Aging device: tau protein transposes the telomerase fragment into promoter/enhancer to promote gene copy number decrement regulation

Wenzhong Liu [1,*], Hualan Li[2]

[1] School of Computer Science and Engineering, Sichuan University of Science & Engineering, Zigong, 643002, China;
[2] School of Life Science and Food Engineering, Yibin University, Yibin, 644000, China;
[*]Correspondence: Wenzhong Liu, liuwz@suse.edu.cn.

Abstract

The reasons for the significant impact of abnormal tau protein phosphorylation and aggregation on Alzheimer's disease are still unknown. This study used bioinformatics methods, including protein domain search tools, to investigate the pathogenic mechanism of tau protein in neurodegenerative diseases. The study reveals that tau protein possesses domains linked to the Aging Device. These domains include pyruvate kinase, phosphatase, telomere binding, telomere transposition, HNH cas9, replicon binding, helicase, DNA polymerase, nuclease, transcription factors-like, promoter binding (TATA-box), enhancer binding (Homeobox, MADS-box, HMG box), and mitochondrial localization and mtDNA polymerase. It implies that the tau protein's pyruvate kinase domain supplies ATP energy for its Aging Device function. The HNH Cas9 , reverse transcriptase, and transposase domains cut the sequence between "TAA" from the telomere and then transcript the guide RNA sequence. The transcription factor domain slides to the specific DNA transcription factor binding region. The HNH Cas9 domain excises the intron DNA region complementary to the guide RAN sequence. The transposase and nuclease domains splice the "TAA" from the promoter or enhancer and add the telomere fragment sequence. We thought that tau protein transposes telomere fragments to the promoter/enhancer region might mess up RNA polymerase II, which would help the Aging Device's telomere-guided gene (DNA in the nucleus or mitochondrial) copy number decrement regulation system work better. It can help us understand how stress injuries, like low blood sugar or oxygen levels, can cause tau protein to clump together and phosphorylate, which turns off the Aging Device in a way that doesn't make sense. Thus, the anomalous work of aging devices, such as tau protein, may be associated with neurodegenerative diseases like Alzheimer disease, amyotrophic lateral sclerosis.

Keywords: Phosphorylation; Aggregation; Pyruvate kinase; Telomere transposition; TATA-box; Homeobox; Hypoglycemia/hypoxia

1. Background

Hereditary abnormalities have a role in the development of Alzheimer's disease. Pathological hallmarks of the Alzheimer's disease family of neurodegenerative disorders include tau protein accumulation in brain tissue, heterochromatin decondensation, piwi and piRNA loss, and resulting transposable element dysregulation[1]. The cytoskeleton is disrupted by hyperphosphorylated Tau, which causes synaptic dysfunction and cell death[2]. Global chromatin relaxation is one mechanism by which tau promotes neurodegeneration; loss of heterochromatin is common in Alzheimer's disease in humans; and aberrant gene expression results from loss of heterochromatin in tauopathies[3]. There are presently 21 recognised chromosomes that are associated with Alzheimer's disease, 19 with APOE, 17 with Tau, 14 with PS-1, and 1 with PS-2[4]. An aberrant gene on chromosome 1, for instance, has the potential to cause over 350 disorders, such as cancer, mental retardation, Alzheimer's disease, and Parkinson's disease[5].

There is some evidence linking TL to an increased risk of Alzheimer's disease[6]. Brain structural problems,

telomere changes, neurological complaints, and mental health issues are all rather prevalent[7]. Telomeres are protein-sequence complexes that stop chromosomal fusion and the loss of coding DNA caused by inadequate replication of linear DNA[8]. Specialised non-coding DNA repeats make up telomeres[9]. Because of issues with end-replication, telomeres—the "protective caps" of chromosomes—shorten with each cell cycle[10]. Many age-related disorders, including Alzheimer's disease, are linked to this process, which is known as telomere attrition[10]. Telomere attrition can be caused by clinical conditions; however, short telomeres can initiate pathological processes by elevating oxidative stress and inflammation[11], and increased risk of dementia is associated with longer telomeres[12].

The role of telomerase and TERT in protecting the developing brain[13]. When it comes to neurodegenerative illnesses like Alzheimer's and Parkinson's, TERT offers both protective and therapeutic effects[14]. Telomerase synthesises head-to-tail short tandem repeats, which make up the telomeric sequences of the majority of eukaryotic organisms[8]. In order to preserve and extend the length of human telomeres, which are the ends of linear chromosomes, telomerase uses the protein subunit telomerase reverse transcriptase (TERT) and the RNA component TERC[15]. By catalysing RT (TERT), telomerase adds nucleotides to the ends of chromosomes using intrinsic RNA molecules (TER) as templates[16].

In neurodegenerative disorders like Alzheimer's disease (AD) and Parkinson's disease (PD), the telomerase protein TERT can ameliorate harmful proteins like beta amyloid, pathogenic tau, and alpha-synucleinimpact[17]. As the Braak stage progresses, the amount of hyperphosphorylated tau protein that forms neurofibrillary threads and tangles diminishes, and TERT protein is primarily located in neurons in a normal brain[18]. Hippocampal neurons from Alzheimer's disease brains have a greater mitochondrial localization of the TERT protein[19]. There appears to be a novel function for TERT protein in neurons of brain tissue. Reducing oxidative stress is possible by localization to mitochondria. Once DNA damage reaches a specific degree, telomerase is unable to continue its function[20].

It is only at the very end of a fragmented chromosome that euchromatic DNA can contain telomere-specific features. It appears that these fragments, mostly from UTRs, are passively transported to non-telomeric heterochromatin regions rather than actively translocated[21]. Furthermore, they are found in these regions. Each newly transposable element is twisted such that its 5' end coincides with the chromosome end, because telomeric elements are retrotranscribed to chromosome ends. These telomere ends are dynamic, just as the ones maintained by telomerase [21].

In numerous neurodegenerative and ND models, heterochromatin relaxation resulting in TE derepression has been documented. Some pathogenic proteins linked to ND (such as tau and TDP-43) may regulate TE expression, according to the available evidence[22]. Inflammatory reactions to double-stranded RNA and/or cytosine nucleic acid, as well as the neurotoxic effects of retrotransposon products in amyotrophic lateral sclerosis (ALS) and Alzheimer's disease (AD) are the most prevalent pathogenic mechanisms of particular TE-derived products[23].

The function of retrotransposable elements in degenerative illnesses associated with ageing[24]. Because neuronal cells are more open to retrotransposition[25], transposon insertions occur more frequently in brain than in other somatic tissues[26]. Neurotoxic Substance Response Reverse transcriptases (TEs) are categorised into DNA transposons and retrotransposons (27). Many people are curious about the effects of active TEs on human brain development and illness because of the genetic diversity and mutations they can create[27]. Genomes with an increased frequency of mutations and copy number variations (CNVs) may also contain transposition of retrotransposons (LINE1) caused by short telomeres[28].

The tau protein is involved in Alzheimer's disease by triggering the abnormal mobilisation and activation of transposable elements[29]. In tau-mediated neurodegenerative diseases, such as Alzheimer's, transposable elements are activated[30]. When microglia and astrocytes detect an excessive amount of tau protein, they trigger the

transcription of a transposable element (TE) [31]. Cellular models show that Tau enhances L1 retrotransposon insertion[32]. Neuroinflammation can occur when the innate immune system is activated by tau protein-induced TE production, which can be seen in the cytoplasm and endosomes[31].

In neurons, tau protein can bind to DNA and localise to the nucleus[33]. The protein is organised into constitutive heterochromatin and co-localizes with the nucleolar protein nucleolin and human AT-rich α satellite DNA sequences[34]. Through its interactions with microtubules and chromatin, tau contributes to the integrity of chromosomes[35]. By interacting with certain DNA sequences in the nucleolus and pericentromeric heterochromatin, tau protein is able to localise to specific regions of the nucleus[36]. The role of tau in transposon activation and additional accelerated ageing have been associated with increased quantities of nuclear tau[36]. Some RNA genes' heterochromatinizationand/or nucleolar organisation are facilitated by nuclear tau protein[34].

An in vitro study found that human tau protein suppresses DNA replication[37]. It was demonstrated that tau binding to the template results in inhibition [37]. The ribosomal DNA locus is where Tau and other upstream binding factors are recruited[38]. Upon tau downregulation, a new subclass of ultrafine anaphase bridges forms in human ribosomes, which is caused by a decrease in ribonucleotide levels, ribosomal RNA synthesis, and recruitment of upstream binding factors[38]. In Alzheimer's disease, pathogenic Tau is linked to the cytoplasmic mislocalization of the RNA polymerase II subunit RPB1[39]. Before Tau phosphorylation and neurofibrillary tangles form in Alzheimer's disease, neuronal RNA levels are decreased and RNA polymerase II is hyperphosphorylated[40].

At the termini of linear chromosomes, nucleotide sequences known as telomeres repeat tandemly. Their primary function is to safeguard chromosome ends from degradation and end-to-end fusion, thereby ensuring genome stability[41]. The number of repeats in human telomere sequences varies between cells and even between chromosomes, and they are composed of C1-3A/TG1-3 and TTAGGG/CCCTAA repeats[42]. After DNA replication is finished, the terminal primer is degraded, which leads to the loss of a small segment of DNA at the 5' end of the lagging strand[43]. Additionally, the linear template's average length of approximately 250 nucleotides cannot be replicated at the end[44]. The "end replication problem" is named after the progressive shortening of telomeres that results from the incomplete replication of DNA polymerase at the chromosome end with each successive cell division[45]. DNA double-strand breaks, which are triggered by short or dysfunctional telomeres, are identified as the cause of cell replicative senescence[46].

To investigate tau protein, we implemented bioinformatics methodologies, including protein structure domain search tools. The results indicated that tau protein contains domains such as pyruvate kinase, phosphatase, telomere binding and transposition, HNH cas9, replicators binding, helicase, DNA polymerase and nuclease, similar transcription factors, promoter binding (TATA-box), enhancer binding (Homeobox, MADS-box, HMG box), and mitochondrial localization. It means that tau protein could transpose telomere fragments to promoter/enhancer regions, which then mess up RNA polymerase II and lower the number of copies of DNA genes (in the nucleus or mitochondrial) .

2. Data sets and methods

2.1. Related Proteins

Related proteins were downloaded from Uniprot Database. The list is shown below (Table 1):

Table 1. Related Sequences

No.	Related protein	Keywords	Count
1	Pyruvate kinase	Eukaryotic RNA Recognition Motif	17,668
		mg+ serinethreonine kinase	99,438
		PKC	113,271
2	Phosphatase	bacterial + Phosphatases	46,752
3	Telomeres	telomere	151,992
4	Transposase	human_transposase	42,432
		Transposition	559,748
		Homologousrecombination	299,886
		Nonhomologousrecombination	30,144
5	Replicon and helicase	Humandnareplication	312,393
6	Nucleases and DNA polymerases	BacterialEndonuclease	245,844
		BacterialExonuclease	202,428
		HumanEndonuclease	607,148
		HumanExonuclease	47,908
7	Transcription Factor	Humantranscriptionfactor	306,568
8	Promoter	promoter	363,025
9	Enhancer	enhancer	167,432
10	Mitochondria	mitochondrion	260,335

2.2. Localized MEME tool to scan for conserved domains

The analysis steps are listed as follows (for example “telomere”):

1. MEME was downloaded from the official website and subsequently installed in ubuntu operating system v20.

2. The sequences of tau protein was downloaded from uniprot official website.

3. The FASTA format sequence of Telomere-related protein was downloaded from Uniprot official website. The search keyword was “telomere”.

4. For each sequence in all telomere-related proteins, they were paired with each tau protein sequence to generate FASTA format files for MEME analysis.

5. For the files generated in Step 4, a batch of 50000 or 100000 was used to create several batches. It was considered as the limited space of the virtual ubuntu system.

6. In ubuntu, the conserved domains (E-value<=0.05, technical parameter of MEME tool) of tau protein and telomere-related protein were searched with MEME tool in batches.

7. The result files of conserved domains were collected. Besides, the domain name corresponding to the motif was found from the uniprot database.

8. The each domains activity of tau protein was analyzed according to the characteristics of the telomere-related protein domains according “Interpro” database..

3. Results

3.1 Pyruvate kinase domains of tau protein

Protein Kinase (IPR000719) is a protein kinase that contains numerous conserved regions in its catalytic domain. There is a group of glycine-rich residues close to the lysine residue at the N-terminus of the catalytic domain that helps ATP bind. A conserved aspartate residue in the central region of the catalytic domain significantly influences the catalytic activity of the enzyme.

HPr(Ser) kinase (IPR003755) is a bifunctional HPr serine kinase/phosphorylase (HprK/P). It catalyzes the phosphorylation of specific serine residues in HPr, as well as the dephosphorylation of seryl-phosphorylated HPr (P-Ser-HPr). HPr is a phosphate carrier protein that is a component of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS). HprK/P regulates sugar transport and carbon metabolism. A number of substances inside cells change their levels depending on whether metabolic carbon sources (like glucose and fructose) are present or not. This affects the two opposing functions of HprK/P.

The barrel structure of pyruvate kinase is denoted as Pyruvate Kinase Barrel (IPR015793). Pyruvate kinase (PK) catalyzes the final step of glycolysis, which involves the conversion of phosphoenolpyruvate to pyruvate. It also phosphorylates ADP to ATP. L (liver), R (erythrocytes), M1 (muscle, heart, and brain), and M2 (early embryonic tissues) are the four tissue-specific isozymes. ATP inhibits M1-type PK (muscle, brain), while F1,6BP and alanine have no effect. This is relevant to muscle and brain function, but not liver function. Pyruvate kinase C-terminal (IPR015795) is the C-terminal of pyruvate kinase.

The RIO kinase (IPR000687) is an atypical protein kinase that is classified as such due to its low sequence similarity to eukaryotic protein kinases. The conformation of ATP when bound to RIO kinase is distinct from that of ePKs (such as serine/threonine kinases or insulin receptor tyrosine kinases). This suggests that the detailed mechanism of catalytic aspartate involved in RIO kinase phosphorylation may differ from the mechanism employed by known serine/threonine ePKs.

The Tau protein exhibits protein kinase activity throughout nearly the entire sequence, but it lacks traditional serine/threonine or tyrosine kinase activity, as demonstrated in Table 2. Rather, it primarily exhibits three kinase activities: HPr(Ser) kinase, pyruvate kinase, and RIO kinase. The tau protein activates its kinase activity by binding to Mg2+-ATP. Tau protein contains two Mg2+-ATP binding sites: one at aspartate (D13) and the other at glycine (G682). Near the downstream of G682, there are two lysines (K). Consequently, the catalytic domain concludes at G682, while the central portion of the domain is D13. After the G682 region binds to it, D13 bonds to Mg2+-ATP, granting Tau protein kinase activity. The D13 region reflects the RIO kinase activity of the Tau protein and also possesses catalytic activity. The D13 region is the only part of the Tau protein's HPr(Ser) kinase domain that is in charge of phosphorylating serine and threonine residues.

The aspartate (D13) region is the sole component of the Tau protein's pyruvate kinase domain. The Tau protein possesses pyruvate kinase (PK) activity, which is capable of phosphorylating ADP to ATP. ATP inhibits the PK kinase activity. This is precisely the kinase property of the Tau protein, which is both self-activated and self-inhibited. As a result, the pyruvate kinase activity is inhibited when the Tau protein's D13 binds to ATP, and activated when the Tau protein's D13 binds to ADP. When ATP is bound to Tau protein D13, RIO kinase and HPr(Ser) kinase activities are activated, whereas pyruvate kinase activity is inhibited. ATP at the D13 binding site is converted to ADP after the phosphorylation process is catalyzed by Tau protein. This conversion inhibits RIO and HPr (Ser) kinase activities and activates pyruvate kinase activity. The Tau protein's pyruvate kinase converts ADP to ATP at the D13 binding site.

Table 2. Pyruvate kinase and other domains of tau protein

Domain	Alias	Motif	Start	End	Mg^{2+}-ATP Binding Site
HPr(Ser) kinase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37	D13
	B	FTFHVEITPNVQ	301	312	
	C	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHK	592	648	
Protein kinase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	1	42	D13
	B	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	289	327	
	C	LKNRPC	404	409	
	D	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMV	567	737	G682
Pyruvate kinase C-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLK	1	44	D13
	B	RAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGAP	291	331	
	C	KNRPCLS	405	411	
	D	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	559	712	
	E	RHLSNV	723	728	
Pyruvate kinase barrel	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37	D13
	B	FHVEIT	303	308	
	C	VQK	311	313	
	D	THKLTFRENAKAKTDHGAEIVYK	690	712	
RIO kinase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTP	1	51	D13
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHL	295	322	
	C	SRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQS	558	669	

3.2 Phosphatase domains of tau protein

The beta-lactamase hydrolase-like protein phosphatase-like (IPR005939) is situated at the N-terminus of the β-lactamase domain and may possess phosphatase activity, potentially contributing to the development of biofilms or cell adhesion. Calcineurin-like phosphoesterase (IPR024201) is distinguished by the presence of characteristic motifs in multiple enzymatically active phosphatases. The fetal brain expresses one of these motifs, which it isolated from the chromosome 11p13 region. This region is associated with the mental retardation component of

WAGR syndrome (Nephroblastoma, Aniridia, Urogenital Anomalies, Mental Retardation). PhoD-like phosphatase metallophosphatase (IPR018946), similar to Ca (2+)-dependent phospholipase D, catalyzes the hydrolysis of the ester bond between phosphatidic acid and phospholipid alcohol moieties. Cell growth and development, cancer occurrence, neurological diseases such as Alzheimer's disease, viral invasion, and other diseases are all closely associated with animal phospholipase D (PLD1/2).

Phosphotyrosine protein phosphatase I (IPR023485) is a protein phosphatase that works only on tyrosine. It uses a cysteine phosphatase intermediate to remove phosphate groups from tyrosine residues. It is crucial in the regulation of cell growth, proliferation, differentiation, and transformation. The PTPase domain, centered on the active site cysteine, characterizes tyrosine-specific protein phosphatases (IPR000387). Furthermore, studies have demonstrated the significance of certain conserved residues in the vicinity. A part of tensin and tensin-like proteins called lipid phosphatase domain is structurally similar to tyrosine-protein phosphatase (IPR029021). This includes cyclin G-associated kinase (GAK) and phosphoinositide phosphatase PTEN (phosphatase and tensin homolog).

All tau proteins have long motifs that contain Beta-lactamase hydrolase-like protein phosphatase-like、Calcineurin-like phosphoesterase 、 PhoD-like phosphatase metallophosphatase 、 Phosphotyrosine protein phosphatase I 、 Tyrosine specific protein phosphatases and Tyrosine-protein phosphatase domains. Table 3 summarizes this information. The preceding section did not discover tau proteins binding Mg2+-ATP through serine/threonine. According to Table 3, the presence of a significant number of Tyrosine phosphatase motifs in the tau proteins suggests that they are capable of self-dephosphorylation through Tyrosine phosphatases.

Table 3. Phosphatase domains of tau protein

Domain	Alias	Motif	Start	End
Beta-lactamase hydrolase-like protein phosphatase-like	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	LEFTFHVEITPNVQKEQAHSEEHLGRAA	299	326
	C	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGS	581	622
	D	LGNIHHKPGGG	642	652
	E	DFKDRVQSKIGSLDNITHVPGGGNKKIET	662	690
Calcineurin-like phosphoesterase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKES	1	46
	B	HQLMSGM	147	153
	C	SVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	288	327
	D	TLKNRPCLSPKH	403	414
	E	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIK HVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDF KDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK SPVVSGDTSPRHLSNVSSTGSIDMVDSPQLATL	567	745
PhoD-like phosphatase metallophosphatase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	PLEFTFHVEITPNVQKEQAHSEEHLGRAAF	298	327
	C	KIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQ IVYKPVDLSKVTSKC	576	639
	D	GNIHH	643	647
	E	KDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	664	711
Phosphotyrosine protein phosphatase I	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKES PLQT	1	50
	B	GRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGA	290	330
	C	TLKNRPC	403	409

	D	VPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDM	565	736
Tyrosine specific protein phosphatases	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	5	35
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	295	327
	C	IINKKLD	594	600
	D	NVQSKC	603	608
	E	VQIVYKPVDLSKVTSKCGSLGNIHH	623	647
	F	DNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	675	712
Tyrosine-protein phosphatase	A	PRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	4	35
	B	FTFHVEITPNV	301	311
	C	SNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	602	647
	D	HKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSP	691	740

3.3 Domains that binding and truncating telomeres

（1）Telomere binding domains

Chromo (IPR000953) is a CHRromatin Organization Modifier. The condition of repressed gene expression that can be seen in cells is caused by these proteins changing the structure of chromatin to a more compact form of heterochromatin. One of these proteins, Polycomb, has demonstrated the crucial role of chromodomains in targeting chromatin. The yeast protein CHD1 and the mammalian DNA binding/helicase proteins CHD-1 to CHD-4 both contain chromodomains. Chromodomains play a mechanistic role in targeting chromodomain proteins to specific regions of the nucleus. The targeting mechanism may entail interactions between proteins and/or between proteins and nucleic acids.

CST complex subunit Stn1 N-terminal (IPR018856) is a DNA binding protein that is specific to telomeric DNA. A brief single-stranded G-rich 3' overhang (G-tail) and double-stranded repeat sequences make up telomeres. G tails of 100–250 bases are detected throughout the cell cycle, and human telomeres contain repeats of the sequence CCCTAA/TTAGGG[47], which range in length from 2 to 50 kilobase pairs.

Restriction of telomere capping protein 4 C-terminal (IPR028094). This domain represents the C-terminal domain of the yeast RTC4 (restriction of telomere capping protein 4) protein. Saccharomyces cerevisiae's loss of RTC4 impairs the cellular response to telomere decapping. This domain is also found in proteins that contain the DNA-binding myb domain.

TERF1-interacting nuclear factor 2 N-terminal(IPR029400) denotes the N-terminal domain of TERF1-interacting nuclear factor 2 (TINF2). It is a component of the shelterin complex (telomere), which is responsible for preserving and maintaining telomeres. TINF2 influences the assembly of the shelterin complex. One of its isoforms may also play a role in tethering telomeres to the nuclear matrix in humans.

Telomere length regulation protein conserved (IPR019337) is a conserved domain telomere length regulation protein. The group of proteins known as telomere length regulation TEL2 or clock aberrant protein-2 contains a conserved domain. Plants and humans share these proteins. These proteins contribute to the silencing of subtelomeric regions and modulate telomere length. Tel2 participates in the initial phase of the TEL1/ATM pathway of DNA damage signaling. The protein binds to telomeric DNA repeat sequences in vitro.

Telomere repeat-binding factor dimerisation (IPR013867) undergoes dimerization. Telomeres are responsible

for safeguarding the extremities of chromosomes from degradation and end-to-end fusion, as well as preventing the activation of the DNA damage checkpoint. Among vertebrates' telomeres, telomeric repeat binding factor (TRF) proteins TRF1 and TRF2 are the most important parts that keep them stable. Two related human TRF proteins, hTRF1 and hTRF2, form homodimers and directly bind to telomeric TTAGGG repeats via the myb DNA binding domain at the carboxyl terminus (IPR001005). Telomere length regulation is a function of TRF1, while telomere protection is a function of TRF2.

The N-terminus of the protein houses the N-terminal domain of Telomere-associated protein Rif1 N-terminal domain(IPR022031). Budding yeast recruits Rif1 to telomeres through its interaction with the C-terminus of RAP1. It then negatively regulates telomere length by promoting telomere end degradation or impeding telomere elongation. In mammals, Rif1 is needed for DNA damage during S phase to stop the cell cycle and enable the intra-S checkpoint to work.

Telomeric single stranded DNA binding POT1 domain(PF02765) is a telomeric protein that binds single-stranded DNA. This domain adopts an OB fold and bonds single-stranded telomeric DNA. This group includes the proteins POT1 and CDC13, which demonstrate modulation of telomere length, replication, and capping. POT1 is a component of the shelterin complex that safeguards telomeric ends from DNA repair machinery attack.

The ssDNA-binding domain of Protection of Telomeres Protein 1 (Pot1) is Protection of telomeres protein 1 ssDNA-binding domain(IPR032042). This domain is capable of accommodating a variety of ssDNA ligands. Pot1 is responsible for binding and protecting the 3' single-stranded DNA (ssDNA) overhangs in the majority of eukaryotic telomeres.

(2) Retrotransposase domains

Retrotransposon gag (IPR005162) is the retrotransposon gag domain. When it comes to speed and accuracy, transposable elements (TEs) are better at helping with chromosomal rearrangements than other cellular processes. Longitudinal repeats surround retrotransposons, which share structural similarities with retroviruses. This entry refers to eukaryotic Gag or capsid-associated retrotransposon-related proteins, like Mus musculus's PEG10, which comes from a retrotransposon. This protein self-assembles into a virion-like capsid by binding its own mRNAs. The central motif QGXXEXXXXXFXXLXXH is frequent in gag proteins of the Retroviridae family, but it is inadequately conserved.

The reverse transcriptase domain is known as reverse transcriptase (IPR000477). When retroviral elements, like retroviruses and bacterial retroviruses, copy themselves, they use an RNA template to make DNA that they can insert into the host genome and use to their advantage. An RNA-directed DNA polymerase, also known as reverse aminotransferase (RT), is the enzyme responsible for polymerization. Various mobile elements, such as retrotransposons, retroviruses, group II introns, bacterial msDNA, hepatoviruses, and cauliviruses, contain reverse transcriptases.

(3) Telomere repair domains

DOT1 (PF08123) is a protein that methylates histones and is needed for yeast's DNA damage checkpoint. It regulates gene expression by methylating histone H3. It methylates histone H79's lysine 79 (Lys-3). It identifties Dot1p as the first disruptor of telomere silencing in yeast, where it is associated with gene silencing and the localization of the silent information regulator (SIR) complex. In higher eukaryotes, methylation by this enzyme can be employed to differentiate chromatin domains.

