Tat basic domain: A "Swiss army knife" of HIV-1 Tat?

Margarita Kurnaeva, Eugene Sheval, Yana Musinova, Yegor Vassetzky

TAT BASIC DOMAIN: A 'SWISS ARMY KNIFE' OF HIV-1 TAT?

Journal:	*Reviews in Medical Virology*
Manuscript ID	RMV-2018-075.R1
Wiley - Manuscript type:	Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Kurnaeva, Margarita; Lomonosov Moscow State University, Moscow, Russia Sheval, Eugene; Lomonosov Moscow State University, Moscow, Russia Musinova, Yana; Lomonosov Moscow State University, Moscow, Russia Vassetzky, Yegor; Gustave Roussy, CNRS UMR8126
Keywords:	HIV, Tat, Transcription
Abstract:	Tat regulates transcription from the human immunodeficiency virus (HIV) provirus. It plays a crucial role in disease progression, supporting efficient replication of the viral genome. Tat also modulates many functions in the host genome via its interaction with chromatin and proteins. Many of the functions of Tat are associated with its basic domain rich in arginine and lysine residues. It is still unknown why the basic domain exhibits so many diverse functions. However, the highly charged basic domain, coupled with the overall structural flexibility of Tat protein itself, makes the basic domain a key player in binding to or associating with cellular and viral components. In addition, the basic domain undergoes diverse post-translational modifications which further expand and modulate its functions. Here we review the current knowledge of Tat basic domain and its versatile role in the interaction between the virus and the host cell.

Tat basic domain: a 'Swiss army knife' of HIV-1 Tat?

Margarita A. Kurnaeva[1,2], Eugene V. Sheval[2,3,4], Yana R. Musinova[2,4,5], Yegor S. Vassetzky[4,5,6]

[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
[2]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
[3]Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia
[4]LIA 1066 LFR2O French-Russian Joint Cancer Research Laboratory, Villejuif, France
[5]Koltzov Institute of Developmental Biology of Russian Academy of Sciences, Moscow, Russia
[6]UMR8126, CNRS, Université Paris-Sud, Institut Gustave Roussy, Villejuif, France

Correspondence

Dr. Yegor Vassetzky, Nuclear Organization and Pathologies CNRS, UMR-8126, Institut Gustave Roussy, 39 rue Camille-Desmoulins, 94805 Villejuif, France

Email: yegor.vassetzky@cnrs.fr

Running head: HIV-1 Tat basic domain

Funding information

Russian Science Foundation, Grant/Award Number: 17-75-20199; Russian Foundation for Basic Research, Grant/Award Number: 18-54-16002; the Plan Cancer; ANRS; la Ligue Contre le Cancer

List of abbreviations: CPP, cell penetrating peptide; HAT, histone acetyltransferase; HIV, human immunodeficiency virus; LANA, latency-associated nuclear antigen; LTR, long terminal repeats; NES, nuclear export signal; NLS, nuclear localization signal; NoLS, nucleolar localization signal; P-TEFb, positive transcription elongation factor, PTD, protein transduction domain; TAR, transactivation response element; Tat, transactivator of transcription

Summary

Tat regulates transcription from the human immunodeficiency virus (HIV) provirus. It plays a crucial role in disease progression, supporting efficient replication of the viral genome. Tat also modulates many functions in the host genome *via* its interaction with chromatin and proteins. Many of the functions of Tat are associated with its basic domain rich in arginine and lysine residues. It is still unknown why the basic domain exhibits so many diverse functions. However, the highly charged basic domain, coupled with the overall structural flexibility of Tat protein itself, makes the basic domain a key player in binding to or associating with cellular and viral components. In addition, the basic domain undergoes diverse post-translational modifications which further expand and modulate its functions. Here we review the current knowledge of Tat basic domain and its versatile role in the interaction between the virus and the host cell.

KEYWORDS

HIV-1, Tat protein, basic domain, transactivation, nuclear localization signal, nucleolar localization signal, protein transduction domain

1 | INTRODUCTION

The HIV-1 genome is composed of nine genes including *tat* (transactivator of transcription) coding for a Tat regulatory protein which plays a pivotal role in regulation of viral transcription.[1-3] Depending on the HIV-1 strain, the length of Tat varies between 86-104 aa. The *tat* gene is composed of two exons: the first exon codes for 72 amino acids, the remaining part of the protein is encoded by the second exon (Fig. 1).[4] Tat protein can be divided into several domains: (i) N-terminal acidic domain (1-21 aa) essential for structural stability and transcription elongation; (ii) Cysteine-rich domain (22-37 aa) required for transcription elongation, Zn-dependent function, and binding to cellular components; (iii) Hydrophobic core domain (38-47 aa) participates in structural stability and transcription elongation; (iv) Basic domain (48-59 aa) essential in binding to trans-activation response (TAR) element and to cellular components; (v) Glutamine-rich region (60-72) required for structural stability; (vi) Domain encoded by the second exon.[5,6] The domain encoded by the second exon is less conserved and less well studied, but it has been implicated in cell adhesion, HIV-1 replication, interactions with integrins and regulation of host cell gene expression.[7-10] Domains encoded by the first exon are believed to be sufficient for the transactivation activity and modulation of numerous cellular components by Tat protein.[4,5,11]

Tat is an intrinsically disordered protein,[12] and therefore, only nuclear magnetic resonance structures are available for Tat alone. Intrinsically disordered proteins are believed to gain a more ordered state upon interaction with their target partners via conformational selection and induced folding.[13,14] Indeed, Tat undergoes induced but limited folding upon binding to specific fragments of antibodies.[15] Conformational changes likely appear in the basic region of Tat protein as it was shown for Tat protein from the equine infectious anemia virus,[16] which shares similarities with the basic domain of HIV Tat.[15] Additionally, the basic domain provides structural stability for Tat protein through electrostatic interactions with its N-terminal part.[17] Other Tat domains can undergo conformational changes as well: the prime example is Tat-positive transcription elongation factor (P-TEFb) complex. X-ray crystallography showed that the first three N-terminal Tat domains (1-49 residues) sustain extended conformation mostly through interactions with cyclin T1 whereas 50-86 residues are not defined.[11] Nuclear magnetic resonance opened the veil on the structural propensity of Tat protein suggesting that the cysteine-rich region tends to fold into α-helixes in contrast to the basic domain with extended or β-sheet conformation.[18] Comparison analysis of X-ray and nuclear magnetic resonance studies suggests that different fragments of Tat protein can employ different folding mechanisms.[18] This flexibility enables Tat to adopt diverse conformations upon interaction with its physiological partners, thus greatly extending its multifunctionality.

One of the most important and well-studied domains of Tat protein is the basic domain. While Tat can tolerate up to 40% of sequence mutations without significant changes in its activity, its basic domain is highly conserved among Tat variants.[6] It is enriched with positively-charged arginine and lysine residues comprising ^{49}RKKRRQRRR^{57} motif. Basic domain confers many properties to Tat such at regulation of viral transcription and manipulation of cellular processes in favor of HIV. In this review, we discuss different aspects of Tat basic domain and its versatile role in the interaction between the host cell and the virus.

2 | BASIC DOMAIN FUNCTIONS AS AN RNA BINDING MOTIF

Absence of Tat causes predominantly short transcript production from the HIV-1 long terminal repeat (LTR).[19] Tat interacts with an RNA enhancer element (TAR) positioned at the 5' end of the viral transcript via its basic domain leading to facilitation of viral transcription.[20-24] Tat interacts with P-TEFb comprised of cyclin T1 and CDK9 and mediates ternary Tat-P-TEFb complex to TAR RNA.[11,25-28] Tat binds the nascent RNA via its basic domain, causing conformational changes of the P-TEFb complex, thus enabling CDK9 to phosphorylate RNA polymerase II resulting in the full-length HIV-1 transcript production.[25-27] Nullbasic Tat mutant, in which the entire basic domain was replaced with glycine/alanine residues, was shown to interact with P-TEFb complex but failed to recruit the ternary complex to the nascent viral RNA confirming that the basic domain plays a pivotal role in the transactivation function of Tat.[29]

Numerous attempts had been made to identify the key residues responsible for TAR RNA binding, though with some discrepancies. An early study by the

Calnan group showed that peptides derived from basic domain of Tat protein could directly bind TAR RNA while their amino acid sequence could be completely rearranged and still retained its high affinity to TAR.[22] Mutations of arginines to alanines significantly reduced the binding capacity while substitution to lysines restored the capacity to nearly wild-type levels, proposing that the overall charge of the basic domain is likely the key factor of Tat-TAR RNA binding.[22] Substitution of arginine residues for glutamine at 52 and 53 positions led to abrogation of transactivation activity.[24] In another experiment, K50 interacted with G34 of TAR RNA loop, indicating that the protein-RNA cross-link occurred at K50 position whereas mutation at G34 to U34 significantly reduced Tat-Cyclin T1 binding capacity.[30,31] Mutations of the first two lysines at 50 and 51 positions to serine and glycine (K50S and K51G), respectively, showed a decrease in Tat activity by 50% *in vivo*.[32] Interesting results assessing transactivation of HIV LTR came from molecular dynamics simulations combined with *in vitro* experiments by the Carloni group. While K50A and K51A mutants were functionally defective for HIV LTR transactivation, K50R and K51R had a functional transactivation capacity although it was lower than that of the wild-type Tat. Still, the K50A and K51A mutants localized to the nucleus, suggesting that these mutations most likely affected Tat interactions with RNAs or nuclear protein complexes.[33]

In addition to viral RNA, Tat is believed to interact with cellular RNAs. The ability of Tat basic domain to associate with human RNAs was examined via immunoprecipitation analysis of the wild-type Tat and its mutated form, K50S-K51G.[32] The mutant form showed a significant decrease in interaction with RNA, in particular Tat-bound FADD and TNFRSF8 RNAs, leading to speculations that, in addition to TAR RNA, Tat was able to associate with the specific set of human RNAs for which an intact basic domain was required.[32] Moreover, Tat basic domain was proposed to specifically target Dicer-dependent RNAi,[34] the innate immune response against the viral infection.[35-37] Besides mammalian RNAs, Tat was shown to impair pre-rRNA processing in *Drosophila melanogaster* cells via association with U3 snoRNA and fibrillarin (nucleolar components necessary for pre-rRNA processing), although it was not explicitly stated that the basic domain was specifically involved in such interactions.[38]

3 | INTERACTION WITH CELLULAR COMPONENTS

Besides its capacity to bind viral and cellular RNAs, the basic domain of Tat also associates with cellular and viral proteins. Tat peptide comprising 48-60 amino acids was shown to block protein kinase C activity by binding to the kinase active site.[39] NPM1 (B23 or nucleophosmin), a ubiquitous protein involved in diverse cellular processes, has been proposed to directly interact with Tat *via* its basic domain.[40] Tat basic domain has been also implemented in recruitment of CIS protein (cytokine-inducible SH2 containing protein) to CD127 surface receptor of CD8 T cells for internalization and subsequent degradation of CD127 which led to reduction in T cells.[41,42] The basic domain along with a conserved tryptophan residue W11 are responsible for unconventional secretion of Tat from cells.[12] The basic domain binds phosphatidylinositol-4,5-biphosphate (a phospholipid of the inner leaflet of the plasma membrane) causing conformational changes which enable insertion of W11 into the membrane with the subsequent secretion of Tat.[12,43] Arginine residues from Tat basic domain target Tat to cell membrane lipid rafts and enhance fibroblast growth factor-2 (FGF-2) signaling in human podocytes isolated from children with HIV-associated nephropathy (HIVAN), whereas alanine substitutions abrogated Tat nuclear localization, association with lipid rafts, and enhancement of FGF-2 signaling.[44] Tat, via its basic domain, binds Tip60, a cellular histone acetyltransferase (HAT) which controls expression of cellular genes capable to interfere with the efficient viral replication and propagation.[45] Additionally, histone chaperone hNAP-1 binds Tat basic domain, stimulating regulation of Tat-mediated viral transcription.[46] Tat basic domain has been recently observed to interfere with the host cell proliferation and induction of apoptosis of HIV-1-infected lymphocytes. In Jurkat cells, Tat disrupts localization of PRS3, which in association with α-tubulin plays a critical role in mitosis, leading to faulty mitotic spindle and chromosome formation.[47] The basic domain of Tat associates with IkB-α (an inhibitor of nuclear factor NF-kB), leading to liberation of p65 from IkB-α/p65 complex and the subsequent transcriptional activation of pro-inflammatory genes.[48] These are just several examples of the interaction of Tat basic domain with cellular proteins. Many other cellular partners of Tat have been recently discovered.[49] Yet, whether the Tat basic domain was involved in these interactions remains to be elucidated.

