (according to Laemmli, 1970). The respective samples were mixed with 2 x SDS loading dye and denatured at 95°C for 2 min. 10 µL of each sample was loaded. As a marker 3 µL of PageRuler unstained protein ladder (Thermo Fisher Scientific) was also loaded on the gel. Electrophoresis was performed in tanks with 1x SDS running buffer at 200 V until the running front reached the bottom of the gel. Thereafter, the gel was stained with Coomassie Brilliant Blue.

6.2.6 Denaturing PAGE

Denaturing PAGE was used to analyze RNA or DNA samples (according to (Summer *et al*, 2009). For RNase protection assays the samples were subjected to 22 % (w/v)

denaturing PAGE with 5 M urea. For RNA:DNA duplex unwinding and *in vitro* transcription termination assays the samples were subjected to 15 % (w/v) denaturing PAGE (8 M urea) and 10% (w/v) denaturing PAGE (8 M urea), respectively.

6.2.7 Western-blot

For Western-blot analysis, 15 µL of samples were loaded on 10 % SDS-PAGE gel to separate proteins by their size. After electrophoresis, the gels were incubated in Transfer buffer (20 mM trizma base, 150 mM glycine, 0.1 % SDS, 20 % methanol) for 15 min and the proteins were transferred onto a polyvinylidene difluoride (PVDF) membrane (Merck Millipore) for 80 min at 200 mA using cold Transfer buffer. To fix the transferred proteins, the membrane was incubated in 10 % HOAc for 15 min and air-dried. Afterwards rehydration in 1x Tris-buffered saline (TBS, 50 mM Tris pH 7.5, 150 mM NaCl) the membrane was blocked with 5 % non-fat dry milk in 1x TBS to prevent non-specific antibodies binding to the membrane by incubating for 1 h at room temperature, washed two times with TBS containing 0.05 % Tween 20, and once with TBS. Then the proteins were detected with the mouse anti-HA-tag antibodies followed by incubation with horseradish peroxidase (HRP) conjugated anti-Mouse secondary antibodies. Alternatively, the membranes were stained with Ponceau S solution (Sigma-Aldrich).

6.2.8 Measurements of protein concentration

Proteins concentrations were determined using a spectrophotometer (Nanodrop) to measure the UV absorbance of proteins at 280 nm. Extinction coefficient was calculated with ProtParam. The protein concentration (mg/mL) was calculated equal to $Absorbance_{280}$ (mg/mL)/extinction coefficient.

6.2.9 Protein storage

For short storage proteins were kept at 4°C or on ice. If needed to store for longer period, the proteins were mix with Sample freezing buffer (40 mM HEPES pH 7.5, 300 mM NaCl, 2 mM $MgCl_2$, 99% (w/v) glycerol) and stored at -80°C or at -20°C.

6.2.10 Mass spectrometry

For mass spectrometry analysis protein sample bands from SDS-gels were cut out or, alternatively, a liquid protein sample was analyzed. The peptide fingerprinting was used to determine the region that was cleaved endogenously during $Sen1_{976-1880}$ purification. The

samples were digested with trypsin; the peptides were analyzed by Orbitrap mass spectrometry, and identified using the Max Quant software. Peptide finger printing was performed by Dr. Nagarjuna Nagaraj, PhD (MPI Biochemistry Core Facility). The total mass spectrometry was used to assess the molecular mass of a protein in solution. ESI-TOF MS was performed by Elisabeth Weyher-Stingl (MPI Biochemistry Core Facility).

6.2.11 Edman sequencing

SDS-PAGE gel with a sample of $Sen1_{976-1880}$ was blotted onto PVDF membrane. The membrane was stained with Amido Black. N-terminal Sequencing of the first four residues was carried out by Dr. Josef Kellermann (MPI Microchemistry Core Facility)

6.2.12 ATP hydrolysis assay

ATPase assays were performed with 5 nM Sen1 proteins at 28°C in 10-µL reactions containing 10 mM Tris–HCl pH 7.5, 75 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT, 25% glycerol, and 50 ng/µL polyU. The reaction started with the addition of a 250 µM ATP solution containing 0.25 µM of 800 Ci/mmol a ^{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 analyzed by phosphor-imaging with Typhoon scanner (Fuji).

6.2.13 Fluorescence anisotropy

Fluorescence anisotropy measurements were performed by Dr. Claire Basquin, MPI Biochemistry. The 5'-end fluorescein-labeled 15-mer 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 (485 nm and 535 nm, respectively) were measured at 20°C in 50-µL reactions on Infinite M1000 Pro (Tecan). Each titration point was measured three times using ten reads. The data were analyzed by nonlinear regression fitting using the BIOEQS software (Royer, 1993).