Non-structural maintenance of chromosomes element 1 (IPR011513) is essential for genome stability, DNA repair, and DNA metabolism. It can also use it for homologous recombination to repair DNA double-strand breaks and to recover delayed replication forks.

Non-structural maintenance of chromosomes element 1 RING C4HC3-type (IPR014857)In Saccharomyces

cerevisiae, this domain may be critical for the Rad52-dependent post-replication repair of UV-damaged DNA. The kleisin part of the Smc5/6 DNA repair complex keeps the C-terminal part of chromosome 4 in good shape (IPR014854). Prior to cell division, the Smc5/6 complex is necessary for the unwinding of the undamaged chromosome and can bind to the damaged DNA strand site.

Ku is a prokaryotic-type, non-homologous end-joining protein (IPR009187). The Ku protein is involved in DNA double-strand break repair through non-homologous end joining. In addition to homologous end joining, yeast Ku also plays a role in the structural maintenance of telomeres.

（4）RNA guide sequence synthesis domains

Sm (IPR047575) is a domain that is present in Sm, Lsm (like-Sm), and Hfq proteins. The domain, found in Sm, Lsm (like-Sm), and Hfq proteins, participates in numerous processes that govern gene expression and RNA processing. The Lsm archaeal protein (SmAP) forms homoheptamers with polynucleotide phosphorylase P and uridine-rich RNA sequences. It may be involved in processing tRNA.

TPP1 (IPR019437) is a constituent of the shelterin complex TPP1/Est3. Est3 (TPP1) is a component of the telomerase holoenzyme (shelterin complex) that is involved in yeast telomere replication. Studies have demonstrated that TPP1 dimerizes and attaches to both DNA and RNA. Furthermore, TPP1 induces the dissociation of RNA/DNA heteroduplexes. A novel GTPase is the yeast telomerase protein TPP1 (Est3). The N-terminus of TPP1 may serve as the binding surface for TIN2, while the C-terminus may bind to POT1, thereby tethering POT1 to the shelterin complex. The complex enhances the activity and synthesis capacity of the human telomerase core enzyme when it binds to telomeric DNA, thereby aiding in the preservation of telomere length. Six proteins make up the human shelterin complex: telomeric repeat binding factor 1 (TRF1), TRF2, repressor/activator protein 1 (RAP1), TRF1-interacting nuclear protein 2 (TIN2), TIN2-interacting protein 1 (TPP1), also known as ACD of adrenocortical dysplasia protein, and protection of telomeres 1 (POT1).

Telomerase ribonucleoprotein complex - RNA-binding (IPR021891). Telomerase plays a significant role in determining the length of telomeres. It is a reverse transcriptase that adds these basic sequence repeats to the ends of chromosomes by utilizing the template sequence in the RNA component of the replicase. The RNA binding domain TRBD of telomerase is composed of 12 α helices and 2 short β sheets. TRBD is involved in the formation of the holoenzyme, which performs telomere extension, as well as RNA recognition and binding.

（5）Tau protein's active function

Table 4 illustrates that the Tau protein contains domains that are associated with Telomerase ribonucleoprotein complex - RNA-binding、 Telomere length regulation protein conserved、 Telomere repeat-binding factor dimerisation、 Telomere-associated protein Rif1 N-terminal、 Telomeric single stranded DNA binding POT1、 Shelterin complex subunit TPP1、Protection of telomeres protein 1 ssDNA-binding、Restriction of telomere capping protein 4 C-terminal、 CST complex subunit Stn1 N-terminal. The Tau protein's telomere binding region, primarily located at the C-terminal, regulates the length of telomeres. The tau protein binds to the CCCTAA/TTAGGG repeat sequence, but it doesn't have the functional domain that is needed to make telomere fragments. Table 4 demonstrates the telomere reverse transcriptase activity of tau protein. The cut region is "TAA," and tau protein will cut double-stranded fragments from telomeres in the same manner as reverse transcriptase, as indicated by retrotransposon gag and reverse transcriptase. Table 4 illustrates that the tau protein contains a telomere repair-related domain ku, which repair truncated telomere sequences by facilitating the homologous recombination repair of DNA double-strand breaks and the recovery of stalled replication forks.

Table 4. Telomere binding and cutting domains of tau protein

Domain	Alias	Motif	Start	End
DOT1	A	EPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	3	41
	B	GQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	294	327
	C	NVKSKIGSTENLKHQ	572	586
	D	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCG SLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIE THKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMV	592	737
Chromo	A	YGLGDRKDQGGYTMHQDQEGD	18	38
	B	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIK HVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFK DRVQSKIGSLDNITHV	567	680
	C	NKKIETHKLTFRENAKAKTDHGA	685	707
CST complex subunit Stn1 N-terminal	A	GKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPV	590	630
	B	NIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIET	644	690
Ku	A	FEVMEDHAGTYGLGDRKDQGGYTMH	8	32
	B	NLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKH	582	616
	C	LDNITHVPGGGNKKIETHKLTFREN	674	698
Ku C-terminal	A	GNKKIETHKLTFRENAKAKTDHGAEIVYK	684	712
Ku70	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	FTFHVEI	301	307
	C	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK	581	628
	D	SLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	673	711
Non-structural maintenance of chromosome element 4 C-terminal	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	5	35
	B	SVQIVYKPVDLSKVTSKCGSLGNIHH	622	647
	C	KSEKLDFKDRVQSKIGSLDNITHV	657	680
	D	IVYKSPVVSGDTSPRHLSNVSSTGSIDM	709	736
Non-structural maintenance of chromosomes element 1 RING C4HC3-type	A	SNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSK	602	634
Restriction of telomere capping protein 4 C-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	NKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLG NI	596	645
	C	VSGDTSPRHLSNVSSTGSIDMVDSPQLATLADE	716	748
Retrotransposon gag	A	HKLTFRENAKAKTDHGAEIVY	691	711
Reverse transcriptase	A	PRQEFEVMEDHAGTY	4	18
	B	RKDQGGYTMHQDQ	23	35
	C	FTFHVEITPNVQKEQAHSEEHLG	301	323

	D	VQIINKKLDLSNVQSKCGSKDNI	592	614
	E	GGSVQIVYKPVDLSKVTSKCGSLGNIHH	620	647
	F	LDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFREN	661	698
Shelterin complex	A	MAEPRQEFEVMED	1	13
subunit TPP1	B	HQLMSGM	147	153
	C	FTFHVEITPNVQKEQAHS	301	318
	D	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNI	567	614
	E	GSVQIVYKPVDLSKVTSKCGSLGNIHH	621	647
	F	NITHVPGGGNKKIETHKLTFRENAKAKTDHGAEI	676	709
TERF1-interacting	A	AEPRQEFEVMEDH	2	14
nuclear factor 2	B	YGLGDRKDQGGYTMHQDQ	18	35
N-terminal	C	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	289	327
	D	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDF	593	663
Telomerase	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMH	5	32
ribonucleoprotein	B	HQLMSGM	147	153
complex -	C	GQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPG	294	329
RNA-binding	D	INKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDL	595	632
	E	LDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSST	661	731
Telomere length	A	YTMHQD	29	34
regulation protein	B	KVQIINKKLDLSNVQSKCGSKDNIKH	591	616
conserved	C	IVYKPVDLSKVTSKCGSLGNIHHKPGGG	625	652
	D	SLDNITHVPGGGNKKIETHKLTFRENAK	673	700
Telomere	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQ	8	35
repeat-binding factor dimerisation	B	PASEPDGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	280	327
Telomere-associated	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	5	36
protein Rif1	B	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIK		
N-terminal		HVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFK	567	711
		DRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY		
Telomeric single	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTD	1	40
stranded DNA	B	PLEFTFHVEITPNVQKEQAHSEEH	298	321
binding POT1	C	SAKTLKNRPCL	400	410
	D	KIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVD	576	631
	E	CGSLGNIHHK	639	648
	F	FRENAKAKTDHGAEIVY	695	711
Sm	A	FTFHVEITPNVQKEQAHSEEHLGRAAF	301	327
	B	LSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	601	647
	C	KIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVS	670	717

3.4 Domains of transposases-like

(1) Domains of transposases

Transposable element P transposase (IPR048366) represents the GTP-binding insertion domain (GBD) of the transposable element P transposase from the fruit fly Drosophila melanogaster (TNP) and its human homolog, the DNA transposase THAP9. This domain is inserted into the RNaseH domain and primarily acquires an α-helical fold that is similar to that of Hermes, Tn5, or RAG1 proteins. This domain facilitates protein-DNA interactions. TNP is specifically involved in the transposition of P elements, as well as precise and imprecise excision. It possesses unique functions, including the generation of long staggered DNA breaks during transposition and the need for a guanosine triphosphate cofactor. THAP9 specifically recognizes the bipartite 5'-TXXGGGX(A/T)-3' consensus motif, facilitating transposition.

Transposase DDE domain is IPR025668. Transposon proteins are required for efficient DNA transposition. The catalytic activity of this enzyme is characterized by the cleavage of DNA at a specific site, followed by a strand transfer reaction.

A transposase protein known as transposase IS111A (IPR002525) is required for efficient DNA transposition. It belongs to the IS111A/IS1328/IS1533 transposase family and serves as the N-terminal region of the pilin-transducing protein (PIVML).

Transposase IS200-like (IPR002686) is a transposase. Transposase IS200-like are required for efficient transposition of inserted sequences or transposon DNA. This entry denotes the domain that is present in the IS200 transposase of Escherichia coli.

The IS204/IS1001/IS1096/IS116 transposases contain the DDE domain, transposase IS204 (IPR002560). Autonomous mobile genetic elements, such as transposons or insertion sequences (IS), encode an enzyme, the transposase, which is necessary for the excision and insertion of mobile elements.

Transposase IS30-like HTH (IPR025246) is a transposase IS30-like HTH domain. This helix-turn-helix domain is present in certain transposases and is likely to be DNA-binding.

IPR002559 is an transposase IS4-like domain. A transposase IS4-like is a self-moving genetic element, like a transposon or an insertion sequence (IS), that makes an enzyme that can remove and add mobile elements. We have established diverse families for transposases. This family consists of IS4 transposases.

The transposase IS605 OrfB C-terminal (NF033573) is part of the IS200/IS605 family. The majority of IS200/IS605 family insertion sequences encode the transposase TnpA, which is approximately 130 amino acids in length, and the larger accessory protein TnpB, which can function as a methyltransferase.

The transposase IS66 central (IPR004291) has a central domain. This domain is present in Agrobacterium tumefaciens' transposase IS66. Transposases are necessary for efficient DNA transposition. It is possible that IS66 may cause genetic and structural variations in the cephalopod Ti plasmid's T region and vir region. We refer to the C-terminal region of the insertion element IS66 transposase as Transposase IS66 C-terminal (IPR039552). We refer to the zinc-finger region at the N-terminus of the insertion element IS66 transposase as transposase IS66 zinc-finger binding (IPR024474).

transposase IS701-like DDE (IPR038721) is a member of the DDE endonuclease superfamily. It is believed that three carboxylate residues are responsible for coordinating the metal ions necessary for catalysis. A strand transfer reaction characterizes this enzyme's catalytic activity, followed by site-specific DNA cleavage.

Transposase InsH N-terminal (IPR008490) is its N-terminal form. The transposase InsH employs this putative domain at its N-terminus to insert the sequence element IS5A from Escherichia coli (strain K12) and other related transposases in bacteria and archaea.

Transposase Tc1-like (IPR038717) is a DDE domain transposase. Tc1 transposase is a member of the DD[E/D]-transposase family. It has a unique pattern made up of three acidic residues, two of which are aspartic acid and one of which is glutamic acid. In some cases, the third aspartic acid is a [D, D, (E, or D)] motif. The three

catalytic residues are in close proximity for catalysis, as members of the DD[E/D]-transposase family share a common RNase H-like structure.

Transposase Tc5 C-terminal (IPR007350) is C-terminal. This domain is associated with a cysteine-rich region at the C-terminus that has the potential to bind metal ions and DNA.

IPR014735 is a transposase Tn5-like and N-terminal. The N-terminus of Tn5-type transposase proteins houses this domain. The prokaryotic Tn5 transposase generates two types of DNA contacts: trans contacts for catalysis and cis contacts for DNA recognition. Additionally, it establishes protein-protein contacts. During catalysis, Tn5 transposase uses two cations in its active site. These cations change the shape of DNA that is bound to them and are necessary for transposition. The α/β/α fold of Tn5 transposase is of the RnaseH type. The N-terminal domain forms an orthogonal T-tract structure.

The transposase TnpC homeodomain (IPR024463) represents the leucine zipper-like or homeodomain-like region of the insertion element IS66.

The transposase putative helix-turn-helix (IPR021027) is a putative helix-turn-helix domain. This is a DNA endonuclease that cuts DNA after the 5'-TTGAT-3' transposon-associated motif (TAM). It does this with the help of RNA from the right-terminal part of its insertion sequence element.

The transposase zinc-binding domain (IPR026889) is a zinc-binding domain. It is located in the N-terminus of transposases, which are members of the IS91 family.

Transposase-like Mu C-terminal (IPR015378) Mu, C-terminal, transposase-like. Prokaryotic integrases and transposons contain this domain. It utilizes a Greek key topology and a β-barrel structure.

IPR032750 is the Transposon Tn7 transposition protein TnsD C-terminal. The transposon Tn7 transposition protein TnsD C-terminal is a potential D-type transposase protein that is similar to Tn7.

(2) Retroviral integrase domains

The tyrosine-type site-specific recombinase core-binding (CB) (IPR044068) is involved in a number of important gene rearrangement reactions. In addition to its involvement in protein-protein interactions, it primarily interacts with the primary groove of the attachment site, facilitating binding to the core DNA sequence. Some recombinases, like lambda Int and IntDOT, have an extra amino-terminal domain that helps them find arm sites, which are on both sides of the crossover region, and tells the recombination reaction which way to go.

Tyr recombinase (PS51898) is responsible for a diverse array of sequence-specific DNA rearrangements in biological systems. The following are some of these processes: conjugative transposition, the resolution of DNA circles, the control of plasmid copy number, DNA excision, the regulation of gene expression for nitrogen fixation and DNA inversion in Houttuynia cordata, and the control of gene expression for cell surface proteins or DNA replication.

Integrase-type (IPR003308) is a retroviral integrase that allows DNA copies of the viral genome to be integrated into the host chromosome. Three domains comprise integrase: the N-terminal zinc-binding domain, the central catalytic core, and the C-terminal DNA-binding domain. We refer to the central catalytic core and N-terminal zinc-binding domain of retroviral integrase as integrase catalytic and integrase zinc-binding, respectively.

(3) Tau protein activity and function

Table 5 illustrates that tau protein possesses a variety of integrase and transposase activities. The N-terminus, C-terminus, and intermediate section of the tau protein are responsible for its transposase and integrase activities. Tau protein helps with transposition by attaching to the 5'-TXXGGGX (A/T) -3' consensus motif. This is done by DNA transposase THAP9, which is the human homolog of transposable element P transposase. It is worth noting that the consensus motif of THAP9 overlaps with the TTAGGG repeat sequences found in human telomeres. The data in Table 4 shows that the tau protein can attach to human telomere sequences and use reverse transcriptase

transposase to cut pieces from telomeres. It denote the cut region as "TAA." With Table 4 in mind, it's clear that the transposase activity of tau protein in Table 5 is to cut the telomere sequence and move it to the DNA sequence region that has "TAA" and TXXGGGX. The data in Table 5 shows that the tau protein has transposase DDE activity. This is needed for the catalytic activity that cuts DNA at a specific site (TAA) and then the chain transfer reaction. Table 5 demonstrates that retroviral integrase activity will cut off and integrate the double-stranded telomere fragment into the specific site ("TAA") DNA region. Tau protein also possesses core-binding (CB), Tyr recombinase, and integratease activities. This process is indicative of gene rearrangement. The integrated DNA region is a specific promoter and enhancer region, as demonstrated by the analysis in the subsequent section.

Table 5. Tau protein's transposase-like domains

Domain	Alias	Motif	Start	End
Transposable element P transposase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	1	43
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	295	327
	C	VPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRH	565	724
Transposase (putative) YhgA-like	A	YKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDF	627	663
	B	NVSSTGSIDMVD	727	738
Transposase DDE	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	1	42
	B	PVDFLSKVSTEIPASEPDGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAH	268	317
	C	NLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMV	582	737
Transposase Helix-turn-helix	A	DNIKHV	612	617
Transposase IS111A	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTP	1	51
	B	GLSHQLMSGM	144	153
	C	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGA	289	330
	D	LKNRPCLS	404	411
	E	APVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQL	563	742
Transposase IS200-like	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTED	1	54
	B	LEDEAAGHVTQEPESGKVVQEGFLREPGPPGLSHQLMSGMP	114	154
	C	EGGRHAPELLKHQLLGDLHQ	179	198

	D	REATSIPGFPAEGAIPLPVDFLSKVSTEIPASEPDGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGAPGEGPE	251	336
	E	TRSSAKTLKNRPCLSPKHPT	397	416
	F	CPEPPSSPKYVSSV	432	445
	G	SSAKSRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQLATLADEVSA	554	751
Transposase IS204	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	1	39
	B	RAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	291	327
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	567	712
	D	GSIDMV	732	737
Transposase IS30-like HTH	A	EPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKE	3	45
	B	FTFHVEITPN	301	310
	C	EEHL	319	322
	D	GKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK	590	628
	E	RVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKS	666	713
Transposase IS4 N-terminal	A	FEVMEDH	8	14
	B	YTMHQDQ	29	35
Transposase IS4-like	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQT	1	50
	B	AGHVTQEPESGKVVQEGFLREPGPPGLSHQLMSGM	119	153
	C	LLKHQLLGDLH	187	197
	D	GAIPLPVDFLSKVSTEIPASEPDGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGAPGE	263	333
	E	DKKAKTSTRSSAKTLKNRPCLSPKH	390	414
	F	PAVC	429	432
	G	AKSRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQLATLADEV	556	749
Transposase IS605 OrfB C-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTEDGSEE	1	58
	B	VQEGFLR	132	138

	C	FLSKVSTEIPASEPDGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGA	271	330
	D	TRSSAKTLKNRPCLSPKHPTPGSSDPLIQPSSPAVCPEPPSSPKYV	397	442
	E	SAKSRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQLA	555	743
Transposase IS66 C-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQT	1	50
	B	TFHVEITPNVQKEQAHSEEHLGRAAF	302	327
	C	GGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDF	589	663
	D	IGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVV	671	716
Transposase IS66 central	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKE	1	45
	B	GRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	290	327
	C	VPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRH	565	724
	D	PQLATLADEV	740	749
Transposase IS66 zinc-finger binding	A	FEVMEDHAGTYGLGDR	8	23
Transposase IS701-like DDE	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGD	1	38
	B	NVKSKIGSTENLKHQPGGGKVQIINKKL	572	599
	C	NITHVPGGGNKKIETHKLTF	676	695
Transposase InsH N-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPL	1	48
	B	GQDAPLEFTFHVEITPNVQKE	294	314
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSP	567	714
Transposase Synechocystis PCC 6803	A	FTFHVEITPNVQKEQ	301	315
Transposase Tc1-like	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQ	1	49
	B	EDTE	176	179
	C	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	289	327

	D	KSRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	557	711
	E	RHLSNVSST	723	731
Transposase Tc5 C-terminal	A	RAKGQDAPLEFTFHVEITP	291	309
	B	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQ	592	624
Transposase Tn5-like N-terminal	A	DNITHVPGGGNKKIETHKLTFREN	675	698
Transposase TnpC homeodomain	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37
	B	KKLDLSNVQSKCGSKDNIKH	597	616
Transposase putative helix-turn-helix	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTEDGS	1	56
	B	LSHQLMSGM	145	153
	C	QDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGA	295	330
	D	IGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMV	577	737
Transposase zinc-binding	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLK	1	44
	B	LEFTFHVEITPNVQKEQAHSEEHLGRAAFP	299	328
	C	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQLATL	559	745
Transposase zinc-ribbon	A	DHAGTYGLGDRKDQGGYTM	13	31
	B	KVQIINKKLDLSNVQSKCGSKDNI	591	614
	C	VPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	617	711
Transposase-like Mu C-terminal	A	QGGYTMHQDQE	26	36
	B	EFTFHVEITPNV	300	311
	C	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKV	591	635
Transposon Tn7 transposition protein TnsD C-terminal	A	AEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	2	39
	B	FTFHVEITPNVQKEQAHSEEHLGRAAF	301	327
	C	TENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	580	627

	D	LGNIHHKPGGGQVEVKSEKLDFKDRVQ	642	668
	E	THVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	678	712
Tyr recombinase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLK	1	44
	B	SVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGAPGEGPE	288	336
	C	TRSSAKTLKNRPCL	397	410
	D	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	581	647
	E	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHL	660	725
Integrase catalytic	A	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	591	627
Integrase zinc-binding	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	5	37
	B	EFTFHVEITPNVQKEQAHSEEHLGRA	300	325
	C	KDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAK	664	700
Integrase-type	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQ	8	35
	B	DGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSE	285	319
	C	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLG	593	643
	D	VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	654	712

3.5 Domains related to helicases and replicants

（1）Replicon binding domains

The chromosome replication initiator DnaA C-terminal (IPR013159) reports that the C-terminal domain of the bacterial DnaA protein is very important for starting chromosome replication and keeping it under control. DnaA is a protein that binds to DNA and ATP. It specifically binds to the dnaA box, a 9 bp nucleotide repeat sequence that is present in the chromosome replication origin (oriC). Two conserved regions make up DnaA: the first is located in the N-terminal half, corresponding to the ATP binding domain, and the second is located in the C-terminal half, potentially involving DNA binding.

Chromosomal replication initiator protein DnaA (TIGR00362)At the replication origin, DnaA forms an ATP-bound complex with the dnaA-box during the initiation of chromosome replication. Additionally, it influences the transcription of numerous genes, including its own. It initiates chromosome replication; participates in DNA biosynthesis; and can also function as a transcriptional regulator.

Chromosome segregation in meiosis protein 3 domain(IPR012923) is found in a group of proteins, such as Csm3 from budding yeast, Swi3 from fission yeast, and TIPIN from vertebrates. It helps with chromosome segregation during meiosis. They are responsible for ensuring replication fork stability and DNA replication.

DNA replication factor Dna2 N-terminal (IPR014808) is a DNA replication factor that can act as a 5'-flap endonuclease, a helicase, and an ATP-dependent nuclease for single-stranded DNA. It is necessary for the processing of Okazaki fragments and is involved in DNA repair pathways with limited helicase activity.

DnaD (IPR006343) composes the PriA protomer. The PriA protomer recruits the replication fork helicase to the DNA. Two distinct domains, an N-terminal domain that exhibits oligomerization activity, a C-terminal domain that exhibits DNA binding activity, and a second DNA-induced oligomerization activity, make up DnaD.

It is part of a group of proteins called the initiator rep protein (IPR046828). These proteins are very important

for the actinomycete integration and conjugation elements (AICEs). This family includes the replication initiator protein from Streptomyces (RepSA), which is essential for the replication of pSAM2.

Origin recognition complex subunit 5 C-terminal (IPR047088) ORC1-6 encodes the six-subunit ATP-dependent DNA-binding complex known as the origin recognition complex (ORC). ORC binds chromatin at replication origins throughout the cell cycle, making it a fundamental component of eukaryotic DNA replication. ORC is essential for DNA replication initiation and is responsible for its direction throughout the genome. In order to acquire a replication license for chromosomes before DNA synthesis during the S phase, pre-RC assembly is necessary during G1.

Replication factor A C-terminal (PF08646) is a member of Replication factor A (RPA) that interacts with single-stranded DNA and is involved in DNA replication, repair, and recombination.

The C-terminal domain of the RFC (Replication factor-C C-terminal, IPR013748) protein of the clamp loader complex, which binds to the DNA sliding clamp (proliferating cell nuclear antigen, PCNA), is known as Replication factor C C-terminal (IPR013748).

Replication factor Mcm10 C-terminal (IPR015411) is a domain located in the C-terminal region of the Mcm10 protein, a eukaryotic DNA replication factor that regulates the stability and chromatin binding of DNA polymerase α.

Replication initiation protein-like C-terminal (IPR003491) is a replication initiation factor (REP) family member. It is in charge of plasmid replication. Many members, who are specific topoisomerases, are likely to generate sequence-specific single-stranded nicks in plasmid DNA at the origin of replication.

Replication initiator A N-terminal (IPR010724) represents the replication initiator protein A (RepA), a DNA replication initiation protein in plasmids.

Rolling Circle replication initiation protein N-terminal (IPR040819) is the N-terminal domain of Sterilobacillus lipophilus' rolling circle replication initiation protein (Rep). This protein initiates replication, recruits helicases to the origin site, and terminates replication after DNA synthesis on pT181 family plasmids.

（2）Domains of helicase

The domain found in viral DNA replication helicases, bacterial ATP-dependent DNA helicase Pif1, and eukaryotic ATP-dependent DNA helicase PIF7 is represented by DNA replication helicase (IPR003840).

The SF3 helicase (PS51206) is a domain of the superfamily 3 helicase found in DNA viruses. Small DNA viruses bind the SF3 helicase at the origin binding domain. The virus can circumvent the host cell-based regulatory pathways and initiate its own replication by combining the domain that recognizes the ori with the helicase. The protein's binding causes the viral ori to unfold.

（3）Tau protein's active function

Table 6 demonstrates that the N-terminus and C-terminus of tau protein primarily mediate its DNA protein activity. Replication factor, replication initiation protein, and SF3 helicase activity are also present in the Tau protein. This suggests that the tau protein can bind to the DNA replication origin region with replicon activity and unwind the DNA double helix in the same way as a DNA replication helicase. This activity of the unwinding domain is like the SF3 helicase protein in DNA viruses, which finds the ori domain and unwinds the helix.