4 | BASIC DOMAIN FUNCTIONS AS A NUCLEAR LOCALIZATION SIGNAL

The primary role of Tat lies in activation of viral transcription, hence Tat must be able to pass the large nuclear pore complexes of the nuclear envelope. Globular proteins with a Mr less than 40,000 to 60,000 or 5-10 nm in diameter can freely diffuse between the cytoplasm and the nucleus whereas larger macromolecules require an energy-driven mechanism to traverse the nuclear pore complexes.[50-52] In most cases, proteins targeted to the nucleus contain specific trafficking motifs such as the nuclear localization (NLS) and nuclear export signals (NES). Since the discovery of the first NLS signal in the SV40 large T-antigen protein containing a short stretch of basic amino acids, similar target sequences have been identified and characterized in a broad range of viral and cellular proteins. While the classical (or canonical) NLS pathway employs an adaptor molecule importin-α for binding to importin-β, a non-classical NLS pathway involves direct binding of the cargo protein to the importin-β. In both pathways, importin-β acts as a carrier by docking cargo-importin(s) complex to the nuclear pore complexed and releasing the cargo into the nucleus upon binding to Ran-GTP.[53-56]

Passive diffusion has been suggested to be a major mechanism of Tat nuclear entry.[57] However, despite its small size (Mr 14,000-16,000) favoring passive diffusion, Tat contains a functional NLS (^{49}RKKRRQRRR^{57}) within its basic domain and was shown to localize preferentially in the nucleoplasm and nucleolus.[4,58,59] Classical and non-classical mechanisms of nuclear entry along with association with nuclear components had been previously proposed for Tat protein. *In vitro* assays suggested that Tat nuclear import was mediated by the direct binding of its basic domain to importin-β, thus competing with importin-α for the same binding site of importin-β.[60] In contrast, a novel mechanism independent of the importin pathway was proposed by another group, indicating the ability of Tat basic domain to interact with nuclear components.[61] Deletion of the basic domain led to cytoplasmic localization[62] and a dramatic decrease in Tat activity.[4] Mutation analysis of amino acids 50, 55 and 56 replaced by uncharged residues revealed a loss in nuclear localization, suggesting the presence of two partially overlapping or juxtaposed NLSs. In other words, mutations in the RKKRR motif or RRR alone had little effect on nuclear localization while mutations in both parts led to the cytoplasmic accumulation.[63] Additionally, the first set of basic domain amino acids was shown to function as NLS while the remaining RRR motif tended to bind to intracellular components.[57,64] Using oriented peptide binding approach, it was proposed that KKKRR, KKKRK, and KKRKK motifs are sufficient for binding importin-α.[65] Structure of importin-α with ^{48}GRKKRRQRRRAPQN^{61} peptide has been recently determined sing X-ray crystallography. It was shown that ^{48}GRKKRRQR^{55} residues mediate a strong association with importin-α.[66] Taken together, these results suggest that Tat may utilize different pathways to enter the nucleus. Ability to combine both classical and non-classical NLSs has been previously demonstrated for the latency-associated nuclear antigen (LANA) of KSHV.[67] Another example is Rev protein, whose basic domain is enriched with positively-charged arginine residues similar to Tat. Depending on the cell type, Rev has been shown to utilize different mutually-exclusive pathways of nuclear entry, either through importin-β or transportin.[68]

5 | BASIC DOMAIN FUNCTIONS AS A NUCLEOLAR LOCALIZATION SIGNAL

Besides its predominant accumulation in the nucleus of different cell lines, Tat is observed to localize to the nucleoli as well. Tat was proposed to possess a nucleolar localization signal (NoLS) ^{48}GRKKRRQRRRAP^{57} which drives Tat accumulation into nucleoli and encompasses Tat basic domain which serves as NLS.[69] However, experimental data showed that only positively-charged amino acids are involved in Tat nucleolar localization while flanking non-charged residues hardly exhibit any effect on its localization; this implies that charge-dependent mechanism is a major force of dynamic accumulation of Tat in the nucleolus.[70] Indeed, the nucleolus is a highly dynamic structure and accumulation of proteins can be achieved through interactions with nucleolar building blocks, such as rDNA or its transcripts.[71] *In vitro* experiments revealed binding of GST-Tat to NPM1, a major nucleolar protein, emphasizing that such interactions occur within Tat basic domain.[40] Additionally, NPM1 has been suggested to be vital for the nuclear entry of Tat and the subsequent nucleolar localization. In this case, NPM1 behaves as a shuttling protein driving Tat through the nuclear pore complexes to the nucleoli.[40] Such shuttling mechanism by NPM1 has been previously observed for NCL (C23 or nucleolin), another major nucleolar protein lacking defined NoLS.[72] Apart from these two studies, there is no other experimental data exist confirming NPM1/Tat interactions *in* *vivo*. Alternatively, based on nucleolar interactome analysis, Tat itself can physically modulate protein nucleolar accumulation involved in HIV-1 pathogenesis; but whether Tat basic domain is directly involved in targeting of these proteins to nucleolus remains to be elucidated.[73] NoLS was also suggested to be a key player in targeting Tat to the nucleoli in *Drosophila melanogaster* cells, thus following localization pattern of mammalian cells[38].

It might be possible that Tat basic domain serves as a NoLS in the first place rather than NLS due to its main function in the regulation of viral transcription and the small size enabling passive diffusion. Nucleoli are one of the main targets of some viral proteins including Tat. Presence of independent sites for nuclear entry and nucleolar accumulation might be evolutionarily unfavorable for Tat. Indeed, HIV-1 genome itself is just under 10 kb with 16 proteins successfully serving its biology.[74] Both nuclear and nucleolar accumulation could be achieved by Tat binding to/associating with other nuclear/nucleolar proteins or RNAs through its highly charged basic domain. It is thus its overall structural plasticity coupled with highly charged basic domain make it the prime example of "minimum complexity - maximum efficiency".

6 | BASIC DOMAIN FUNCTIONS AS A PROTEIN TRANSDUCTION DOMAIN

Tat is actively secreted by infected cells and can traverse plasma membranes of various eukaryotic cells,[75-77] affecting their gene expression and cellular functions. Concentration of extracellular Tat can reach up to 40 ng/ml in blood of acutely infected patients.[78] Tat has been detected in sera of patients undergoing antiretroviral therapy, showing that modern anti-HIV drugs are not able to block the cellular release of Tat protein.[79] The ability to penetrate cell plasma membranes has been attributed to Tat basic domain thus termed as protein transduction domain (PTD) or cell penetrating peptide (CPP) if used as a peptide. Its potency to mediate cellular uptake has been widely exploited by a large number of laboratories for transcellular protein transduction not only in mammalian cells,[80] but in plants as well,[81] thus making Tat basic domain a promising tool for transcellular drug delivery to a wide variety of cells.

Various experiments with short peptides spanning the Tat basic domain fused to different cargoes showed the ability of these peptides to enter cells,[82-84] while peptides with a truncated or mutated basic domain failed to translocate through cell membranes.[83,85] One of the first mechanisms of cell penetration proposed that ionic interactions between the highly dense positively-charged basic domain of Tat and negatively-charged phospholipids of the plasma membrane prompted an invagination of the membrane.[83] In contrast, two other studies suggested an adsorptive-mediated endocytosis as a way for internalization.[76,86] Further work performed by several research groups demonstrated that Tat basic domain fused to different cargoes could bind heparin,[87-89] a structural homolog of heparan sulfate glycosaminoglycan, abundantly present on the cell surface. Heparan sulfate proteoglycans (HSPGs) have been proposed to be highly versatile receptors responsible for the mechanism of cellular entry.[90] It has been shown that Tat internalization can be achieved via HSPG binding following subsequent active caveolar endocytosis through cell membrane lipid rafts.[90,92] Full-length Tat protein can utilize HSPG receptors whereas unconjugated Tat peptides can be internalized by cells that lack these receptors. Alternatively, the study performed on T cells demonstrated that a full-length Tat can use a clathrin/AP-2-dependent endocytosis; however, whether the basic domain plays any roles in this pathway remains unclear.[93] These results demonstrate that different internalization pathways can be employed, depending on the cell type and specificity of the cargo.[94]

7 | POST-TRANSLATIONAL MODIFICATIONS OF THE BASIC DOMAIN REGULATE TAT ACTIVITY

Post-translational modifications play a key role in the heterogeneity of protein functions. Disordered protein domains have been suggested to be a subject for many post-translational modifications.[13] The disordered state is advantageous in that in can provide greater accessibility to the sites for post-translational modifications.[13] Post-translational modifications include acetylation, methylation, phosphorylation, to name but a few. Each modification involves specific enzymes that recognize distinct amino acids within the polypeptide chain. Tat undergoes post-translational modifications which greatly expands its functions. The basic domain of Tat undergoes acetylation and methylation, affecting its capacity to facilitate viral transcription and modulate a broad range of cellular processes.

7.1 | ACETYLATION

Acetylation is mediated by HATs catalyzing the transfer of acetyl groups from acetyl coenzyme A to the ϵ-amino group of lysine. It has been speculated that HIV-1 transcription involves at least two phases. Defined as early TAR-dependent and late TAR-independent, these phases are equally important for the functions of Tat whereas Tat acetylation has been proposed to act as a regulatory switch between them.[95,96] Tat transactivation activity has been shown to depend upon lysine acetylation at K50 and K51.[98,99-102] The critical role in K50 acetylation has been attributed to p300,[98,100] a HAT responsible for regulation of gene expressions via chromatin remodeling. K50 acetylation leads to liberation of Tat from TAR RNA and cyclin T1 and activation of Tat-mediated transcriptional elongation of HIV-1 through binding to RNA polymerase II.[98,100,101] Further direct binding of acetylated K50 to the bromodomain of PCAF, a p300/CBP-associated factor, has been proposed to be essential for Tat transactivation since the site-directed mutation of K50A led to termination of Tat transactivation activity whereas substitution to arginine K50R did not affect the interaction with TAR, cyclin T1 or PCAF (Fig. 2).[102-105]

Acetylation of K50 has been further shown to facilitate Tat interaction with SWI/SNF chromatin remodeling complex containing BRG-1 and its subsequent recruitment to the viral LTR.[106-108] This interaction is achieved via direct binding of acetylated Tat to the bromodomain of BRG-1, permitting SWI/SNF to alter the structure of downstream nucleosomes and enabling further viral transcription.[108]

In addition to p300, the hGCN5 HAT acetylates both K50 and K51 residues *in vitro* and significantly enhances Tat-mediated transcription of HIV LTR.[109] Indeed, mutational analysis aiming to neutralize the electrical charge of lysine and block acetylation by substitution with the alanine residue demonstrated that acetylation of K50 exclusively regulates Tat transactivation activity (Fig. 2).[110]

Proteomic analysis and *in vivo* experiments showed that Tat acetylated at K50 and K51 residues preferentially binds p32, an inhibitor of splicing factor ASF/SF-2, and mediates its transport to the viral promoter, thus regulating the splicing pattern of HIV-1.[111] Acetylation also decreases cellular uptake of Tat-derived peptides acetylated at either K50 or K51 positions.[112] Dysregulation of expression of C5, APBA1, BDNF, and CRLF2 genes associated with inflammation and damage by the K50A mutant has been recently identified in human macrophages.[113]

Acetylation of Tat can be reversed by sirtuin 1 (SIRT1) class III deacetylase. Acetylation and deacetylation cycles are believed to be necessary for the pursuit of complete HIV transcription (Fig. 2).[101] In particular, acetylation results in the release of Tat from TAR leading to translocation of Tat and chromatin-modifying transcriptional coactivators to elongating RNA polymerase and recruitment of chromatin-remodeling complexes while deacetylation by SIRT1 restores Tat basic domain to its initial form so that Tat can interact with P-TEFb and bind TAR RNA, leading to the new transcriptional cycle.[101,114]

7.2 | METHYLATION

In addition to acetylation, Tat basic domain undergoes methylation, a process in which methyl groups from S-adenosylmethionine are transferred to proteins regulating many protein functions. Unlike acetylation, methylation does not result in neutralization of residual electrical charge. Because methylation primarily affects arginines and lysines, Tat basic domain serves a prime substrate for post-translational modifications by different methyltransferases.