6.2.14 RNase protection assay

Proteins (10 pmol each) were mixed with 5 pmol ^{32}P body-labeled RNA (kindly provided by Fabien Bonneau) 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 at 4°C for 1 h, the reaction mixtures were digested with 1 µg RNase A/T1 mix and 2.5 U RNase T1 (Fermentas) at 20°C for 20 min. Protected RNA fragments were then extracted twice with 400 µL phenol:chloroform:isoamyl alcohol (25:24:1 (v/v), Invitrogen), precipitated with 1 mL 100% ethanol. The RNA pellet was dried out in speed-vac at 45°C for 10min and then resuspended in 10 µL 10% bromphenol blue buffer. The samples were subjected to 22% (w/v) denaturing PAGE (4 M urea), and gels were visualized by phosphor-imaging (Fuji).

6.2.15 *In vitro* transcription termination assay

The termination assays were performed by Dr. Zhong Han (Institut Jacque Monod in Paris) as described in (Porrua & Libri, 2015a) (Figure 6-5). A ternary Pol II elongation complexes were assembled in a promoter-independent manner by incubating the RNA:DNA hybrid with purified RNA Pol II. Next, the non-template DNA strand and recombinant Rpb4/7 heterodimer were sequentially added to the mixture to form a artificial transcription bubble and complete elongation complex. Used RNA is FAM labeled and DNA is biotin labeled at the 5'-ends. The ternary Pol II elongation complexes 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 $MgCl_2$, 10 µM $ZnCl_2$, 10% glycerol, and 2 mM DTT. The experiment reactions were performed at 28°C in TB in a final volume of 20 µL in the absence or in the presence of 20–80 nM of Sen1 proteins. Transcription was initiated after addition of a 1 mM mixture of ATP, UTP, and CTP to allow transcription through the G-less cassette up to a G-stretch in the non-template strand. After 15 min the reactions were stopped with 1 µL of 0.5 M EDTA and the supernatant was separated from the beads. The beads fractions were resuspended in 8 µL of loading buffer (1× 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 µL of loading buffer. Transcripts were

subjected to 10% (w/v) denaturing PAGE (8 M urea), and gels were scanned with a Typhoon scanner (Fuji).

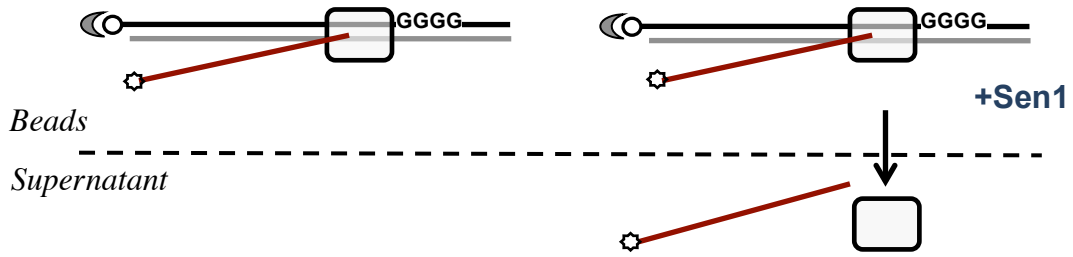


Figure 6-5. Schematic representation of *in vitro* transcription termination assay
The ternary Pol II elongation complex contains a biotinylated DNA that has a G-rich site to pause the transcription in the absence of GTP in the reaction. Biotinylated DNA allows immobilizing the complex on the streptavidin beads. If the helicase Sen1 is added to the reaction, it dissociates the elongation complex and the 5'-end FAM labeled transcripts are found in supernatant fraction. The transcription termination efficiency is calculated as percentage of released. FAM labeled RNA is shown in a dark red bar line and the Pol II shown in a rectangular box.

6.2.16 RNA:DNA duplex unwinding assay

Duplex unwinding assays were performed by Dr Zhong Han (Institut Jacque Monod in Paris). The RNA:DNA annealed using 44-mer RNA oligonucleotide with a 20-mer DNA to form a 5'-overhang of RNA. The unwinding reactions were mixed to a final 20 µL volume in 10 mM Tris–HCl pH 7.5, 50 mM NaCl, 7.5 µM $ZnCl_2$, 0.5 mM DTT, 10% (w/v) glycerol, and 0.1 mg/ml BSA at 28°C. Sen1 proteins were preincubated with the corresponding duplex substrate, and the reaction was initiated by adding a 2 mM ATP, 2 mM $MgCl_2$, and 0.1 µM of unlabeled DNA oligonucleotide 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% (w/v) SDS, and 20% (w/v) glycerol. Samples were subjected to electrophoresis on a 15% (w/v) native PAGE, and gels were directly scanned using Typhoon scanner (Fuji).