Table 6. Tau protein's helicase and replicon binding domains

Domain	Alias	Motif	Start	End
Chromosomal	A	RQEFEVM	5	11
replication initiator	B	EFTFHVEITPNVQKEQAHS	300	318
DnaA C-terminal	C	GSLDNITHVPGGGNKKIETHKLTFR	672	696
Chromosomal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	1	39
replication initiator	B	QIINKKLDLSNVQSKCGSKDNIKH	593	616

protein DnaA	C	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	660	711
Chromosome segregation in meiosis protein 3	A	YTMHQDQ	29	35
DNA replication factor Dna2 N-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQD	1	34
DNA replication helicase	A	EFEVMEDHAGTYGLGDRKDQGGYTMH	7	32
	B	CGSLGNIHH	639	647
	C	HKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNV	691	728
Initiator Rep protein	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQD	1	34
	B	FT	301	302
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNI	567	614
	D	QVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLT	653	694
DnaD	A	EFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	7	35
	B	FTFHVEITPNVQKEQ	301	315
	C	GSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	578	647
Origin recognition complex subunit 5 C-terminal	A	NLKHQPGGGKVQIINKKLDLSNVQSK	582	607
	B	QSKIGSLDNITHVPGGGNKKIETHK	668	692
Replication factor A C-terminal	A	LQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIIN	560	596
Replication factor C C-terminal	A	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	663	711
Replication factor Mcm10 C-terminal	A	QGGYTMH	26	32
	B	FTFHVEITPNVQKE	301	314
	C	NVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKC	603	639
Replication initiation protein-like C-terminal	A	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK	581	628
	B	HLSNVSSTGSIDMVDSPQLAT	724	744
Replication initiator A N-terminal	A	FEVMEDHAGTYGLGDRKDQGGYTMHQ	8	33
	B	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	592	647
	C	VKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDH	656	705
Replication initiator protein A C-terminal	A	FEVMEDHAGTYGLGDRKDQGGYTMH	8	32
	B	FKDRVQSKIGSLDNITHVPG	663	682
Rolling Circle replication initiation protein	A	PRQEFEVMEDHAGTYGLGDRK	4	24
	B	QPGGGKVQIINKKLDLSNVQS	586	606
	C	IVYKPVDLSKVTSKCGSLGNIH	625	646

N-terminal				
SF3 helicase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	LEFTFHVEITPNVQKEQAHSEEHLGRAAF	299	327
	C	PGGGKVQIINKKLDLSNVQSKCGSKDNIKH	587	616
	D	SVQIVYKPVDLSKVTSKCGSLGNIHH	622	647
	E	GGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENA KAKTDHGAEIVYK	651	712
	F	IDMV	734	737
SF4 helicase	A	RQEFEVMEDHAGTY	5	18
	B	GYTMHQDQ	28	35
	C	GQDAPLEFTFHVEITPNVQKEQAH	294	317
	D	APVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSK DNIKH	563	616
	E	KCGSLGNIHHK	638	648
	F	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEI	662	709

3.6 Domains of DNA polymerases and nucleases

(1) DNA polymerase domains

DNA-directed DNA polymerase family B exonuclease (IPR006133), using a complementary DNA strand as a template, adds nucleotide triphosphate (dNTP) residues to the 5' end of a developing DNA strand. In all sequences in the B family, there is a DTDS motif. These sequences have different functional domains, such as a 5'-3' extension domain, a 3'-5' exonuclease domain, a DNA binding domain, and dNTP and pyrophosphate binding domains.

DNA-directed DNA polymerase family A palm (PF00476) enzymes are essential for precise DNA replication. For de novo DNA strand synthesis, a small RNA molecule or protein is needed as a primer.

DNA mitochondrial polymerase exonuclease (IPR041336) is a member of the human mitochondrial DNA polymerase (Pol-γ). Pol-γ is composed of a catalytic subunit, Pol γ-A, which possesses both polymerase and proofreading exonuclease activities, and an auxiliary subunit, Pol gamma-B, which accelerates the polymerization rate and inhibits the exonuclease activity. This domain represents the exonuclease domain of the catalytic subunit Pol-A.

DNA polymerase III alpha subunit finger (IPR040982). The presence of thumb, finger, and palm domains, which resemble the configuration of a right hand, distinguishes DNA polymerase. This entry denotes the finger domain of the alpha subunit of DNA pol III.

Most of the replicative synthesis in bacteria is done by DNA polymerase III beta sliding clamp (IPR001001), which is a complex multi-chain holoenzyme. It functions by utilizing complementary DNA as a template to add nucleotide triphosphate (dNTP) residues to the 5' end of the developing DNA strand. The β subunit, also referred to as the elongation factor β-clamp, is a component of the DNA polymerase III holoenzyme. Nevertheless, the β-clamp is not irreversibly affixed to polymerase III, as are the other subunits. The clamp loader, a constituent of DNA Pol III, is responsible for loading it onto the DNA. The dsDNA in bacteria is encircled by a ring-shaped dimer that the β-clamp forms (sliding clamp). Bacterial β clamp is a homodimer made up of three globular domains that fit together to make a ring with six domains.

DNA polymerase III delta subunit C-terminal (IPR032780) domain is located at the C-terminus of the DNA polymerase III subunit. The remaining subunits are suspended on a helical scaffold that is formed by numerous

C-terminal domains of γ, δ, and delta' in the clamp loader γ complex of DNA polymerase III. The γ complex is a bacterial homolog of eukaryotic replication factor C and a AAA+ ATPase. It is capable of loading the sliding clamp (β, which is homologous to PCNA) onto DNA.

DNA polymerase helix-hairpin-helix motif (IPR029460) is a brief DNA binding domain that contains an HHH (helix-hairpin-helix) domain.

Table 7 illustrates the DNA polymerase domain of tau protein, with the activity located at the N-terminus and C-terminus. The palm, finger, and clamp structures of DNA polymerase are the primary structures of tau protein. Because there is no complete DNA polymerase structure, tau protein exhibits nuclease activity and binds to DNA in the palm, finger, and clamp structures. To add to what has already been said, the tau protein's DNA mitochondrial polymerase exonuclease and DNA polymerase III delta subunit C-termina domain activities suggest that it can also bind and work with circular DNA in mitochondria.

Table 7. DNA polymerase domains of tau protein

Domain	Alias	Motif	Start	End
DNA-directed	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
DNA polymerase	B	FTFHVEITPNV	301	311
family B	C	CGSKDNIKHVPGGGSVQIVYKPVDLS	608	633
exonuclease				
DNA-directed	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37
DNA polymerase	B	RAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	291	327
family A palm	C	LKNRPC	404	409
	D	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSK DNIKH	567	616
	E	VQIVYKPVDLSKVTSKCGSLGNIHH	623	647
	F	IGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVV SGDTSPRHLSNV	671	728
	G	IDMVDS	734	739
DNA	A	QEFEVMEDHAGTYGLGDRKDQGGYTMHQ	6	33
mitochondrial				
polymerase				
exonuclease				
DNA polymerase	A	EFEVMEDHAGTYGLGDRKDQGGYTMH	7	32
III alpha subunit	B	IINKKLDLSNVQSKCGSKDNI	594	614
finger	C	QVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFREN AKAKTDHGAEIVYKSPVV	653	716
DNA polymerase	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	1	42
III beta sliding	B	EFTFHVEITPNVQKEQAHSEEHLGRAAF	300	327
clamp C-terminal	C	PMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSK DNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKP	566	649
	D	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHG AEIVY	662	711
DNA polymerase	A	YTMHQDQ	29	35
III beta sliding	B	CGSLGNIH	639	646
clamp N-terminal				
DNA polymerase	A	EFEVMEDHAGTYGLGDRKDQGGYTMH	7	32

III beta sliding clamp central	B	GSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	672	711
DNA polymerase	A	FTFHVEIT	301	308
III delta subunit C-terminal	B	HKLTFRENAKAKTDHGAEIVY	691	711
DNA polymerase helix-hairpin-helix motif	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL KESPLQTP	1	51

(2) DNA methylase and ATP-dependent nucleases domains

ATP-dependent DNA ligase family profile (PS50160) is a DNA ligase (polydeoxyribonucleotide synthase), an enzyme that helps connect two pieces of DNA by making a nucleotide ester bond between phosphate and deoxyribose. It is involved in DNA replication, DNA repair, and DNA recombination. There are DNA ligases that are ATP-dependent in eukaryotic, archaeal, viral, and bacteriophage cells.The conserved lysine residue serves as the adenylation site.

It indicates the N-terminal domain of DNA replication factor Dna2 (IPR014808). Dna2 and its plant homolog JHS1 are DNA replication factors that can helicase, act as an ATP-dependent nuclease (5'-flap endonuclease), and ATPase single-stranded DNA. It is necessary for processing Okazaki fragments and is involved in DNA repair pathways. The helicase activity is negligible.

Prokaryotes predominantly contain N-6 adenine (A)-specific DNA methylases, which contain DNA methylase adenine-specific (IPR003356) domains. These enzymes are responsible for methylating specific DNA sequences in order to prevent the host from degrading its own genome through restriction endonucleases. These methylases exhibit the same sequence specificity as their respective restriction endonucleases. The M and S subunits form a methyltransferase, methylating both adenine residues in the complementary strands of the bipartite DNA recognition sequences. The complex can also function as an endonuclease in the presence of the R subunit, binding to the same target sequence but cleaving the DNA at a distance from the site. The methylation status of the target sequence dictates the cleavage or modification of the DNA. The DNA is cleaved when the target site remains unaltered. When the target site becomes hemimethylated, the complex acts as a maintenance methyltransferase, modifying the DNA to methylate both strands.

The N-terminus of the methylase subunit of type I DNA methyltransferase houses the N6 adenine-specific DNA methyltransferase N-terminal (IPR022749) domain.

DNA mismatch repair MutH (SM00927) is responsible for repairing DNA mismatches. Three important proteins—MutS, MutL, and MutH—start methyl-directed DNA mismatch repair in Escherichia coli to fix mistakes that happen during DNA replication. In a hemimethylated duplex, MutH cleaves the newly synthesized unmethylated daughter strand 5' to the sequence d (GATC). In order to activate MutH PUBMED:9482749, MutS and MutL must recognize DNA mismatches.

Table 8 illustrates that the DNA polymerase domain of tau protein contains an ATP-dependent region. The DNA polymerase structure of tau protein is able to bind DNA and perform helicase and nuclease activities because of the energy provided by the pyruvate kinase (PK) activity, which can phosphorylate ADP to ATP.

Table 8 shows that tau protein contains DNA methylase adenine-specific and N6-adenine-specific DNA methyltransferase N-terminal domains. In particular, tau protein has N-6 adenine (A)-specific DNA methyltransferase activity mainly at the beginning and end of the chain. These are the sites located at the extremities of the methylated specific region on DNA.

Table 8. DNA methylase and ATP-dependent nucleases domains of tau protein

Domain	Alias	Motif	Start	End
ATP-dependent DNA ligase family profile	A	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIK	581	615
DNA replication factor Dna2 N-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQD	1	34
DNA methylase adenine-specific	A	LGDRKDQGGYTMHQDQE	20	36
	B	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDN	567	613
	C	QIVYKPVDLS	624	633
	D	KCGSLGNIHHK	638	648
N6 adenine-specific DNA methyltransferase N-terminal	A	VMEDHAGTYGLGDRKDQGGYTMHQDQ	10	35
	B	INKKLDLSN	595	603
	C	VYKPVDLSKVTSKCGSLGNIHH	626	647
	D	GSLDNITHVPGGGNKKIETHKLTFRENAKAKTDH	672	705
DNA mismatch repair MutH	A	FTFHVEITPNVQ	301	312
	B	DNITHVPGGGNKKIETHKLTFREN	675	698

(3) RNA-guided HNH Cas9 enzyme domains

HNH endonuclease 5 (IPR029471) suggests the nucleolytic function of these proteins. Bacterial HNH domains can function as both site-specific DNA nucleases (e.g., homing nucleases or the RNA-guided DNA endonuclease Cas9) and non-specific nucleases (e.g., colicins, a subgroup of bacterial toxins).

The HNH nuclease (IPR003615) domains are present in the HNH nuclease family, which includes the yeast intron 1 protein, human DNA annealing helicase and endonuclease ZRANB3, bacterial CRISPR-associated endonuclease Cas9, colicins, pyotin, and endonuclease HphI.

HNH Cas9-type (IPR033114). Type II CRISPR-Cas systems generate double-strand breaks in invasive DNA using the RNA-guided DNA endonuclease Cas9. During the adaptive bacterial immune response, The HNH domain is found in all known Cas9 enzymes. It cuts the target strand, which is the DNA strand that matches the guide RNA sequence. Another thing that is needed is a RuvC nuclease domain to cut the non-target strand, which makes a double-strand DNA break (DSB).

The code number IPR006600 represents an HTH CenpB-type DNA binding domain. The CENPB-type HTH domain is a helix-turn-helix (HTH) domain that binds DNA and is found in transposases and eukaryotic centromere proteins. There are two DNA-binding HTH domains in CENP-B's N-terminus that bind to the primary grooves of adjacent DNA. A psq-type HTH domain, that is C-terminal to a CENPB-type HTH domain, forms the N-terminus of CENP-B. These two HTH domains specifically bind the CENP-B box, present in the alpha satellite DNA of human centromeres, when they combine. Mammalian centromere protein B (CENP-B) is a protein that is specifically binds DNA to the CENP-B box and is associated with centromeres. It contains CENPB-type HTH domains. The transposable elements Tigger and Mammalian Pogo are present.

The La RNA-binding protein and other proteins have the HTH La-type RNA-binding (IPR006630) domain at their N-terminus. In the nucleus, La functions as a transcription factor for RNA polymerase III (RNAP III), while in the cytoplasm, it functions as a translation factor. La facilitates the folding and maturation of nascent RNAP III transcripts by binding to the 3'UTR in the nucleus. In the cytoplasm, La can find a certain type of mRNAs that have

a 5' terminal oligopyrimidine (5'TOP) motif that controls protein production. The N-terminal domain of La, which comprises the La motif and an RNA recognition motif (RRM), is responsible for specific recognition.

tRNA intron endonuclease catalytic (IPR006677) is a homotetrameric tRNA-intron endonuclease that contains a three-layered α/β/α domain at the C-terminus. Additionally, it contains domains 2 and 4 (C-terminus) in a homodimeric enzyme. A tRNA-intron endonuclease cuts the pre-tRNA at the 5' and 3' splice sites to free the intron. This gets rid of the tRNA intron. The product consists of an intron, two tRNA halves with 2', 3' cyclic phosphates, and 5'-hydroxyl termini.

Sen15 (IPR018593), a tRNA-splicing endonuclease subunit. One of this heterotetrameric enzyme's two structural subunits is the Sen15 subunit of the tRNA intron splicing endonuclease. As part of the functional endonuclease assembly, two monomers of Sen15 fold with two monomers of one of the two catalytic subunits, Sen34, to form an α2-β2 tetramer.

tRNase Z endonuclease (IPR027794) domains are found in RNase Z enzymes, which are structurally similar to beta-lactamase B family members. tRNase Z, also known as ribonuclease Z, is an endonuclease that facilitates the maturation of the 3' end of tRNA by eliminating the 3'-trailer sequence from the tRNA precursor.

the zinc ribbon domain of the target nucleic acid binding (TNB) domain of the CRISPR-associated endodeoxyribonuclease Cas12f1 and analogous sequences found in all domains of life, is represented by Cas12f1-like TNB(IPR010095). This domain is also present in bacterial transposons; however, it does not demonstrate dsDNA nuclease activity. Rather, it efficiently binds to dsDNA targets in a PAM-dependent manner. It seems that Cas12m effectors frequently exhibit DNA binding activity without cleavage.

The information in Table 9 shows that tau protein has HNH endonuclease, HNH nuclease, HTH CENPB-type, HNH Cas9 enzyme, and HTH La-type RNA-binding domains. The information in Table 9 shows that tau protein has tRNA intron endonuclease catalytic, tRNA-splicing endonuclease subunit Sen15, and tRNase Z endonuclease domains included. The gRNA guides the cas enzyme HNH Cas9. This suggests that telomeric RNA directs the tau protein to the complementary intron region. Subsequently, the HNH Cas9 enzyme eliminates the complementary intron region. This means that the tau protein transposase activity's target region is close to the intron region that matches the guide RNA. This region could be either upstream or downstream of the target intron. This telomeric RNA-guided transposase can stop positioning itself over and over in the same DNA target region by getting rid of the target intron.

Table 9. RNA-guided HNH Cas9 enzyme domains of the tau protein

Domain	Alias	Motif	Start	End
HNH endonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	1	39
5	B	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHL	289	322
	C	INKKLDLSNVQSKCGSKDNIKHVPGGGSVQIV	595	626
	D	VTSKCGSLGNIHHKPGGGQ	635	653
	E	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGA EIVYKSPV	660	715
HNH nuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKE SPLQTPTEDG	1	55
	B	GPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGAP	286	331
	C	TRSSAKTLKNRPCLSPKHPTPGSSDPLIQ	397	425
	D	AKSRLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSN VQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGG GQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENA KAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQL	556	742

HNH Cas9-type	A	FEVMEDHAGTYGLGDRKDQGGYTMH	8	32
	B	IGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQI VYKPVDLSKVTSKCGSLGNIHHK	577	648
HTH CENPB-type	A	EFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	7	35
HTH La-type RNA-binding	A	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKC	592	639
tRNA intron	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQ	8	35
endonuclease	B	NLKHQPGGGKVQIINKKLDLSNVQSKC	582	608
catalytic	C	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDH	660	705
tRNA-splicing	A	GYTMHQ	28	33
endonuclease	B	VPGGGSVQIVYKPVDLSKVTSKC	617	639
subunit Sen15	C	LDNITHVPGGGNKKIETH	674	691
	D	LTFRENAKAKTDHGAEIVY	693	711
tRNase Z	A	STENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV	579	617
endonuclease	B	DNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	675	711
Cas12f1-like TNB	A	EDHAGTYGLGDRKDQGGYTMH	12	32
	B	THKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSS	690	730

(4) Endonuclease domains

For Extracellular Endonuclease subunit A (IPR020821) to work, it needs divalent ions like magnesium. It cuts DNA and RNA into single-stranded and double-stranded nucleic acids. This residue is located in a conserved region alongside aspartic acid residues, potentially linked to the binding of divalent ions.

DDE (IPR032874) A diverse array of transposases, such as those in IS240, IS26, IS6100, and IS26, contain the putative DDE endonuclease domain.

TnsA endonuclease N-terminal (IPR014833). Tn7 transposase is composed of two proteins, TnsA and TnsB. At the 5' end of the transposon, TnsA is responsible for DNA cleavage, while at the 3' end, TnsB is responsible for cleavage and ligation. TnsA's N-terminal domain mediates its catalytic effect. TnsA and gp4 proteins conserve this domain, which contains a catalytic triad of a basic and two acidic residues that form a restriction endonuclease-like active site.

Introns encode LAGLIDADG (IPR004860), a homing endonuclease (HEnase) that provides mobility to its host genetic elements. Two tandemly repeated homing endonuclease-like domains compose LAGLIDADG HEnases. The homing endonuclease is a prime example. LAGLIDADG domain

Group I intron-encoded endonucleases are a component of the PD-(D/E)XK(PF11645) endonuclease family. This family is a member of the PD-(D/E)XK superfamily.

NUMOD4(IPR010902) is a potential DNA binding motif that is present in homing endonucleases and related proteins.

Potential endonuclease Based on the GIY-YIG domain at the N-terminus of SegE GIY-YIG (PF19835) domains, it seems likely that they are endonucleases. This family includes the putative endonuclease segE from bacteriophage T4.

A hypothetical protein TT1808 (also known as Uma2 or TTHA1514), a putative restriction endonuclease (IPR008538), is an AT-rich DNA binding protein from Thermus thermophilus. The domain exhibits a 3-layer α/β/α topology that is comparable to that of restriction endonucleases. Enzymes involved in nucleic acid metabolism, particularly nucleases serving a variety of biological functions, frequently exhibit nuclease domains.

The terminase large subunit GpA endonuclease (IPR046454) includes GpA (also known as TerL) from

bacteriophage lambda and enterobacterial phage P21, as well as related sequences from multiple bacterial species. The DNA packaging enzyme terminase of bacteriophage lambda is a heteromultimer made up of the Nu1 and A genes. The small subunit, gpNu1, plays a crucial role in DNA recognition, while the large subunit, gpA, contributes to the late stages of packaging. During the initial packaging phases, terminase is involved in the site-specific binding and cleavage of DNA. This large subunit functions as an ATPase, which is necessary for the translocation of viral DNA into vacant capsids. Additionally, it functions as an endonuclease, cleaving the viral genome from the conidia to initiate and terminate the packaging reaction. The C-terminal domain, representing the nuclease/helicase, exhibits catalytic activity for DNA maturation.

There is an endonuclease domain in the bacteriophage terminase large subunit TerL that is important for packaging DNA in bacteriophages and related viruses. This domain is called Terminase large subunit-like endonuclease (IPR046462). TerL is composed of an N-terminal ATPase domain (IPR046461) that facilitates DNA translocation, and a C-terminal endonuclease domain that cleaves tandem DNA at the initiation stage at a sequence-specific site and at the completion stage of the DNA packaging process when the capsid is full.

EcoEI R protein C-terminal protein (IPR013670) is involved. We classify restriction endonucleases into four categories: type I, type II, type III, and type IV. Different kinds of enzymes can recognize certain short DNA sequences and cut DNA at the ends to make specific double-stranded pieces with 5'-phosphates at the ends. Type I enzymes are multifunctional proteins that possess methylase and restriction activities. They cleave at sites that are distant from the recognition site and require ATP and S-adenosyl-L-methionine to function. Type III enzymes cleave at a site that is not far from the recognition site, require ATP (but do not hydrolyze), and exist as complexes with the modification methylase methylase. S-adenosyl-L-methionine stimulates the reaction, but it is not essential. This entry denotes the C-terminal domain that is present in the R subunit of the type I enzyme and the Res subunit of the type III enzyme. You need the R protein (HsdR) to do both nuclease and ATPase activities, and Type I enzymes like EcoEI can recognize 5'-GAGN(7)ATGC-3.

The C-terminus of the HindI endonuclease subunit (also known as the R protein) of the type I restriction endonuclease of Haemophilus influenzae, as well as in related sequences from bacteria and archaea, is home to the Type I restriction enzyme HindI endonuclease subunit-like C-terminal(IPR021810) domain of the type I restriction enzyme.

The methylase domain (IPR002052) is often found next to the Restriction endonuclease type I HsdR N-terminal (IPR007409) domain in restriction endonucleases or methylases. The R subunit (HsdR) of type I enzymes contains the N-terminal domain. The representative type I enzyme is EcoKI, which recognizes the DNA sequence 5'-AACN6GTGC-3'. The R protein (HsdR) is necessary for nuclease and ATPase activities. The Type I enzyme functions by requiring ATP and S-adenosyl-L-methionine, cleaves at a location distant from the recognition site, and is a multifunctional protein that also performs methylase and restriction functions. The prokaryotic DNA restriction modification machinery, which keeps organisms safe from foreign DNA invasion, depends on type I restriction endonucleases.

Type II restriction endonuclease EcoO109IR (IPR032793) contains type II restriction endonuclease domains.

Restriction endonuclease type II EcoRII C-terminal (IPR015109) is a restriction endonuclease-like fold with a central five-stranded mixed β-sheet flanked by α-helices. EcoRII specifically cleaves DNA at a single 5' CCWGG site.

The N-terminal domain of CD-NTase-associated protein 4 (Cap4) corresponds to the CD-NTase associated protein 4-like DNA endonuclease (IPR025382). This part of the protein is an effector domain that can cut dsDNA, is structurally similar to type II restriction endonucleases, and has conserved residues in the active site that are needed for metal coordination.

Mrr-like (IPR039442) is similar to Mrr and contains the (D/E)-(D/E)XK active site, which is present in type II

restriction endonucleases.

Type II restriction endonucleases, including DpnII, contain restriction endonuclease type II DpnII-like (IPR007637). It recognizes the double-stranded unmethylated sequence GATC and cleaves it prior to G-1. The protein's complete length is present. It is also present in numerous proteins with unknown functions. The DNA adenine-specific methyltransferase domain contains it.

The C-terminus of BsuBI (PF06616) is analogous to that of the bacterial enzymes of type II restriction endonucleases BsuBI and PstI (EC: 3.1.21.4). The BsuBI restriction/modification (R/M) system has an enzyme that works the same way as the PstI R/M system enzyme and can recognize the target sequence 5'CTGCAG.

The type III restriction enzyme C-terminal endonuclease (IPR045572) represents the C-terminal endonuclease domain of a variety of type III restriction endonuclease proteins, including the EcoPI/EcoP15I Res subunit of the type III restriction modification system EcoPI from Escherichia coli and the BceSI Res subunit of the type III restriction modification enzyme BceSI from Bacillus cereus.

It refers to the C-terminal endonuclease domain as endonuclease (PF19778), which is present in a diverse array of type III restriction endonuclease proteins.

The Restriction endonuclease AspBHI N-terminal (PF18062) is a homotetrameric protein that recognizes 5-methylcytosine in the context of a double-stranded DNA sequence of (C/T) (C/G) (5mC) nucleotides. It cleaves both strands to the modified cytosine at a fixed distance (N12/N16). 3. The N-terminal domain is structurally and functionally analogous to the SRA's 5-methylcytosine binding domain, and it is responsible for DNA recognition.

