Emerging Roles of SPT5 in Transcription

Vivek Pandey    Shirani Punniyamoorthy    Yuba Raj Pokharel

Faculty of Life Sciences and Biotechnology, South Asian University, New Delhi, Delhi, India

Keywords
SPT5 • DSIF • RNAP I • Cancer • Neurodegenerative disease

Abstract
Suppressor of Ty homolog-5 (SPT5) discovered in the yeast mutant screens as a suppressor of mutation caused by the insertion of the Transposons of yeast (Ty) element along with SPT4, with which it forms a holoenzyme complex known as DRB sensitivity-inducing factor (DSIF) and plays an essential role in the regulation of transcription. SPT5 is a highly conserved protein across all three domains of life and performs critical functions in transcription, starting from promoter-proximal pausing to termination. We also highlight the emerging role of SPT5 in other non-canonical functions, such as the regulation of post-translational modifications (PTM) and the transcriptional regulation of non-coding genes. Also, in brief, we highlight the clinical implications of SPT5 dysregulation.

Introduction

Transcription is a highly orchestrated event divided into three phases: initiation, elongation, and termination [1][2][3][4]. The RNAP II interacts with several transcription and accessory factors during the above phases. These factors perform functions ranging from transcription to pre-mRNA processing [5]. The interaction of proteins with the RNAP II is dynamic and is coordinated spatially and temporally with the transcription machinery driving signals from intrinsic and extrinsic cellular signalling. Different proteins involved in different phases of transcription associate with the RNAP in a transitional manner. However, a few of them remain associated throughout the process. SPT5/SUPT5H is amongst some of the few proteins that remain associated with the RNAP throughout the transcription process.

SPT5/SUPT5H is one of the few universally conserved proteins across all three domains of life. Suggestive of the essential function it codes for even before the different life forms diverged. Our understanding of proteins and their role has increased in recent years, due to molecular, biophysical, and structural advances in proteins. In this review, we describe the Suppressor of Ty homolog-5 (SPT5/SUPT5H) as one of the few proteins that remain attached to the RNAP II and RNAP I from initiation to termination.

© 2023 The Author(s). Published by Cell Physiol Biochem Press GmbH&Co. KG
Discovery

SPT5/SUPT5H is a nuclear protein that regulates the transcription machinery of the cells. It is among the few proteins that are conserved across all three domains of life, suggesting its indispensability [6]. In bacteria, the SPT5 homolog is called NusG and functions as a monomer [7]. However, in archaea and eukaryotes, it functions as a heterodimer where it forms complexes with SPT4, another protein isolated in the same genetic screens and regulates the transcription [8]. SPT5 together with SPT4 forms a heterodimeric complex referred to as the DRB sensitivity-inducing factor (DSIF) that plays a role in regulating transcription [9][10][11].

SUPT5H is the human homolog of the yeast SPT5 protein. It was discovered in the yeast mutant screens as the mutant repressing mutations caused by the insertion of Ty element (His4-912δ) in the 5'noncoding region of the His4 gene in the Saccharomyces cerevisiae genetic screens [12][13]. Since the gene family was identified as suppressing the Ty element, the family was named Spt [9] [14]. Several other studies reported similar suppression of multiple histone mutations. Chromosome 13 codes for SPT5 in S. cerevisiae, while its human homolog is coded by chromosome 19[15]. SPT5 plays an essential role in transcription and regulates it in various steps, from promoter-proximal pausing to elongation and termination of RNAPII [16][17][18].

Structure

Structurally, SPT5 can be divided into five broad regions, each having a unique function. These include the N-terminal acidic region, a NusG N terminal (NGN) domain, multiple Kyprides-Ouzounis-Woese (KOW) domains, a C-terminal region (CTR) heptapeptide repetitive motif and the C-terminal [19].

The N-terminal region is a 150-170 amino acid long region composed of acidic amino acids and containing the nuclear localization signal at the 66th position. The signal is essential for the nuclear localization of the protein since its removal leads to the cytoplasmic localization of the protein [20]. A recent study by Park et al. highlighted the importance of the SPT5 N-terminal domain for the control of divergent antisense transcription at the transcription start site (TSS) [21]. For the first time, the study highlights the importance of N-terminal phosphorylation event of the SPT5 by the ZWC complex in the regulation of antisense transcription[22].