6.3 X-ray crystallography

6.3.1 Protein crystallization

Purified $Sen1_{Hel}$ was diluted to 3-4 mg/mL (30-35 μM) concentration in SEC buffer containing 20 mM HEPES pH 7.5, 300 mM NaCl, 2 mM $MgCl_2$, and 0.5 mM TCEP. Protein was then mixed with a 10-fold molar excess of freshly prepared ATP analogues and a 1.2 molar excess of RNA. Initial screening was carried out using both commercially available screens and in-house screens. The 100 - 500 nL drops were set up 96-well plates by mixing to a 1:1 volume ration of protein and crystallization solution using a Phoenix pipetting robot. The plates were stored in the incubators at 4°C, 10°C, or 18°C. The crystal grow was monitored using an XtalFocus robot (ExploraNova, La Rochelle, France) or under a microscope. Once a crystallization hit was observed, the grid screening of initial precipitant solution (salt, pH, PEGs, additives) was pursued. The Crystallization facility of the MPI Biochemistry carried all automated screenings and drop visualization.

The best crystals were obtained at 4°C by hanging-drop vapor diffusion from 2-μL 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). The drops formed a film on a surface and many of the crystals were attached to it, hence for crystal mounting the drops had to be opened on the side and the crystals were scooped with a cryo loop (Hampton Research), equilibrated in a range of cryoprotectants and flash frozen in liquid nitrogen. Screening on the in-house D8 VENTURE (Bruker) crystallography system showed 25-28% (w/v) ethylene glycol to be the most effective cryoprotectant.

6.3.2 Data collection

All diffraction data was collected at the super-bending magnet beamline X06DA (PXIII) at the Swiss Light Source (Villigen, Switzerland). Native data was collected at 1.0 Å wavelength using 0.1° oscillations, exposure time of 0.1sec/frame.

The S-SAD data collection was performed at 2.095 Å as a compromise to maximize the anomalous signal while minimizing absorption. However, collecting many low-dose data sets (<0.5 MGy per 360°) from a single crystal and merging the data sets leads to the enhanced anomalous signal-to-noise ratio. The beamline is set up with a multi-axis goniometer PRIGo, which has an inversed head that is mounted on an air-bearing stage and can rotate with a high precision at two angles (χ and φ) (Waltersperger *et al*, 2015). By

collecting data in different orientations the systematic errors are reduced as the same reflections are measured in different diffraction geometries on different areas of the detector. Multi-orientation data collection was carried out on a single crystal. The 4 x 360° ω scans were collected at 5° χ increments from 0 to 15° while keeping the φ orientation constant, using 0.1° ω oscillations, 0.1sec/frame exposure and using a noise-free pixel-array PILATUS 2M-F detector (Henrich *et al*, 2009) at a sample-to-detector distance of 120 mm.