Methylation of position 50 and 51 lysines of the Tat basic domain can be generally accomplished by the action of SETDB1 and Set7/9-KMT7 methyltransferases (Fig. 3). Both are the members of a broad SET protein family that have been originally shown to specifically methylate lysines of histone H3 tail.[115,116] Depending on the position of histone H3 lysine, methylation can prompt either transcriptionally active or transcriptionally repressed state of chromatin.[117-119] Later, nonhistone proteins methylated by SETDB1 and Set7/9-KMT7 were discovered,[115,120-129] one of them is Tat protein.[130,131]

In vitro experiments, using SETDB1, wild-type Tat, and variation of Tat-derived mutant peptides have determined that both K50 and K51 can be methylated by SETDB1, whereas K51 showed an increased amount of methyl groups.[131] SETDB1 knockdown resulted in the increase of viral transcription, implying that methylation of K50 and/or K51 can attenuate HIV-1 transcription[131] similarly to the general assumption that methylation of histone H3 lysine 9 corresponds to the assembly of transcriptionally-silent heterochromatin. Set7/9-KMT7 has been also found to monomethylate K51 residue both *in vitro* and *in vivo*.[130] However, contrary to SETDB1, this monomethylation increased Tat affinity to TAR and Tat transcriptional activity.[130] Interestingly, Set7/9 has been identified as one of the methyltransferases that drives monomethylation of histone H3 lysine 4, initiating formation of transcriptionally-active euchromatin.[132-134] Such strikingly different outcomes in methylation of Tat lysines could favor the notion that methylation can influence HIV-1 biology[135] similar to histone methylation that is linked to both transcriptional activation and repression (Fig. 3).[136]

Lysines can be demethylated by LSD1 demethylase[137] which in complexes with different cofactors demethylates histone H3K4/9 [136,138] and nonhistone proteins such as p53,[139] E2F1,[140] and Tat.[141] *In vitro* experiments demonstrated that Tat K51 residue was specifically targeted by LSD1, resulting in HIV-1 transactivation.[141]

Besides lysine, Tat arginine residues can be methylated by arginine methyltransferase PRMT6. Early experiments by Boulanger and colleagues demonstrated that HA-tagged Tat expressed in HEK293T cell line was subjected to methylation by endogenous PRMT6; cotransfection with PRMT6 increased the level of Tat methylation.[142] Knockdown of PRMT6 led to an increase in HIV-1 production, demonstrating that methylation of the basic domain exerted a negative effect on Tat transactivation function.[142] Further experiments determined position 52 and 53 arginine residues (R52 and R53) to be specifically methylated by PRMT6. [143] Consistent with the previous observation, *in vitro* methylation of R52 and R53 triggered a decrease in Tat interaction with TAR RNA and complex formation with cyclinT1, thus affecting Tat function, whereas *in vivo* experiments augmented the role of Tat in transactivation by downregulating PRMT6.[143] PRMT6 has been proposed to be a restriction factor of HIV-1[142-144] as an innate cellular response to the viral replication.[143] Yet, this restrictive effect is counterbalanced by recently observed downregulation of PRMT6 expression in CD4+ T cell of HIV-infected patients, suggesting that the virus can indeed control expression of cellular genes to benefit its replication (Fig. 3).[145]

While methylation of Tat is generally linked to the attenuation of Tat transactivation activity, the fate of the methylated Tat remains largely obscure. Compelling results came from the study where instead of downregulating Tat transactivation function, overexpression of PRMT6 led to increased Tat stability by protection from proteasome-dependent degradation.[146] This strikingly different outcome has been explained as a way to fulfill multifunctional role of Tat apart from its transactivation function, while methylation serves as a molecular switch between Tat functions.[146] Further investigation by the same group demonstrated that methylation of R52 and R53 residues of Tat basic domain resulted in the exclusion of Tat-GFP fusion protein from the nucleolus of COS cells; thus it could also modulate Tat localization.[147]

Methylation is a complex post-translational modification with pleiotropic effects on protein functions. Arginine contains three nitrogen atoms (one ε and two η) in its side chain each of which can be monomethylated, symmetrically dimethylated or asymmetrically dimethylated. Depending on the state of methylation, thermodynamics of Tat-derived peptides binding capacity to TAR RNA has been recently assayed. Monomethylation of arginine R52 or R53 at ε-nitrogen atom enhanced binding affinity whereas monomethylation or asymmetric dimethylation at η-nitrogen resulted in reduced binding capacity.[148] Methylation was further suggested to influence Tat-TAR RNA interaction in a position- and state-dependent manner. Asymmetric dimethylation of R52 or R53 severely affected Tat-TAR RNA binding while dimethylation of flanking arginines, such as R49 or R57, slightly increased Tat-TAR RNA affinity.[149]

8 | CONCLUSIONS

Tat is a regulatory protein encoded by the HIV-1 viral genome; it plays a crucial role in regulation of viral and host gene expression. Tat can exhibit multiple functions which are required for viral pathogenesis. It can also enter uninfected cells and modulate cellular gene expression according to the viral needs, thus leading to oncogenesis or cellular death through apoptosis. Many functions of Tat are attributed to its basic domain (also designated as arginine-rich motif RNA-binding domain, nuclear localization signal, nucleolar localization signal, and protein transduction domain), which is highly conserved among different Tat variants though Tat itself is prone to mutations. This multifunctionality of the Tat basic domain is linked to its high charge and flexible structure. These can lead to interactions with many physiological partners including glycoproteins, proteins or protein/RNA complexes, chromatin of both viral and cellular origin, thus allowing Tat to accomplish various tasks. Tat basic domain is also subjected to post-translational modifications which may expand and modify its functionality. All these features make the basic domain the key component of Tat protein. In this review, we have summarized the current knowledge on Tat basic domain and its role in Tat functions, but most probably new functions of this viral "Swiss army knife" will be discovered in the near future.

ACKNOWLEDGEMENTS

The work was supported by the Russian Science Foundation (grant 17-75-20199) to Y.R.M., by the Russian Foundation for Basic Research (18-54-16002) to E.V.S. and by the Plan Cancer (ENVIBURKITT), ANRS and la Ligue Contre le Cancer to Y.S.V.

FINANCIAL AND COMPETING INTEREST DISCLOSURE

None declared.

ORCID

Eugene V. Sheval http://orcid.org/0000-0003-1687-1321

Yegor S. Vassetzky https://orcid.org/0000-0003-3101-7043

REFERENCES

1. Ne E, Palstra R-J, Mahmoudi T. Transcription: insights from the HIV-1 promoter. *Int Rev Cell Molecular Biol*. 2018;335:191–243.

2. Faust TB, Binning JM, Gross JD, Frankel AD. Making sense of multifunctional proteins: human immunodeficiency virus type 1 accessory and regulatory proteins and connections to transcription. *Annu Rev Virol.* 2017;4:241–260.

3. Das AT, Harwig A, Berkhout B. The HIV-1 Tat protein has a versatile role in activating viral transcription. *J Virol.* 2011;85:9506–9516.

4. Kuppuswamy M, Subramanian T, Srinivasan A, Chinnadurai G. Multiple functional domains of Tat, the trans-activator of HIV-1, defined by mutational analysis. *Nucleic Acids Res*. 1989;17:3551–3561.

5. Fulcher AJ, Jans DA. The HIV-1 Tat transactivator protein: a therapeutic target? *IUBMB* Li*fe.* 2003;55:669–680.

6. Campbell GR, Loret EP. What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an AIDS vaccine? *Retrovirology.* 2009;6:50.

7. Li L, Dahiya S, Kortagere S, Aiamkitsumrit B, Cunningham D, Pirrone V, et al. Impact of Tat genetic variation on HIV-1 disease. *Adv Virol.* 2012;2012:123605.

8. López-Huertas MR, Callejas S, Abia D, Mateos E, Dopazo A, Alcamí J, et al. Modifications in host cell cytoskeleton structure and function mediated by intracellular HIV-1 Tat protein are greatly dependent on the second coding exon.

Nucleic Acids Res. 2010;38:3287–3307.

9. Neuveut C, Scoggins RM, Camerini D, Markham RB, Jeang K-T. Requirement for the second coding exon of Tat in the optimal replication of macrophage-tropic HIV-1. *J Biomed Sci.* 2003;10:651–660.

10. Barillari G, Gendelman R, Gallo RC, Ensoli B. The Tat protein of human immunodeficiency virus type 1, a growth factor for AIDS Kaposi sarcoma and cytokine-activated vascular cells, induces adhesion of the same cell types by using integrin receptors recognizing the RGD amino acid sequence. *Proc Natl Acad Sci U S A*. 1993;90:7941–7945.

11. Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH. Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature.* 2010;465:747–751.

12. Debaisieux S, Rayne F, Yezid H, Beaumelle B. The ins and outs of HIV-1 Tat. *Traffic*. 2012;13:355–363.

13. Berlow RB, Dyson HJ, Wright PE. Functional advantages of dynamic protein disorder. *FEBS Lett*. 2015;589:2433–2440.

14. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature.* 2007;447:1021–1025.

15. Foucault M, Mayol K, Receveur-Bréchot V, Bussat M-C, Klinguer-Hamour C, Verrier B, et al. UV and X-ray structural studies of a 101-residue long Tat protein from a HIV-1 primary isolate and of its mutated, detoxified, vaccine candidate. *Proteins.* 2010;78:1441–1456.

16. Anand K, Schulte A, Vogel-Bachmayr K, Scheffzek K, Geyer M. Structural

insights into the Cyclin T1–Tat–TAR RNA transcription activation complex from EIAV. *Nat Struct Mol Biol.* 2008;15:1287.

17. Pantano S, Tyagi M, Giacca M, Carloni P. Molecular dynamic simulations on HIV-1 Tat. *Eur Biophys J.* 2004;33:344–351

18. To V, Dzananovic E, McKenna SA, O'Neil J. The dynamic landscape of the full-length HIV-1 transactivator of transcription. *Biochemistry*. 2016;55:1314–1325.

19. Kao S-Y, Calman AF, Luciw PA, Matija Peterlin B. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature.* 1987;330:489–493.

20. Pumfery A, Deng L, Maddukuri A, la Fuente C, Li H, Wade J, et al. Chromatin remodeling and modification during HIV-1 Tat-activated transcription. *Curr HIV Res.* 2003;1:343–362.

21. Berkhout B, Gatignol A, Rabson AB, Jeang KT. TAR-independent activation of the HIV-1 LTR: evidence that Tat requires specific regions of the promoter. *Cell.* 1990;62:757–767.

22. Calnan BJ, Biancalana S, Hudson D, Frankel AD. Analysis of arginine-rich peptides from the HIV Tat protein reveals unusual features of RNA-protein recognition. *Genes Dev*. 1991;5:201–210.

23. Weeks KM, Crothers DM. RNA recognition by Tat-derived peptides: interaction in the major groove? *Cell.* 1991;66:577–588.

24. Delling U, Roy S, Sumner-Smith M, Barnett R, Reid L, Rosen CA, et al. The number of positively charged amino acids in the basic domain of Tat is

critical for trans-activation and complex formation with TAR RNA. *Proc Natl Acad Sci U S A.* 1991;88:6234–6238.

25. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA. A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell*. 1998;92:451–462.

26. Bieniasz PD, Grdina TA, Bogerd HP, Cullen BR. Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *Proc Natl Acad Sci U S A*. 1999;96:7791–7796.

27. Mancebo HS, Lee G, Flygare J, Tomassini J, Luu P, Zhu Y, et al. P-TEFb kinase is required for HIV Tat transcriptional activation in vivo and in vitro. *Genes Dev*. 1997;11:2633–2644.

28. Zhu Y, Pe'ery T, Peng J, Ramanathan Y, Marshall N, Marshall T, et al. Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev*. 1997;11:2622–2632.

