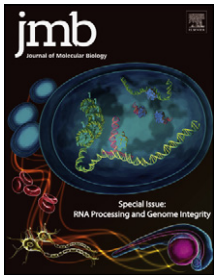
Review



Senataxin: Genome Guardian at the Interface of Transcription and Neurodegeneration

Matthias Groh, Laura Oana Albulescu, Agnese Cristini and Natalia Gromak

Sir William Dunn School of Pathology, *University of Oxford, South Parks Road, OX1 3RE, UK*

Correspondence to Natalia Gromak: *natalia.gromak@path.ox.ac.uk*
http://dx.doi.org/10.1016/j.jmb.2016.10.021
Edited by de Almeida Sergio Fernandes

Abstract

R-loops comprise an RNA/DNA hybrid and a displaced single-stranded DNA. They play crucial biological functions and are implicated in neurological diseases, including ataxias, amyotrophic lateral sclerosis, nucleotide expansion disorders (Friedreich ataxia and fragile X syndrome), and cancer. Currently, it is unclear which mechanisms cause R-loop structures to become pathogenic. The RNA/DNA helicase senataxin (SETX) is one of the best characterised R-loop-binding factors *in vivo.* Mutations in SETX are linked to two neurodegenerative disorders: ataxia with oculomotor apraxia type 2 (AOA2) and amyotrophic lateral sclerosis type 4 (ALS4). SETX is known to play a role in transcription, neurogenesis, and antiviral response. Here, we review the causes of R-loop dysregulation in neurodegenerative diseases and how these structures contribute to pathomechanisms. We will discuss the importance of SETX as a genome guardian in suppressing aberrant R-loop formation and analyse how SETX mutations can lead to neurodegeneration in AOA2/ALS4. Finally, we will discuss the implications for other R-loop-associated neurodegenerative diseases and point to future therapeutic approaches to treat these disorders.

Introduction

Neurodegenerative diseases are conventionally associated with the accumulation of mutated toxic proteins, which can misfold or aggregate in human cells, interfering with normal cellular function. Despite the unifying characteristic of protein aggregation, the affected proteins, the pathomechanism, and the resulting impairment of different pathways differ widely amongst these diseases. Proteins most commonly found in aggregates in neurodegenerative diseases are amyloid-β, associated with Alzheimer's disease (AD), tau (AD and Pick's disease), α-synuclein (AD, Parkinson's disease, and others), and superoxide dismutase 1 (SOD1), mutated in amyotrophic lateral sclerosis (ALS) type 1 [1]. The physiological functions of these proteins are associated with cytoskeletal organisation (tau) or regulation of reactive oxygen species (SOD1) but remain poorly understood for others (amyloid-β and α-synuclein). Intriguingly, mutations in proteins involved in RNA biology such as FUS and TDP43 can also lead to severe ALS (ALS6 and ALS10, respectively) and ALS with frontotemporal lobar degeneration (ALS-FTD) [2–5]. Growing evidence suggests that not only the proteins but the RNA itself may contribute to human pathologies, which is clearly demonstrated by the trinucleotide repeat expansion diseases [6,7]. Very recently, unusual RNA/DNA structures called R-loops have drawn much attention due to their ambivalent nature and their high pathogenic potential.

R-loops consist of RNA/DNA hybrids, formed during transcription when nascent RNA hybridises to the DNA template strand, displacing the non-template DNA strand [8]. R-loops are implicated in many cellular processes in a wide range of organisms including bacteria, yeast, plants, mice, and humans [9]. Amongst others, R-loops play beneficial roles in regulating chromatin architecture, transcription termination, and generation of antibody diversity [10–12]. However, R-loop levels are under tight control, and their dysregulation leads to transcriptional defects, altered chromatin structure, and genome instability—processes that are potentially contributing to an emerging class of R-loop-associated diseases (reviewed in Ref. [13]). These include several neurological diseases such as Friedreich ataxia

(FRDA), fragile X syndrome (FXS), C9orf72-associated ALS-FTD, and Prader–Willi syndrome. However, it remains largely unclear how R-loops become pathogenic and what are the protein factors that control cellular R-loop levels. In this regard, senataxin (SETX) is an exception, since this is one of the best functionally characterised R-loop-associated proteins.

Interest in SETX emerged when it was discovered that mutations in the *SETX* gene are linked to two disorders: the recessive ataxia with oculomotor apraxia type 2 (AOA2; OMIM 606002) and the autosomal dominant ALS type 4 (OMIM 602433) [14–17]. AOA2 is characterised by progressive cerebellar ataxia, oculomotor apraxia, peripheral neuropathy, and increased levels of serum alpha-fetoprotein [14,15]. AOA2 is considered the second most common autosomal recessive cerebellar ataxia after Friedreich ataxia [18]. The pathology of ALS4 substantially differs from AOA2, and it is associated with the progressive degeneration of motor neurons in the brain and spinal cord, resulting in muscle weakness and atrophy [16]. SETX is a putative RNA/DNA helicase, and ALS4/AOA2 diseases have been linked either to the mutations in the helicase domain or the N-terminal putative protein–protein interaction domain [14–16,19–24] (Fig. 1a).