6.3.3 Data processing and structure solution

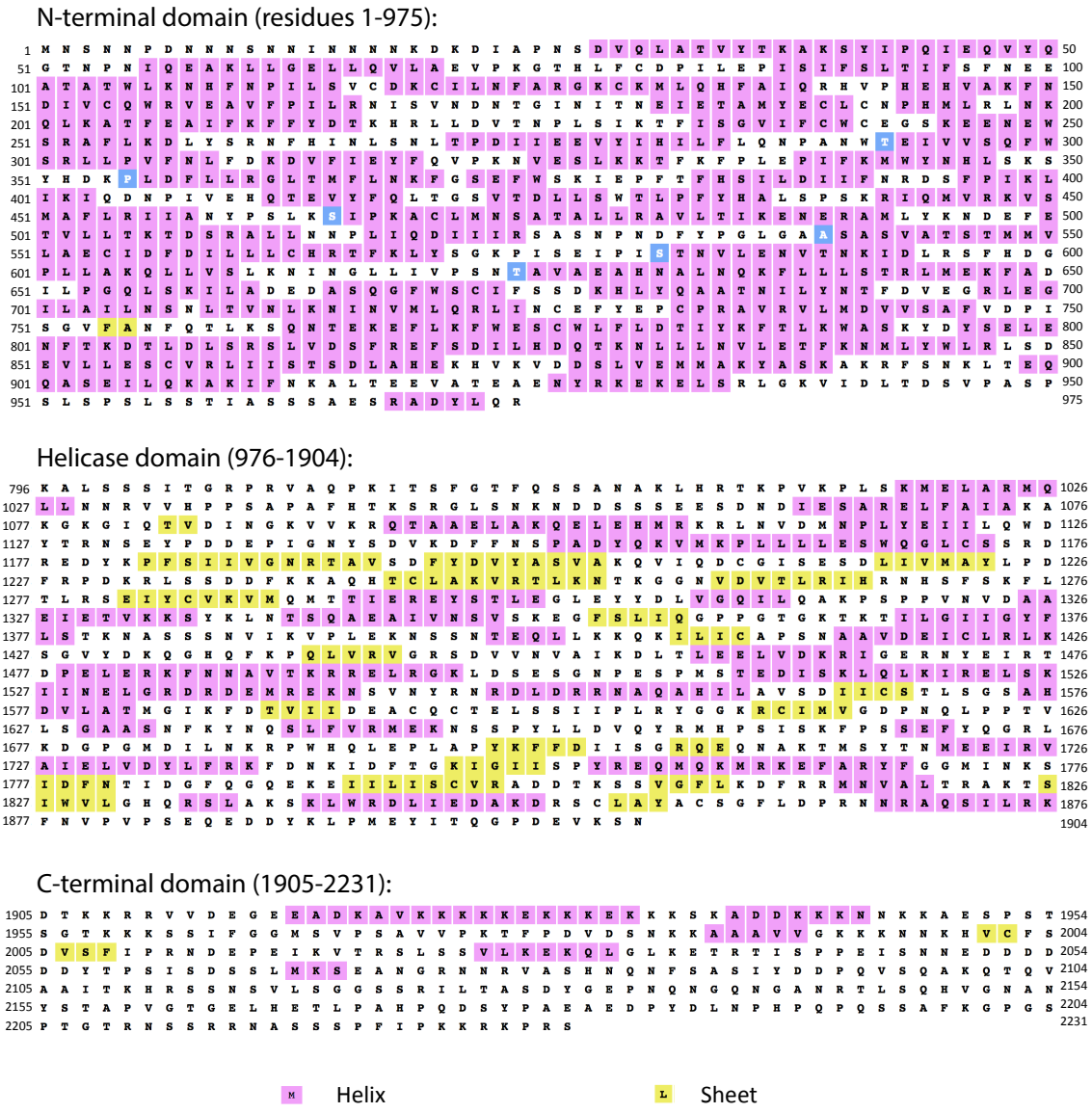
The data was processed using XDS and scaled and merged with XSCALE (Kabsch, 2010). The high-resolution data cutoff was based on the statistical indicators CC1/2 and CC* (Karplus & Diederichs, 2012). Substructure determination and phasing were performed with SHELXC/D/E (Sheldrick, 2010) using the HKL2MAP 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.

6.3.4 Model building and refinement

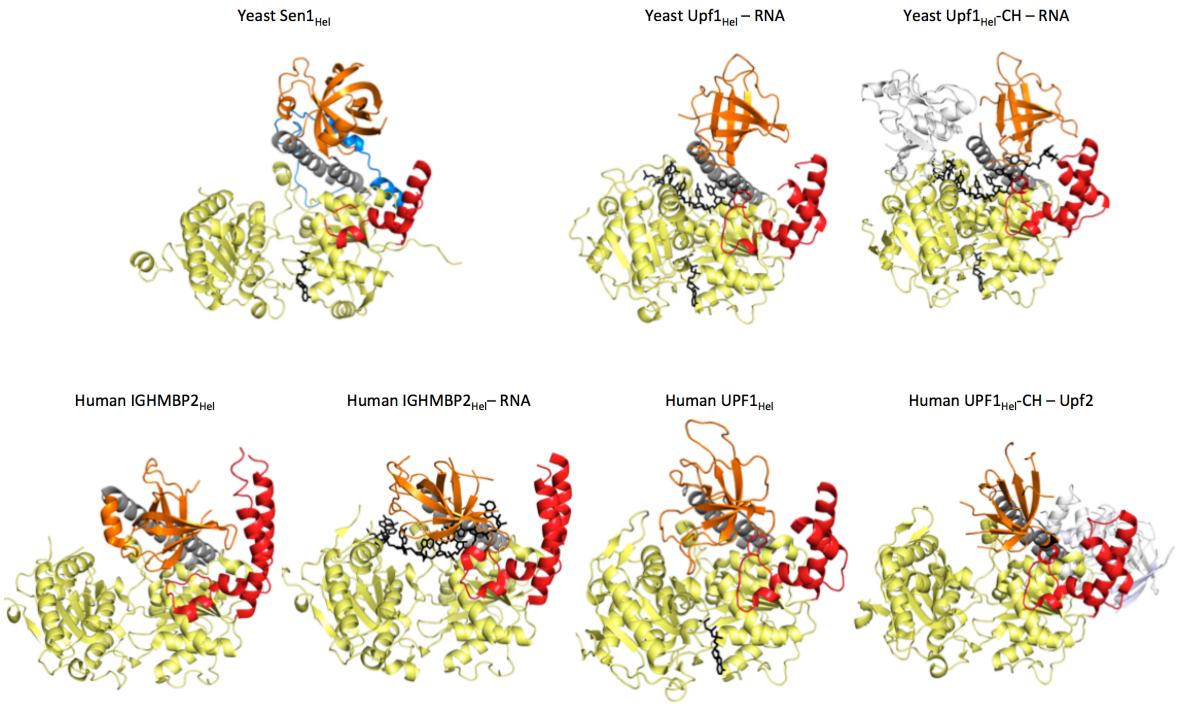
An initial model was built automatically with BUCCANEER (Cowtan, 2006) and completed with iterative rounds of manual in the experimental electron density in COOT (Emsley & Cowtan, 2004) and refinement against the native data with PHENIX.refine (Adams *et al*, 2010). The model was validated using PHENIX or in wwPDB online tools.