29. Rustanti L, Jin H, Lor M, Lin MH, Rawle DJ, Harrich D. A mutant Tat protein inhibits infection of human cells by strains from diverse HIV-1 subtypes. *Virol J*. 2017;14:52.

30. Richter S, Ping Y-H, Rana TM. TAR RNA loop: a scaffold for the assembly of a regulatory switch in HIV replication. *Proc Natl Acad Sci U S A.* 2002;99:7928–7933.

31. Richter S, Cao H, Rana TM. Specific HIV-1 TAR RNA loop sequence and functional groups are required for human Cyclin T1−Tat−TAR ternary complex

formation. *Biochemistry.* 2002;41:6391–6397.

32. Bouwman RD, Palser A, Parry CM, Coulter E, Rasaiyaah J, Kellam P, et al. Human immunodeficiency virus Tat associates with a specific set of cellular RNAs. *Retrovirology*. 2014;11:53.

33. Pantano S, Tyagi M, Giacca M, Carloni P. Amino acid modification in the HIV-1 Tat basic domain: insights from molecular dynamics and in vivo functional studies. *J Mol Biol.* 2002;318:1331–1339.

34. Ponia SS, Arora S, Kumar B, Banerjea AC. Arginine rich short linear motif of HIV-1 regulatory proteins inhibits dicer dependent RNA interference. *Retrovirology*. 2013;10:97.

35. Haasnoot J, Berkhout B. RNAi and cellular miRNAs in infections by mammalian viruses. *Methods Mol Biol.* 2011;721:23–41.

36. Voinnet O. Induction and suppression of RNA silencing: insights from viral infections. *Nat Rev Genet.* 2005;6:206–220.

37. Saumet A, Lecellier C-H. Anti-viral RNA silencing: do we look like plants? *Retrovirology*. 2006;3:3.

38. Ponti D, Troiano M, Bellenchi GC, Battaglia PA, Giuliani F. The HIV Tat protein affects processing of ribosomal RNA precursor. *BMC Cell Biol*. 2008;9:32.

39. Ekokoski E, Aitio O, Törnquist K, Yli-Kauhaluoma J, Tuominen RK. HIV-1 Tat-peptide inhibits protein kinase C and protein kinase A through substrate competition. *Eur J Pharm Sci.* 2010;40:404–411.

40. Li YP. Protein B23 is an important human factor for the nucleolar

localization of the human immunodeficiency virus protein Tat. *J Virol*. 1997;71:4098–4102.

41. Sugden S, Ghazawi F, MacPherson P. HIV-1 Tat protein recruits CIS to the cytoplasmic tail of CD127 to induce receptor ubiquitination and proteasomal degradation. *Virology*. 2016;498:192–200.

42. Sugden S, MacPherson P. The N-terminal region of HIV-1 Tat protein binds CD127 in human CD8 T Cells to target the receptor for down regulation through Tat's basic region. *Curr HIV Res*. 2015;13:226–243.

43. Mele A, Marino J, Chen K, Pirrone V, Janetopoulos C, Wigdahl B, et al. Defining the molecular mechanisms of HIV-1 Tat secretion: PtdIns(4,5)P_2 at the epicenter. *Traffic*. 2018;19:655–665

44. Xie X, Colberg-Poley AM, Das JR, Li J, Zhang A, Tang P, et al. The basic domain of HIV-tat transactivating protein is essential for its targeting to lipid rafts and regulating fibroblast growth factor-2 signaling in podocytes isolated from children with HIV-1-associated nephropathy. *J Am Soc Nephrol.* 2014;25:1800–1813.

45. Creaven M, Hans F, Mutskov V, Col E, Caron C, Dimitrov S, et al. Control of the histone-acetyltransferase activity of Tip60 by the HIV-1 transactivator protein, Tat. *Biochemistry*. 1999;38:8826–8830.

46. Vardabasso C, Manganaro L, Lusic M, Marcello A, Giacca M. The histone chaperone protein Nucleosome Assembly Protein-1 (hNAP-1) binds HIV-1 Tat and promotes viral transcription. *Retrovirology.* 2008;5:8.

47. Kim J, Kim Y-S. Effect of HIV-1 Tat on the formation of the mitotic spindle

by interaction with ribosomal protein S3. *Sci Rep.* 2018;8:8680.

48. Fiume G, Vecchio E, De Laurentiis A, Trimboli F, Palmieri C, Pisano A, et al. Human immunodeficiency virus-1 Tat activates NF-kB via physical interaction with IkB-α and p65. *Nucleic Acids Res.* 2012;40:3548–3562

49. Gautier VW, Gu L, O'Donoghue N, Pennington S, Sheehy N, Hall WW. In vitro nuclear interactome of the HIV-1 Tat protein. *Retrovirology.* 2009;6:47.

50. Kabachinski G, Schwartz TU. The nuclear pore complex--structure and function at a glance. *J Cell Sci*. 2015;128:423–429.

51. Rout MP, Aitchison JD. The nuclear pore complex as a transport machine. *J Biol Chem.* 2001;276:16593–16596.

52. Görlich D, Kutay U. Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol.* 1999;15:607–660.

53. Cook A, Bono F, Jinek M, Conti E. Structural biology of nucleocytoplasmic transport. *Annu Rev Biochem*. 2007;76:647–671.

54. Stewart M. Molecular mechanism of the nuclear protein import cycle. *Nat Rev Mol Cell Biol.* 2007;8:195–208.

55. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH. Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J Biol Chem.* 2007;282:5101–5105.

56. Fried H, Kutay U. Nucleocytoplasmic transport: taking an inventory. *Cell Mol Life Sci.* 2003;60:1659–1688.

57. Cardarelli F, Serresi M, Bizzarri R, Giacca M, Beltram F. In vivo study of HIV-1 Tat arginine-rich motif unveils its transport properties. *Mol Ther.*

2007;15:1313–1322.

58. Musinova YR, Sheval EV. The accumulation of the basic domain of HIV-1 Tat protein in the nuclei and the nucleoli is different from the accumulation of full-length Tat proteins. *Biopolym. Cell.* 2015;31:154–157.

59. Ruben S, Perkins A, Purcell R, Joung K, Sia R, Burghoff R, et al. Structural and functional characterization of human immunodeficiency virus Tat protein. *J Virol.* 1989;63:1–8.

60. Truant R, Cullen BR. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. *Mol Cell Biol.* 1999;19:1210–1217.

61. Efthymiadis A, Briggs LJ, Jans DA. The HIV-1 Tat nuclear localization sequence confers novel nuclear import properties *J Biol Chem.* 1998;273:1623–1628.

62. Siomi H, Shida H, Maki M, Hatanaka M. Effects of a highly basic region of human immunodeficiency virus Tat protein on nucleolar localization. *J Virol.* 1990;64:1803–1807.

63. Hauber J, Malim MH, Cullen BR. Mutational analysis of the conserved basic domain of human immunodeficiency virus Tat protein. *J Virol.* 1989;63:1181–1187.

64. Cardarelli F, Serresi M, Albanese A, Bizzarri R, Beltram F. Quantitative analysis of Tat peptide binding to import carriers reveals unconventional nuclear transport properties. *J Biol Chem*. 2011;286:12292–12299.

65. Yang SNY, Takeda AAS, Fontes MRM, Harris JM, Jans DA, Kobe B.

Probing the specificity of binding to the major nuclear localization sequence-binding site of importin-α using oriented peptide library screening. *J Biol Chem.* 2010;285:19935–19946.

66. Smith KM, Himiari Z, Tsimbalyuk S, Forwood JK. Structural basis for importin-α binding of the human imunodeficiency virus Tat. *Sci Rep.* 2017;7:1650.

67. Howard K, Cherezova L, DeMaster LK, Rose TM. ORF73 LANA homologs of RRV and MneRV2 contain an extended RGG/RG-rich nuclear and nucleolar localization signal that interacts directly with importin β1 for non-classical nuclear import. *Virology.* 2017;511:152–164.

68. Gu L, Tsuji T, Jarboui MA, Yeo GP, Sheehy N, Hall WW, et al. Intermolecular masking of the HIV-1 Rev NLS by the cellular protein HIC: novel insights into the regulation of Rev nuclear import. *Retrovirology.* 2011;8:17.

69. Dang CV, Lee WM. Nuclear and nucleolar targeting sequences of c-erb-A, c-myb, N-myc, p53, HSP70, and HIV tat proteins. *J Biol Chem*. 1989;264:18019–18023.

70. Musinova YR, Kananykhina EY, Potashnikova DM, Lisitsyna OM, Sheval EV. A charge-dependent mechanism is responsible for the dynamic accumulation of proteins inside nucleoli. *Biochim Biophys Acta.* 2015;1853:101–110.

71. Carmo-Fonseca M, Mendes-Soares L, Campos I. To be or not to be in the nucleolus. *Nat Cell Biol.* 2000;2:E107–12.

72. Li Y-P, Busch RK, Valdez BC, Busch H. C23 interacts with B23, a putative

nucleolar-localization-signal-binding protein. *Eur J Biochem*. 1996;237:153–158.

73. Jarboui MA, Bidoia C, Woods E, Roe B, Wynne K, Elia G, et al. Nucleolar protein trafficking in response to HIV-1 Tat: rewiring the nucleolus. *PLoS One.* 2012;7:e48702.

74. Li G, De Clercq E. HIV genome-wide protein associations: a review of 30 years of research. *Microbiol Mol Biol Rev.* 2016;80:679–731.

75. Frankel AD, Pabo CO. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell*. 1988;55: 1189–1193.

76. Mann DA, Frankel AD. Endocytosis and targeting of exogenous HIV-1 Tat protein. *EMBO J.* 1991;10:1733–1739.

77. Li W, Huang Y, Reid R, Steiner J, Malpica-Llanos T, Darden TA, et al. NMDA receptor activation by HIV-Tat protein is clade dependent. J *Neurosci.* 2008;28:12190–12198.

78. Xiao H, Neuveut C, Tiffany HL, Benkirane M, Rich EA, Murphy PM, et al. Selective CXCR4 antagonism by Tat: implications for in vivo expansion of coreceptor use by HIV-1. *Proc Natl Acad Sci U S A.* 2000;97:11466–11471.

79. Mediouni S, Darque A, Baillat G, Ravaux I, Dhiver C, Tissot-Dupont H, et al. Antiretroviral terapy does not block the secretion of the human immunodeficiency virus Tat protein. *Infect Disord Drug Targets.* 2012;12:81–86.

80. Zou L, Peng Q, Wang P, Zhou B. Progress in research and application of HIV-1 TAT-derived cell-penetrating peptide. *J Membr Biol.* 2017;250:115–122.

81. Chugh A, Eudes F. Translocation and nuclear accumulation of monomer and dimer of HIV-1 Tat basic domain in triticale mesophyll protoplasts. *Biochim*

Biophys Acta. 2007;1768:419–426.

82. Green M, Loewenstein PM. Autonomous functional domains of chemically synthesized human immunodeficiency virus Tat trans-activator protein. *Cell*. 1988;55:1179–1188.

83. Vivès E, Brodin P, Lebleu B. A truncated HIV-1 Tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *J Biol Chem.* 1997;272:16010–16017.

84. Ignatovich IA, Dizhe EB, Pavlotskaya AV, Akifiev BN, Burov SV, Orlov SV, et al. Complexes of plasmid DNA with basic domain 47-57 of the HIV-1 Tat protein are transferred to mammalian cells by endocytosis-mediated pathways. *J Biol Chem*. 2003;278:42625–42636.

85. Rusnati M, Tulipano G, Urbinati C, Tanghetti E, Giuliani R, Giacca M, et al. The basic domain in HIV-1 Tat protein as a target for polysulfonated heparin-mimicking extracellular Tat antagonists. *J Biol Chem.* 1998;273:16027–16037.

86. Huang L, Farhood H, Serbina N, Teepe AG, Barsoum J. Endosomolytic activity of cationic liposomes enhances the delivery of human immunodeficiency virus-1 trans-activator protein (TAT) to mammalian cells. *Biochem Biophys Res Commun.* 1995;217:761–768.

87. Rusnati M, Tulipano G, Spillmann D, Tanghetti E, Oreste P, Zoppetti G, et al. Multiple interactions of HIV-I Tat protein with size-defined heparin oligosaccharides. *J Biol Chem.* 1999;274:28198–28205.