Adjacent to the N-terminal region is the NGN domain of the SPT5. It is the most conserved domain of the protein, remaining conserved throughout evolution in all three domains of life over the course of evolution [23]. The NGN domain harbours the site required for interaction with the RNAP II and SPT4 to form the DSIF complex [24][25]. The domain interacts with both the nascent RNA and the non-coding strand of the DNA. The hydrophobic concave depression on the surface of SPT5 interacts with the tip of the clamped coil domain of the RNAPII [26]. This interacting region remains largely intact in all domains of life (Fig 1A, B). In archaea, alanine-4, tyrosine-42, and leucine-44 line in these hydrophobic pockets. Mutagenesis studies revealed that mutation of these residues significantly alters the binding to the RNAP.

Furthermore, replacing the ten residues of the clamp-coil motif with the polyglycine abolished the ability of RNAP II to interact with the NGN domain, thereby establishing conserved complementary binding. The NGN domain binds to the above DNA-RNA in the RNAP active site, thus locking the cleft for the nucleotide, resulting in increased elongation and decreased disassociation [26]. Furthermore, the negatively charged DNA also interacts with the positively charged surface of the Spt5, preventing the collapse of the transcription bubble. The NGN domain alters RNAP activity by altering the interaction with the clamp domain, thereby allosterically regulating the polymerase.
Linked to the NGN domain is the C-terminal KOW domain [27]. It is the other motif of SPT5, which remains relatively conserved across different domains. However, the number of KOW domains varies in each domain of life. The bacteria and archaea have a single KOW domain; however, comparatively, eukaryotes have multiple KOW domains ranging from four to six (Fig. 1A) [28][29][30]. The KOW domain is essential for survival since deletion leads to survival defects. KOW deletion mutants in the yeast showed severe growth defects.[31]

The KOW domains of SPT5 adopt a fold similar to the β barrel of the Tudor domain, which known to interact with nucleic acid. The Tudor domain is structurally compact (~50 amino acids) and, facilitates protein-protein and protein-nucleic acid interactions [32]. In addition to the interactions with the nucleic acid, KOW domains of SPT5 could also be of essential importance for the coupling of the transcriptional modifications and the RNAP (Fig. 1C).

In eukaryotes, the first KOW domain folds together with the linker 1, the polypeptide sequence linking KOW1 to KOW2. KOW1-Linker1 (K1L1) folds into a bilobed rigid structure that serves as a DNA clamp. The N-terminal lobe is formed by the K1, which resembles the β barrel of the Tudor domain, along with some C-terminal sequences of L1 that form the final β strand of the β sheets. In contrast, the C-terminal lobe is formed by inserting sequences in the Tudor barrel. K1 and L1 have highly conserved interface residues in all eukaryotes, stabilized by five hydrogen bonds, van der Wall interaction, and a cation–π electron interaction [31]. High-resolution X-ray structures revealed a positively charged patch in K1L1 that binds to the nucleic acid in a sequence-independent manner. In vitro studies showed that K1L1 prefers dsDNA over ssDNA and ssRNA. The function of K1L1 is complementary to the SPT4, which is also positively charged. Double mutation analysis of K1L1 and Spt4 caused severe growth arrest, which was rescued in the presence of wild-type Spt4, confirming the functional overlap [31]. K1L1 is followed by the KOW2-3 tandem domain, having two Tudor domain β barrel structures separated by a single residue in the KOWX-4-KOW5 domain [33]. KOWX refers to pre-KOW4 region that forms the RNA clamp domain with KOW4-KOW5. KOWX-4