Figures were generated with PyMOL (Schrödinger). Structural superpositions were performed with the structure matching function of COOT or alignment function of PyMOL.

Appendix



Appendix 1. PSIPRED secondary structure prediction of *S. cerevisiae* Sen1.
Because of too large protein size, the secondary structure prediction was done in portions: the N-terminal domain (1-975), the helicase domain (976-1904), and the C-terminal domain (1905-2231).



Appendix 2. All currently available structures of Upf1-like helicases.
The helicases are colored the same according their domains. The CH-domain of Upf1 is shown in light grey and Upf2 is shown in light purple. PDB codes (left to right): 5MZN, 2XZO and 2XZL (top), 4B3F, 4B3G, 2XZO and 2WJV (bottom).

	Sen1 variants	Expression	RNA binding	ATPase	Unwinding	Termination
Test-expressions	1-2231	✗				
	28-937	✓				
	28-983	✓				
	28-1859	✗				
	28-1877	✗				
	28-1880	✗				
	43-1880	✗				
	43-937	✗				
	43-983	✓				
	43-1859	✗				
	43-1877	✗				
	1095-1859	✓				
	976-1880	✓				
	1055-1880	✓				
	1084-1880	✗				
	1095-1880	✓				
	1095-1910	✓				
	1105-1904	✓				
wt	1095-1904 (=Sen1$_{Hel}$):	✓	++	++	++	++
"Brace"	1148-1904	✗				
	1135-1904	✗				
	1128-1904	✗				
	1148-1904,W1166S, L1669N	✗				
	*L1121R	✗				
	*L1121R, W1132R	✗				
	*Y1117R	✗				
	*Y1117R, L1121R	✗				
	*Y1117R, L1121R, W1132R	✗				
"Prong"	*Δ1470-GSGS-1539 (=ΔUP)	✓	++	+	++++	+
	*Δ1460-SGG-1555 (=ΔLP)	✓	++	++++	-	-
	*ΔUP, L1549D	✓			++	+/-
	*L1549D	✓		+++	+	+/-
"Barrel"	*K1246E	✓				++
	*R1248E	✓				++
	*R1263E	✓				++
	*R1188D, Y1195F	✓				++
Loops	*Δ1381-GG-1403	✓	++	++	++	+++
	*Δ1443-GGG-1452	✓	++	++	+	+
	*Δ1426-GGG-1439	✓	++	++	+	+
Disease mutations	*E1591Q	✓		-		-
	*R1820Q	✓	+	-		-
	*T1779P	✓	-		+/-	-
	*T1779P, R1813A	✓	-			-
	*P1622L	✓	+		-	-
	*N1413S	✓	-		+/-	-
	*N1413S, T1568A	✓	-			
	*T1289A, R1293A	✓	-		-	-
	*D1616V	✗				

Appendix 3. List of Sen1 variants.
Asterisk (*) marks mutations that were introduced to Sen1$_{Hel}$ construct. The constructs that did not expressed are marked in red; insoluble – in orange; light green and green indicate low and good expression levels, respectively. The helicase activity is scaled from inactive (-) to enhanced activity (++++) according to wild-type Sen1$_{Hel}$.

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

Alzu A, Bermejo R, Begnis M, Lucca C, Piccini D, Carotenuto W, Saponaro M, Brambati A, Cocito A, Foiani M & Liberi G (2012) Senataxin associates with replication forks to protect fork integrity across RNA-polymerase-II-transcribed genes. *Cell* **151:** 835–846

Andersen PR, Domanski M, Kristiansen MS, Storvall H, Ntini E, Verheggen C, Schein A, Bunkenborg J, Poser I, Hallais M, Sandberg R, Hyman A, LaCava J, Rout MP, Andersen JS, Bertrand E & Jensen TH (2013) The human cap-binding complex is functionally connected to the nuclear RNA exosome. *Nat. Struct. Mol. Biol.* **20:** 1367–1376

Bacikova V, Pasulka J, Kubicek K & Stefl R (2014) Structure and semi-sequence-specific RNA binding of Nrd1. *Nucleic Acids Res.* **42:** 8024–8038