88. Ziegler A, Seelig J. Interaction of the protein transduction domain of HIV-1 TAT with heparan sulfate: binding mechanism and thermodynamic parameters.

Biophys J. 2004;86:254–263.

89. Hakansson S, Jacobs A, Caffrey M. Heparin binding by the HIV-1 tat protein transduction domain. *Protein Sci. 2001*;10:2138–2139.

90. Christianson HC, Belting M. Heparan sulfate proteoglycan as a cell-surface endocytosis receptor. *Matrix Biol.* 2014;35:51–55.

91. Fittipaldi A, Ferrari A, Zoppé M, Arcangeli C, Pellegrini V, Beltram F, et al. Cell membrane lipid rafts mediate caveolar endocytosis of HIV-1 Tat fusion proteins. *J Biol Chem.* 2003;278:34141–34149.

92. Tyagi M, Rusnati M, Presta M, Giacca M. Internalization of HIV-1 Tat requires cell surface heparan sulfate proteoglycans. *J Biol Chem.* 2001;276:3254–3261.

93. Vendeville A, Rayne F, Bonhoure A, Bettache N, Montcourrier P, Beaumelle B. HIV-1 Tat enters T cells using coated pits before translocating from acidified endosomes and eliciting biological responses. *Mol Biol Cell.* 2004;15(5):2347–2360

94. Richard JP, Melikov K, Brooks H, Prevot P, Lebleu B, Chernomordik LV. Cellular uptake of unconjugated TAT peptide involves clathrin-dependent endocytosis and heparan sulfate receptors. *J Biol Chem.* 2005;280:15300–15306.

95. Kaehlcke K, Dorr A, Hetzer-Egger C, Kiermer V, Henklein P, Schnoelzer M, et al. Acetylation of Tat defines a cyclinT1-independent step in HIV transactivation. *Mol Cell.* 2003;12:167–176.

96. Ott M, Dorr A, Hetzer-Egger C, Kaehlcke K, Schnolzer M, Henklein P, et al.

Tat acetylation: a regulatory switch between early and late phases in HIV transcription elongation. *Novartis Foundation Symposia*. 2004;259:182–193.

97. Kiernan RE, Vanhulle C, Schiltz L, Adam E, Xiao H, Maudoux F, et al. HIV-1 tat transcriptional activity is regulated by acetylation. *EMBO J.* 1999;18:6106–6118.

98. Ott M, Schnölzer M, Garnica J, Fischle W, Emiliani S, Rackwitz H-R, et al. Acetylation of the HIV-1 Tat protein by p300 is important for its transcriptional activity. *Curr Biol.* 1999;9:1489–1493.

99. Kumar S, Maiti S. The effect of N-acetylation and N-methylation of lysine residue of Tat peptide on its interaction with HIV-1 TAR RNA. *PLoS One*. 2013;8:e77595.

100. Mujtaba S, He Y, Zeng L, Farooq A, Carlson JE, Ott M, et al. Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. *Mol Cell.* 2002;9:575–586.

101. Pagans S, Pedal A, North BJ, Kaehlcke K, Marshall BL, Dorr A, et al. SIRT1 regulates HIV transcription via Tat deacetylation. *PLoS Biol*. 2005;3:e41.

102. Dormeyer W, Dorr A, Ott M, Schnolzer M. Acetylation of the HIV-1 Tat protein: an in vitro study. *Anal Bioanal Chem.* 2003;376:994–1005.

103. Brès V, Tagami H, Péloponèse J-M, Loret E, Jeang K-T, Nakatani Y, et al. Differential acetylation of Tat coordinates its interaction with the co-activators cyclin T1 and PCAF. *EMBO J.* 2002;21:6811–6819.

104. Dorr A, Kiermer V, Pedal A, Rackwitz H-R, Henklein P, Schubert U, et al. Transcriptional synergy between Tat and PCAF is dependent on the binding of

acetylated Tat to the PCAF bromodomain. *EMBO J*. 2002;21:2715–2723.

105. Deng L, de la Fuente C, Fu P, Wang L, Donnelly R, Wade JD, et al. Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones. *Virology.* 2000;277:278–295.

106. Tréand C, du Chéné I, Brès V, Kiernan R, Benarous R, Benkirane M, et al. Requirement for SWI/SNF chromatin-remodeling complex in Tat-mediated activation of the HIV-1 promoter. *EMBO J.* 2006;25:1690–1699.

107. Mahmoudi T, Parra M, Vries RGJ, Kauder SE, Peter Verrijzer C, Ott M, et al. The SWI/SNF chromatin-remodeling complex is a cofactor for Tat transactivation of the HIV promoter. *J Biol Chem.* 2006;281:19960–19968.

108. Agbottah E, Deng L, Dannenberg LO, Pumfery A, Kashanchi F. Effect of SWI/SNF chromatin remodeling complex on HIV-1 Tat activated transcription. *Retrovirology.* 2006;3:48.

109. Col E, Caron C, Seigneurin-Berny D, Gracia J, Favier A, Khochbin S. The histone acetyltransferase, hGCN5, interacts with and acetylates the HIV transactivator, Tat. *J Biol Chem.* 2001;276:28179–28184.

110. He M, Zhang L, Wang X, Huo L, Sun L, Feng C, et al. Systematic analysis of the functions of lysine acetylation in the regulation of Tat activity. *PLoS One*. 2013;8:e67186.

111. Berro R, Kehn K, de la Fuente C, Pumfery A, Adair R, Wade J, et al. Acetylated Tat regulates human immunodeficiency virus type 1 splicing through its interaction with the splicing regulator p32. *J Virol.* 2006;80:3189–3204.

112. Liu M-C, Chen C-Y, Chiang C-H, Wang W-M, Cheng RP. Effect of lysine

methylation and acetylation on the RNA recognition and cellular uptake of Tat-derived peptides. *Bioorg Med Chem*. 2016;24:5047–5051.

113. Carvallo L, Lopez L, Fajardo JE, Jaureguiberry-Bravo M, Fiser A, Berman JW. HIV-Tat regulates macrophage gene expression in the context of neuro AIDS. *PLoS One.* 2017;12:e0179882.

114. Ott M, Geyer M, Zhou Q. The control of HIV transcription: keeping RNA polymerase II on track. *Cell Host Microbe*. 2011;10:426–435.

115. Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P, et al. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev*. 2002;16:479–489.

116. Schultz DC, Ayyanathan K, Negorev D, Maul GG, Rauscher FJ 3rd. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev*. 2002;16:919–932.

117. Rice JC, Briggs SD, Ueberheide B, Barber CM, Shabanowitz J, Hunt DF, et al. Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains. *Mol Cell.* 2003;12:1591–1598.

118. Fischle W, Wang Y, David Allis C. Histone and chromatin cross-talk. *Curr Opin Cell Biol.* 2003;15:172–183.

119. Keating ST, El-Osta A. Transcriptional regulation by the Set7 lysine methyltransferase. *Epigenetics.* 2013;8:361–372.

120. He S, Owen DR, Jelinsky SA, Lin L-L. Lysine methyltransferase SETD7

(SET7/9) regulates ROS signaling through mitochondria and NFE2L2/ARE pathway. *Sci Rep.* 2015;5:14368.

121. Li Y, Reddy MA, Miao F, Shanmugam N, Yee JK, Hawkins D, Ren B, Nataraja. Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent inflammatory genes. Relevance to diabetes and inflammation. *J Biol Chem*. 2008;283:26771-26781.

122. Subramanian K, Jia D, Kapoor-Vazirani P, Powell DR, Collins RE, Sharma D, et al. Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. *Mol Cell.* 2008;30:336–347.

123. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol*. 2005;6:227.

124. Kouskouti A, Scheer E, Staub A, Tora L, Talianidis I. Gene-specific modulation of TAF10 function by SET9-mediated methylation. *Mol Cell.* 2004;14:175–182.

125. Grewal SIS, Moazed D. Heterochromatin and epigenetic control of gene expression. *Science*. 2003;301:798–802.

126. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*. 2001;410:116–120.

127. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature*. 2001;410:120–124.

128. Nakayama J, Rice JC, Strahl BD, Allis CD, Grewal SI. Role of histone H3

lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*. 2001;292:110–113.

129. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*. 2000;406:593–599.

130. Pagans S, Kauder SE, Kaehlcke K, Sakane N, Schroeder S, Dormeyer W, et al. The cellular lysine methyltransferase Set7/9-KMT7 binds HIV-1 TAR RNA, monomethylates the viral transactivator Tat, and enhances HIV Transcription. *Cell Host Microbe*. 2010;7:234–244.

131. Van Duyne R, Easley R, Wu W, Berro R, Pedati C, Klase Z, et al. Lysine methylation of HIV-1 Tat regulates transcriptional activity of the viral LTR. *Retrovirology*. 2008;5:40.

132. Wang H, Cao R, Xia L, Erdjument-Bromage H, Borchers C, Tempst P, et al. Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. *Mol Cell.* 2001;8:1207–1217.

133. Herz H-M, Mohan M, Garruss AS, Liang K, Takahashi Y-H, Mickey K, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes Dev.* 2012;26:2604–2620.

134. Kwon T, Chang JH, Kwak E, Lee CW, Joachimiak A, Kim YC, et al. Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO J*. 2003;22:292–303.

135. Yedavalli VRK, Jeang K-T. Methylation: a regulator of HIV-1 replication? *Retrovirology*. 2007;4:9.

136. Zhang Y. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev*. 2001;15:2343–2360.

137. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell*. 2004;119:941–953.

138. Klose RJ, Zhang Y. Regulation of histone methylation by demethylimination and demethylation. *Nat Rev Mol Cell Biol*. 2007;8:307–318.

139. Huang J, Sengupta R, Espejo AB, Lee MG, Dorsey JA, Richter M, et al. p53 is regulated by the lysine demethylase LSD1. *Nature*. 2007;449:105–108.

140. Kontaki H, Talianidis I. Lysine methylation regulates E2F1-induced cell death. *Mol Cell.* 2010;39:152–160.

141. Sakane N, Kwon H-S, Pagans S, Kaehlcke K, Mizusawa Y, Kamada M, et al. Activation of HIV transcription by the viral Tat protein requires a demethylation step mediated by lysine-specific demethylase 1 (LSD1/KDM1). *PLoS Pathog.* 2011;7:e1002184.

142. Boulanger M-C, Liang C, Russell RS, Lin R, Bedford MT, Wainberg MA, et al. Methylation of Tat by PRMT6 regulates human immunodeficiency virus type 1 gene expression. *J Virol.* 2005;79:124–131.

143. Xie B, Invernizzi CF, Richard S, Wainberg MA. Arginine methylation of the human immunodeficiency virus type 1 Tat protein by PRMT6 negatively affects Tat Interactions with both cyclin T1 and the Tat transactivation region. *J Virol.* 2007;81:4226–4234.

144. Boehm D, Ott M. Host Mmethyltransferases and demethylases: potential new epigenetic targets for HIV cure strategies and beyond. *AIDS Res Hum Retroviruses.* 2017;33:S8–S22.

145. Bogoi RN, de Pablo A, Valencia E, Martín-Carbonero L, Moreno V, Vilchez-Rueda HH, et al. Expression profiling of chromatin-modifying enzymes and global DNA methylation in CD4+ T cells from patients with chronic HIV infection at different HIV control and progression states. *Clin Epigenetics*. 2018;10:20.

146. Sivakumaran H, van der Horst A, Fulcher AJ, Apolloni A, Lin M-H, Jans DA, et al. Arginine methylation increases the stability of human immunodeficiency virus type 1 Tat. *J Virol.* 2009;83:11694–11703.

147. Fulcher AJ, Sivakumaran H, Jin H, Rawle DJ, Harrich D, Jans DA. The protein arginine methyltransferase PRMT6 inhibits HIV-1 Tat nucleolar retention. *Biochim Biophys Acta*. 2016;1863:254–262.

148. Kumar S, Maiti S. Effect of different arginine methylations on the thermodynamics of Tat peptide binding to HIV-1 TAR RNA. *Biochimie*. 2013;95:1422–1431.

149. Li J-H, Chiu W-C, Yao Y-C, Cheng RP. Effect of arginine methylation on the RNA recognition and cellular uptake of Tat-derived peptides. *Bioorg Med Chem*. 2015;23:2281–2286.