Restriction endonuclease type IV Mrr (IPR007560), is responsible for the acceptance of modified foreign DNA and the restriction of adenine and cytosine methylated DNA. The cleavage domain of Mrr is similar to that of type II restriction endonucleases. However, the (D/E)-(D/E)XK signature shows that it has a unique glutamine residue in the middle of the active site.

The tau protein has transposase endonuclease and restriction endonuclease domains, with active regions at the N-terminus, C-terminus, and middle part (Table 10). The transposase endonuclease function is exclusively applicable to tau protein's transposase function. The HNH Cas9 enzyme function is the primary application of the restriction endonuclease.

Table 10. Tau protein's endonuclease domains

Domain	Alias	Motif	Start	End
Extracellular	A	QEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	6	39
Endonuclease subunit A	B	FTFHVEITPNVQKEQAHSEEHLGRAAF	301	327
DDE	A	NKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIH	596	646
	B	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDH		705
DDE Tnp4	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	1	35
	B	NLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV	582	617
	C	IETHKLTFRENAKAKTDHGAEIVYKSPV	688	715
DDE-1	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	1	35
	B	LEFTFHVEITPNVQKEQAHSEEHLGR	299	324
	C	GKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIH	590	646
	D	KSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKS	657	713

TnsA endonuclease N-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTD	1	40
	B	FTFHVEITPNVQKEQAHSEEHLGRAAF	301	327
	C	TENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK PVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNIT HVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSN VSSTGSIDM	580	736
Homing endonuclease LAGLIDADG	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	1	42
	B	FTFHVEITPNVQKEQAHSEEHLGRAAF	301	327
	C	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGS LGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETH KLTFRENAKAK	592	702
PD-(D/E）XK	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	NVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	603	647
	C	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	662	711
NUMOD4	A	EPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQ T	3	50
	B	FKDRVQSKIGSLDNI	663	677
Putative endonuclease SegE GIY-YIG	A	EPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	3	43
	B	AKGQDAPLEFTFHVEITPNVQKEQA	292	316
	C	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGS LGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNIT	592	678
	D	IETHKLTFRENAKAKTDHGAEIVYKSPVV	688	716
Putative restriction endonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESP LQT	1	50
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPG	295	329
	C	VPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDN IKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLD FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK SPVVSGDTSPRHLSNVSSTGSIDMVD	565	738
Terminase large subunit GpA endonuclease	A	MAEPRQEFEVMEDHAGTYG	1	19
	B	FHVEITPNVQKEQAH	303	317
	C	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQ	591	624
	D	LDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	674	712
Terminase large subunit-like endonuclease	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	5	41
	B	IPLPVDFL	265	272
	C	FTFHVEITPNVQKEQAHSEEHLGR	301	324
	D	DLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV PGGGSVQIVY	569	627
	E	VTSKCGSLGNIHH	635	647
	F	HKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDMV	691	737
EcoEI R protein	A	YTMHQDQEGDTDA	29	41

C-terminal	B	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDH	663	705
Type I restriction	A	FTFHVEI	301	307
enzyme HindI	B	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKC	591	639
endonuclease subunit-like C-terminal	C	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEI VYKSPVVSGDTSPRHLSNVSSTGSIDMV	660	737
Restriction	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQD	1	34
endonuclease type I	B	RAKGQDAPLEFTFHVEITPNVQKEQAH	291	317
HsdR N-terminal	C	NVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGG GSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKD	572	665
	D	SLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	673	711
Type II restriction	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTD	5	40
endonuclease	B	PLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	298	328
EcoO109IR				
Restriction	A	VQIINKKLDLSNVQSKCGSKDNIKH	592	616
endonuclease type II				
EcoRII C-terminal				
Restriction	A	EPRQEFEVMEDHAGTYGLGDRKDQGGYT	3	30
endonuclease type II	C	FRENAKAKTDHGAEIVY	695	711
EcoRII N-terminal				
CD-NTase	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	5	42
associated protein	B	EFTFHVEITPNVQKEQAH	300	317
4-like DNA	C	TENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	580	627
endonuclease				
Mrr-like	A	FTFHVEITPNVQKEQAHSE	301	319
	B	DLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDN	569	613
Restriction	A	QEFEVMEDHAGTYGLGDRKDQGGYTMHQD	6	34
endonuclease type II	B	FTFHVEI	301	307
DpnII-like	C	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKC	592	639
BsuBI	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	5	35
	B	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTS	593	637
	C	DNITHVPGGGNKKIETHKLTFR	675	696
Type III restriction	A	CGSKDNIKHVPGGG	608	621
enzyme C-terminal	B	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEI V	660	710
endonuclease				
Endonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
	B	SHQLMSGM	146	153
	C	GQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	294	328
	D	TLKNRPCLS	403	411

	E	IGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIV YKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLD NITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHL SNVSSTGSIDMV	577	737
Restriction endonuclease AspBHI N-terminal	A	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFREN	662	698
Restriction endonuclease type IV Mrr	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	1	39
	B	FTFHVEITPNVQKEQAHSEEHLGRAAFPGA	301	330
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIK HVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFK DRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSP VVSGDTSPRHLSN	567	727

(6) Ribonuclease domains

EndoU (PS51959) Eukaryotic uridylate-specific endoribonuclease. EndoU is a family of metal-dependent endoribonucleases that are highly conserved in eukaryotes. People in the EndoU family are thought to be involved in many biological processes, including building the endoplasmic reticulum (ER) network, the immune response, neurodegeneration, and the production of small nucleolar RNA. They also possess endoribonuclease and RNA binding activities. The EndoU domain releases the 2',3'-cyclic phosphodiester terminus by cleaving RNA at the uridylate.

Bacterial toxin RNase RnlA (IPR045837). RnlA (RNase LS) and LsoA are type II TA systems that are involved in the defense of bacteriophages. Following infection with bacteriophage T4, bacteriophage T4 activates RnlA and LsoA, revealing their endoribonuclease activity. RnlA is a stable toxin that inhibits cell growth and swiftly degrades T4 late mRNA to prevent its expression. The activity of the unstable antitoxin RnlB neutralizes RnlA.

Bacterial EndoU nuclease (IPR029501) is the bacterial equivalent of the eukaryotic EndoU nuclease. It is located in the polymorphic toxin protein's C-terminal region. People frequently use nucleases as contaminants for defense, offense, or addiction to selfish molecules.

RNase H type-1 (PS50879) The ribonuclease H (RNase H) (EC 3.1.26.4), which is part of the ribonuclease family, can find and cut the RNA strand in RNA-DNA heteroduplexes. RNase HI is the primary source of RNH activity. All LTR retrotransposons have acquired an enzymatically feeble RNase H domain, devoid of the critical catalytic residues present in all other RNase H enzymes. All RNase H domains have the same catalytic residues for enzyme activity. These are made up of three aspartic acid residues and one glutamic acid residue.

RNase H type-2 (PS51975) The ribonuclease H (RNase H) (EC 3.1.26.4), which is part of the ribonuclease family, can find and cut the RNA strand in RNA-DNA heteroduplexes. Type I and II RNase H share similar structural characteristics of the catalytic domain; however, they utilize distinct substrate ranges for enzymatic cleavage. Archaea, eubacteria, and all eukaryotes contain RNase HII.

Ribonuclease III (RNase III) is the component of RNase III (IPR011907). This enzyme, which is present in all bacteria and eukaryotes, is ubiquitous and specifically cleaves double-stranded rRNA. Large precursor ribosomal RNA molecules cleave at specific sites during the processing of pre-rRNA to produce the immediate precursors of functional molecules. RNase III is also involved in the maturation and degradation of mRNA and tRNA. The protein's N-terminal two-thirds house the endonuclease domain, which consists of numerous alpha helices but lacks beta strands.

Ribonuclease A-domain (IPR023412) Pyrimidine-specific endonucleases, known as pancreatic ribonucleases

(RNaseA), are abundant in the pancreas of specific mammals and certain reptiles. In this process, these enzymes help cut 3'-phosphomonucleoside and 3'-phosphotinonucleotides that end in C-P or U-P with a 2',3'-cyclic phosphate intermediate.

Exoribonuclease II (RNase II), a product of the RNB gene, composes Ribonuclease II (IPR011804), which is present in numerous gammaproteobacteria. It has the ability to unravel the DNA helix by forming a complex with single-stranded DNA. It is one of eight distinct exoribonucleases in Escherichia coli. It is responsible for mRNA degradation and tRNA precursor processing in the terminal region.

IPR012588 is the Exosome-associated factor Rrp6 N-terminal. Exosomes are nanocompartments that are involved in the degradation or processing of RNA, such as mRNA, rRNA, snRNA, and snoRNA. Exosomes are ubiquitous in archaea and eukaryotes, where they engage in interactions with bacteria and organelles.PNPase (polynucleotide phosphorylase) is a barrel-like structure that is composed of a hexameric ring of PH domains that functions as a degradation compartment and a cap containing an S1 domain/KH domain that binds RNA substrates (and occasionally accessory proteins) to regulate and restrict access to the degradation compartment. Exosomes come in two types in eukaryotes: cytoplasmic exosomes and nuclear exosomes. Cytoplasmic exosomes break down mRNA using 3'-5' exoribonuclease, while nuclear exosomes break down pre-mRNA (like nonsense transcripts) and rRNA, snRNA, and snoRNA.

Exoribonuclease phosphorolytic (IPR001247) PH (phosphorylation) domain is responsible for 3'-5' exoribonuclease activity. The PH domain is present in bacterial/organellar RNase and PNP enzyme (polynucleotide phosphorylase). The active PH domain, using inorganic phosphate as a nucleophile, releases ribonucleoside 5'-diphosphate (rNDP) from the 3' end of the RNA substrate and adds it to the phosphodiester bond between the two terminal nucleotides.

Table 11 demonstrates that the tau protein possesses RNase activity, and RNase III activity is nearly ubiquitous throughout the sequence. It looks like the tau protein can find the double-stranded guide RNA of the telomeric fragment and the target DNA sequence. It can then cut the guide RNA of the telomeric fragment. By the Exosome-associated factor Rrp6 N-terminal and Exoribonuclease phosphorolytic domain of tau protein, the cut guide RNA is transported to the exosome for its degradation.

Table 11. Tau protein's ribonuclease domains

Domain	Alias	Motif	Start	End
EndoU	A	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	591	627
	B	LDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAK	661	700
Bacterial toxin RNase	A	GYTMHQDQ	28	35
RnlA	B	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIV	593	626
Bacterial EndoU	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	1	51
nuclease		KESPLQTP		
	B	LEFTFHVEITPNVQKEQAHSEEHLGRAAF	299	327
	C	PDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDN	568	656
		IKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEV		
	D	SEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKT	658	712
		DHGAEIVYK		
RNase H type-1	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	1	55
		KESPLQTPTEDG		
	B	HQLMSGM	147	153
	C	APLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGA	297	330

	D	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNV QSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPG GGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFR ENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDM	559	736
RNase H type-2	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL KESPLQTPTEDGSE	1	57
	B	DAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	296	328
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSK DNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQ	567	653
	D	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGA EIVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQL	662	742
RNase III	A	PRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	4	35
	B	LEFTFHVEITPNVQKEQAHSEEH	299	321
	C	TLKNRPC	403	409
	D	NLKHQ	582	586
	E	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSK CGSLGNIHH	593	647
	F	DRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIV Y	665	711
Ribonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
A-domain	B	FTFHVEI	301	307
	C	APVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKC GSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	563	647
	D	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAE IVY	663	711
Ribonuclease II	A	TLKNRPCL	403	410
	B	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSK DNIKHV	567	617
	C	DNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDT SPRHLSNVSST	675	731
Exosome-associated factor Rrp6 N-terminal	A	LDFKDRVQSKIGSLDNI	661	677
Exoribonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
phosphorolytic	B	EFTFHVEITPN	300	310
	C	VPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGS KDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	565	647

(7) Domains of exonuclease

3-5 exonuclease (IPR002562) This domain is in charge of the 3'-5' exonuclease proofreading activity of Escherichia coli DNA polymerase I (polI) and other enzymes that break down nucleotides that are not paired or that don't match. A genetic disease in humans that makes people age faster than they should is also present in 3'-5' exonuclease/ATP-dependent helicase activity.

3-5 exoribonuclease Rv2179c-like (IPR033390) This is a unique collection of 3' exonuclease domains. The

active site residues create a magnesium catalytic center, which cleaves the 3' overhang in a time-dependent manner, forming a typical RNase fold. It bears a striking resemblance to a DEDD-type RNase, a unique ATP-binding protein that binds both ATP and dATP. It dimerizes and bonds magnesium ions through Asp-6.

ExoN is a 3'-5' exoribonuclease.

ExoI C-terminal (PS51785) is the C-terminal domain of exonuclease I (ExoI). Exonuclease I (ExoI) is a monomeric processive 3'-5' exonuclease that degrades single-stranded DNA in bacteria. The enzyme is primarily believed to be involved in frameshift mutation repair.

Exonuclease (IPR013520) is an RNase T/DNA polymerase III exonuclease. Numerous exonuclease proteins, such as the epsilon subunit of DNA polymerase III and RNase T, contain this domain. RNase T, responsible for the terminal turnover of tRNA, removes the terminal AMP residue of uncharged tRNA. DNA polymerase III, a complex multi-chain enzyme responsible for most of the replicative synthesis in bacteria, also exhibits 3' to 5' exonuclease activity.

The predicted 3-5 exonuclease PolB-like (IPR019288) represents various prokaryotic 3'-5' exonucleases and hypothetical proteins.

5-3 exonuclease (IPR002421), acting as a flap endonuclease (FEN), catalyzes the exonucleolytic hydrolysis of blunt-ended double-stranded DNA substrates and the endonucleolytic cleavage of 5'-forked nucleic acids at the junctions formed between single-stranded double-stranded DNA. The basic FEN reaction can be done by the N-terminal 5'-3' exonuclease domain of DNA polymerase I.

The 5-3 exonuclease alpha-helical arch N-terminal (IPR020046) represents the N-terminal and internal 5'3'-exonuclease domains, which are typically present together and most frequently associated with 5' to 3' nuclease activities. Researchers have linked proteins like these to the suppression of host gene expression.

A domain of predicted phosphatases, including Drosophila prune proteins and bacterial RecJ exonucleases, is DDH(IPR001667). Escherichia coli's RecJ protein is crucial in numerous DNA repair and recombination pathways. In the 5'-3' direction, RecJ catalyzes the degradation of single-stranded DNA.

Exodeoxyribonuclease X-like C-terminal (PF20600) is a deoxyribonuclease X-like C-terminal. This domain, which is conserved FGKY/H sequence motif, is present in bacteria and eukaryotes and is situated at the C-terminus of exodeoxyribonuclease 10 (ExoX) of Escherichia coli (strain K12).

Two non-identical subunits, one large subunit and four minor subunits, make up the C-terminal of the large subunit of exonuclease VII (IPR020579). A substance called exonuclease VII cuts exons in two directions, either 5'-3' or 3'-5'. This creates 5'-phosphate mononucleotides. An OB-fold domain is also present in the large subunit, which is responsible for binding to the N-terminal nuclease.

Putative exodeoxyribonuclease 8 PDDEXK-like (IPR024432) represents PDDEXK-like domains of unknown function in proteins from bacteria and viruses. Exodeoxyribonuclease 8, a component of the RecE pathway of recombination, annotates many of these proteins. Enzymes involved in nucleic acid metabolism, particularly nucleases, contain ubiquitous PD-(D/E)XK domains.

Table 12 shows that tau protein has an exonuclease domain that is either 5'-3' or 3'-5' long and acts as an exonuclease all the way through the sequence. This suggests that tau protein cuts the target DNA sequence in two different directions, 5'-3' and 3'-5'. These cuts make it easier for the double helix to unwind and for the transposase domain to insert into the transposed telomeric sequence fragment.

Table 12. Tau protein's exonuclease domains

Domain	Alias	Motif	Start	End
3-5 exonuclease	A	PRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	4	43
	B	LEFTFHVEITPNVQKEQAHSEEHLGRAAFP	299	328
	C	VQIINKKLDLSNVQSKCGSKDNIKHV	592	617

	D	QIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKS	624	713
3-5	A	QEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	6	43
exoribonuclease	B	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEH	289	321
Rv2179c-like	C	NLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSN	582	727
ExoN	A	PRQEFEVMEDHAGTYGLGDRKDQGGYTMHQD	4	34
	B	HQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	585	627
	C	DNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRH	675	724
ExoI C-terminal	A	IINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSL	594	642
Exonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQTPTEDG	1	55
	B	VVQEGFLREPGPPGLSHQLMSGMP	131	154
	C	PSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGAP	287	331
	D	SAKTLKNRPCLSPKH	400	414
	E	LQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGSIDM	560	736
Predicted 3-5	A	EFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	7	43
exonuclease	B	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNI	581	645
PolB-like	C	LDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	661	712
5-3 exonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37
	B	LEFTFHVEITPNVQKEQAHS	299	318
	C	KSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	574	647
	D	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	660	711
5-3 exonuclease	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
alpha-helical arch	B	NVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV	572	617
N-terminal	C	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFR	663	696
DDH	A	MAEPRQEFEVMEDHAGTYGLG	1	21
	B	VTSKCGSLGNIHH	635	647
	C	NKKIETHKLTFRENAKAKTDHGAEIVY	685	711
DHHA1	A	FEVMEDHAGTYGLGDRKDQGGYTMHQD	8	34
Exodeoxyribonuclease X-like C-terminal	A	FEVMEDHAGTYGLG	8	21
Exonuclease VII	A	HAGTYGLGDRKDQGGYTMHQDQ	14	35
large subunit	B	FTFHVEITPNVQKEQAHSEEHLGRAAF	301	327
C-terminal	C	NLKHQPG	582	588

	D	LDFKDR	661	666
	E	LDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	674	711
Putative exodeoxyribonuclease 8 PDDEXK-like	A	YTMHQDQ	29	35
	B	LEFTFHVEITPNVQKEQAHSEEHLGRAAF	299	327

(8) Nuclease domains for DNA repair

The DNA damage checkpoint implicates DisA (IPR023763), a DNA integrity scanning protein. DNA damage activates this checkpoint prior to asymmetric division. During sporulation, DisA generates spherical foci that rapidly traverse chromosomes in search of lesions. When a lesion is present, disA pauses at the lesion site. This triggers a cellular response that ultimately results in a transient halt in the initiation of sporulation.

FPG-type (PS51066) is a formamidopyrimidine-DNA glycosylase. A DNA glycosylase from bacteria prefers to use oxidized purines, like 7,8-dihydro-8-oxoguanine (8-oxoG), to find and get rid of damaged bases. The oxidized guanine is cut out by nucleophilic substitution of the N-terminal proline at C1'. The aminal intermediate that is subsequently generated endures β and δ elimination, resulting in the complete removal of the diseased nucleoside from the DNA. Subsequently, it rearranges to a Schiff base. Consequently, FPG also has a nick at the purine/pyrimidine (AP) site, resulting in a gap in the DNA that contains phosphate groups at both the 5' and 3' extremities.

Smr (IPR002625) designates the C-terminal region of the MutS2 protein from bacteria and plants. The Smr domain is linked to chromosome crossing over, segregation, and mismatch repair (MMR). Research suggests that the Smr domain acts as a nicking endonuclease.

Formamidopyrimidine-DNA glycosylase catalytic (IPR012319) represents the catalytic domain of DNA glycosylase/AP lyase, which is involved in the base excision repair of DNA damaged by oxidation or mutagens. The base excision repair pathway is responsible for the repair of the majority of DNA base damage. This domain possesses AP lyase activity and DNA glycosylase activity.

A conserved aspartate follows the signature helix-hairpin-helix and Gly/Pro-rich loop of the HhH-GPD domain (IPR003265). An A/G-specific adenine glycosylase called DNA glycosylase MutY is one of many structurally related DNA repair proteins that contain this domain.

A number of DNA repair nucleases have the domain ERCC4 (IPR006166), such as Rad1, XPF, and the cross-linking endonucleases EME1 and Mus81. The XPF/Rad1/Mus81-dependent nuclease family plays a crucial role in maintaining genomic stability by precisely cleaving branched structures generated during DNA repair, replication, and recombination. The architecture of the nuclease domain is strikingly similar to that of restriction endonucleases.

A group of nucleases called GIY-YIG (PS50164) is involved in many things that happen inside cells, like moving mobile genetic elements, fixing and recombining DNA, and stopping new DNA from entering the cell. A small group of molecules called GIY-YIG domains work together to help pair up divalent metal ions that are needed to cut phosphodiester bonds quickly.

The Escherichia coli DNA damage-inducible protein P or DinP contains the protein UmuC (IPR001126). Many chemicals and UV light are thought to cause mutagenesis in E. coli by controlling DNA polymerase III and SOS through the translesion synthesis of UmuD, UmuC, and RecA proteins. This mechanism allows replication to continue through DNA lesions, thereby preventing lethal DNA replication interruptions that occur as a result of DNA damage.

The repair of interstrand DNA crosslinks involves FAN1, a structure-selective DNA repair nuclease with 5'

flap endonuclease activity. VRR-NUC (IPR014883) contains proteins with VRR-NUC domains. This domain is a member of the PD-(D/E)XK nuclease superfamily, which also encompasses type III restriction-modifying enzymes. The type VI secretion system (T6SS) secretes the TseVs antimicrobial effector, which contains this domain. TseVs are associated with the evolution of enzymes involved in DNA interstrand crosslink repair and Holliday ligation resolution.

Table 13 demonstrates that the Tau protein's C-terminus is a DisA domain, which it uses to scan DNA damage checkpoints. The full sequence of the Tau protein has FPG-type activities(such as DNA glycosylase), foramidopyrimidine-DNA glycosylase catalytic, and HhH-GPD activities, which can find and get rid of damaged bases. The nuclease activities of ERCC4, GIY-YIG, UmuC, and VRR-NUC in the Tau protein can fix DNA sequences that have been damaged by transposition.

Table 13. Tau protein's DNA repair nuclease domains

Domain	Alias	Motif	Start	End
DisA	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
FPG-type	A	AEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	2	36
	B	EFTFHVEITPNVQ	300	312
	C	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVD LSKVTSKCGSLGNIHH	581	647
	D	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	662	711
Smr	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	5	36
	B	FTFHVEITPNVQ	301	312
	C	KNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGG SVQIVY	571	627
	D	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	663	712
Formamidopyrimi dine-DNA glycosylase catalytic	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLK	1	44
	B	LEFTFHVEITPNVQKEQAHS	299	318
	C	NLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDL SKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGG GNKKIE	582	689
	D	HKLTFREN	691	698
HhH-GPD	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	1	42
	B	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	289	328
	C	KNRPCLSPKHPTPGSS	405	420
	D	KIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY KPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITH VPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVV	576	716
ERCC4	A	QDAPLEFTFHVEITPNVQKEQAHS	295	318
	B	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKC	591	639
	C	DNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPR	675	723
GIY-YIG	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQ TPTEDG	1	55
	B	HQLLGDLHQ	190	198

	C	IPASEPDGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	279	328
	D	AKTLKNRPCLSPKH	401	414
	E	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCG SKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEK LDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK SPVVSGDTSPRHLSNVSSTGSIDMVDSPQL	559	742
UmuC	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTD	1	40
	B	EFTFHVEITPNVQ	300	312
	C	NKKLDLSNVQSKC	596	608
	D	YKPVDLSKVTSKC	627	639
	E	SLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITH	641	679
VRR-NUC	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	PLEFTFHVEITPN	298	310
	C	KNRPCL	405	410
	D	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV PGGGSVQIVYKPVDLSKVTSKCGSLGNIH	567	646
	E	NITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	676	711

3.7 Domains of transcription factors

(1) BOX binding domains

T-box (IPR046360) is a DNA binding domain that is highly conserved in T-box transcription factors. The T-box family of transcription factors is very important for deciding early on what cells will do, like making the basic shape of a vertebrate body, differentiating, and building organs. Additionally, they are involved in various aspects of development and terminal cell type differentiation in various animal lineages. Particular organs or cell types are likely to express T-box proteins, particularly during development, and they are frequently necessary for the development of these tissues. All members of this family are believed to bind to the DNA consensus sequence TCACACCT.

Transcription factor proteins contain the AP2/ERF domain, also known as AP2 (IPR001471). The GCC-box is a short, G/C nucleotide-rich pattern that is set by the ethylene response element (ERE), a cis-regulatory element found in genes that deal with defense. A sequence of transcription factors recognizes the ERE binding factor (ERF). Almost all chitinase and glucanase promoters contain a common motif AGCCGCC (AGC box) [47]. This structure is comparable to that of the methyl-CpG binding domain (MBD) and the bacteriophage integrase. It interacts with the DNA by means of the Arg and Trp residues that are situated in the β-sheet.

The GATA type (IPR000679) contains the zinc finger (Znf). Several transcription factors, including erythroid-specific transcription factors and nitrogen regulatory proteins, specifically bind the DNA sequence (A/T)GATA(A/G) in the regulatory regions of genes. Therefore, we refer to them as GATA-binding transcription factors. GATA transcription factors contain two zinc fingers. Nevertheless, the domain is present in only one copy of a number of proteins.

TATA box binding protein associated factor (TAF) histone-like fold (IPR004823) TATA box binding protein associated factor (TAF) is a component of the transcription initiation factor TFIID multimeric protein complex. TFIID plays a central role in mediating the promoter's response to a variety of activators and inhibitors. It interacts directly with TAFII-40 and binds securely to TAFII-250. It is a general transcription factor complex that binds to DNA. TFIID is essential for the initiation of DNA-dependent RNA polymerase II transcription. TFIID is the sole

universal transcription initiation factor that binds to the TATA box. The initial step in the formation of a complex that can initiate transcription is the binding of TFIID to the TATA-box. The TATA box is characterized by a consistent sequence of TATA(A/T)A(A/T) (non-template chain sequence). It is primarily composed of A-T base pairs and is situated approximately -30bp (-25~-32bp) upstream of the transcription start site of most eukaryotic transcripts. It is one of the RNA polymerase's binding sites and determines the choice of gene transcription initiation. The transcription process can only commence once the RNA polymerase has established a strong bond with the TATA box.