Bennett CL & La Spada AR (2015) Unwinding the role of senataxin in neurodegeneration. *Discov Med* **19:** 127–136

Bennett CL, Chen Y, Vignali M, Lo RS, Mason AG, Unal A, Huq Saifee NP, Fields S & La Spada AR (2013) Protein interaction analysis of senataxin and the ALS4 L389S mutant yields insights into senataxin post-translational modification and uncovers mutant-specific binding with a brain cytoplasmic RNA-encoded peptide. *PLoS ONE* **8:** e78837

Berretta J & Morillon A (2009) Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* **10:** 973–982

Borggrefe T, Davis R, Bareket-Samish A & Kornberg RD (2001) Quantitation of the RNA polymerase II transcription machinery in yeast. *J. Biol. Chem.* **276:** 47150–47153

Buratowski S (2009) Progression through the RNA polymerase II CTD cycle. *Mol. Cell* **36:** 541–546

Butler JS & Mitchell P (2011) Rrp6, rrp47 and cofactors of the nuclear exosome. *Adv. Exp. Med. Biol.* **702:** 91–104

Cakiroglu SA, Zaugg JB & Luscombe NM (2016) Backmasking in the yeast genome: encoding overlapping information for protein-coding and RNA degradation. *Nucleic*

Acids Res. **44:** 8065–8072

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

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:** 6241–6252

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, Müller 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:** 577–590

Chen X, Poorey K, Carver MN, Müller U, Bekiranov S, Auble DT & Brow DA (2017a) Transcriptomes of six mutants in the Sen1 pathway reveal combinatorial control of transcription termination across the Saccharomyces cerevisiae genome. *PLoS Genet.* **13:** e1006863

Chen X, Poorey K, Carver MN, Müller U, Bekiranov S, Auble DT & Brow DA (2017b) Transcriptomes of six mutants in the Sen1 pathway reveal combinatorial control of transcription termination across the Saccharomyces cerevisiae genome. *PLoS Genet.* **13:** e1006863

Cheng Z, Muhlrad D, Lim MK, Parker R & Song H (2007) Structural and functional insights into the human Upf1 helicase core. *EMBO J.* **26:** 253–264

Chinchilla K, Rodriguez-Molina JB, Ursic D, Finkel JS, Ansari AZ & Culbertson MR (2012) Interactions of Sen1, Nrd1, and Nab3 with multiple phosphorylated forms of the Rpb1 C-terminal domain in Saccharomyces cerevisiae. *Eukaryotic Cell* **11:** 417–429

Clerici M, Mourao A, Gutsche I, Gehring NH, Hentze MW, Kulozik A, Kadlec J, Sattler M & Cusack S (2009) Crystal structure of the complex between human nonsense mediated decay factors UPF1 and UPF2 Orthorhombic form.

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 Buccaneersoftware for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62:** 1002–1011

Creamer TJ, Darby MM, Jamonnak N, Schaughency P, Hao H, Wheelan SJ & Corden JL (2011) Transcriptome-Wide Binding Sites for Components of the Saccharomyces cerevisiae Non-Poly(A) Termination Pathway: Nrd1, Nab3, and Sen1. *PLoS Genet.* **7:** e1002329

Darby MM, Serebreni L, Pan X, Boeke JD & Corden JL (2012) The Saccharomyces cerevisiae Nrd1-Nab3 transcription termination pathway acts in opposition to Ras signaling and mediates response to nutrient depletion. *Molecular and Cellular Biology* **32:** 1762–1775

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW & Steinmetz LM (2006) A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. U.S.A.* **103:** 5320–5325

DeMarini DJ, Papa FR, Swaminathan S, Ursic D, Rasmussen TP, Culbertson MR & Hochstrasser M (1995) The yeast SEN3 gene encodes a regulatory subunit of the 26S proteasome complex required for ubiquitin-dependent protein degradation in vivo. *Molecular and Cellular Biology* **15:** 6311–6321

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:** 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, Guenther U-P & Jankowsky E (2010) SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **20:** 313–324

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 Genet.* **11:** e1005044

Franco-Echevarría E, González-Polo N, Zorrilla S, Martínez-Lumbreras S, Santiveri CM, Campos-Olivas R, Sánchez M, Calvo O, González B & Pérez-Cañadillas JM (2017) The structure of transcription termination factor Nrd1 reveals an original mode for GUAA recognition. *Nucleic Acids Res.* **45:** 10293–10305

Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425:** 737–741

Grzechnik P, Gdula MR & Proudfoot NJ (2015) Pcf11 orchestrates transcription

termination pathways in yeast. *Genes Dev.* **29:** 849–861

Gudipati RK, Neil H, Feuerbach F, Malabat C & Jacquier A (2012a) The yeast RPL9B gene is regulated by modulation between two modes of transcription termination. *EMBO J.* **31:** 2427–2437