Figure legends

FIGURE 1 Structure of the Tat protein. Tat contains the following domains: N-terminal acidic domain (1-21 aa), cysteine-rich domain (22-37 aa), hydrophobic core domain (38-47 aa), basic domain (48-59 aa), glutamine-rich domain (60-72), and domain encoded by the second exon.

FIGURE 2 Post-translational modifications of the Tat basic domain by different acetyltransferases. Acetylation of K50 by P300 and pGCN5 leads to association with RNA polymerase II, bromodomain of PCAF, and with SWI/SNF chromatin remodelling complex resulting in liberation of Tat protein from TAR and its subsequent translocation along with the chromatin-remodelling complexes to elongating RNA polymerase II. Deacetylation by SIRTI enables Tat to return to its non-acetylated form and begin a new transactivation cycle.

FIGURE 3 Methylation of Tat basic domain. SETDBI methyltransferase methylates K50 and K51 residues leading to a decrease in transactivation of transcription. Contrary to SETDBI, methylation of the same residues by SET7/9 methyltransferase prompts an increased affinity to TAR RNA resulting in upregulation of transactivation of transcription. PRMT6 (an important host factor) methylates R52 and R53 residues liberating Tat basic domain from the TAR RNA and leading to attenuation of Tat transcriptional activity. PRMT6 modification also allows Tat protein to escape from proteasome-dependent degradation and increases Tat stability.

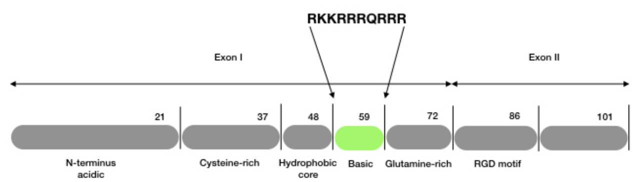

Figure 1. Structure of the Tat protein. Tat contains the following domains: N-terminus, cysteine-rich, hydrophobic core, basic, glutamine-rich, RGD motifs.

361x270mm (72 x 72 DPI)

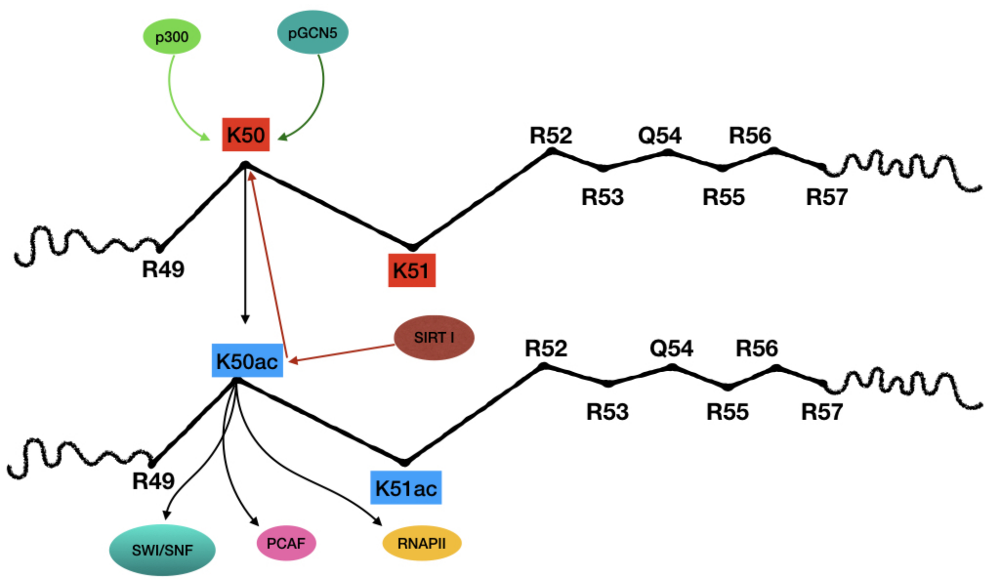

Figure 2. Post-translational modifications of the Tat basic domain by different acetyltransferases. Acetylation of K50 by P300 and pGCN5 leads to association with RNA polymerase II (RNAPII), bromodomain of PCAF, and with SWI/SNF chromatin remodelling complex resulting in liberation of Tat protein from TAR and its subsequent translocation along with the chromatin-remodelling complexes to elongating RNA polymerase II. Deacetylation by SIRTI enables Tat to return to its non-acetylated form and begin a new transactivation cycle.

361x270mm (72 x 72 DPI)

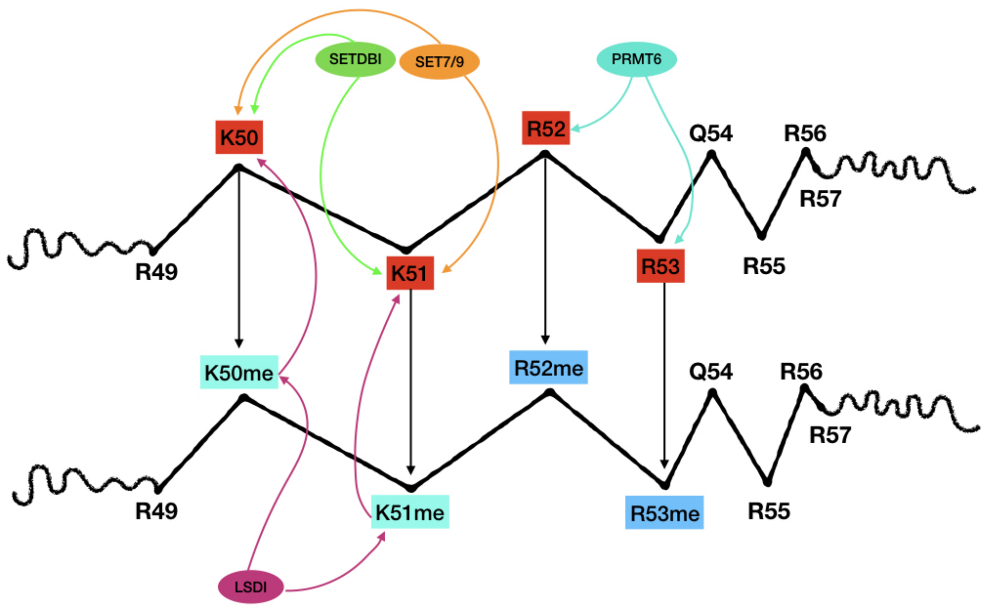

Figure 3. Methylation of Tat basic domain. SETDBI methyltransferase methylates K50 and K51 residues leading to a decrease in transactivation of transcription. Contrary to SETDBI, methylation of the same residues by SET7/9 methyltransferase prompts an increased affinity to TAR RNA resulting in upregulation of transactivation of transcription. PRMT6 (an important host factor) methylates R52 and R53 residues liberating Tat basic domain from the TAR RNA and leading to attenuation of Tat transcriptional activity. PRMT6 modification also allows Tat protein to escape from proteasome-dependent degradation and increases Tat stability.

361x270mm (72 x 72 DPI)

Tat basic domain: a 'Swiss army knife' of HIV-1 Tat?

Margarita A. Kurnaeva[1,2], Eugene V. Sheval[2,3,4], Yana R. Musinova[2,4,5], Yegor S. Vassetzky[4,5,6]

[1]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia~~,~~
[2]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia~~,~~
[3]Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia
[4]LIA 1066 LFR2O French-Russian Joint Cancer Research Laboratory, ~~94805~~ Villejuif, France
[5]Koltzov Institute of Developmental Biology of Russian Academy of Sciences, ~~Vavilov str. 26, 119334~~ Moscow, Russia
[6]UMR8126, CNRS, Université Paris-Sud, Institut Gustave Roussy, Villejuif, France~~,~~

~~*Corresponding author:~~

~~Yegor Vassetzky~~

Correspondence

Dr. Yegor Vassetzky, Nuclear Organization and Pathologies CNRS, UMR-8126, Institut Gustave Roussy, 39 rue Camille-Desmoulins, 94805 Villejuif, France

Email: yegor.vassetzky@cnrs.fr

~~**Word count excluding referecnces:4600**~~

Formatted: Header

Style Definition: Normal: Font: Font color: Black, Border: : (No border), Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border), Bar : (No border)

Style Definition: Balloon Text: Font: (Default) Segoe UI, 9 pt, Font color: Black, Border: : (No border), Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border), Bar : (No border)

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Font: Not Italic

Formatted: Left, Indent: Left: 0 cm, Right: 0 cm, Line spacing: 1,5

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic, Pattern: Clear (White), Not Highlight

Formatted: Font: Not Italic, Pattern: Clear (White), Not Highlight

Formatted: Font: Not Italic

Formatted: Font: Not Italic, French (France)

Formatted: French (France)

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Header

Funding information

Russian Science Foundation, Grant/Award Number: 17-75-20199; Russian Foundation for Basic Research, Grant/Award Number: 18-54-16002; the Plan Cancer; ANRS; la Ligue Contre le Cancer

Formatted: Header

List of abbreviations: CPP, cell penetrating peptide; HAT, histone acetyltransferase; HIV, human immunodeficiency virus; LANA, latency-associated nuclear antigen; LTR, long terminal repeats; NES, nuclear export signal; NLS, nuclear localization signal; NoLS, nucleolar localization signal; P-TEFb, positive transcription elongation factor, PTD, protein transduction domain; TAR, transactivation response element; Tat, transactivator of transcription

Summary

Tat regulates transcription ~~of~~from the human immunodeficiency virus (HIV~~).~~) provirus. It plays a crucial role in disease progression, supporting efficient replication of the viral genome. Tat also modulates many functions in the host genome *via* its interaction with chromatin and proteins. Many of ~~Tat's~~the functions of Tat are associated with its basic domain rich in arginine and lysine residues. It is still unknown why the basic domain exhibits so many diverse functions. However, the highly charged basic domain, coupled with the overall structural flexibility of Tat protein itself, makes the basic domain a key player in binding to or ~~association~~associating with cellular and viral components. In addition, the basic domain ~~has been shown to undergo~~undergoes diverse post-translational modifications which further expand and modulate its functions. Here we review the current knowledge of Tat basic domain and its versatile role in the interaction between the virus and the host cell.

Formatted: Header

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Header

Formatted: Font: Not Bold

KEYWORDS

HIV-1, Tat protein, basic domain, transactivation, nuclear localization signal, nucleolar localization signal, protein transduction domain

Formatted: Header

1 | INTRODUCTION

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: English (U.S.)

The HIV-1 genome is composed of nine genes including *tat* (transactivator of transcription) coding for a Tat regulatory protein ~~(Tat)~~ which plays a pivotal role in regulation of viral transcription.[1-3] Depending on the HIV-1 strain, ~~Tat's~~the length of Tat varies between 86-104 aa. The *tat* gene is composed of two exons: the first exon codes for 72 amino acids, the remaining part of the protein is encoded by the second exon (Fig. 1).[4] Tat protein can be divided into several domains: (i) N-~~terminus~~terminal acidic domain (1-21 aa) ~~is~~ essential for structural stability and transcription elongation; (ii) Cysteine-rich domain (22-37 aa) ~~is~~ required for transcription elongation, Zn-dependent function, and binding to cellular components; (iii) Hydrophobic core domain (38-47 aa) participates in structural stability and transcription elongation; (iv) Basic domain (48-59 aa) ~~is~~ essential in ~~TAR binding, NLS, PTD, and~~ binding to trans-activation response (TAR) element and to cellular components; (v) Glutamine-rich region (60-72) ~~is~~ required for structural stability~~.~~[5,6] ~~Domains~~; (vi) Domain encoded by the second exon ~~are .~~[5,6] The domain encoded by the second exon is less conserved and less well studied, but ~~they have~~it has been implicated in cell adhesion, ~~better~~ HIV-1 replication, interactions with integrins~~,~~ and regulation of host cell gene expression.[7-10] Domains encoded by the first exon are believed to be sufficient for the transactivation activity and modulation of numerous cellular components by Tat protein.~~[4,5,11]~~[4,5,11]

Formatted: Indent: Left: 0 cm, Right: 0 cm, No widow/orphan control, Don't adjust space between Latin and Asian text, Don't adjust space between Asian text and numbers

Formatted: Font: 11 pt, English (U.S.)