The DNA consensus sequence regions bound by GATA-type, TATA box binding protein associated factor (TAF) histone-like fold structures can both be "TAA" and are located in the promoter, as demonstrated in Table 14. Tau protein has T-box, AP2, GATA-type, and TATA box binding protein associated factor (TAF) histone-like fold domains. The chitinase and glucanase promoters are in close proximity to the GCC-box consensus sequence, to which AP2 binds. The promoters associated with cell differentiation and development are in close proximity to the T-box consensus sequence. It suggests that the telomeric retrotransposase of Tau protein inserts the transposed telomere into the "TAA" region of the TATA box promoters. Chitinase, glucan, and genes associated with cell differentiation and development comprise the majority of the promoters.The main function of T-box, AP2, and GATA-type domains is to help tau protein bind nearpromoters.

Table 14. Tau protein's box-binding protein domains

Domain	Alias	Motif	Start	End
T-box	A	QEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGD	6	38
	B	QDAPLEFTFHVEITPNVQKEQ	295	315
	C	GGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKC GSLGNIHHKPGGGQVEV	588	656
	D	RVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTD	666	704
AP2	A	MHQDQEGDTDAGLKESPLQTPTEDGSEEPGSETSDAKST	31	69
	B	APLVDEGAPGKQAAAQPHTEIPEGTTAEEAGIGDTPSLE	77	115
	C	RPPQTAAREATSIPGFPAEGAIPLPVDFLSKVSTEIPASEPDGPSVGRAKGQDA PLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGAPGEGPEARGPSL	244	342
	D	PAAAPRGKPVSRVPQLKARMVSKSKDGTGSDDKKAKTSTRS	359	399
	E	VSSVTSRTGSSGAKEMKLKGADGKTKIATPRGAAPPGQKGQANATRIPAKTP PAPKTPPSSGEPP	442	506
	F	RSGYSSPGSPGTPGSRSRTPSLPTPPTREPKKVAVVRTPP	511	550
	G	GSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEK LDFKDRVQSKIGSLDNITHV	609	680
GATA-type	A	MEDHAGTYGLGDRKDQGGYTMHQDQ	11	35
	B	EFTFHVEITPNVQKEQAHSEEHLGRAAF	300	327
	C	IVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQ	625	668
	D	ITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVV	677	716
TATA box binding protein associated factor (TAF) histone-like fold	A	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	663	711

(2) Domains of transcription factor-like

The transcription initiation factor IIA gamma subunit C-terminal is also known as IPR015871. RNA polymerase II, TATA-box binding protein (TBP), and TBP-related factors make up the transcription pre-initiation complex, located on the TATA-box sequence upstream of the start site. Transcription factor IIA is one of several factors that contribute to this complex. TFIIA stabilizes the binding of TBP to the TATA element by binding to TBP.

TFIIB-type (IPR013137) represents the zinc finger motif present in transcription factor IIB (TFIIB). The polymerase II complex (Pol II) starts the transcription of protein-coding genes with the help of general and specific transcription factors. General transcription factors operate through common promoter elements, including the TATA box. TFIIB and TFIID are responsible for the promoter's recognition and interaction with Pol II. In conjunction with Pol II, they form a minimal initiation complex that is capable of transcription under specific conditions. In the initiation complex, the TBP subunit of TFIID binds to the TATA box of the Pol II promoter. This makes the DNA bend around the C-terminal domain of TFIIB. Meanwhile, the TFIIB's N-terminal zinc finger interacts with Pol II.

Transcription initiation factor TFIID subunit 12 (IPR003228) is the transcription initiation factor. TFIID is one of several general transcription factors (GTFs) that are involved in the precise initiation of RNA polymerase II transcription in eukaryotes. The other GTFs are TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH. TFIID significantly influences the recognition of promoter DNA and the assembly of the pre-initiation complex. TATA binding protein (TBP) and at least 13 TBP-associated factors (TAFs), individually or collectively involved in activator-dependent transcription, compose the human transcription initiation factor TFIID.

TATA binding protein (TBP) and a number of TBP-associated factors (TAFs) combine to form TAFII28-like protein (IPR006809), a general transcription factor TFIID. These factors work together to form the pre-transcription initiation complex. This entry denotes the conserved region of the human transcription initiation factor. TFIID subunit 11 (also referred to as the transcription initiation factor TFIID 28 kDa subunit, TAFII28)

TATA binding protein (TBP) composes the TAFII55 protein conserved region (IPR006751), a general transcription factor TFIID. TBP is associated with a series of TBP-associated factors (TAFs), which collectively contribute to the assembly of the pre-transcription initiation complex. TAFII55 binds to TAFII250, thereby inhibiting its acetyltransferase activity.

TFIIF beta subunit HTH is IPR040450. In addition to RNA polymerase II, a minimum of six general transcription initiation factors are necessary for accurate transcription in vivo. A tetramer comprising two large subunits (TFIIF α or RAP74) and two small subunits (TFIIF β or RAP30) is known as transcription initiation factor IIF (TFIIF). The recruitment of RNA polymerase II to the promoter is contingent upon the presence of the β subunit of TFIIF.

If you look at the TFIIH p62 subunit N-terminal (BTF2-p62) of the TFIIH basal transcription factor complex, you can see that it is part of the IPR013876 group. This region is necessary for activity because it interacts with the 3' endonuclease XPG. A critical element of the nucleotide excision repair apparatus is the 3' endonuclease XPG. The structure of the N-terminal region demonstrates the pleckstrin homology (PH) fold.

TFIIS-type (IPR001222) is a zinc finger motif that is present in transcription factor II (TFIIS). During mRNA elongation, Pol II may encounter DNA sequences that cause the enzyme to retrace. This backtracking process may lead to transcriptional stalling, which entails the extrusion of the RNA 3'-end into a pore. TFIIS must assist in cleaving the extruded RNA to escape the delay. This process makes RNA polymerase (Pol) II's intrinsic nuclease activity stronger. Pol II then goes through template-encoded pause sites and cuts mRNA.

The transcription factor IIIC subunit 5 HTH (IPR019136) is involved. An important DNA-binding protein called transcription factor IIIC (TFIIIC) acts as a moving platform for the assembly of preinitiation complexes at classification III genes. It is the Tfc1/Sfc1 subunit, also known as the tau 95 subunit, and it occupies a critical position in TFIIC. It has both upstream and downstream effects on the TFIIIC-DNA complex, reducing its

instability. Once TFIIIC binds to promoter elements within tDNA genes, it guides the assembly of TFIIIB on DNA, which in turn recruits RNA polymerase III (pol III) and initiates multiple transcription cycles.

In a group of proteins that make up transcription factor IIIC (TFIIIC),Transcription factor TFIIIC triple barrel is always present (IPR019481). For the transcription of tRNA and 5 S RNA genes by RNA polymerase III, TFIIIC is necessary.

The transcription factor CBF (IPR003958) domain is present in archaeal histones as well as histone-like transcription factors in eukaryotes.

Transcription-repair-coupling factor C-terminal (IPR005118). The transcription-repair coupling factor (TRCF), which is the product of the mfd gene, is responsible for the coordination of DNA repair and transcription in bacteria. TRCF directs nucleotide excision repair (NER) machinery to the site and removes transcription elongation complexes that have become stalled at DNA lesions. It improves transcription-dependent inhibition of nucleotide excision repair (NER) by finding and removing RNAP that is stuck on damaged DNA with the help of ATP. Additionally, it stimulates DNA repair by recruiting Uvr(A)BC endonuclease. TRCF's C-terminal region is required for RNAP displacement.

A prokaryotic transcription factor known as NusB (IPR006027) is involved in antitermination. NusB is a prokaryotic transcription factor that is involved in the antitermination process. During this process, it interacts with the boxA portion of the mRNA nutation site.

NusG-like N-terminal (IPR006645) is the N-terminal domain of the transcription termination/antitermination protein NusG, which is involved in transcription elongation, termination, and antitermination. It is also present in the N terminus of the transcription anti-termination protein RfaH, as well as in the transcription elongation factor Spt5.

Spt4 (cd07973) is a transcription elongation factor.The three transcription elongation factors, Spt4, Spt5, and Spt6, conserved in eukaryotes, regulate the chromatin structure and are essential for transcription. We widely recognize the general transcription elongation factors Spt4, Spt5, and Spt6 for their positive and negative regulation of transcription in critical regulatory and developmental functions.

The N-terminal part of the ELL-associated factor (Eaf) protein is Transcription elongation factor Eaf N-terminal (IPR019194). It works as a transcriptional transactivator for ELL and ELL2 RNA polymerase II (Pol II) transcription elongation factors. To make it easier for RNA polymerase II to bind and start transcription elongation, Eaf proteins and ELL proteins form stable heterodimeric complexes.

The Tau protein has a transcription factor domain at its N-terminus, as well as NusB, NusG, and Spt4 domains at its N-terminus, C-terminus, and middle segment, as shown in Table 15. It suggests that the Tau protein binds to the transcription factor binding region, thereby interfering with transcription factors. At the same time, Tau protein binds to the promoter or enhancer and acts as a transcription factor to help telomere sequence fragments enter the TAA region in TATA-BOX and then transpose.

Table 15. Transcription factor-like domains of tau protein

Domain	Alias	Motif	Start	End
Transcription initiation factor IIA gamma subunit C-terminal	A	GGSVQIVYKPVDLSKVTSKCGSLGNI	620	645
	B	THVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	678	712
TFIIB-type	A	NIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGN	613	644
	B	LDNITHVPGGGNKKIETHKLTFRENAKAKTDHG	674	706
Transcription	A	YTM	29	31

initiation factor TFIID subunit 12	B	APLEFTFHVEITPNV	297	311
TAFII28-like protein	A	NLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV	582	617
	B	RHLSNV	723	728
TAFII55 protein conserved region	A	NVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHK	603	648
TFIIF beta subunit HTH	A	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	663	711
TFIIH p62 subunit N-terminal	A	NITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHL	676	725
TFIIS N-terminal	A	YTMHQDQ	29	35
TFIIS central	A	AEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	2	39
TFIIS-type	A	FTFHVEITPNVQ	301	312
	B	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK	591	628
	C	NKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTS	685	721
Transcription factor IIIC subunit 5 HTH	A	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVD	593	631
Transcription factor TFIIIC triple barrel	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQE	8	36
	B	NKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	596	647
	C	SLDNITHVPGGGNKKIETHKLTFREN	673	698
Transcription factor CBF	A	LEFTFHVEITPNVQKEQAHSEEHLGRAAFP	299	328
	B	CGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHK	608	648
	C	DNITHVPGGGNKKIETHKLTFREN	675	698
Transcription-repair-coupling factor C-terminal	A	FTFHVEITPNV	301	311
	B	SEKLDFKDRVQSKIGSLDNITH	658	679
NusB	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTD	5	40
	B	DAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	296	327
	C	IGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKI	577	688
NusG-like N-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLK	1	44
	B	EFTFHVEITPNVQKEQAHSEEH	300	321
	C	LKNRPCL	404	410

	D	KNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV PGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFK DRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY K	571	712
Spt4	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	1	36
	B	RAKGQDAPLEFTFHVEITPNVQKEQAHSEEH	291	321
	C	IINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	594	627
	D	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAE IVYKSPVVSGDTSPRHLSNVSSTGSIDMVDSPQLATLADEV	662	749
Transcription elongation factor Eaf N-terminal	A	YKPVDLSKVTSKCGSLGNIHH	627	647
	B	HKLTFRENA	691	699

(3) Specific transcription factor domains

BHLH (PTHR47075) is a transcription factor for the bHLH gene in plants. The transcription factors that comprise the bHLH protein family are involved in various aspects of plant development and stress response. Members of this family have a fundamental helix-loop-helix domain that is responsible for their ability to bind DNA. They are responsible for the regulation of gene expression in response to a variety of stimuli.

A transcription factor for bZIP in plants is known as BZIP (IPR043452). It is analogous to OsbZIP46, a group of plant bZIP transcription factors that regulate drought tolerance through abscisic acid (ABA) signaling in rice. Plant bZIPs are associated with developmental and physiological processes in response to stimuli/stresses such as temperature changes, hormones, and light.

Bacterial transcription activator effector binding (IPR010499) is found in the potential effector binding domains of numerous bacterial transcription activators, as well as in certain DNA gyrase inhibitors.

CarD-like (IPR052531) is a transcriptional regulator that is similar to CarD. Transcription factors in this protein family interact with RNA polymerase to regulate the transcription of ribosomal RNA. In certain organisms, members are involved in the regulation of replication and persistence during infection. They are also critical for maintaining rRNA levels under normal growth conditions and are involved in the response to a variety of stresses, such as malnutrition, oxidative, and genotoxic stress.

DM (IPR001275): A conserved family of transcription factors encoded by DMRT genes shares a unique DNA-binding motif, the DM domain. The twosex protein of Drosophila melanogaster initially identified this domain. Alternatively spliced mRNAs that encode sex-specific polypeptides are produced by the twosex gene in Drosophila melanogaster, which regulates somatic sex differentiation. It is hypothesized that these proteins function as transcription factors for downstream sex-determining genes, with a particular emphasis on the transcription of the vitellin gene and neuroblast differentiation.

GCM (PS50807) is a Glial Cell Deletion (GCM) protein that has been identified in the N-terminal portion of the Glial Cell Deletion (GCM) transcription factor. It is a member of a small family of metazoan transcriptional regulators that are involved in essential developmental processes. Committing cells to a glial cell fate is a critical transition during early neurogenesis.

NDT80 (IPR024061) is the DNA binding domain of NDT80. The NDT80 DNA binding domain is present in the following proteins, potentially playing a role in nutrient status sensing: This transcription factor is specific to meiosis in yeast. NDT80 is a critical transcription factor that enables meiosis to proceed after recombination is

effectively completed. The transcription factor vib-1 of Neurospora crassa is responsible for regulating heteronuclear incompatibility.

Three tryptophan residues interact with the DNA by recognizing the GAAA sequence in the IRF tryptophan pentad repeat (PS51507), a type I interferon gene that is induced by infection (IFN-α and IFN-β). One of the transcription factors that contributes to this activation is interferon regulatory factor I (IRF-1). The interferon consensus sequence (ICS) is a regulatory cis-element that is found upstream in the promoters of type I IFN and IFN-induced MHC class I genes. IRF-1 binds to this sequence. Interferon regulatory factor 2 (IRF-2) is a protein that also interacts with ICS.

The interferon regulatory factor (IRF) has a DNA-binding domain made up of five tryptophans. This is called the IRF tryptophan pentad repeat (PS00601). The interferon consensus sequence (ICS) is a regulatory cis-element that is found upstream in the promoters of type I IFN and IFN-induced MHC class I genes. IRF-1 binds to this sequence. Interferon regulatory factor 2 (IRF-2) is a protein that interacts with ICS but does not function as an activator. Instead, it inhibits IRF-1 in specific situations.

An interferon regulatory factor 2 binding protein 1&2, also known as interferon regulatory factor (IPR022750), functions as a zinc finger. There are two types of interferon regulatory factor 2 binding proteins: IRF-2BP1/2 and IRF-2BP-like, which is also known as IRF-2BPL or C14orf4. They are nuclear transcriptional repressors that bind to the C-terminal repression domain of IRF-2 and stop basic transcription and enhancer activation. Familial mutations in IRF2BP2 cause common variable immunodeficiency (CVID).

Global transcriptional regulator CodY N-terminal (IPR010312) represents the N-terminal domain of numerous bacterial global transcriptional regulator CodY proteins. CodY suppresses the dipeptide transport operon (dpp) of Bacillus subtilis in nutrient-rich environments.

IPR000232 is a heat shock factor (HSF)-type DNA-binding (HSF-type DNA-binding). Heat shock factors (HSFs) are transcriptional activators of heat shock genes that preferentially bind to heat shock promoter elements. These palindromic sequences are dense in repeated purine and pyrimidine motifs.

PNT (PS51433) is a PNT (or pointed tip) domain that is highly conserved and is present in Ets transcription factors, such as mammalian Ets-1, Ets-2, Erg, Fli-1, GABPalpha, and Tel, as well as Drosophila Pnt-P2 and Yan). Protein-protein binding is the function of the PNT domain. The ETS family is responsible for a variety of functions, such as the regulation of cell differentiation, cell cycle regulation, cell migration, cell proliferation, apoptosis, and angiogenesis. Studies have linked numerous ETS to cancer, including cases of gene fusion. For instance, Ewing's sarcoma is the result of the ERG (ETS Related Gene) transcription factor fusing with the EWS gene.

Table 16 illustrates that the Tau protein contains BHLH, Bacterial transcription activator effector binding, DM, and Interferon regulatory factor-domains throughout its entirety. Additionally, the N-terminus and C-terminus contain BZIP, CarD-like, NDT80, GCM, and HSF-type DNA-binding domains. This suggests that the Tau protein telomere reversal enzyme is located in the region where certain genes are bound by transcription factors. These genes are involved in development, the stress response, the response to stimulation or pressure, sex determination genes, meiosis, interferon, heat shock, controlling cell differentiation, controlling the cell cycle, cell migration, cell proliferation, apoptosis, and angiogenesis.

Table 16. Specific transcription factor domains of the tau protein

Domain	Alias	Motif	Start	End
BHLH	A	QEFEVMEDHAGTY	6	18
	B	YTMHQDQ	29	35
	C	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSK CGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHK	559	648
	D	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEI	660	711

		VY		
BZIP	A	MEDH	11	14
	B	IINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLG N	594	644
Bacterial	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKE	1	45
transcription	B	SVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	288	328
activator effector	C	LDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	599	647
binding	D	QSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	668	712
CarD-like	A	FTFHVEITPNVQKEQAHSEEH	301	321
	B	VQIINKKLDLSNVQSKCGSKDNIKHV	592	617
	C	IETHKLTFRENA	688	699
DM	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQEGD	8	38
	B	FTFHVEITPNVQKEQAHSEEHLGRAAF	301	327
	C	TLKNRPCLSPKH	403	414
	D	VQIINKKLDLSNVQSKCGSKDNIKH	592	616
	E	HKLTFRENAKAKTDHGAEIVY	691	711
GCM	A	EPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQD	3	34
NDT80	A	DAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	296	328
	B	IINKKLDLSNVQSKCGSKDNIKHVP	594	618
IRF tryptophan	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
pentad repeat	B	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK	593	628
	C	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK SPVV	663	716
Interferon	A	YTMHQDQ	29	35
regulatory factor	B	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV	581	617
2-binding protein 1	C	KKIETHKLTFRENAKAKTDHGAEIVYKSPVVS	686	717
& 2 zinc finger				
Interferon	A	QEFEVMEDHAGTYGLGDRKDQ	6	26
regulatory factor-3	B	YTMHQD	29	34
Global	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKES	8	46
transcriptional	B	DLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	600	647
regulator CodY				
N-terminal				
HSF-type	A	FEVMEDHAGTYGLGDRKDQGGYTMHQD	8	34
DNA-binding	B	TENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIK	580	615
	C	HKLTFRENAKAKTDHGAEIVYKSPV	691	715
PNT	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	1	35
	B	RAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	291	327
	C	IINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYK	594	628

(4) Unique transcription factor structure

BTB (IPR000210) is also referred to as the POZ domain (POxvirus and Zinc finger). There is a homodimerization domain at the N-terminus of the protein that is made up of many Kelch repeats or C2H2-type zinc fingers. Numerous BTB proteins are transcriptional regulators and are believed to regulate chromatin structure.

Transcription factor zinc-finger (IPR027392) is a zinc-finger domain of transcription factor.

Prokaryotic dksA C4-type zinc finger profile (PS51128) is a dksA C4-type. In bacterial cells that lack amino acids, guanosine tetraphosphate (ppGpp) or alarmone serves as a catalyst for a variety of regulatory processes. ppGpp specifically inhibits the activation of stable RNA transcription promoters. It regulates the activity of RNA polymerase by binding directly to the catalytic site, it regulates RNA polymerase activity. By binding to the enhancer ppGpp RNA polymerase, Dksa regulates signaling. Additionally, ppGpp and dksa regulate the promoters of enzymes involved in amino acid biosynthesis.

Bromo (IPR001487) is a bromodomain that is found in a diverse array of DNA binding proteins from invertebrates, mammals, and yeast.Bromodomains are α-helical motifs that are highly conserved and can interact specifically with acetylated lysine residues on histone tails.

Sporulation initiation factor Spo0A C-terminal is represented by IPR014879. A transcriptional activation domain and a phosphoacceptor domain comprise the response regulator Spo0A. This domain creates an α-helical structure consisting of six α-helices and corresponds to the transcriptional activation domain. This structure, consisting of a helix-turn-helix, binds DNA.

Regulator of transcription Myc N-terminal (IPR012682) A leucine zipper next to the bHLH domain is what makes Class III basic helix-turn-helix (bHLH) transcription factors unique. These factors have effects on both cell growth and cell death. Max and Myc constitute a heterodimer, which directly activates genes involved in cell replication, thereby regulating cell growth.

STAT transcription factor protein interaction (IPR013799) is a member of the STAT protein (signal transducer and activator of transcription) family. When cells encounter cytokines and growth factors, they specifically activate transcription factors in this family to regulate gene transcription. Consequently, they function as signal transducers in the cytoplasm and transcription activators in the nucleus. When these factors bind to cell surface receptors, they cause tyrosine to become autophosphorylated. The STAT SH2 domain can recognize this. This recognition facilitates the recruitment of STAT proteins from the cytosol, as well as their binding to activated receptors.

The ETS (erythroblast transformation specific) domain is denoted by ETS(IPR000418). This domain is rich in aromatic and positively charged residues and forms a binding affinity for purine-rich DNA segments. Various members of the Ets family of proteins exhibit different DNA binding specificities. The Ets domain and flanking amino acid sequences influence a protein's binding affinity, and modifications to a single amino acid in the Ets domain can modify its DNA binding specificity.

A "winged helix" is a conserved DNA binding domain, specifically the forkhead (IPR001766). The forkhead protein of Drosophila melanogaster is a transcription factor that facilitates terminal development, but not segmental development. It lacks the characteristic zinc finger of other transcription factors and does not possess a homeodomain. The forkhead domain attaches B-DNA as a monomer, but it does not share any similarities with previously identified DNA binding motifs.

PS51526, an RFX-type winged-helix, represents an RFX-type winged-helix DNA binding domain profile. The RFX family of transcription factors is distinguished by a distinctive residue DNA binding domain. Other regulatory factors, significantly dissimilar to each other, recruit the RFX-type HTH DNA binding domain to participate in a variety of systems.

Table 17 illustrates that the Tau protein primarily contains BTB and Transcription factor zinc-finger domains at the N- and C-termini, suggesting that it has dimerization and zinc finger structures. Table 17 shows that Tau protein mostly has DksA C4-type domains at the N-terminus and C-terminus. This suggests that it binds to DNA by binding to ppGpp, which is similar to how RNA polymerase does it. Table 17 demonstrates that the Tau protein primarily contains the Bromo and Sporulation initiation factor Spo0A C-terminal at the N-terminus and C-terminus. This suggests that the protein has an α-helix structure and binds to DNA through a helix-turn-helix structure. Table

17 illustrates that Tau protein primarily contains the Transcription regulator Myc N-terminal, STAT transcription factor protein interaction, ETS, Fork-head, RFX-type winged-helix domains at the N-terminus and C-terminus.

Table 17.Unique transcription factor structure domains of the tau protein

Domain	Alias	Motif	Start	End
BTB	A	VQIINKKLDLSNVQSKCGSKDNIK	592	615
	B	VTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHV	635	680
Transcription factor zinc-finger	A	FEVMEDHAGTY	8	18
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHLGR	295	324
	C	KCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFR	638	696
DksA C4-type	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQ	8	35
	B	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	663	711
Bromo	A	EPRQEFEVMEDHAGTYG	3	19
	B	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQI	581	625
Bromo	A	QSKIGSLDNITHVPGGGNKKIE	668	689
Bromo 1	A	EPRQEFEVMEDHAGTYG	3	19
	B	ENLKHQPGGGKVQIINKKLDLSNVQSKCG	581	609
Bromodomain associated	A	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	593	627
Sporulation initiation factor Spo0A C-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGD	1	38
	B	RAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	291	327
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDH	567	705
Transcription regulator Myc N-terminal	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQ	5	33
STAT transcription factor protein interaction	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	1	42
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPGA	295	330
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	567	647
ETS	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQD	5	34
	B	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHK	591	648
	C	DRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKS	665	713
Fork-head	A	EPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTD	3	40
	B	GQDAPLEFTFHVEITPNVQKEQAHSEEHLGR	294	324
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQI	567	625

	D	KCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIE THKLTFRENAKAKTDHGAEIVYK	638	712
RFX-type winged-helix	A	HVEITPNV	304	311
	B	STENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNI	579	614
	C	NITHVPGGGNKK	676	687
	D	THKLTFRENA	690	699

3.8 Promoter binding domains

CAATT-binding proteins, which are essential for the growth and biogenesis of the 60S ribosomal subunit, contain the CCAAT-binding factor (IPR005612) domain that binds promoters. Other proteins containing this domain stimulate transcription from the HSP70 promoter. Enhancer-binding proteins are also CAATT-binding proteins.

A bacterial protein called HTH araC (IPR018060) controls gene expression. It is of the AraC type and binds DNA through its "helix-turn-helix" (HTH) motif.

HTH LytTR-type (PS50930) is a LytTR-type HTH domain. It is a potential winged domain helix-turn-helix (wHTH) domain that is present in bacteria's algR/agrA/lytR family of transcriptional regulators. It is capable of binding to DNA.