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

Gudipati RK, Xu Z, Lebreton A, Séraphin B, Steinmetz LM, Jacquier A & Libri D (2012b) Extensive degradation of RNA precursors by the exosome in wild-type cells. *Mol. Cell* **48:** 409–421

Halbach F, Rode M & Conti E (2012) The crystal structure of S. cerevisiae Ski2, a DExH helicase associated with the cytoplasmic functions of the exosome. *RNA* **18:** 124–134

Hamperl S & Cimprich KA (2014) The contribution of co-transcriptional RNA:DNA hybrid structures to DNA damage and genome instability. *DNA Repair (Amst.)* **19:** 84–94

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

Hazelbaker DZ, Marquardt S, Wlotzka W & Buratowski S (2013) Kinetic Competition between RNA Polymerase II and Sen1-Dependent Transcription Termination. *Mol. Cell* **49:** 55–66

Heidemann M, Hintermair C, Voß K & Eick D (2013) Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. *Biochim. Biophys. Acta* **1829:** 55–62

Henrich B, Bergamaschi A, Broennimann C, Dinapoli R, Eikenberry EF, Johnson I, Kobas M, Kraft P, Mozzanica A & Schmitt B (2009) PILATUS: A single photon counting pixel detector for X-ray applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **607:** 247–249

Heo D-H, Yoo I, Kong J, Lidschreiber M, Mayer A, Choi B-Y, Hahn Y, Cramer P, Buratowski S & Kim M (2013) The RNA polymerase II C-terminal domain-interacting domain of yeast Nrd1 contributes to the choice of termination pathway and couples to RNA processing by the nuclear exosome. *J. Biol. Chem.* **288:** 36676–36690

Hobor F, Pergoli R, Kubicek K, Hrossova D, Bacikova V, Zimmermann M, Pasulka J, Hofr C, Vanacova S & Stefl R (2011) Recognition of transcription termination signal

by the nuclear polyadenylated RNA-binding (NAB) 3 protein. *J. Biol. Chem.* **286:** 3645–3657

Jamonnak N, Creamer TJ, Darby MM, Schaughency P, Wheelan SJ & Corden JL (2011) Yeast Nrd1, Nab3, and Sen1 transcriptome-wide binding maps suggest multiple roles in post-transcriptional RNA processing. *RNA* **17:** 2011–2025

Jankowsky E & Fairman ME (2007) RNA helicases--one fold for many functions. *Curr. Opin. Struct. Biol.* **17:** 316–324

Jenks MH, O'Rourke TW & Reines D (2008) Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Molecular and Cellular Biology* **28:** 3883–3893

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

Kilchert C, Wittmann S & Vasiljeva L (2016) The regulation and functions of the nuclear RNA exosome complex. *Nat. Rev. Mol. Cell Biol.* **17:** 227–239

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

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

Kim K-Y & Levin DE (2011) Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. *Cell* **144:** 745–756

Kim M, Suh H, Cho E-J & Buratowski S (2009) Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2, 5, and 7. *J. Biol. Chem.* **284:** 26421–26426

Komarnitsky P, Cho EJ & Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* **14:** 2452–2460

Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M &

Yanagawa H (2009) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.* **284:** 478–485

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

Kuehner JN, Pearson EL & Moore C (2011) Unravelling the means to an end: RNA polymerase II transcription termination. *Nat. Rev. Mol. Cell Biol.* **12:** 283–294

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

Lenstra TL, Tudek A, Clauder S, Xu Z, Pachis ST, van Leenen D, Kemmeren P, Steinmetz LM, Libri D & Holstege FCP (2013) The role of Ctk1 kinase in termination of small non-coding RNAs. *PLoS ONE* **8:** e80495

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

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 Res.* **44:** 6794–6802

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

Loya TJ, O'Rourke TW & Reines D (2013a) Yeast Nab3 protein contains a self-assembly domain found in human heterogeneous nuclear ribonucleoprotein-C (hnRNP-C) that is necessary for transcription termination. *J. Biol. Chem.* **288:** 2111–2117

Loya TJ, O'Rourke TW, Degtyareva N & Reines D (2013b) A network of interdependent molecular interactions describes a higher order Nrd1-Nab3 complex involved in yeast transcription termination. *J. Biol. Chem.* **288:** 34158–34167

Makino DL, Schuch B, Stegmann E, Baumgärtner M, Basquin C & Conti E (2015) RNA degradation paths in a 12-subunit nuclear exosome complex. *Nature* **524:** 54–58

Marquardt S, Hazelbaker DZ & Buratowski S (2011) Distinct RNA degradation pathways and 3' extensions of yeast non-coding RNA species. *Transcription* **2:** 145–154