Tat is an intrinsically disordered protein~~.~~,[12] and therefore, only nuclear magnetic resonance structures are available for Tat alone. Intrinsically disordered proteins are believed to gain a more ordered state upon interaction with their target partners via ~~two previously proposed extreme mechanism:~~ conformational selection and induced folding.[13,14] ~~In UV and X-Ray structural studies~~Indeed, Tat ~~has been shown to be subjected to~~undergoes induced but limited folding upon binding to specific fragments of antibodies~~, Fab'~~.[15] ~~It was also noted that conformational~~Conformational changes likely appear in the basic region of Tat protein as it was ~~proven~~ shown for ~~EIAV~~ Tat~~[16]~~, protein from the equine infectious anemia virus,[16] which shares similarities with the basic domain of HIV Tat~~.~~~~[15]~~.[15] Additionally, the basic domain provides structural stability for Tat protein through electrostatic interactions with its N-terminal part.[17] Other Tat domains can undergo conformational changes as well~~-~~; the prime example is Tat-positive transcription elongation factor (P-TEFb) complex. X-~~Ray~~ray crystallography showed that the first three N-~~terminus~~terminal Tat domains (1-49 residues) sustain extended conformation mostly through interactions with cyclin T1 whereas 50-86 residues are not defined.[11] Nuclear magnetic resonance ~~(NMR) experiments~~ opened the veil on the structural propensity of Tat protein suggesting that the cysteine-rich region tends to fold into α-helixes ~~while~~in contrast to the basic ~~and RGB domains - to~~ domain with extended or β-sheet ~~conformations.~~~~[17]~~conformation.[18] Comparison analysis of X-~~Ray~~ray and ~~NMR~~nuclear magnetic resonance studies suggests that different fragments of Tat protein can employ different folding mechanisms.~~[17]~~[18] This flexibility enables Tat to adopt diverse

Formatted: Header
Formatted: Right: 0 cm
Formatted: Font: 11 pt, Russian
Formatted: Indent: Left: 0 cm, Right: 0 cm
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian, Not Superscript/ Subscript
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian, Superscript
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted ...
Formatted: Font: 11 pt, Russian

conformations upon interaction with its physiological partners, thus greatly extending its multifunctionality.

One of the most important and well-studied domains of Tat protein is the basic domain. While Tat can tolerate up to 40% of sequence mutations without significant changes in its activity, its basic domain is highly conserved among Tat variants.6 It is enriched with positively-charged arginine and lysine residues comprising 49RKKRRQRRR57 motif. Basic domain confers many properties to Tat such at regulation of viral transcription and manipulation of cellular processes in favor of HIV. ~~Therefore in different studies it is referred to as the RNA-binding domain (ARM), the nuclear localization signal (NLS), and the protein transduction domain (PTD).~~ In this review, we discuss different aspects of Tat basic domain and its versatile role in the interaction between the host cell and the virus.

2 | BASIC DOMAIN FUNCTIONS AS AN RNA BINDING MOTIF

Absence of Tat causes predominantly short ~~transcripts~~transcript production from the HIV-1 long terminal repeat (LTR).[~~18~~19] Tat interacts with an RNA enhancer element (TAR) positioned at the 5' end of the viral ~~transcriptvia~~transcript via its basic domain ~~and is termed the arginine-rich RNA-binding motif (ARM) due~~leading to ~~its binding capacity to the bulge~~facilitation of ~~the stem-loop of TAR RNA, thus facilitating the~~ viral transcription.[~~19-21,22,23~~20-24] Tat interacts with ~~the positive transcription elongation factor (~~P-TEFb~~)~~ comprised of ~~CycT1~~cyclin T1 and CDK9 and mediates ternary Tat-P-TEFb complex to TAR RNA.11,[24-27] ~~Via its basic domain,~~[25-28] Tat binds the nascent

Formatted: Header

Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian, Not Superscript/ Subscript

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian, Not Superscript/ Subscript

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian, Not Superscript/ Subscript

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Font: 11 pt, Russian

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian, Not Superscript/ Subscript

Formatted: Font: 11 pt, Russian

RNA via its basic domain, causing conformational changes of the P-TEFb complex, thus enabling CDK9 to phosphorylate RNA polymerase II resulting in the full-length HIV-1 transcript production.[~~24-26~~25-27] Nullbasic Tat mutant, in which the entire basic domain was replaced with glycine/alanine residues, was shown to interact with P-TEFb complex but failed to recruit the ternary complex to the nascent viral RNA confirming that the basic domain plays a pivotal role in ~~Tat's~~the transactivation function~~.~~[~~28~~] of Tat.[29]

Numerous attempts had been made to identify the key residues responsible for TAR RNA binding, though with some discrepancies. ~~Early~~An early study by the Calnan group showed that peptides derived from basic domain of Tat ~~ARM~~protein could directly bind TAR RNA while their amino acid sequence could be completely rearranged and still retained its high affinity to TAR.[~~21~~22] Mutations of arginines to alanines significantly reduced the binding capacity while substitution to lysines restored the capacity to nearly wild-type levels, proposing that the overall charge of the basic domain is likely the key factor of Tat-TAR RNA binding.[~~21~~22] Substitution of arginine residues for glutamine at 52 and 53 positions led to abrogation of transactivation activity.[~~23~~24] In another experiment, K50 interacted with G34 of TAR RNA loop, indicating that the protein-RNA cross-link occurred at K50 position whereas mutation ~~of~~at G34 to ~~alanine~~U34 significantly reduced Tat-~~CycT1~~Cyclin T1 binding capacity.[~~29,~~30,31] Mutations of the first two lysines at 50 and 51 positions to serine and glycine (K50S and K51G), respectively, showed a decrease in Tat activity by 50% in vivo.[~~31~~32] Interesting results assessing transactivation of HIV LTR came from molecular dynamics simulations combined with in vitro experiments by ~~Pantano~~the Carloni group. While K50A and K51A mutants were functionally defective for HIV LTR transactivation, K50R and K51R had a functional transactivation capacity although it was lower than that of the wild-type

Formatted: Header
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, English (U.K.)
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Border: : (No border), Pattern: Clear (White)
Formatted: Border: : (No border), Pattern: Clear (White)
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian, Not Superscript/ Subscript
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Not Italic, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian, Not Superscript/ Subscript
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Not Italic, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian

Tat. Still, the K50A and K51A mutants localized to the nucleus, suggesting that these mutations most likely affected Tat interactions with RNAs or nuclear protein complexes.[~~32~~33]

In addition to viral RNA, Tat is believed to interact with cellular RNAs. ~~Ability~~The ability of Tat basic domain to associate with human RNAs was examined via immunoprecipitation analysis of the wild-type Tat and its mutated form, K50S-K51G.[~~31~~] ~~Mutant~~[32] The mutant form showed a significant decrease in interaction with RNA, in particular Tat-bound ~~RNAs~~ FADD and TNFRSF8 RNAs, leading to speculations that, in addition to TAR RNA, Tat was able to associate with the specific set of human RNAs for which an intact ~~RNA-binding motif~~basic domain was required.[~~31~~32] Moreover, Tat basic domain was proposed to specifically target Dicer-dependent RNAi ~~[33],~~[34] the innate immune response against the viral infection.[~~34-36~~35-37] Besides mammalian RNAs, Tat was shown to impair pre-rRNA processing in *Drosophila melanogaster* cells via association with U3 snoRNA and fibrillarin (nucleolar components necessary for pre-rRNA processing), although it was not explicitly stated that the basic domain was specifically involved in such interactions.[38]

3 | INTERACTION WITH CELLULAR COMPONENTS

Besides its capacity to bind viral and cellular RNAs, ~~Tat's~~the basic domain of Tat also associates with cellular and viral proteins. Tat peptide comprising 48-60 amino acids was shown to block protein kinase C ~~(PKC)~~ activity by binding to the kinase active site.[~~37~~] ~~Nucleophosmin~~[39] NPM1 (B23 or ~~NPM1~~nucleophosmin), a ubiquitous protein

Formatted: Header
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Not Bold, English (U.S.)
Formatted: Font: 11 pt, Russian
Formatted: Indent: Left: 0 cm, Right: 0 cm
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian

involved in diverse cellular processes, has been proposed to directly interact with Tat *via* its basic domain.[~~38~~40] Tat basic domain has been also implemented in recruitment of CIS protein (cytokine-inducible SH2 containing protein) to CD127 surface receptor of CD8 T cells for internalization and subsequent degradation of CD127 which led to reduction in T cells.[~~39,40~~41,42] The basic domain along with a conserved tryptophan residue W11 are responsible for unconventional secretion of Tat from cells.[12] The basic domain binds phosphatidylinositol-4,5-biphosphate (a phospholipid of the inner leaflet of the plasma membrane) causing conformational changes which enable insertion of W11 into the membrane with the subsequent secretion of Tat.[12,43] Arginine residues from Tat basic domain ~~were found to be responsible for targeting~~target Tat to cell membrane lipid rafts ~~(LRs)~~ and ~~Tat-mediated enhancement of~~enhance fibroblast growth factor-2 (FGF-2) signaling in human podocytes isolated from children with HIV-associated nephropathy (HIVAN), whereas alanine substitutions abrogated Tat nuclear localization, association with ~~LRs~~lipid rafts, and enhancement of FGF-2 signaling.[~~41~~44] Tat, via its basic domain, binds Tip60, a cellular histone~~-~~ acetyltransferase (HAT) which controls expression of cellular genes capable to interfere with the efficient viral replication and propagation.[~~42~~45] Additionally, histone~~-~~ chaperone hNAP-1 ~~has been shown to bind~~binds Tat basic domain, stimulating regulation of Tat-mediated viral transcription.[~~43~~46] Tat basic domain has been recently observed to interfere with the host cell proliferation and induction of apoptosis of HIV-1-infected lymphocytes. In Jurkat cells, Tat ~~inhibits α-tubulin~~disrupts localization of PRS3, which in ~~Jurkat cells~~association with α-tubulin plays a critical role in mitosis, leading to faulty mitotic spindle and ~~chromosomes~~chromosome formation.[47] The basic domain of Tat associates with IkB-α (an inhibitor of nuclear factor NF-

Formatted: Header
Formatted: Font: 11 pt, Italic, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted: Font: 11 pt, Russian, Superscript
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Not Italic, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian

kB), leading to liberation of p65 from IkB-α/p65 complex and ~~reduction through binding to ribosomal protein PRS3, which plays a critical role in mitosis via association with α-tubulin.~~[~~44~~]the subsequent transcriptional activation of pro-inflammatory genes.[48] These are just ~~some~~several examples of the interaction of Tat basic domain with cellular proteins. Many other cellular partners of Tat have been recently discovered.[~~45~~49] Yet, whether the Tat basic domain was involved in these ~~interaction~~interactions remains to be elucidated.