Fig. 1. SPT5 organization in different domains of life. (A) Domain organization of SPT5-like proteins present in Bacteria, Archaea, and Eukaryotes modified from Decker et al. Bacterial NusG protein and archaeal SPT5 proteins are comparatively small, consisting of either an N-terminal domain (NTD) or a NusG N-terminal domain (NGN) and a single KOW domain. In contrast, yeast SPT5 consists of an additional N-terminal acidic region followed by an NGN domain, 1-5 KOW domains, and a C-terminal repeat (CTR). Human SUPT5H consists of all the elements on the yeast SPT5 and an additional CTR followed by two additional KOW domains. Domain boundaries are drawn to proportionally according to their polypeptide lengths. (B) A cladogram showing evolutionary relatedness between different SPT5/NusG belonging to the different domains of life. (C) Structure of SPT5 (orange) bound to SPT4 (green) at transcription paused state. Other entities removed for better understanding (PDB ID: 6GML)
is a positively charged region that contacts the negatively charged RNA coming out of the ternary complex and blocks it from the action of RNase. The KOW4-KOW5 linker features an essential phosphorylation site at S666 responsible for the productive elongation of the transcription complex. Acute loss or substitution at S666 results in proteasomal degradation to impaired pause release [34].

In addition to all of the mentioned domains, which are common to all eukaryotes, metazoans are characterized by an additional KOW domain at the end of the C-terminal repeat, i.e. KOW6-7 [35]. These domains are essential for maintaining SPT5's negative regulatory role, as mutations result in defects in negative regulatory function, resulting in compromised growth [36].

Adjacent to the KOW domain lies the C-terminal repeat, which contains one to two repeating heptapeptide domains essential for transcription elongation [19][37]. The CTR mutant SPT5 is defective in transcription induction and causes lethality in the diploid state in S.cerevisiae; however, loss of CTR did not have such a lethal effect on inflammatory response genes in humans apart from slow growth [38][39]. CTR is the phosphorylation site for positive transcription elongation factor b (P-TEFb), which is responsible for phosphorylation of RNAP II CTD and recruitment of multiple elongation factors [40]. In addition, the CTR is the landing platform for several other proteins involved in the various transcription-coupled activities such as 5’capping, phosphorylation, 3’processing, and many others [41].

The C-terminal repeat terminates into the C-terminal domain; much is unknown. However, together with the CTR, it is thought to play a role in recruitment of proteins such as Paf-1, which is responsible for transcription elongation [42].

**Biological Functions of SPT5**

As a part of the DSIF complex, SPT5 associates RNAP II and regulates promoter-proximal pausing (Fig. 2A,B) [43]. The complex makes multiple contacts with the RNAP II around the active cleft and exit channel, thereby regulating its moment. In addition, several studies reported that SPT5 shares the same site as the transcription factors TFIIB, TFIIE, and TFIIF (Fig. 2C). In a validation study, it was suggested that the spt5-spt4 complex replaces the TFIIE, resulting in the promoter clearance of the pre-elongation complex (Fig. 2D) [44]. Therefore, it is evident that binding of SPT5 to the complex stabilizes the stalled complex. The stabilization of the stalled complex is further enhanced by the interaction of SPT5 with a protein known as the negative elongation factor (NELF)[45]. NELF binding alters the DNA-RNA hybrid in the transcription complex and impairs the NTP binding site, leaving the transcription complex in the paused state (Fig. 2E) [46][47]. The paused state is overcome upon the binding of P-TEFb to the complex [48]. P-TEFb is a heterodimeric complex of cyclin T1 and Cdk9 [49]. P-TEFb phosphorylates the SPT5, primarily at the two phosphorylation hot spot loci: the KOW4-KOW5 linker region and the C-terminal repeat domain [50]. Phosphorylation at the above site results in the release of NELF, resulting in transition from paused elongation complex to the active elongation complex. Removal of NELF alters the conformation, allowing entry of the nucleotides and progression of the transcription complexes (Fig. 2F) [51]. In addition to regulating the transcription, SPT5 ensures a constant rate of transcription and uniform distribution of RNAP II across the gene. In the absence of SPT5, a reduction in RNAP II levels and accumulation at the 5’ end of the gene occurs, further validated by the reduced level of mRNA post-SPT5 inhibition. A recent study by Aoi et al. reported an acute loss of SPT5 that led to the degradation of RPB1 of the RNAP II by the action of Cullin 3, VCP/p97 and CDK9 complex [52]. In addition, SPT5 was found to play an essential role in repressing antisense transcription. SPT5 inhibits two classes of antisense transcription, including convergent transcription downstream of the TSS, and the divergent transcription, upstream of the TSS [53][54].
SPT5 plays essential roles by interacting with various RNA processing and modification systems enroute to transcription [55]. It has been reported to resemble the Pol II CTD; the SPT5-CTD can also recruit the capping complex during transcription onto the newly formed RNA (Fig. 3A) [56]. In contrast to RNAP II CTD, in which phosphorylation of Ser 5 plays an essential role in the Guanylyl transferase (GTase) recruitment to the capping site, an unphosphorylated CTD of SPT5 is crucial to recruit the capping complex [57]. Upon phosphorylation at the Thr 1 position of SPT5-CTD, uncoupling of the capping complex from RNAP II occurs, resulting in release of the capping apparatus. Another independent study performed on yeast showed similar results and reported co-purification of capping enzymes Cet1 and Ceg1. Additionally, Abd1 with additional methyltransferase activity was co-purified along with the capping enzymes cet1 and ceg1(Fig. 3B) [58]. In addition to playing a role in 5' processing activity, SPT5 is also critical in regulating the 3' processing. Mayer et al., reported in their study on the importance of SPT5-CTR in the recruitment of pre-mRNA cleavage factor I (CFI) at the 3'end of S. cerevisiae [59]. In addition, the group also reported CFI in the SPT5-CTR pull-down in vitro. The mutant SPT5 Δ CTR showed a reduced occupancy of the elongation factor Paf1 and the CFI. NusG is reported to have a similar function in the Rho-dependent termination of transcription by the bacteria [60]. In addition to recruitment of various pre-mRNA processing intermediates onto RNA, SPT5 also plays a role in recruitment of complexes such as Paf1, involved in histone modifications, including methylation [61].