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

Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, Kremmer E, Eick D & Cramer P (2012) CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II. *Science* **336:** 1723–1725

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

Nedea E, Nalbant D, Xia D, Theoharis NT, Suter B, Richardson CJ, Tatchell K, Kislinger T, Greenblatt JF & Nagy PL (2008) The Glc7 phosphatase subunit of the cleavage and polyadenylation factor is essential for transcription termination on snoRNA genes. *Mol. Cell* **29:** 577–587

O'Rourke TW, Loya TJ, Head PE, Horton JR & Reines D (2015) Amyloid-like assembly of the low complexity domain of yeast Nab3. *Prion* **9:** 34–47

Pape T & Schneider TR (2004) HKL2MAP: a graphical user interface for macromolecular phasing with SHELXprograms. *Journal of Applied Crystallography* **37:** 843–844

Pearson E & Moore C (2014) The evolutionarily conserved Pol II flap loop contributes to proper transcription termination on short yeast genes. *Cell Rep* **9:** 821–828

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) Characterization of the mechanisms of transcription termination by the helicase Sen1. *Methods Mol. Biol.* **1259:** 313–331

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

Porrua O, Hobor F, Boulay J, Kubicek K, D'Aubenton-Carafa Y, Gudipati 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. *Annual Review of Biophysics* **37:** 317–336

Richard P & Manley JL (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.* **23:** 1247–1269

Rondón AG, Mischo HE, Kawauchi J & Proudfoot NJ (2009) Fail-safe transcriptional termination for protein-coding genes in S. cerevisiae. *Mol. Cell* **36:** 88–98

Saikrishnan K, Powell B, Cook NJ, Webb MR & Wigley DB (2009) Mechanistic basis of 5'-3' translocation in SF1B helicases. *Cell* **137:** 849–859

Sariki SK, Sahu PK, Golla U, Singh V, Azad GK & Tomar RS (2016) Sen1, the homolog of human Senataxin, is critical for cell survival through regulation of redox homeostasis, mitochondrial function, and the TOR pathway in Saccharomyces cerevisiae. *FEBS J.* **283:** 4056–4083

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

Schuch B, Feigenbutz M, Makino DL, Falk S, Basquin C, Mitchell P & Conti E (2014) The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *EMBO J.* **33:** 2829–2846

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

Singh V, Azad GK, Sariki SK & Tomar RS (2015) Flocculation in Saccharomyces cerevisiae is regulated by RNA/DNA helicase Sen1p. *FEBS Lett.* **589:** 3165–3174

Singleton MR, Dillingham MS & Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu. Rev. Biochem.* **76:** 23–50

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

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

Summer H, Grämer R & Dröge P (2009) Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *J Vis Exp*: e1485–e1485

Sydow JF, Brueckner F, Cheung ACM, Damsma GE, Dengl S, Lehmann E, Vassylyev D

& Cramer P (2009) Structural Basis of Transcription: Mismatch-Specific Fidelity Mechanisms and Paused RNA Polymerase II with Frayed RNA. *Mol. Cell* **34:** 710–721

Tan-Wong SM, Zaugg JB, Camblong J, Xu Z, Zhang DW, Mischo HE, Ansari AZ, Luscombe NM, Steinmetz LM & Proudfoot NJ (2012) Gene loops enhance transcriptional directionality. *Science* **338:** 671–675

Terzi N, Churchman LS, Vasiljeva L, Weissman J & Buratowski S (2011) H3K4 Trimethylation by Set1 Promotes Efficient Termination by the Nrd1-Nab3-Sen1 Pathway. *Molecular and Cellular Biology* **31:** 3569–3583

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 (1997) The yeast SEN1 gene is required for the processing of diverse RNA classes. *Nucleic Acids Res.* **25:** 4778–4785

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

Ursic D, DeMarini DJ & Culbertson MR (1995) Inactivation of the yeast Sen1 protein affects the localization of nucleolar proteins. *MGG Molecular & General Genetics* **249:** 571–584

Vasiljeva L & Buratowski S (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell* **21:** 239–248

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. *Nat. Struct. Mol. Biol.* **15:** 795–804

Waltersperger S, Olieric V, Pradervand C, Glettig 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

Whitehouse I, Rando OJ, Delrow J & Tsukiyama T (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450:** 1031–1035

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

Wyers F, Rougemaille M, Badis G, Rousselle J-C, Dufour M-E, Boulay J, Régnault B, Devaux F, Namane A, Séraphin B, 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

Xu Z, Wei W, Gagneur J, Perocchi F, Clauder-Münster S, Camblong J, Guffanti E, Stutz F, Huber W & Steinmetz LM (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457:** 1033–1037