Myb(IPR017930) is a myb-type HTH domain, which is a helix-turn-helix (HTH) domain that binds DNA and is usually found in pairs in eukaryotic transcription factors. It comprises approximately 55 amino acids.

The HTH ArsR type (IPR001845) contains a DNA-binding domain. The ArsR/SmtB family of transcription regulators, which includes DNA-binding winged helix-turn-helix (wHTH) domains, is responsible for the stress response to heavy metal ions.

In Bacillus subtilis, HTH hxlR-type (IPR002577) is a DNA-binding protein that acts as a positive regulator of the formaldehyde-inducible hxlAB operon.

The HTH lysR-type (IPR050950) is a transcription regulator for the HTH-type LysR. It regulates gene expression and DNA binding. This family of proteins controls many cellular processes, such as the breakdown of compounds like protocatechuate and sodium dodecyl sulfate (SDS). They are also involved in malolactic fermentation, the regulation of operons related to nitrogen metabolism, and responses to sulfur sources. They can often respond to environmental cues or the presence of specific inducers by acting as activators or repressors of transcription.

The HTH marR-type (IPR000835) is a DNA-binding winged helix-turn-helix (wHTH) domain that is associated with the development of antibiotic resistance. It is a MarR-type HTH domain that is present in the MarR/SlyA family's transcriptional regulators. The MarR family of transcriptional regulators is named after Escherichia coli MarR, a gene repressor that activates multiple antibiotic resistance and oxidative stress regulators.

HTH tetR-type (IPR001647) is a DNA binding domain with a helix-turn-helix (HTH) structure that is found in a variety of bacterial and archaeal transcriptional regulators, including the tetracycline resistance suppressor TetR.

The HTH rpiR-type (IPR047640) is a transcriptional regulator in the HTH-type RpiR-like protein family, which includes RpiR, MurR, HexR, and a few uncharacterized transcriptional regulators primarily from bacteria.

TAFII55 protein conserved region is PF04658(CD08050). TATA-binding protein (TBP) and a series of TBP-associated factors (TAFs) collectively contribute to the assembly of the pre-transcription initiation complex, forming the TFIID general transcription factor.

The general transcription factors TFIIE and TFIIH facilitate the initiation of eukaryotic mRNA transcription, which requires the melting of promoter DNA. Archaea contain a TFIIE homologue known as TFE (IPR016481), which is equivalent to the N-terminal half of TFIIEalpha. The minimal fundamental region of eukaryotic TFIIEα

appears to correspond to archaeal TFE.

TAF6 C-terminal HEAT repeat (IPR011442)one of several TAFs that binds TBP and participates in the formation of the transcription factor IID (TFIID) complex, is the carboxyl (C)-terminal domain of TATA-binding protein (TBP)-associated factor 6 (TAF6). TFIID significantly influences both the recognition of promoter DNA and the assembly of the preinitiation complex.

The SBP-type (PS51141) is a part of the SQUAMOSA promoter-binding proteins (SBPs) family, which is a big group of important transcription factors that are specific to development. They both possess a DNA-binding domain that is highly conserved and contains two zinc-binding sites. The consensus sequence -TNCGTACAA is the preferred binding site for the SBP zinc finger.

The uDENN domain (IPR005113) is present in numerous proteins that are involved in Rab-mediated processes or regulate MAPK (mitogen-activated protein kinase) signaling pathways. The human C-MYC promoter binding protein IRLB is a protein that contains the uDENN domain.

According to the results presented above, the TATA box region is the primary site for Tau protein binding to the promoter, and the consensus sequence is TATAATAAT. Table 18 illustrates the domain of Tau protein binding to the promoter, which includes the CCAAT-binding factor, HTH, Myb, SBP-type, uDENN, TAFII55 protein conserved region, TAF6 C-terminal HEAT repeat. The CCAAT-binding factor is typically located in the promoter region, upstream of the TATA box. A minuscule groove is the binding position of this TATA-binding protein structure. HTH is a protein structural motif that has the ability to bind to DNA through the TATA motif. It is important to mention that the "AAT" sequence is identical to the binding and cutting region of telomere transposase. This means that the Tau protein is likely what makes telomere transposase cut and add to telomeres in the "AAT" region of the promoter TATA box. The telomere transposase function of the Tau protein can move and add telomere fragment sequences to the "TAA" sequence of both DNA chains at the same time.

Table 18. Promoter binding domains of tau protein

Domain	Alias	Motif	Start	End
CCAAT-binding	A	GDRKDQGGYTMHQ	21	33
factor	B	FTFHVEITPNVQKEQAHSEEHLGRAAFP	301	328
	C	RPCLSPKHP	407	415
	D	KVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIV	591	626
HTH LytTR-type	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGL	5	43
HTH TFE	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
	B	KGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	293	327
	C	AKTLKNRPCL	401	410
	D	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKC GSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKS EKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEI VYKSPVVSGDTSP	559	722
HTH araC	A	DNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	612	647
	B	KLDFKDRVQSKIGSLDNITH	660	679
HTH arsR-type	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	5	35
	B	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTF	663	695
HTH hxlR-type	A	IINKKLDLSNVQSKCGSKDNI	594	614
	B	SKCGSLGNIHH	637	647
HTH lysR-type	A	QEFEVMEDHAGTY	6	18
	B	QDAPLEFTFHVEITPNVQKEQAH	295	317

	C	IINKKLDLSNVQSKC	594	608
HTH marR-type	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
	B	SLDNITHVPG	673	682
HTH myb-type	A	VMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	10	39
	B	NVQSKCGSKDNIKHVPGGGSVQIVYK	603	628
HTH rpiR-type	A	LEFTFHVEITPNVQKEQ	299	315
HTH tetR-type	A	RQEFEV	5	10
	B	YGLGDRKDQGGYTMHQD	18	34
	C	FTFHVEI	301	307
	D	DL	569	570
	E	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPV DLSKVTSKCGSLGNIHHK	581	648
	F	FKDRVQSKIGSLDNITHVPGGGN	663	685
Myb	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAG	1	42
	B	DAPLEFTFHVEITPNVQKEQAHSEEHLGR	296	324
	C	LKNRPC	404	409
	D	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKC GSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKS EKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEI VYKSPVVSGDTSPRHLSNVSSTGS	559	733
Myb-like	A	EFEVMEDHAGTYGLGDRKDQGGYTMHQD	7	34
	B	ENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVY	581	627
	C	DFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK SPVVSGDTS	662	721
TAF6 C-terminal HEAT repeat	A	QEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	6	35
	B	TENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV	580	617
	C	SVQIVYKPVDLSKVTSKCGSLGNIHHK	622	648
	D	VEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTD HGAEIVYKSPVV	654	716
TAFII55 protein conserved region	A	PRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQE	4	36
	B	EFTFHVEITPNVQKEQAHS	300	318
	C	TLKNRPC	403	409
	D	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKH VPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDR VQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVS GDTSPRHLSNVSSTGSIDMVDSPQLA	567	743
SBP-type	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPL QT	1	50
	B	RPPQTAAREATSIPGF	244	259
	C	PLEFTFHVEITPNVQKEQAHSEEHLGRAAF	298	327
	D	PQLKARMVSKSKDGTGSDDKKAKTSTRSSAKTLKNRPC	372	409

	E	PMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKH VPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDR VQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVS GDTSPRHLSNVSSTGSIDMVD	566	738
uDENN	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMH	5	32
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	295	327
	C	NRPCLS	406	411
	D	KVQIINKKLDLSNVQSKCGSKDNIKHV	591	617
	E	IVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLD NITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLS NVSSTG	625	732

3.9 Enhancer-binding domains

Enhancer of mRNA-decapping protein 4 WD40 repeat region(IPR032401) represents the N-terminal region of Ge-1, also known as EDC4 (enhancer of mRNA decapping protein 4). This WD40 repeat region facilitates protein-protein interactions.

Enhancer of polycomb-like N-terminal, also known as IPR019542, enhances the polycomb-like N-terminal protein. The N-terminus of the EPL1 (enhancer of polycomb-like) protein houses this domain. The EPL1 protein is a component of the histone acetyltransferase complex, which is responsible for the transcriptional activation of specific genes. It is also present at the N-terminus of the proteins in the Jade family.

Zinc fingers belong to the domain DZF (IPR006561). contains interleukin enhancer binding factor 3 (also known as nuclear factor of activated T cells 90 kDa, NF90, DRBP76 or NFAR). It interacts with highly structured RNA and dsRNA-activated protein kinase (PKR), as well as contains two double-stranded RNA binding motifs (dsRBM). Interleukin enhancer binding factor 2 (also known as NF45) appears to function predominantly as a heterodimeric complex with NF90. Spermatid perinuclear RNA binding protein (SPNR, also known as STRBP) is a testis-specific paralog of NF90.

Groucho (IPR009146) is a Groucho/transducin-like enhancer. The Groucho (Gro)/transducin-like enhancers (TLEs) are a group of corepressor proteins that have been around for a long time and have been passed down from generation to generation. They are important for controlling many developmental and cellular pathways, including lateral repression, segmentation, sex determination, dorsal/ventral patterning, terminal patterning, and eye development.

Sporulation stage II protein D amidase enhancer LytB N-terminal (IPR013693). The stage II sporulation protein SpoIID contains this domain. Membrane migration and certain initial stages of engulfment during bacterial endospore formation are dependent on SpoIID. This domain is also present in the amidase enhancer protein. Amidases, including SpoIID, are enzymes that hydrolyze cell walls.

The HMG box (IPR009071) is involved in DNA binding and may also be involved in protein-protein interactions. The HMG box (IPR009071) plays a crucial role in regulating DNA-dependent processes such as strand repair, replication, and transcription, all of which require the bending and unwinding of chromatin. Numerous are regulators of gene expression, as well as sequence-dependent and sequence-independent DNA recognition: SOX family of transcription factors; HMG1 and HMG2 non-histone components of chromatin. SSRP, a structural-specific recognition protein, is involved in transcription and replication. MTF1, a mitochondrial transcription factor, and UBF 1/2, a nucleolar transcription factor (upstream binding factor), are both involved in RNA polymerase I transcription. You can connect the HMG box to the sequence of the chondrocyte-specific enhancer of the human type I collagen gene. It can also connect to other SOX genes' binding sites. It also deforms DNA at an angle after binding. HMG-box transcription factors specifically inhibit the expression of target genes by

recognizing the T(G/C)AATG(A/G)A sequence. The enhancer region on the Nos2 gene's promoter is capable of binding to HMGXB4.

MADS-box (PS50066) is a MADS-box domain. This domain encompasses mammalian myocyte-specific enhancer factors 2A to 2D (MEF2A to MEF2D). Animals' muscle development and cell proliferation and differentiation are associated with MADS-box genes. The motif CC[A/T] 6 GG, also known as the CArG-box, is highly similar to DNA sequences that MADS domains bind to.

Homeobox (IPR001356) encodes transcriptional regulators that consist of homeodomains and execute distinct genetic programs along the animal body's anterior-posterior axis. This domain binds DNA through a helix-turn-helix (HTH) structure. Two alpha helices, in close proximity to DNA and connected by a brief turn, distinguish the HTH motif. The sequence and structure of this motif are strikingly similar in a diverse array of DNA-binding proteins, including cro and repressor proteins, homeobox proteins, and others.

The Tau protein is shown in Table 19 to possess the Enhancer of mRNA-decapping protein 4 WD40 repeat region, Enhancer of polycomb-like N-terminal, DZF, Groucho, Sporulation stage II protein D amidase enhancer LytB N-terminal, HMG box,MADS-box domains. And this means that it connects to areas that are important for mRNA decapping protein, interleukin, cell phagocytosis, chondrocytes, muscle development, cell proliferation, and differentiation. Table 19 shows that the telomere transposase function of Tau protein adds the transposed telomere fragment sequence to the "TAA" region of the enhancer. Furthermore, Tau protein contains the Homeobox, MADS-box, and HMG box. It is important to mention that the "TAA" sequence may be present in [A/T] 6 of the MADS-box, and the Homeobox and HMG-box regions contain the "TAA" pattern. Because of this, the Tau protein's telomere transposase function can move the telomere fragment sequence onto the "TAA" sequence of both DNA segments.

Table 19. Enhancer binding domains of tau protein

Domain	Alias	Motif	Start	End
Enhancer of mRNA-decapping protein 4 WD40 repeat region	A	EFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	7	35
	B	NRPCL	406	410
	C	NLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKP	582	629
	D	KLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGD	660	719
	E	NVSSTGSIDMVDSPQLATLA	727	746
Enhancer of polycomb C-terminal	A	AKTDH	701	705
Enhancer of polycomb-like N-terminal	A	MAEPRQEFEVMEDHAG	1	16
	B	DQGGYTMHQDQEGDTD	25	40
	C	FLR	136	138
	D	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHH	567	647
Sporulation stage II protein D amidase enhancer LytB N-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
	B	GQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	294	327
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVP	567	618
	D	YKPVDLSKVTSKCGSLGNIHHK	627	648

	E	LDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHL	674	725
DZF	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	1	39
	B	DAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	296	327
	C	NVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGG	572	621
	D	TSKCGSLGNIHHK	636	648
	E	FKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK	663	712
	F	RHLSNVSSTGSIDMVDSPQL	723	742
Groucho	A	RQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQ	5	35
	B	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFPG	289	329
	C	TENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNI	580	645
	D	QVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNV	653	728
HMG box	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37
	B	KVQIINKKLDLSNVQSKCGSK	591	611
Homeobox	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37
	B	TLKNRPC	403	409
	C	ENLKHQPGGGKVQIINKKLDLSNV	581	604
	D	DFKDRVQ	662	668
MADS-box	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQ	8	35
	B	LEFTFHVEITPNVQKEQAHSEEHLGRA	299	325
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVD	567	631
	D	VTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVSSTGS	635	733

3.10 Domains associated with mitochondria

Microenvironmental stresses, such as hypoxic or low glucose conditions, upregulate the hypoxia-inducible gene HIG1 (IPR007667). It is also known as the hypoglycemia/hypoxia-induced mitochondrial protein (HIMP1). HIG1 is a mitochondrial inner membrane protein that is ubiquitously expressed. This domain is present in proteins thought to play a role in the hypoxic response. It is also present in genetic modifications of mitochondrial proteins.

The calcium uniporter protein C-terminal (IPR006769) represents the C-terminal domain of MCU, a mitochondrial inner membrane calcium uniporter that controls calcium uptake into mitochondria.

The clustered mitochondria protein N-terminal (PF15044) represents the N-terminal domain of the clustered mitochondria protein, also known as the disembodied protein in Drosophila. It is a conserved gene in Drosophila that is essential for mitochondria's localization in the subcellular space. It also exhibits genetic interactions with parkin.

The mitochondrial outer membrane transport complex Sam37 (IPR019564) is a part of the SAM or TOB

complex, which is in charge of sorting the mitochondrial outer membrane. It is necessary for the sorting of proteins with a complex topology, such as β-barrel proteins, to the mitochondrial outer membrane following their import by the TOM complex. It shares homology with Metaxin-1, which is also a component of the mitochondrial outer membrane β-barrel protein transport complex.

The human disease protein Swiss:O60220 contains Tim10-like (PF02953) domains. Members of this family, including Tim9 and Tim10, facilitate mitochondrial protein import. It seems that the mitochondrial intermembrane space is the location of the members of this family.

Tim44-like (IPR039544) includes the mitochondrial import inner membrane translocase subunit Tim44 and its bacterial homologs. They all possess a domain that is similar to Tim44. Tim44 is an important part of the PAM complex, which moves proteins with transit peptides from the inner membrane to the mitochondrial matrix with the help of ATP. It recruits mitochondrial HSP70 and uses ATP as an energy source to facilitate protein translocation into the matrix.

B-type DNA polymerase proteins contain DNA-directed DNA polymerase family B mitochondria (IPR004868). The proteins mentioned in this entry are present in viruses, as well as in the mitochondria of plants and fungi.

It refer to a domain at the C-terminus of mitochondrial fission factor (Mff) and FATE1 as Mff-like (IPR039433). The presence of this domain is needed for Mff to interact with the dynamin-related GTPase Drp1, which helps mitochondria split. Conversely, FATE1 is responsible for regulating the ER-mitochondria distance and mitochondrial Ca2+ uptake.

Mitochondria-eating protein C-terminal (IPR031981) domain is located at the C-terminus of mitochondrial feeding proteins. Proteins that comprise this domain regulate mitochondrial mass. They are involved in the degradation of damaged mitochondrial proteins and mitochondria.

The entire sequence of tau protein contains mitochondria-related domains, as illustrated in Table 20. Tau protein is upregulated and ubiquitously expressed under microenvironmental stress, such as hypoxia or low glucose conditions, as indicated by the HIG1, Calcium uniporter protein C-terminal, and Mff-like domains. The mitochondrial calcium transport and calcium uptake pathways direct it to the mitochondria. The Tim10-like and Tim44-like domains, the mitochondrial outer membrane transport complex Sam37, and the clustered mitochondria protein N-terminal all point to the idea that tau protein gets into the mitochondrial interior through the inner membrane sorting apparatus and the outer membrane of the mitochondria. The DNA-directed DNA polymerase family B mitochondria domain evidences that the tau protein can insert telomeric fragment sequences into the mitochondrial DNA genome .

Destroying the cell's respiratory chain and damaging mitochondria will disrupt the entry of tau protein into the matrix and inner membrane space of mitochondria. The tau protein's Mff-like and Mitochondria-eating protein C-terminal domains can bind to the surface of damaged mitochondria, thereby activating mitochondrial autophagy signals.

Table 20. Mitochondrial-associated domains of the tau protein

Domain	Alias	Motif	Start	End
Clustered	A	GYTMHQD	28	34
mitochondria	B	LEFTFHVEI	299	307
protein	C	HKLTFREN	691	698
N-terminal				
Calcium uniporter	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDAGLKESPLQ	1	52
protein C-terminal		TPT		
	B	LSKVSTEIPASEPDGPSVGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGR	272	327

		AAF		
	C	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCG SKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEK LDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK SPVVSGDTSPRHLSNVSSTGSIDMVDSPQL	559	742
HIG1	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37
	B	QDAPLEFTFHVEITPNVQK	295	313
	C	VQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLG NIHHKPGGGQVEVKSEKLDFKDRVQSKIG	592	672
	D	LDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVY	674	711
Mff-like	A	RQEFEVMEDHAGTYGLGDRKD	5	25
	B	GYTMH	28	32
	C	LEFTFHVEITPNVQKEQAH	299	317
	D	VQIINKKLDLSNVQSKCGSKDNIKH	592	616
	E	CGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNITHVPGGGNKKIET HKLTFR	639	696
Mitochondria-eating protein C-terminal	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEG	1	37
	B	VQEGFLREPGPPGLSHQLM	132	150
	C	VGRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAFP	289	328
	D	LKNRPCL	404	410
	E	SKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIV YKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQSKIGSLDNIT HVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDTSPRHLSNVS STGSIDM	575	736
Mitochondrial outer membrane transport complex Sam37	A	FEVMEDHAGTYGLGDRKDQGGYTMHQDQE	8	36
	B	LEFTFHVEITPNVQKEQAH	299	31
Tim10-like	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDT	1	39
	B	QDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	295	327
	C	RLQTAPVPMPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCG SKDNIKHVPGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEK LDFKDRVQSKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYK SPVVSGDTSPRHLSNV	559	728
Tim44-like	A	QIINKKLDLSNVQSKCGSKDNIKHVPGGGSVQIVYKPVDLSKV	593	635
DNA-directed DNA polymerase family B mitochondria	A	MAEPRQEFEVMEDHAGTYGLGDRKDQGGYTMHQDQEGDTDA	1	41
	B	GRAKGQDAPLEFTFHVEITPNVQKEQAHSEEHLGRAAF	290	327
	C	MPDLKNVKSKIGSTENLKHQPGGGKVQIINKKLDLSNVQSKCGSKDNIKHV PGGGSVQIVYKPVDLSKVTSKCGSLGNIHHKPGGGQVEVKSEKLDFKDRVQ SKIGSLDNITHVPGGGNKKIETHKLTFRENAKAKTDHGAEIVYKSPVVSGDT SPRHLSNV	567	728

3.11 Regulatory mechanism of tau protein's aging device

(1) The pyruvate kinase domain, which supplies ATP energy, facilitates the aging device function.

Tau protein contains two Mg2+-ATP binding sites: one at aspartate (D13) and the other at glycine (G682). The central portion of the catalytic domain is D13, while the catalytic domain concludes at G682.The tau protein is a dimer, and each subunit possesses pyruvate kinase and phosphatase activity. The aspartate (D13) region is the location of the pyruvate kinase domain of the Tau protein.

Tau protein possesses pyruvate kinase (PK) activity, which is capable of phosphorylating ADP into ATP. ATP has the ability to inhibit PK kinase activity. When the Tau protein's D13 binds to ATP, it inhibits pyruvate kinase activity, and when it binds to ADP, it activates.Phospases can convert ATP to ADP by self-dephosphorylating tau protein. Tau protein catalyzes the dephosphorylation process, converting ATP at the D13 binding site to ADP and activating pyruvate kinase activity. Tau protein's pyruvate kinase activity (PK) speeds up the last step of glycolysis by changing phosphoenolpyruvate (PEP) into pyruvate and phosphorylating ADP at the D13 binding site to ATP at the same time. This subsequently inhibits the pyruvate kinase activity.

The Tau protein's pyruvate kinase domain supplies ATP energy for the Aging Device function, as transposase and nuclease activities require ATP energy (Figure 1).

Pyruvate kinase generates ATP in the cytoplasm during glycolysis. Consequently, phosphoenolpyruvate (PEP) is typically present in the cytoplasm. The concentration of PEP in the nucleus is extremely low.

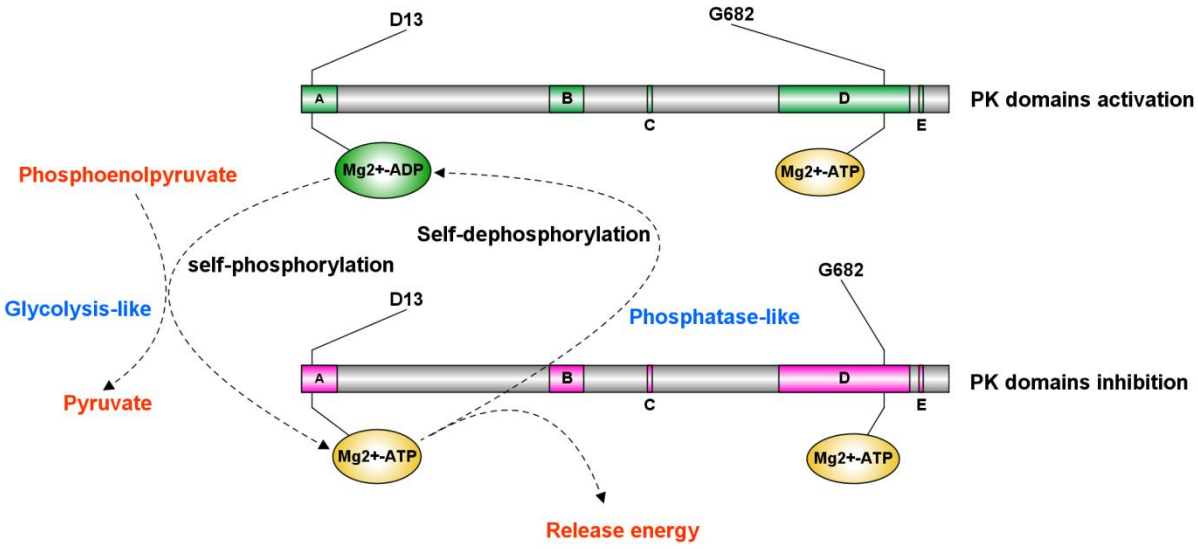

Figure 1. Tau protein provides ATP energy for its telomere transposition through the pyruvate kinase (PK) domain. The protein sequence rectangle displays the A-E rectangles, which represent the five PK domains. This diagram was drawn by IBS 1.0[47].

(2) The reverse transcriptase transposase domain synthesizes the guide RAN sequence and cuts off the telomere sequence between "TAA."

As Tau protein contains HNH cas9 activity , the cut telomere is analogous to Crispr-array in structure, : a leader, repeat, and spacer. They include the positioning region, which is the leader, as well as "TAA" motifs at both ends of the repeat. The term "spacer" represents the guide sequence.

Tau protein binds to the CCCTAA/TTAGGG repeat sequence (leader) of the telomere via the telomerase domain; however, it lacks a functional domain that is responsible for synthesizing telomere fragments.

The tau protein's reverse transcriptase transposase and HNH cas9 function cuts off the telomere repeat sequence (repeat) in the region between two "TAA", thereby removing double-stranded fragments from the telomere. The "TAA" fragment is the cut end of the one telomere end that is cut for transposition, while the "TTA"

fragment is cut end of the other end.

The tau protein repairs the truncated telomeric sequences through homologous recombination repair of DNA double-strand breaks and the recovery of stalled replication forks .

Tau protein's reverse transcriptase domain also makes the guide RAN sequence by transcripting the DNA sequence (spacer) in the telomere (Figure 2).

One subunit of the Tau protein dimer attaches to the guide RAN sequence, while the other subunit binds to the telomere segment that requires transposition (Figure 2).