Eukaryotic DNA is packed as a nucleosome, with the DNA tightly coiled around histones and the nucleosome acting as a transcription barrier [62]. SPT5 has been reported to facilitate transcription across the nucleosome barrier by acting as a histone chaperone similar to the function of FACT and SPT6 [63][64][65]. This function is attributed to the intrinsically disorganized region in the N-terminus of SPT5. The N-terminal is thought to initiate the development and binding of DNA to H2A and H2B, thus facilitating the transcription through the nucleosome (Fig 3B) [66].
In addition to serving as a recruitment platform for several RNA processing intermediates, SPT5 is also known to play a significant role in suppressing transcription-coupled DNA repair (TCR) [67]. TCR is one of the two pathways involved in nucleotide excision repair (NER), the other being global genome repair (GGR) [68]. It has been reported that the SPT5-CTR is essential in suppressing the Rad26-independent TCR. The yeast cell with a mutant SPT5-CTR showed an enhanced transcriptional activity compared to the wild-type in the yeast strain lacking Rad16 and Rad26 involved in the GGR and TCR, respectively [69]. Also, SPT4, part of the DSIF complex, showed a similar phenotype. Similar to SPT5-CTR, SPT4 also showed the ability to suppress the Rad26-independent CTR (Fig. 3C) [70].

Several other studies suggest the crucial role of SPT5 and its partner SPT4 in non-homologous end joining (NHEJ) and homologous recombination (HR), two major DNA repairing mechanisms. Inhibition of SPT5 resulted in a severely compromised NHEJ repair pathway of the damaged DNA in H1299d-Ad3. PCR analysis of the damaged DNA region revealed significant resurrection and lengthy insertion. A similar effect was seen in CH12F3-2A, which showed an increased sensitivity towards the ionizing radiation. A similar decline in the potential to repair DSBs was seen in GM7166VA cell lines, which further validates its
role in TCR. These roles played by SPT5 in DNA repair mechanisms are essential for class switch recombination (CSR) in B cell, required for switching antibodies from IgM to IgA.[71] [72]. CH12F3-2A upon induction with CIT (CD40L, IL4, and TGFβ) resulted in switching from IgM to IgA. Reduced antibody switching was the outcome of SPT5/SPT4 inhibition, which is associated with a decline in the S region cleavage following SPT5 inhibition (Fig. 3D) [73].