Saccharomyces cerevisiae Sen1 Protein

Given the high conservation within the helicase domain between human SETX and its *S. cerevisiae* homologue splicing endonuclease 1 (Sen1), extensive

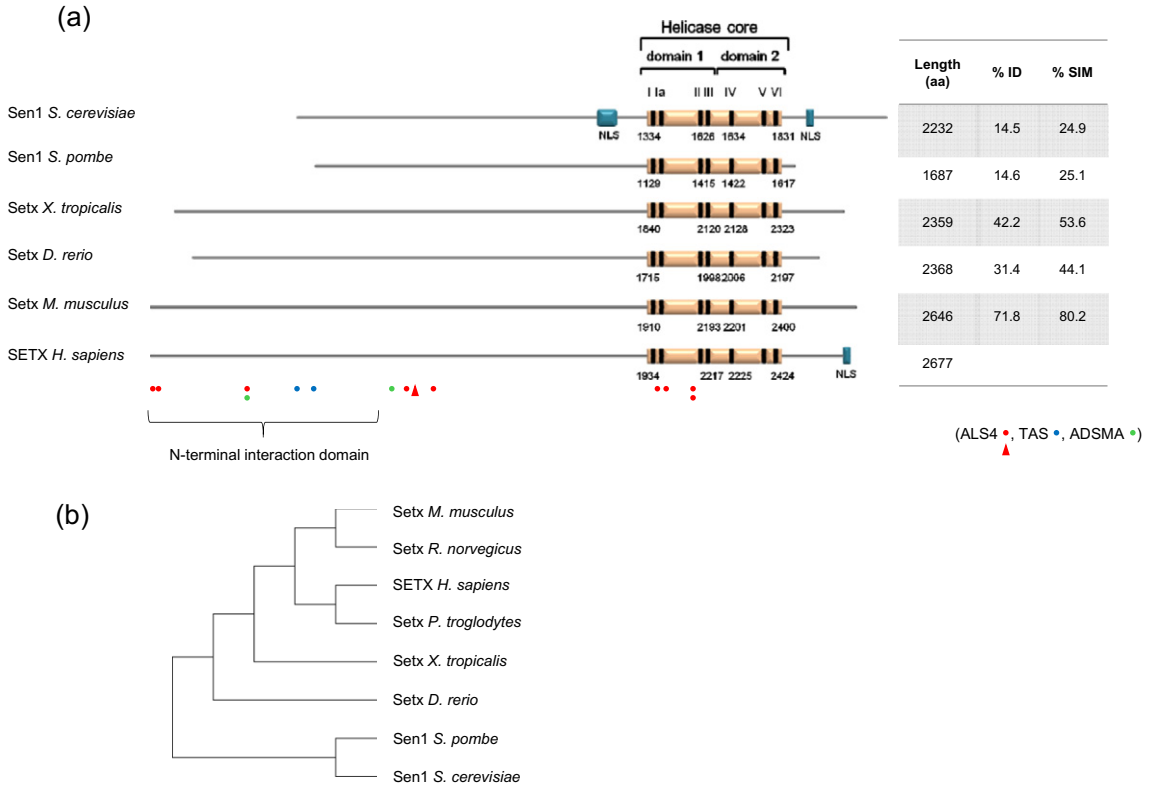


Fig. 1. Conservation of SETX across different species. (a) Multiple sequence alignment of SETX homologues from six species: *S. cerevisiae* sen1 (NCBI accession number Q00416), *S. pombe* sen1 (Q92355), *X. tropicalis* setx (F6ZZC8), *D. rerio* setx (E7FBJ2), *M. musculus* setx (A2AKX3), and *H. sapiens* SETX (Q7Z333). The conserved helicase domains and their subdomains are highlighted by coloured boxes in orange and black. Numbers indicate amino acid residues. The positions of NLS sequences (*S. cerevisiae* Sen1 aa 1004–1089 and 1908–1929, and Human SETX aa 2661–2677) are indicated in blue box. Percentage of identity (% ID, defined as percentage of amino acids that are identical between the two compared species) and similarity (% SIM, defined as percentage of amino acids that share the same charge or are hydrophobic or hydrophilic as defined in Sequence Manipulation Suite: Ident&Sim at www.bioinformatics.org) with human SETX are provided in the table on the right. The location of disease mutations in SETX is shown in coloured dots: ALS4 (red), TAS (blue), and ADSMA (green; based on Leiden Open Variation Database). AOA2 patient mutations identified throughout SETX protein are not depicted on this diagram due to their extensive number (~120 mutations, including missense, nonsense, splice site, frameshift, and insertion/deletion mutations). The triangle represents the insertion. Details of sequence alignment are in the Supplementary material. (b) Phylogenetic alignment for the senataxin gene. Phylogenetic tree of SETX homologues across eight taxa, generated by alignment for the conserved helicase domain, as specified in the Supplementary material. The phylogenetic tree was generated using the Maximum Likelihood method.