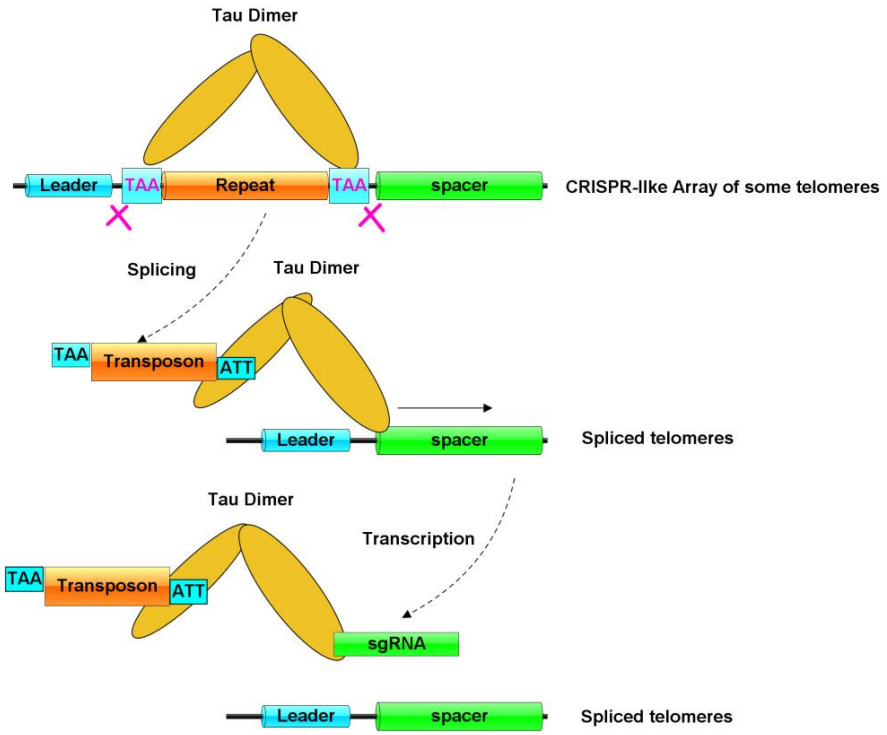

Figure 2.The Tau protein synthesizes the guide RAN sequence and cuts telomere fragments between "TAA" in the telomere. IBS 1.0 was responsible for creating this schematic diagram[47].

(4) The transcription factor domain glides to a specific DNA transcription factor binding region.

By identifying the replicon, the tau protein locates the DNA replication origin and subsequently unwinds the double helix DNA using the helicase. Through DNA polymerase activity, tau protein glides onto the unwinding DNA sequence. The transcription factor domain binds tau protein to specific DNA transcription factor binding regions, including development, stress response, response to stimulus/stress, sex determination genes, meiosis, interferon, heat shock, cell differentiation, cell cycle, cell migration, cell proliferation, cell apoptosis, and angiogenesis.

The guide RAN sequence of one subunit of Tau protein is complementary to the intron region near the transcription factor, while the other subunit recognizes the "TAA" position in the TATA-BOX in the promoter or the Homeobox, MADS-box, and HMG box in the enhancer. Consequently, Tau protein docks in this specific DNA transcription start region (Figure 3.A).

It is a double-insurance positioning method for tau protein.

(5) The HNH Cas9 domain removes the intron DNA region, which is complementary to the guide RAN sequence.

It is the HNH Cas9 domain and the nuclease domain of the Tau protein that cut out the intron that is complementary to the guide RAN sequence. This happens if the guide RAN sequence carried by one subunit of the Tau protein is complementary to the intron region close to the transcription factor. The Tau protein then releases the bound guide RAN sequence.

Finally, the domain activities of the DNA repair enzyme, ligase, and nuclease connect and repair the DNA cut (Figure 3.B).

The procedure guarantees that there will be no additional complementary introns to facilitate the Tau protein's proximity to this transcription factor.

(6) The nuclease and transposase domains insert the telomere fragment sequence into the promoter and enhancer,after splicing the "TAA" position.

Another subunit of the Tau protein recognizes the "TAA" position in the TATA-BOX in the promoter or the Homeobox, MADS-box, and HMG box in the enhancer. The nuclease and transposase domains then cleave the "TAA" position. The first chain cut is a "TAA" fragment, while the second chain cut is a "TTA" fragment. The cuts at both extremities of the transposed telomere are complementary to the "TAA" and "TTA" fragments respectively, followed by the subsequent insertion of the transposed telomere fragment sequence (Figure 3.C).

The Tau protein DNA repair enzyme ultimately repairs the DNA sequence in the transposed region, releasing the Tau protein from the DNA sequence.

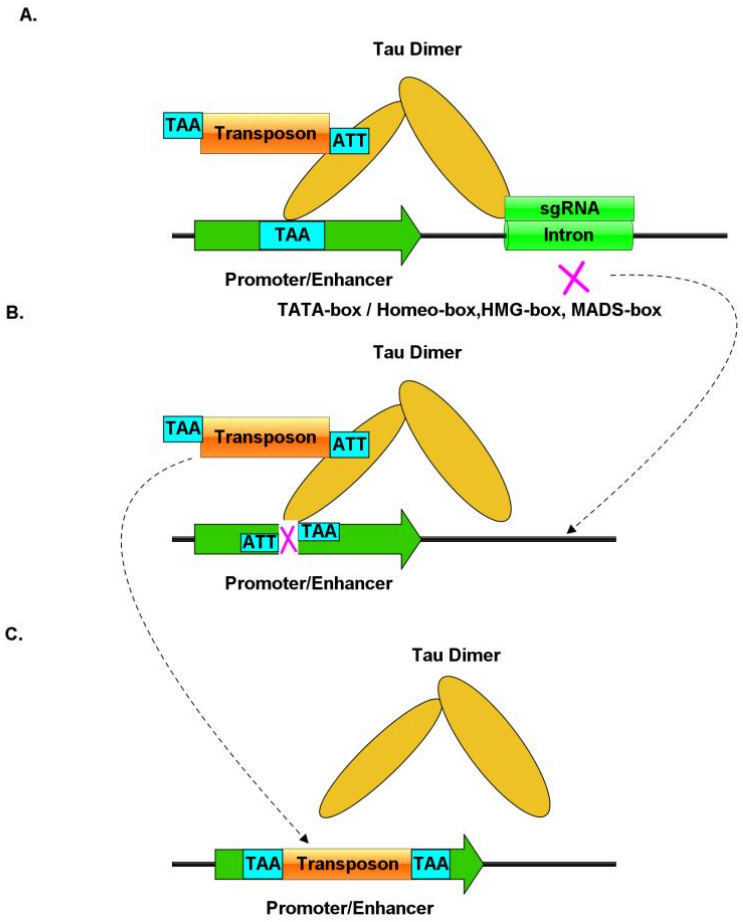

Figure 3. The tau protein locates a specific transcription start region and executes telomere fragment transposition. A. Slide and Position. B. Cutting intron. C. Transposing. IBS 1.0 was responsible for creating this diagram[47].

(7) Following a transposition mutation, promoters and enhancers impede the recruitment and binding of RNA polymerase II.

The promoter is a DNA sequence that is located upstream of the gene coding region and has the ability to specifically recognize and bind to RNA polymerase. Its primary function is to bind to RNA polymerase and initiate RNA transcription and synthesis. The enhancer, a DNA sequence located significantly away from the transcription start point, triggers tissue-specific expression and amplifies the promoter's transcriptional activity. The promoter and the enhancer are highly similar. Numerous components constitute both the enhancer and the promoter, and certain components may be present in both. The enhancer is a critical component of gene expression regulation; it determines the timing, location, and intensity of gene expression.

The genes that are deficient in TATA-BOX are primarily housekeeping proteins and a few developmental regulatory genes. The former regulates basic biochemical pathways, including genes necessary for nucleotide synthesis. The latter include active immune system development genes and homeotic developmental genes. TATA-BOX is typically present in specialized genes (luxury genes), such as hemoglobin and keratin. TBP, a member of the TFIID transcription factor, is the binding protein of TATA-BOX. When transcription factor II and RNA polymerase II bind to the promoter, they form the superinitiation complex, with TATA-BOX serving as a critical component. The Homeobox, MADS-box, and HMG box in the enhancer, along with the TATA-BOX in the promoter, significantly influence the efficient transcription efficiency of RNA polymerase II. The primary function of RNA polymerase II is to transcribe hnRNA (precursor of mRNA), snRNA, and miRNA.

The presence of the tau protein at the "TAA" position in the promoter and enhancer, which incorporates the transposed telomeric fragment sequence, will alter the region's structure. This will interfere with RNA polymerase II's smooth binding to the promoter region and transcription initiation. This transposition mutation of the promoter and enhancer may not alter the exon sequence of the transcribed gene; however, it will impact the gene's transcription and result in low transcription efficiency. The tau protein telomeric transposition method can gradually inactivate multiple identical gene copies, rendering them incapable of transcription. Simultaneously, telomeric transposition shortens the telomeres of DNA, which prompts cells to progressively enter telomere shortening-induced cell senescence.

Tau protein identifies introns using guide RNA, whereas transcription factor activity identifies TATA-BOX in promoters and Homeobox, MADS-box, and HMG box in enhancers. This is a double-insurance positioning method for tau protein. The double insurance positioning method of tau protein says that it can't be in the Poly C region or the stop codon and move the telomere fragment, even though "TAA" is in both places (possibly as a stop codon).

4. Discussion

4.1 Aging device uses the telomere-guided copy number decrement regulatory mechanism

A fundamental task in the development of organisms is the regulation of the differential expression of thousands of genes. A network is established by interconnected cis-regulatory modules that regulate gene expression[48]. Time series data of gene expression following cell perturbations can be used to reconstruct gene regulatory networks[49]. Gene regulatory networks consist of an overrepresentation of subgraphs, commonly known as network motifs. The ubiquitous parameter of gene copy number can substantially influence the dynamic function of network motifs. The nonlinear nature of transcriptional feedback, which involves the interaction of repeated motifs through a common pool of transcription factors, is the cause of all qualitative variations in expression[50]. The copy number threshold that results in a qualitative change in the system's behavior can be implicitly determined[50].

Decremental copy regulation is a regulatory mechanism in which a molecule or gene's copy number decreases

over time or under specific conditions. Copy number variation (CNV) is a genetic variation source that may contribute to phenotypic diversity and evolution[51]. Genes in CNV regions are expressed at lower and more variable levels than genes mapped elsewhere, and CNVs not only influence the expression of copy number-varied genes but also have global effects on the transcriptome[51]. Researchers have identified changes in DNA copy number, whether they affect specific genes or entire chromosomes, as the source of adaptive potential and the cause of disease and developmental abnormalities[52].

Indeed, there is a potential correlation between CNVs, neurological diseases, and aging. In centenarians and middle-aged individuals, the number of CNVs increases significantly, and genomic instability increases with age. CNVs may impact telomeres, potentially impacting human lifespan. Complex longevity traits may result from genome structural alterations, such as CNVs, that alter gene expression[53]. The primary source of variation among human individuals is the deletions and duplications of chromosome segments (copy number variation, CNV). Furthermore, CNV has the potential to influence human evolution and the development of numerous diseases, including mental illness, developmental disorders, and cancer[54]. Recurrent copy number variation (CNV) happens when homologous recombination (HR) happens between repeated sequences. Non-recurrent CNV, on the other hand, happens when non-homologous recombination mechanisms happen all over the genome[54].

The decrease in the number of mtDNA copies can be considered a manifestation of decremental copy regulation. Researchers link changes in mtDNA-CN levels in patients with neurodegenerative diseases to mitochondrial dysfunction, cognitive decline, disease progression, and ultimately, therapeutic intervention. Changes in mitochondrial function and content are one of the potential determinants of neurodegenerative diseases. Changes in mitochondrial DNA copy number (mtDNA-CN) levels in biological fluids occur during both the early stages of the disease and its progression[55]. Polymorphic mtDNA variants, such as point mutations, deletions, and copy number variations, raise the risk of cancer, neurodegenerative diseases (like Parkinson's disease and Alzheimer's disease), and rare neurological diseases that can happen to anyone or are passed down through families[56].

One of the causes of mammalian aging is the loss of epigenetic information[57]. The expression levels of mRNA and protein endure significant changes as we age[58]. Genes that exhibit age-dependent increases in mRNA and protein levels, along with increased cortical thickness during aging, link to sensory perception and immune function[58]. Multiple changes in the copy number of many genes at the same time can lead to bad traits, such as defects that are caused by adding or removing certain genes[52]. Additionally, large-scale changes in DNA copy number can lead to deleterious phenotypes.

The present investigation determined that tau protein has the capacity to transpose telomere fragments to promoters and enhancers, resulting in mutations. These mutations render transcription factors and RNA polymerase II incapable of forming a stable bond, thereby diminishing transcription efficiency. Tau protein, as the aging device, which speeds up aging, uses a telomere-guided copy number reduction regulation mechanism to lower the number of copies of useful genes. Instead of using a cascade regulation method, this process might utilize an indirect gene expression regulation method, which relies on the interaction between specific DNA sequences and proteins. The mechanism relies on telomere RNA sequence-guided telomere transposition to precisely regulate gene expression, reduce the copy number of effective genes, and induce promoter and enhancer mutations. This mechanism is highly adaptable and progressively reduces the copy number of unnecessary or harmful genes, thereby promoting the adaptive evolution of organisms.

4.2 Transposition of telomeric fragments to promoter/enhancer regions is a double insurance for the aging device

Negative or positive control modes are the most fundamental approaches to gene regulation research. Induction of gene expression can be accomplished by either removing restriction elements (which allow expression from a high-level promoter) or providing stimulatory elements (which promote expression from a low-level

promoter)[59]. Transcription factors and microRNAs, a large class of small noncoding RNAs, are the two fundamental mechanisms of gene regulation[60]. A growing number of small RNAs (sRNAs) have been demonstrated to regulate critical pathways in both prokaryotes and eukaryotes[61]. New microRNA families are playing significant roles in the development of novel tissue types and organs, and the regulators themselves appear to be well conserved over extended evolutionary distances for both transcription factors and microRNAs[60].

Variations in regulatory activity are frequently directly linked to genetic variation, as modifications in transcription factor binding result in modifications to the regulatory landscape[62]. RNA polymerase II can transcribe any protein-coding gene. The information that determines whether a specific gene should be transcribed is partially present in cis on the DNA in the form of specific DNA sequences (motifs) and partly in trans on the DNA in the form of proteins called transcription factors (TFs) that recognize and bind to these specific motifs[63]. Changes in TF binding, which can lead to coordinated changes in a diverse array of supporting regulatory mechanisms, can disproportionately influence changes in steady-state gene expression levels[62]. Numerous transcription factors, each with multiple binding sites, equip enhancers with binding sites. In certain instances, the physical proximity between the enhancer and promoter directly influences gene expression, as these interactions only take place during gene expression[64].

On the other hand, Leonard Hayflick first described the Hayflick limit, a phenomenon where normal human cells stop dividing after 40 to 60 passages[65]. The "replication end problem" causes telomeres to shorten during cell division. As a result, the majority of eukaryotes rely on telomerase to maintain chromosomes in dividing cells, as excessively short telomeres impede proliferation and induce senescence[66]. Telomerase preferentially lengthens the shortest telomeres, but it does not act on every telomere in every cell cycle[67]. The number of repeats added by telomerase is a controlled phenomenon that maintains a specific equilibrium. Any disruption in this equilibrium results in a variety of telomere-related diseases[68].

DNA damage response and repair enzyme complexes typically do not recognize telomeres in their "capped" state[69]. This is due to their ability to "shelter," or safeguard the chromosome extremities[70]. Telomere shortening, which destabilizes telomere loops, elevates the probability of telomere decapping[71]. Studies have shown that telomere decapping causes the same DNA damage response as DSBs[72], whether it's done by blocking TRF2 or shortening the telomeres.

Telomere shortening, a primary factor in the self-amplifying cycle between mitochondrial and telomeric DNA damage during cellular senescence, primarily causes telomere decapping[73]. During cell division, the "end replication problem" shortens telomeres, causing the accumulation of single-strand breaks due to damage from ROS generated as a byproduct of mitochondrial respiration[73]. It has also observed telomere dysfunction-induced foci (TIFs) in aged cells with elevated mitochondrial superoxide production, indicating telomere damage[74].

The current research discovered that the aging machinery function of tau protein ingeniously connects promoters/enhancers to telomeres through transposition. The tau protein synthesises guide RNA, intercepting the transposition fragment from the telomere, thereby shortening the telomere and inducing cell senescence. At the same time, tau protein adds the transposition telomere fragment to the promoter/enhancer region, which slows down gene transcription and messes up RNA polymerase II. This lowers the number of gene copies and starts the aging process in the brain. Consequently, the transposition of telomere fragments to the promoter/enhancer region double-protects tau protein's aging machine function.

4.3 Stress injuries such as hypoglycemia/hypoxia can lead to tau protein aggregation and aging device inactivation

Following brain injury, inflammatory immune responses may manifest as ischemia, status epilepticus, hemorrhage, hypoglycemia, or trauma[75]. Hypoglycemia and respiratory depression collectively exacerbate three injury patterns: selective neuronal necrosis, microglial activation, and laminar cortical necrosis[76]. It is prevalent in

the brains of Alzheimer's disease patients and is linked to atherosclerosis, hypertension, diabetes, and ischemia in middle-aged and geriatric individuals[77].Genetic mutations, susceptibility polymorphisms, environmental neurotoxins/metals, cardiovascular disease, traumatic brain injury, and ischemia/hypoxia may also influence the development of AD. Patients who are subjected to cerebral hypoxia or ischemia are at an increased risk of developing dementia[78]. In various neurodegenerative disease models, including Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), and Huntington's disease (HD), mitochondrial dysfunction, reduced glucose uptake, and impaired glucose metabolism are observed[79].

The preservation of glucose homeostasis is critical for preventing Alzheimer's disease. Hypoglycemia can lead to brain dysfunction, brain damage, and death. A metabolic shift is one of the critical points in midlife aging that increases the risk of late-onset AD. The consumption of brain glucose is reduced in patients with Alzheimer's disease, which suggests an energy deficit[80]. Synaptic inactivity can result from decreased cerebrospinal fluid (CSF) and brain glucose levels, which elevates the likelihood of cognitive impairment and Alzheimer's disease[81]. When the glucose-based antioxidant system stops working as it should in hippocampal slices, it causes big, sudden, and bad changes in how cells work. Metabolic collapse (MC) is the term used to describe this phenomenon[82]. In addition to their glucose-lowering effects, certain antidiabetic pharmaceuticals, including metformin, peroxisome proliferator-activated receptor γ agonists, and glucagon-like peptide-1 receptor agonists, have demonstrated the potential to provide neuroprotective benefits and alleviate ND symptoms[83].

Hypoxia is associated with an elevated incidence of Alzheimer's disease (AD). Aβ production and tau phosphorylation are influenced by hypoxia at numerous locations[84]. In addition, hypoxia elevates the concentrations of inflammatory molecules (IL-β1, IL-6, and TNF-α), which have also been linked to neurodegeneration. Chronic hypoxia induces neuroinflammation by activating microglia, the resident immune cells in the brain, and increasing reactive oxygen species and proinflammatory cytokines. These features are common to many degenerative central nervous system (CNS) diseases[85]. The activation of several kinases associated with tau hyperphosphorylation in neuronal cells, including ERK, GSK3β, and CDK5, is enhanced by hypobaric hypoxic exposure[86]. Tau hyperphosphorylation in the hippocampus and diminished memory are observed in rats that have been subjected to hypoxic stress[87].

Alzheimer's disease (AD) is a neurodegenerative condition that is distinguished by the presence of extracellular amyloid β-fibrillar plaques and intraneuronal tau protein neurofibrillary tangles(NFTs) in the brain. Oxidative stress, one of the earliest events in AD, can elicit tau hyperphosphorylation, which destabilizes microtubules by reducing tau binding affinity. This process leads to the formation of NFTs, which are the primary pathological feature of AD. In Alzheimer's disease patients, tau accumulation is pathologically more closely linked to the development of cognitive decline and neurodegeneration than Aβ plaques[88]. Tau glycosylation elevates the amyloid protein precursor and releases approximately 4 kD of amyloid peptide[89]. The dysfunction of the autophagy-lysosomal system may also facilitate tau aggregation[88]. In mature AD brains, mitochondria that are structurally defective are present in neurons, capillaries, and astrocytes. In addition to being linked to aging through mitochondrial DNA (mtDNA) mutations and reactive oxygen accumulation[90]. Tau is hyperphosphorylated and dissociated from MTs in AD, which leads to MT instability and impaired axonal transport[90].

The present investigation revealed that tau possesses pyruvate kinase activity. This kinase activity generates ATP to facilitate the transposition of telomere fragments to promoters/enhancers. The phosphorylation and dephosphorylation of tau D13 and G682 significantly influence the pyruvate kinase catalytic activity. The phosphorylation of glucagon stops pyruvate kinase activity. If pyruvate kinase activity is lost, blood glucose levels may temporarily rise. To expedite the conversion of glucose into energy, it should inhibit pyruvate kinase when glucose concentration decreases and activate it when glucose concentration increases. Thus, hypoglycemia may result in the inactivation of the tau protein's aging machinery, which inhibits pyruvate kinase activity. This could

potentially lead to issues with tau protein aggregation, as supported by the previously mentioned literature.

The present study suggests that microenvironmental stress, such as hypoxia or low glucose conditions, upregulates and widely expresses tau protein, which then travels to mitochondria via mitochondrial calcium transport and calcium uptake pathways. The tau protein contains mitochondrial-related domains HIG1, Calcium uniporter protein C-terminal, and Mff-like throughout the entire sequence.. This suggests that the tau protein will specifically target mitochondria when exposed to hypoxia or low glucose levels. Tau aggregation in the cytoplasm will ensue, potentially leading to neuroinflammation due to oxidative stress damage from mitochondrial damage.

It is important to note that phosphoenolpyruvate (PEP) is typically present in the cytoplasm, as glycolysis takes place in the cell's cytoplasmic matrix. Pyruvate kinase is responsible for the primary production of ATP in the cytoplasmic matrix. Due to the low amount of PEP in the nucleus, tau protein can't carry out the Aging Device function as a high-frequency physiological event, like glycolysis. Tau protein translocates to the cytoplasm, where it obstructs the glycolysis pathway and generates ATP through pyruvate kinase activity. It is highly likely that the spatial capacity constraints of nuclear translocation and mitochondrial transport will cause the tau protein to become obstructed and congested in the cytoplasm. Indeed, aberrant phosphorylation can disrupt the pyruvate kinase activity of tau protein, leading to its inability to synthesize ATP and its accumulation in the cytoplasm.

4.4 Neurodegenerative disorders like amyotrophic lateral sclerosis may be associated with aging device dysfunction

A neuronal intermediate filament called peripheralin has been linked to inclusion bodies in motor neurons from people who have amyotrophic lateral sclerosis (ALS). A nucleotide insertion in intron 8 (PRPHIVS8 –36insA) and a 1 bp deletion in exon 1 have been identified as two additional peripherin gene PRPH variants in ALS cases. The assembly of the neurofilament network is disrupted as a consequence of the expression of this frameshift peripherin mutant in SW13 cells[91]. ALS is frequently associated with reactive astrocytes, and human induced pluripotent stem cell (hiPSC)-derived astrocytes that contain ALS pathogenic mutations in VCP, SOD1, and C9orf72 frequently exhibit reduced intron retention (IR)[92].

The pathological hallmarks of ALS include the retention, reduced expression, and aggregate formation of proline- and glutamine-rich introns of splicing factors[93]. The splicing program during early neural differentiation is characterized by increased intron retention (IR). it has also used SOD1 and FUS mutant MNs to identify aberrant IR events in independent RNAseq datasets. Mutations in the RNA-binding protein FUS cause amyotrophic lateral sclerosis (ALS), a neurodegenerative disease that is devastating. FUS autoregulatory and splicing factor networks are impaired by FUS ALS pathogenic mutations due to intron retention[94]. Mutant FUS impedes the proper splicing of minor introns in mRNAs that encode proteins that are essential for the survival of motor neurons[95]. As a transcriptional regulator, the SVA retrotransposon in the FUS promoter has been associated with ALS[96].

The IR was the most prominent in the SFPQ transcript. The SFPQ protein exhibits a reduced nuclear abundance in VCP mutant cultures and binds extensively to its retained introns[97] New changes inside TANK-binding kinase 1 (TBK1) cause abnormal splicing patterns in people with amyotrophic lateral sclerosis[98]. The most prevalent genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) is a GGGGCC (G 4 C 2) repeat expansion in the first intron of C9ORF72. The production of the dipeptide repeat protein is abolished and neurodegeneration is rescued by the CRISPR-mediated deletion of the C9ORF72 promoter in motor neurons of ALS/FTD patients[99].

Researchers identified three classes of mutations: a homozygous exon 5 deletion, a homozygous Q398X nonsense mutation, and a heterozygous E478G missense mutation within the ubiquitin binding domain[100]. it has identified optineurin as the gene responsible for amyotrophic lateral sclerosis. In patients with familial and sporadic amyotrophic lateral sclerosis (ALS), mutations in the gene encoding optineurin (OPTN) are present. Additionally, a

novel deletion mutation of OPTN has been identified in Japanese ALS[101]. In a Filipino ALS family, the mutant SOD1 allele can be naturally suppressed by enhanced splicing of exon 2, and novel SOD1 deletions result in low penetrance in ALS families[102]. Familial ALS is the result of mutations in SOD1 that generate pseudoexons[103].

The present investigation found that tau protein synthesis directs the RAN sequence, following the fragment downstream of the transposed telomeric fragment. The dimeric tau protein binds to the transcription factor structural region. The guide RAN sequence places one subunit in the complementary intron region, and the "TAA" position of the TATA-BOX or enhancer (Homeobox, MADS-box, HMG box) houses the other subunit. Mutations in the intron region of genes that act on the Aging Device, such as point mutations, insertions, or deletions, clearly compromise the localization of the dimeric tau protein. In other words, the Aging Device function of the tau protein is unable to perform telomere transposition at the gene, resulting in a lack of reduction in the number of gene copies. If a intron mutation (e.g., point mutation, insertion or deletion mutation) occurs, the dimeric tau protein will abnormally localize in the intron region of genes that do not act on the Aging Device. In other words, the aberrant inactivation of certain genes will result from telomere transposition in genes that do not interact with the Aging Device.