Further, SPT5 plays a critical role in the PTM of several histones and modifications of the nucleosome complex. SPT5 knock-down (KD) is known to downregulate the basal and induced level of H3K4me3 in the NFkB-regulated gene and downregulate the level of H4K5Ac at the promoter-proximal site (Fig. 4D)[74]. Downregulation in the H3K4me3 & H4K5Ac results in an increased promoter-proximal pausing, suggesting SPT5 is essential for acetylation and methylation of the nucleosome at the promoter-proximal site. These PTM regulated by the SPT5 are essential in recruiting TFIIID and releasing TFIIIE and the mediator complex from RNAP II [75]. In addition to controlling the methylation and acetylation of the histone markers, SPT5 is also methylated by a number of arginine methyltransferases, including PRMT1 and PRMT5, which affects its interaction with RNAP II and its potential for further elongation [76].

**SPT5 in non-coding RNA synthesis**

Non-coding RNA refers to the large class of non-protein coding RNAs that are known to activate and regulate a variety of protein-coding genes and active genes. Since non-coding RNA is involved in a variety of cellular and developmental processes, it has attracted a lot of interest in recent years [77]. Many non-coding RNAs are known to be involved in multiple disease pathogenesis, while others are engaged in the regulation of diseases. Therefore, it is very interesting to study the biology of non-coding RNA.

![Fig. 4. Roles of SUPT5H/SPT5 in disease pathogenesis.](image-url)

(A) SPT5 supports the continuous cell proliferation of cancer cells by activating hTERT, an essential element for telomere maintenance. (B) Haploinsufficient mutation in the SUPT5H gene results in the downregulation of the HBB gene, causing β-thalassemia. (C) SPT4/SPT5 complex plays a role in synthesizing proteins with long stretches of CAG, producing insoluble proteins that cause neurodegenerative disorders, including Huntington’s disease. (D) SPT5 plays a crucial role in the transcription of inflammation and stress-induced genes. SPT5 (PDB ID: 6GML).
The SPT5-SPT4 complex is also reported to regulate the expression level of non-coding RNA in a manner similar to how it controls the synthesis of protein-coding RNA. The role of SPT4 and SPT5 in controlling the expression of the germline genome was demonstrated by a study by Owain et al. in the Paramecium tetraurelia. SPT4-SPT5 complex played a significant role in the genome rearrangement, necessary for the development of P.tetraurelia, by being required for the synthesis of scnRNA that is required for generating sRNA [78]. Another group in an independent research revealed that SPT5 suppression reduced the expression of eRNA in both Drosophila and humans [79][72]. The non-coding RNA known as enhancers is crucial for several regulatory processes connected to the activation of genes [80]. SPT5 plays a critical role here by providing much-needed stability to the RNAP II at the promoter region of enhancers. Due to decreased stability, RNAPII is significantly more susceptible to premature termination at the enhancer location of the promoter than it is at the site of the protein-coding gene. In addition to SPT5 directly controlling eRNA synthesis, it is also known that eRNA indirectly controls SPT5 activity by downregulating or upregulating various of inhibitors and activators. eRNAs play an important role in releasing NELF, resulting in a transition elongation complex [81]. SPT5 is crucial for chromatin accessibility, as was described in the preceding section [82]. Inhibition of SPT5 restricted access to enhancers and lowered SWI/SNF occupancy, both of which are necessary for chromatin remodelling [83].

**SPT5 Role in Disease Pathogenesis**

The human analog of SPT5, SUPT5H has been linked to a number of diseases. According to Chen et al., SUPT5H acts as an activator for hTERT and increases its reverse transcriptase activity [84]. Telomere maintenance in cancer cells depends on increased hTERT activity [85]. The proliferative and metastatic capacity of colon cancer cells was reduced by SPT5 suppression by shRNA. A similar study by Bilal et al. reported the role of SUPT5H in breast cancer proliferation. The metastatic potential was reduced and the survival-related MAPK signalling was significantly downregulated when SUPT5H was inhibited [86]. Additionally, SUPT5H was interacting with the Paf complex, which is known to be involved in the development of tumours. According to a recent study, SUPT5H plays a prognostic function in predicting the likelihood of developing intrahepatic cholangiocarcinoma (ICC) (Fig. 4A) [87].

SPT5 not only plays essential role in cancer progression, but also plays a crucial function in HIV transcription. It interacts with HIV-Tat1 and preserves its stability at the transcription site. SPT5 knockdown lowered pol II and Cdk9 occupancy and accelerated HIV entry into latency [88][89]. Additionally, it is known that SUPT5H controls the HBB gene, which codes for the beta-globin gene. The HBB protein, which is accountable for β-thalassemia, was downregulated as a result of a haploinsufficient mutation on the SUPT5H (Fig. 4B, D) [90].

Recent research reports the SPT4-SPT5 complex's involvement in neurodegenerative diseases. They have reportedly contributed to the production of insoluble proteins that cause neurodegenerative diseases by synthesising proteins with lengthy CAG stretches [91]. Frontotemporal dementia (FTD), Huntington's disease (HD), and amyotrophic lateral sclerosis (ALS) have all been linked to the SPT4/SPT5 complex [92]. RNA interference-mediated inhibition of Spt4/Spt5 reduced the amount of mutant Htt protein having extensive CAG stretches The short CAG protein, however, is unaffected by the aforementioned interference(Fig. 4C)[93].

**Discussion**

SPT5 performs important functions in the transcription process from the promoter-proximal pause to the termination, making it clear that SPT5 is a crucial part of the transcription machinery in all three domains of life. It also associates with different RNA
modifiers, causing varying RNA modifications during the course of transcription. However, there are still a number of unanswered concerns regarding the precise function of SPT5 at various stages of disease aetiology and regulation. The global and gene-specific roles of SPT5 in transcription regulation is the most important issues that require further study. It has been discovered that blocking SPT5 in yeast reduced the expression of only few sets of essential genes that are important for its development. However, blocking SPT5 in human and Drosophila caused to have reduced expression of majority of their genes [94]. Therefore, studies focusing on SPT5’s overall and gene-specific roles in transcriptional control are required.

It is well known that SPT5 is essential for controlling the both RNAP I, which is in charge of rRNA synthesis, and RNAP II[95][96]. Although it is well known that SPT5 affects RNAP I in both a positive and a negative manner, further research is required to define the molecular mechanism and the protein interactome [97].

Additionally, more research needs to be done on the SPT5 phosphorylation landscape. Therefore, it is necessary to look into how phosphorylation at various sites affect the transcriptional machinery in order to better our understanding. The role of SPT5 in various metabolic stressors, how it responds to them, and the crucial factors that cause it to switch from a regulator to an activator at the promoter-proximal site are all things that need to be understood in order for this process to proceed. Numerous studies have emphasised the crucial function of SPT5 and the DSIF complex in controlling non-coding RNA involved in a variety of processes, such as development, gene regulation, and the activation of the target gene via enhancer function, in addition to controlling transcription at the protein-coding genes. Some studies have highlighted the role of SPT5 in the epigenetic landscape, as SPT5 inhibition resulted in altered expression of acetylation and the methylation markers. Therefore, thorough research showing the precise part SPT5 plays in altering the epigenetic landscape is required.

Several studies highlighted the oncogenic role of SUPT5H in the activation of MAPK, Myc, and hTERT proteins [84][98][99]. Apoptosis, cell cycle arrest, and senescence are just a few of the pathways that are activated when SUPT5H is knocked down that are important in preventing cell proliferation [84] [99][100]. Different SUPT5H knocked-down cell lines showed differential outcomes regarding apoptosis or senescence upon SUPT5H inhibition. Therefore, it is necessary to investigate the cause of either one or both pathways becoming engaged. SPT5 is also involved in the pathogenesis of a number of other illnesses, including inherited illnesses like thalassemia and neurological diseases. However, research aimed at novel paradigms in this direction is still needed to fully understand its role as a transcriptional regulator in these disorders.

**Acknowledgements**

We gratefully acknowledge South Asian University, Rajpur Road, Maidan Garhi, New Delhi, India, for the institutional support, Vivek Pandey would like to acknowledge the Department of Biotechnology, INDIA, for providing Senior Research Fellowship to accomplish the study.

**Author Contributions**

VP and YRP conceptualized the review, VP did the review writing, VP, SP & YRP performed the editing, VP and SP made figures.

**Funding Sources**

We thank the Department of Biotechnology, Government of India & South Asian University.
Statement of Ethics

The authors have no ethical conflicts to disclose.

Disclosure Statement

The authors have no conflicts of interest to declare.

References