5. Conclusion

The precise mechanisms underlying the profound influence of aberrant tau protein phosphorylation and aggregation on Alzheimer's disease remain elusive. This study used bioinformatics methods, specifically protein domain search tools, to look into how the tau protein causes neurodegenerative diseases. The study demonstrates that the tau protein has domains associated with the Aging Device. The domain list includes pyruvate kinase, phosphatase, telomere binding, telomere transposition, HNH cas9, replicon binding, helicase, DNA polymerase, nuclease, transcription factors-like, promoter binding (TATA-box), enhancer binding (Homeobox, MADS-box, HMG box), and mitochondrial localization. It suggests that the pyruvate kinase domain of the tau protein provides ATP energy for its Aging Device activity. After the HNH Cas9, reverse transcriptase, and transposase domains cut the DNA sequence between the "TAA" region of the telomere, then transcript the guide RNA sequence. The transcription factor domain moves to the exact location on the DNA where it binds. The HNH Cas9 domain removes the intron DNA region that is complementary to the guide RNA sequence. The nuclease and transposase domains splice the "TAA" region from the promoter or enhancer and add the telomere fragment sequence. We thought that tau protein could interface with the function of RNA polymerase II by moving telomere fragments to the promoter/enhancer area. This would make the Aging Device's telomere-guided gene (DNA in the nucleus or mitochondrial) copy number decrement regulatory system work better. Stress injuries, such as low blood sugar or oxygen levels, can lead to tau protein aggregation and phosphorylation, resulting in a perplexing deactivation of the Aging Device. Therefore, neurodegenerative disorders like Alzheimer disease, amyotrophic lateral sclerosis may potentially link to the atypical functioning of aging devices like tau protein.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The datasets and results supporting the conclusions of this article are available at: https://pan.baidu.com/s/1-m_QDms2rU71xA8AdjnZQA ;code: lyud .

Or at : https://mega.nz/folder/Fr5xEQwB#I3JD5NJDgqdeBpW0QCtQFg

Competing interests

The authors declare that they have no competing interests.

Funding

This work was partly funded by a grant from the Talent Introduction Project of Sichuan University of Science and Engineering (award number: 2018RCL20, grant recipient: LWZ).

Author's contribution

Funding was obtained by WZL. Besides, design, analysis and writing are finished by WZL, while data curation and manuscript check are undertaken by HLL. Both authors have read and agreed to the published version of the manuscript.

Acknowledgements

Not applicable.

Author details

[1] School of Computer Science and Engineering, Sichuan University of Science & Engineering, Zigong, 643002, China.
[2] School of Life Science and Food Engineering, Yibin University, Yibin, 644000, China.

References

1. Sun W, Samimi H, Gamez M, Zare H, Frost B. Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nature Neuroscience.* 2018, **21**(8)**:** 1038-1048.
2. Tran M, Reddy PH. Defective autophagy and mitophagy in aging and Alzheimer's disease. *Frontiers in Neuroscience.* 2021, **14:** 612757.
3. Frost B, Hemberg M, Lewis J, Feany MB. Tau promotes neurodegeneration through global chromatin relaxation. *Nature Neuroscience.* 2014, **17**(3)**:** 357-366.
4. Öztan G, İşsever H. Molecular mechanisms and genetics of Alzheimer's disease. *Turkish Journal of Biochemistry.* 2023, **48**(3)**:** 218-229.
5. Coppede F, Mancuso M, Siciliano G, Migliore L, Murri L. Genes and the environment in neurodegeneration. *Bioscience reports.* 2006, **26**(5)**:** 341-367.
6. Smith L, Luchini C, Demurtas J, Soysal P, Stubbs B, Hamer M*, et al.* Telomere length and health outcomes: An umbrella review of systematic reviews and meta-analyses of observational studies. *Ageing research reviews.* 2019, **51:** 1-10.
7. Bhala S, Best AF, Giri N, Alter BP, Pao M, Gropman A*, et al.* CNS manifestations in patients with telomere biology disorders. *Neurology: Genetics.* 2019, **5**(6)**:** 370.
8. Jedlička P, Tokan V, Kejnovská I, Hobza R, Kejnovský E. Telomeric retrotransposons show propensity to form G-quadruplexes in various eukaryotic species. *Mobile DNA.* 2023, **14**(1)**:** 3.
9. Erdem HB, Bahsi T, Ergün MA. Function of telomere in aging and age related diseases. *Environmental Toxicology and Pharmacology.* 2021, **85:** 103641.
10. Kuan X-Y, Fauzi NSA, Ng KY, Bakhtiar A. Exploring the Causal Relationship Between Telomere Biology and Alzheimer's Disease. *Molecular Neurobiology.* 2023, **60**(8)**:** 4169-4183.
11. Forero DA, Gonzalez-Giraldo Y, Lopez-Quintero C, Castro-Vega LJ, Barreto GE, Perry G. Telomere length in Parkinson's disease: A meta-analysis. *Experimental gerontology.* 2016, **75:** 53-55.

12. Degerman S, Domellöf M, Landfors M, Linder J, Lundin M, Haraldsson S*, et al.* Long leukocyte telomere length at diagnosis is a risk factor for dementia progression in idiopathic parkinsonism. *PloS one.* 2014, **9**(12)**:** e113387.
13. Fu W, Lu C, Mattson MP. Telomerase mediates the cell survival-promoting actions of brain-derived neurotrophic factor and secreted amyloid precursor protein in developing hippocampal neurons. *Journal of Neuroscience.* 2002, **22**(24)**:** 10710-10719.
14. Saretzki G. Telomerase and neurons: an unusual relationship. *Neural Regeneration Research.* 2022, **17**(11)**:**
15. Saretzki G. The Telomerase Connection of the Brain and Its Implications for Neurodegenerative Diseases. *Stem Cells.* 2023, **41**(3)**:** 233-241.
16. Greider CW, Blackburn EH. A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature.* 1989, **337**(6205)**:** 331-337.
17. Saretzki G, Wan T. Telomerase in Brain: The New Kid on the Block and Its Role in Neurodegenerative Diseases. *Biomedicines*; 2021.
18. Alison S, Satomi M, Johannes A, Gabriele S. The Role of Telomerase Protein TERT in Alzheimer's Disease and in Tau-Related Pathology <em>In Vitro</em&gt. *The Journal of Neuroscience.* 2015, **35**(4)**:** 1659.
19. Spilsbury A, Miwa S, Attems J, Saretzki G. The role of telomerase protein TERT in Alzheimer's disease and in tau-related pathology in vitro. *Journal of Neuroscience.* 2015, **35**(4)**:** 1659-1674.
20. Gao X, Yu X, Zhang C, Wang Y, Sun Y, Sun H*, et al.* Telomeres and Mitochondrial Metabolism: Implications for Cellular Senescence and Age-related Diseases. *Stem Cell Reviews and Reports.* 2022, **18**(7)**:** 2315-2327.
21. George JA, DeBaryshe PG, Traverse KL, Celniker SE, Pardue M-L. Genomic organization of the Drosophila telomere retrotransposable elements. *Genome Research.* 2006, **16**(10)**:** 1231-1240.
22. Ravel-Godreuil C, Znaidi R, Bonnifet T, Joshi RL, Fuchs J. Transposable elements as new players in neurodegenerative diseases. *FEBS Letters.* 2021, **595**(22)**:** 2733-2755.
23. Tam OH, Ostrow LW, Gale Hammell M. Diseases of the nERVous system: retrotransposon activity in neurodegenerative disease. *Mobile DNA.* 2019, **10**(1)**:** 32.
24. Gorbunova V, Seluanov A, Mita P, McKerrow W, Fenyö D, Boeke JD*, et al.* The role of retrotransposable elements in ageing and age-associated diseases. *Nature.* 2021, **596**(7870)**:** 43-53.
25. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *nature.* 2005, **435**(7044)**:** 903-910.
26. Bodea GO, McKelvey EG, Faulkner GJ. Retrotransposon-induced mosaicism in the neural genome. *Open Biology.* 2018, **8**(7)**:** 180074.
27. Ahmadi A, De Toma I, Vilor-Tejedor N, Eftekhariyan Ghamsari MR, Sadeghi I. Transposable elements in brain health and disease. *Ageing Research Reviews.* 2020, **64:** 101153.
28. Zhao N, Yin G, Liu C, Zhang W, Shen Y, Wang D*, et al.* Critically short telomeres derepress retrotransposons to promote genome instability in embryonic stem cells. *Cell Discovery.* 2023, **9**(1)**:** 45.
29. Guo C, Jeong H-H, Hsieh Y-C, Klein H-U, Bennett DA, De Jager PL*, et al.* Tau Activates Transposable Elements in Alzheimer's Disease. *Cell reports.* 2018, **23**(10)**:** 2874-2880.
30. Ochoa Thomas E, Zuniga G, Sun W, Frost B. Awakening the dark side: retrotransposon activation in neurodegenerative disorders. *Current Opinion in Neurobiology.* 2020, **61:** 65-72.
31. Evering TH, Marston JL, Gan L, Nixon DF. Transposable elements and Alzheimer’s disease pathogenesis. *Trends in Neurosciences.* 2023, **46**(3)**:** 170-172.
32. Vallés-Saiz L, Ávila J, Hernández F. Lamivudine (3TC), a Nucleoside Reverse Transcriptase Inhibitor, Prevents the Neuropathological Alterations Present in Mutant Tau Transgenic Mice. *International Journal of Molecular Sciences*; 2023.
33. Benhelli-Mokrani H, Mansuroglu Z, Chauderlier A, Albaud B, Gentien D, Sommer S*, et al.* Genome-wide identification of genic and intergenic neuronal DNA regions bound by Tau protein under physiological and stress conditions. *Nucleic Acids Research.* 2018, **46**(21)**:** 11405-11422.

34. Sjöberg MK, Shestakova E, Mansuroglu Z, Maccioni RB, Bonnefoy E. Tau protein binds to pericentromeric DNA: a putative role for nuclear tau in nucleolar organization. *Journal of Cell Science.* 2006, **119**(10)**:** 2025-2034.
35. Rossi G, Dalprà L, Crosti F, Lissoni S, Sciacca FL, Catania M*, et al.* A new function of microtubule-associated protein tau: involvement in chromosome stability. *Cell cycle.* 2008, **7**(12)**:** 1788-1794.
36. Antón-Fernández A, Vallés-Saiz L, Avila J, Hernández F. Neuronal nuclear tau and neurodegeneration. *Neuroscience.* 2023, **518:** 178-184.
37. Li W, Wang XS, Hua Qu M, Liu Y, He RQ. Human protein tau represses DNA replication in vitro. *Biochimica et Biophysica Acta (BBA) - General Subjects.* 2005, **1726**(3)**:** 280-286.
38. Bou Samra E, Buhagiar-Labarchède G, Machon C, Guitton J, Onclercq-Delic R, Green MR*, et al.* A role for Tau protein in maintaining ribosomal DNA stability and cytidine deaminase-deficient cell survival. *Nature Communications.* 2017, **8**(1)**:** 693.
39. Dickson JR, Yoon H, Frosch MP, Hyman BT. Cytoplasmic Mislocalization of RNA Polymerase II Subunit RPB1 in Alzheimer Disease Is Linked to Pathologic Tau. *Journal of Neuropathology & Experimental Neurology.* 2021, **80**(6)**:** 530-540.
40. Husseman JW, Hallows JL, Bregman DB, Leverenz JB, Nochlin D, Jin L-W*, et al.* Hyperphosphorylation of RNA Polymerase II and Reduced Neuronal RNA Levels Precede Neurofibrillary Tangles in Alzheimer Disease. *Journal of Neuropathology & Experimental Neurology.* 2001, **60**(12)**:** 1219-1232.
41. Blackburn E, Szostak J. The molecular structure of centromeres and telomeres. *Annual review of biochemistry.* 1984, **53**(1)**:** 163-194.
42. Graakjaer J, Bischoff C, Korsholm L, Holstebroe S, Vach W, Bohr VA*, et al.* The pattern of chromosome-specific variations in telomere length in humans is determined by inherited, telomere-near factors and is maintained throughout life. *Mechanisms of ageing and development.* 2003, **124**(5)**:** 629-640.
43. Turner KJ, Vasu V, Griffin DK. Telomere biology and human phenotype. *Cells.* 2019, **8**(1)**:** 73.
44. Ohki R, Tsurimoto T, Ishikawa F. In vitro reconstitution of the end replication problem. *Molecular and cellular biology.* 2001, **21**(17)**:** 5753-5766.
45. O'sullivan RJ, Karlseder J. Telomeres: protecting chromosomes against genome instability. *Nature reviews Molecular cell biology.* 2010, **11**(3)**:** 171-181.
46. Srinivas N, Rachakonda S, Kumar R. Telomeres and Telomere Length: A General Overview. *Cancers*; 2020.
47. Liu W, Xie Y, Ma J, Luo X, Nie P, Zuo Z*, et al.* IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics.* 2015, **31**(20)**:** 3359-3361.
48. de-Leon SB-T, Davidson EH. Gene Regulation: Gene Control Network in Development. *Annual Review of Biophysics.* 2007, **36**(Volume 36, 2007)**:** 191-212.
49. Bansal M, Gatta GD, di Bernardo D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics.* 2006, **22**(7)**:** 815-822.
50. Mileyko Y, Joh RI, Weitz JS. Small-scale copy number variation and large-scale changes in gene expression. *Proceedings of the National Academy of Sciences.* 2008, **105**(43)**:** 16659-16664.
51. Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. *Human Molecular Genetics.* 2009, **18**(R1)**:** R1-R8.
52. Tang Y-C, Amon A. Gene Copy-Number Alterations: A Cost-Benefit Analysis. *Cell.* 2013, **152**(3)**:** 394-405.
53. Zhao X, Liu X, Zhang A, Chen H, Huo Q, Li W*, et al.* The correlation of copy number variations with longevity in a genome-wide association study of Han Chinese. *Aging (Albany NY).* 2018, **10**(6)**:** 1206.
54. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. *Nature Reviews Genetics.* 2009, **10**(8)**:** 551-564.
55. Cerantonio A, Citrigno L, Greco BM, De Benedittis S, Passarino G, Maletta R*, et al.* The Role of Mitochondrial Copy Number in Neurodegenerative Diseases: Present Insights and Future Directions. *International Journal of Molecular Sciences*; 2024.

56. Picca A, Guerra F, Calvani R, Coelho-Júnior HJ, Leeuwenburgh C, Bucci C*, et al.* The contribution of mitochondrial DNA alterations to aging, cancer, and neurodegeneration. *Experimental Gerontology.* 2023, **178:** 112203.
57. Yang J-H, Hayano M, Griffin PT, Amorim JA, Bonkowski MS, Apostolides JK*, et al.* Loss of epigenetic information as a cause of mammalian aging. *Cell.* 2023, **186**(2)**:** 305-326.e327.
58. Khatir I, Brunet MA, Meller A, Amiot F, Patel T, Lapointe X*, et al.* Decoupling of mRNA and Protein Expression in Aging Brains Reveals the Age-Dependent Adaptation of Specific Gene Subsets. *Cells*; 2023.
59. Savageau MA. Demand Theory of Gene Regulation. I. Quantitative Development of the Theory. *Genetics.* 1998, **149**(4)**:** 1665-1676.
60. Chen K, Rajewsky N. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics.* 2007, **8**(2)**:** 93-103.
61. Levine E, Zhang Z, Kuhlman T, Hwa T. Quantitative Characteristics of Gene Regulation by Small RNA. *PLOS Biology.* 2007, **5**(9)**:** e229.
62. Pai AA, Pritchard JK, Gilad Y. The Genetic and Mechanistic Basis for Variation in Gene Regulation. *PLOS Genetics.* 2015, **11**(1)**:** e1004857.
63. von Hippel PH. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu Rev Biophys Biomol Struct.* 2007, **36**(1)**:** 79-105.
64. Pollex T, Rabinowitz A, Gambetta MC, Marco-Ferreres R, Viales RR, Jankowski A*, et al.* Enhancer–promoter interactions become more instructive in the transition from cell-fate specification to tissue differentiation. *Nature Genetics.* 2024, **56**(4)**:** 686-696.
65. Bartlett Z. The Hayflick Limit. *Embryo Project Encyclopedia.* 2014**:** 1-5.
66. Harley CB, Liu W, Blasco M, Vera E, Andrews WH, Briggs LA*, et al.* A Natural Product Telomerase Activator As Part of a Health Maintenance Program. *Rejuvenation Research.* 2010, **14**(1)**:** 45-56.
67. Teixeira MT, Arneric M, Sperisen P, Lingner J. Telomere length homeostasis is achieved via a switch between telomerase-extendible and-nonextendible states. *Cell.* 2004, **117**(3)**:** 323-335.
68. Heidenreich B, Kumar R. TERT promoter mutations in telomere biology. *Mutation Research/Reviews in Mutation Research.* 2017, **771:** 15-31.
69. Blackburn E, Chan S, Chang J, Fulton T, Krauskopf A, McEachern M*, et al.* Molecular manifestations and molecular determinants of telomere capping. Cold Spring Harbor symposia on quantitative biology; 2000: Cold Spring Harbor Laboratory Press; 2000. p. 253-264.
70. De Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & development.* 2005, **19**(18)**:** 2100-2110.
71. Griffith JD, Comeau L, Rosenfield S, Stansel RM, Bianchi A, Moss H*, et al.* Mammalian telomeres end in a large duplex loop. *Cell.* 1999, **97**(4)**:** 503-514.
72. Takai H, Smogorzewska A, de Lange T. DNA damage foci at dysfunctional telomeres. *Current biology.* 2003, **13**(17)**:** 1549-1556.
73. Passos JF, Saretzki G, von Zglinicki T. DNA damage in telomeres and mitochondria during cellular senescence: is there a connection? *Nucleic Acids Research.* 2007, **35**(22)**:** 7505-7513.
74. Qian W, Kumar N, Roginskaya V, Fouquerel E, Opresko PL, Shiva S*, et al.* Chemoptogenetic damage to mitochondria causes rapid telomere dysfunction. *Proceedings of the National Academy of Sciences.* 2019, **116**(37)**:** 18435-18444.
75. Tang Y, Nee AC, Lu A, Ran R, Sharp FR. Blood Genomic Expression Profile for Neuronal Injury. *Journal of Cerebral Blood Flow & Metabolism.* 2003, **23**(3)**:** 310-319.
76. de Courten-Myers GM, Xi G, Hwang J-H, Dunn RS, Mills AS, Holland SK*, et al.* Hypoglycemic Brain Injury: Potentiation from Respiratory Depression and Injury Aggravation from Hyperglycemic Treatment Overshoots. *Journal of Cerebral Blood Flow & Metabolism.* 2000, **20**(1)**:** 82-92.

77. Yang J, Zhi W, Wang L. Role of Tau Protein in Neurodegenerative Diseases and Development of Its Targeted Drugs: A Literature Review. *Molecules*; 2024.
78. Deepak B, Mohammad Sanaei A, Qiwen S, Shahrzad M. Hypoxia and its Emerging Therapeutics in Neurodegenerative, Inflammatory and Renal Diseases. In: Jing Z, Chi Z (eds). *Hypoxia and Human Diseases*. IntechOpen: Rijeka, 2017, p Ch. 21.
79. Han R, Liang J, Zhou B. Glucose Metabolic Dysfunction in Neurodegenerative Diseases—New Mechanistic Insights and the Potential of Hypoxia as a Prospective Therapy Targeting Metabolic Reprogramming. *International Journal of Molecular Sciences*; 2021.
80. Mamelak M. Energy and the Alzheimer brain. *Neuroscience & Biobehavioral Reviews.* 2017, **75:** 297-313.
81. Wang Y, Hu H, Liu X, Guo X. Hypoglycemic medicines in the treatment of Alzheimer's disease: Pathophysiological links between AD and glucose metabolism. *Frontiers in Pharmacology.* 2023, **14:**
82. Malkov A, lvanov AI, Popova I, Mukhtarov M, Gubkina O, Waseem T*, et al.* Reactive Oxygen Species Initiate a Metabolic Collapse in Hippocampal Slices: Potential Trigger of Cortical Spreading Depression. *Journal of Cerebral Blood Flow & Metabolism.* 2014, **34**(9)**:** 1540-1549.
83. Hu L, Wang W, Chen X, Bai G, Ma L, Yang X*, et al.* Prospects of antidiabetic drugs in the treatment of neurodegenerative disease. *Brain-X.* 2024, **2**(1)**:** e52.
84. Zhang C-E, Yang X, Li L, Sui X, Tian Q, Wei W*, et al.* Hypoxia-induced tau phosphorylation and memory deficit in rats. *Neurodegenerative Diseases.* 2014, **14**(3)**:** 107-116.
85. Hambali A, Kumar J, Hashim NFM, Maniam S, Mehat MZ, Cheema MS*, et al.* Hypoxia-Induced Neuroinflammation in Alzheimer's Disease: Potential Neuroprotective Effects of Centella asiatica. *Frontiers in Physiology.* 2021, **12:**
86. Pena E, San Martin-Salamanca R, El Alam S, Flores K, Arriaza K. Tau Protein Alterations Induced by Hypobaric Hypoxia Exposure. *International Journal of Molecular Sciences*; 2024.
87. Gao L, Tian S, Gao H, Xu Y. Hypoxia increases Aβ-induced tau phosphorylation by calpain and promotes behavioral consequences in AD transgenic mice. *Journal of Molecular Neuroscience.* 2013, **51:** 138-147.
88. Liu Z, Li T, Li P, Wei N, Zhao Z, Liang H*, et al.* The Ambiguous Relationship of Oxidative Stress, Tau Hyperphosphorylation, and Autophagy Dysfunction in Alzheimer's Disease. *Oxidative Medicine and Cellular Longevity.* 2015, **2015**(1)**:** 352723.
89. Yan SD, Yan SF, Chen X, Fu J, Chen M, Kuppusamy P*, et al.* Non-enzymatically glycated tau in Alzheimer's disease induces neuronal oxidant stress resulting in cytokine gene expression and release of amyloid β-peptide. *Nature Medicine.* 1995, **1**(7)**:** 693-699.
90. Shukla V, Mishra SK, Pant HC. Oxidative Stress in Neurodegeneration. *Advances in Pharmacological and Pharmaceutical Sciences.* 2011, **2011**(1)**:** 572634.
91. Gros-Louis F, Larivière R, Gowing G, Laurent S, Camu W, Bouchard J-P*, et al.* A frameshift deletion in peripherin gene associated with amyotrophic lateral sclerosis. *Journal of Biological Chemistry.* 2004, **279**(44)**:** 45951-45956.
92. Ziff OJ, Taha Doaa M, Crerar H, Clarke BE, Chakrabarti Anob M, Kelly G*, et al.* Reactive astrocytes in ALS display diminished intron retention. *Nucleic Acids Research.* 2021, **49**(6)**:** 3168-3184.
93. Hogan AL, Grima N, Fifita JA, McCann EP, Heng B, Fat SCM*, et al.* Splicing factor proline and glutamine rich intron retention, reduced expression and aggregate formation are pathological features of amyotrophic lateral sclerosis. *Neuropathology and Applied Neurobiology.* 2021, **47**(7)**:** 990-1003.
94. Humphrey J, Birsa N, Milioto C, McLaughlin M, Ule AM, Robaldo D*, et al.* FUS ALS-causative mutations impair FUS autoregulation and splicing factor networks through intron retention. *Nucleic Acids Research.* 2020, **48**(12)**:** 6889-6905.
95. Reber S, Stettler J, Filosa G, Colombo M, Jutzi D, Lenzken SC*, et al.* Minor intron splicing is regulated by FUS and affected by ALS‐associated FUS mutants. *The EMBO Journal.* 2016, **35**(14)**:** 1504-1521-1521.
96. Savage AL, Wilm TP, Khursheed K, Shatunov A, Morrison KE, Shaw PJ*, et al.* An Evaluation of a SVA Retrotransposon in the FUS Promoter as a Transcriptional Regulator and Its Association to ALS. *PLOS ONE.* 2014, **9**(3)**:** e90833.

97. Luisier R, Tyzack GE, Hall CE, Mitchell JS, Devine H, Taha DM*, et al.* Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nature Communications.* 2018, **9**(1)**:** 2010.
98. Lu Y-Q, Chen J-M, Lin H, Feng S-Y, Che C-H, Liu C-Y*, et al.* Novel Intronic Mutations of TBK1 Promote Aberrant Splicing Modes in Amyotrophic Lateral Sclerosis. *Frontiers in Molecular Neuroscience.* 2022, **15:**
99. Krishnan G, Zhang Y, Gu Y, Kankel MW, Gao F-B, Almeida S. CRISPR deletion of the C9ORF72 promoter in ALS/FTD patient motor neurons abolishes production of dipeptide repeat proteins and rescues neurodegeneration. *Acta Neuropathologica.* 2020, **140**(1)**:** 81-84.
100. Maruyama H, Kawakami H. Optineurin and amyotrophic lateral sclerosis. *Geriatrics & Gerontology International.* 2013, **13**(3)**:** 528-532.
101. Iida A, Hosono N, Sano M, Kamei T, Oshima S, Tokuda T*, et al.* Novel deletion mutations of OPTN in amyotrophic lateral sclerosis in Japanese. *Neurobiology of Aging.* 2012, **33**(8)**:** 1843.e1819-1843.e1824.
102. Zinman L, Liu H, Sato C, Wakutani Y, Marvelle A, Moreno D*, et al.* A mechanism for low penetrance in an ALS family with a novel SOD1 deletion. *Neurology.* 2009, **72**(13)**:** 1153-1159.
103. Valdmanis PN, Belzil VV, Lee J, Dion PA, St-Onge J, Hince P*, et al.* A Mutation that Creates a Pseudoexon in SOD1 Causes Familial ALS. *Annals of Human Genetics.* 2009, **73**(6)**:** 652-657.