4 | BASIC DOMAIN FUNCTIONS AS A NUCLEAR LOCALIZATION SIGNAL

The primary role of Tat lies in activation of viral transcription, hence Tat must be able to pass the large nuclear~~-~~ pore ~~complex (NPC)~~complexes of the nuclear envelope. Globular proteins ~~smaller~~with a Mr less than 40~~-~~,000 to 60 ~~kDa~~,000 or 5-10 nm in diameter can freely diffuse between the cytoplasm and the nucleus whereas larger macromolecules ~~exceeding the NPC size limit~~ require an energy-driven mechanism to traverse the ~~NPC.~~[~~46-48~~]nuclear pore complexes.[50-52] In most cases, proteins targeted to the nucleus contain specific trafficking motifs such as the nuclear localization (NLS) and nuclear export signals (NES). Since the discovery of the first NLS signal in the SV40 large T-antigen protein containing a short stretch of basic amino acids, similar target sequences have been identified and characterized in a broad range of viral and cellular proteins. While the classical (or canonical) NLS pathway employs an adaptor molecule importin-α for binding to importin~~-β,~~-β, a non-classical NLS pathway involves direct binding of the cargo protein to the importin~~-β.~~-β. In both pathways, importin~~-β~~-β

Formatted: Header

Formatted: Font: 11 pt, Russian

Formatted

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted

Formatted

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Font: 11 pt, Russian

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

acts as a carrier by docking cargo-importin(s) complex to the ~~NPC~~nuclear pore complexed and releasing the cargo into the nucleus upon binding to Ran-GTP.[~~49-52~~53-56]

Passive diffusion has been suggested to be a major mechanism of Tat nuclear entry.[~~53~~57] However, despite ~~the~~ its small size (Mr 14,000-16 ~~kDa) for~~,000) favoring passive diffusion, Tat contains a functional NLS (49RKKRRQRRR57) within its basic domain and ~~has been~~was shown to localize preferentially in the nucleoplasm and nucleolus.[4,~~54,55~~58,59] Classical and non-classical mechanisms of nuclear entry along with association with ~~cellular~~nuclear components had been previously proposed for Tat protein. In vitro assays suggested that Tat nuclear import ~~is~~was mediated by the direct binding of its basic domain to importin-~~β,~~-β, thus competing with importin-α for the same binding site of importin-~~β.~~[~~56~~]-β.[60] In contrast, a novel mechanism independent of ~~importins~~the importin pathway was proposed by another group, indicating the ability of Tat basic domain to interact with nuclear components.[~~57~~61] Deletion of the basic domain led to cytoplasmic ~~localization~~[~~58~~]localization[62] and a dramatic decrease in Tat activity.4 ~~Mutational~~Mutation analysis of amino acids 50, 55 and 56 replaced by uncharged residues~~,~~ revealed a loss in nuclear localization, suggesting the presence of two partially overlapping or juxtaposed NLSs. In other words, mutations in the RKKRR motif or RRR alone had little effect on nuclear localization while mutations in both parts led to the cytoplasmic accumulation.[~~59~~63] Additionally, the first set of basic domain amino acids was shown to function as NLS while the remaining RRR motif tended to bind to intracellular components.[~~53,60~~57,64] Using oriented peptide binding approach, it was proposed that KKKRR, KKKRK, and KKRKK motifs are sufficient for binding importin-α.[~~61~~] ~~Crystal structure~~[65] Structure of ~~Tat:NLS/CPP~~importin-α with

Formatted: Header

Formatted

Formatted

^{48}GRKKRRQRRRAPQN^{61} peptide has been recently determined~~.~~ sing X-ray crystallography. It was shown that ~~^{48}GRKKRR^{53}~~^{48}GRKKRRQR^{55} residues mediate a strong association with importin-α.[~~61,62~~66] Taken together, these results suggest that Tat may utilize different pathways to enter the nucleus. Ability to combine both classical and non-classical NLSs has been previously demonstrated for the latency-associated nuclear antigen (LANA) of KSHV.[~~63~~67] Another example is Rev protein, whose basic domain is enriched with positively-charged arginine residues similar to Tat. Depending on the cell type, Rev has been shown to utilize different mutually-exclusive pathways of nuclear entry, either through importin-~~β~~-β or transportin.[~~64~~68]

5 | BASIC DOMAIN FUNCTIONS AS A NUCLEOLAR LOCALIZATION SIGNAL

Besides its predominant accumulation in the nucleus of different cell lines, Tat is observed to localize to the nucleoli as well. Tat was proposed to possess a nucleolar localization signal (NoLS~~),~~) 48GRKKRRQRRRAP57 which drives Tat accumulation into nucleoli and encompasses Tat basic domain which serves as NLS.[~~65~~69] However, experimental data showed that only positively-charged amino acids are involved in Tat nucleolar localization while flanking non-charged residues hardly exhibit any effect on its localization; ~~thus implying~~this implies that charge-dependent mechanism is a major force of dynamic accumulation of Tat in the nucleolus.[~~66~~70] Indeed, the nucleolus is a highly~~-~~ dynamic structure and accumulation of proteins can be achieved through interactions with nucleolar building blocks, such as rDNA or its transcripts.[~~67~~71] In vitro experiments revealed binding of GST-Tat to NPM1 ~~(B23)~~, a major nucleolar protein, emphasizing that such interactions occur within Tat basic domain.[~~38~~40] Additionally, NPM1 has been suggested to be vital for ~~Tat's~~the nuclear entry of Tat and the subsequent nucleolar localization ~~- in~~. In this case, NPM1 behaves as a shuttling protein driving Tat through the nuclear pore ~~complex~~complexes to the nucleoli.[~~38~~40] Such shuttling mechanism by NPM1 has been previously observed for NCL (C23 or nucleolin ~~(C23~~), another major nucleolar protein lacking defined NoLS.[~~68~~72] Apart from these two studies, there is no other experimental data exist confirming NPM1/Tat interactions in vivo. Alternatively, based on nucleolar interactome analysis, Tat itself can physically modulate ~~proteins~~protein nucleolar accumulation involved in HIV-1 pathogenesis; but whether Tat basic domain is directly involved in targeting of these proteins to nucleolus remains to be elucidated.[69–73] NoLS was also suggested to be a key player in targeting Tat to the nucleoli in *Drosophila melanogaster* cells, thus following localization pattern of mammalian cells[38].

It might be possible that Tat basic domain serves as a NoLS in the first place rather than NLS due to its main function in the regulation of viral transcription and the small size enabling passive diffusion. Nucleoli are one of the main targets of some viral proteins including Tat. Presence of independent sites for nuclear entry and nucleolar accumulation might be evolutionarily unfavorable for Tat. Indeed, HIV-1 genome itself is just under 10 kb with 16 proteins successfully serving its biology.[~~70~~] ~~In addition, Tat lacks a~~[74] Both nuclear ~~export signal even though over the course of HIV biology around 2/3 of produced Tat is released into the extracellular milieu.~~[~~71~~] ~~Such processes~~and nucleolar accumulation could be achieved ~~via Tat's~~ by Tat binding to/~~associations~~associating with other nuclear/nucleolar proteins or RNAs through its highly charged basic domain. It is thus its overall structural plasticity coupled with highly charged basic domain make it the prime example of "minimum complexity - maximum efficiency".

6 | BASIC DOMAIN FUNCTIONS AS A PROTEIN TRANSDUCTION DOMAIN

Tat is actively secreted by infected cells and can traverse plasma membranes of various eukaryotic cells,[~~72-74~~75-77] affecting their gene expression and cellular functions. Concentration of extracellular Tat can reach up to 40 ng/ml in blood of acutely infected patients.[~~75~~78] Tat has been detected in sera of patients undergoing antiretroviral therapy, showing that modern anti-HIV drugs are not able to block the cellular release of Tat protein.[~~76~~79] The ability to penetrate cell plasma membranes has been attributed to Tat basic domain thus termed as protein transduction domain (PTD~~).~~) or cell penetrating peptide (CPP) if used as a peptide. Its potency to mediate cellular uptake has been widely exploited by a large number of laboratories for transcellular protein transduction not only in mammalian cells,[~~77~~80] but in plants as well,[~~76,78~~81] thus making Tat basic domain a promising tool for transcellular drug delivery to a wide variety of cells.

Various experiments with short peptides spanning the Tat basic domain fused to different cargoes showed the ability of these peptides to enter ~~the~~ cells,[~~79-81~~82-84] while peptides with a truncated or mutated basic domain failed to translocate through cell membranes.[~~80,82~~83,85] One of the first mechanisms of cell penetration proposed that ionic interactions between ~~Tat's~~the highly dense positively-charged basic domain of Tat and negatively-charged phospholipids of the plasma membrane prompted an invagination of the membrane.[~~80~~83] In contrast, two other studies suggested an adsorptive-mediated

Formatted: Header
Formatted: Font: Bold
Formatted: Font: 11 pt, Not Bold, Russian
Formatted: Font: 11 pt, Russian
Formatted: Indent: Left: 0 cm, Right: 0 cm
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: English (U.K.)
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian

endocytosis as a way for internalization.[~~73,83~~76,86] Further ~~works~~work performed by several research groups demonstrated that Tat basic domain fused to different cargoes could bind heparin,[~~84-86~~87-89] a structural homolog of heparan sulfate glycosaminoglycan ~~(GAG)~~, abundantly present on the cell surface. Heparan sulfate proteoglycans (~~HSPG~~HSPGs) have been proposed to be highly versatile receptors responsible for the mechanism of ~~the~~ cellular entry.[~~87~~90] It has been shown that Tat internalization can be achieved via HSPG binding ~~[87-90]~~ following subsequent active caveolar endocytosis through cell membrane lipid rafts ~~(LRs).~~[~~87,88,90~~ 90,92]. Full-length Tat protein can utilize HSPG receptors whereas unconjugated Tat peptides can be internalized by cells that lack these receptors. ~~This observation demonstrated that the~~Alternatively, the study performed on T cells demonstrated that a full-length Tat can use a clathrin/AP-2-dependent endocytosis; however, whether the basic domain plays any roles in this pathway remains unclear.[93] These results demonstrate that different internalization pathways can be employed, depending on the cell type and specificity of the cargo.[~~91~~94]

7 | POST-TRANSLATIONAL MODIFICATIONS OF THE BASIC DOMAIN REGULATE TAT ACTIVITY

Post-translational modifications play a key role in the heterogeneity of protein functions. Disordered protein domains have been suggested to be a subject for ~~much of~~many post-translational modifications.[13] ~~Disordered~~The disordered state is advantageous in that in can provide greater accessibility to the sites for post-translational modifications.[13] ~~Indeed, being an intrinsically~~

~~disordered, Tat with its basic domain is able to bind or associate with a wide array of physiological partners and to regulate their functional activities.~~

Post-translational modifications include acetylation, methylation, phosphorylation, to name but a few. Each modification involves specific enzymes that recognize distinct amino acids within the polypeptide chain. Tat ~~basic domain~~ undergoes post-translational modifications which greatly expands its functions. The basic domain of Tat undergoes acetylation and methylation, affecting its capacity to facilitate viral transcription and modulate a broad range of cellular processes. ~~The most common and increasingly studied post-translational modifications favoring diversification of Tat's basic domain functions are acetylation and methylation.~~

7.1 | ACETYLATION

Acetylation is mediated by ~~histone acetyltransferases (~~HATs~~)~~ catalyzing the transfer of acetyl groups from acetyl coenzyme A to the ϵ-amino group of lysine. It has been speculated that HIV-1 transcription involves at least two phases. Defined as early TAR-dependent and late TAR-independent, these phases are equally important for ~~Tat's~~the functions of Tat whereas Tat acetylation has been proposed to act as a regulatory switch between them.~~[92,93]~~[95,96] Tat transactivation activity has been shown to depend upon lysine acetylation at K50 and K51.~~[92,94-96]~~[98,99-102] The critical role in K50 acetylation has been attributed to p300,~~[95,97]~~[98,100] a HAT responsible for regulation of gene expressions via chromatin remodeling. K50 acetylation leads to liberation of Tat from TAR RNA and ~~CycT1~~cyclin T1 and activation of Tat-mediated transcriptional elongation of HIV-1 through binding to ~~RNAP~~RNA polymerase II.~~[95,97,]~~[98][,100,101] Further direct

Formatted: Header

Formatted: Font: 11 pt, Russian

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Not Bold, English (U.S.)

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian, Pattern: Clear (White), Not Highlight

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted

Formatted

Formatted: Font: 11 pt, Russian

Formatted

Formatted

Formatted: Font: 11 pt, Russian

Formatted

Formatted

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted

Formatted

Formatted

Formatted: Font: 11 pt, Russian

binding of acetylated K50 to the bromodomain of PCAF, a p300/CBP-associated factor, has been proposed to be essential for Tat transactivation since the site-directed mutation of K50A led to termination of Tat transactivation activity whereas substitution to arginine K50R did not affect the interaction with TAR, ~~CycT1~~cyclin T1 or PCAF (Fig. 2).[~~99-~~102-105]

Acetylation of K50 has been further shown to facilitate Tat interaction with SWI/SNF chromatin remodeling complex~~,~~ containing BRG-1~~,~~ and its subsequent recruitment to the viral LTR.[~~103-105~~106-108] This interaction is achieved via direct binding of acetylated Tat to the bromodomain of BRG-1, permitting SWI/SNF to alter the structure of downstream nucleosomes and enabling further viral transcription.[~~105~~108] In addition to p300, the hGCN5 HAT ~~has been demonstrated to acetylate~~acetylates both K50 and K51 residues in vitro and significantly ~~enhance~~enhances Tat-mediated transcription of HIV LTR.[~~106~~-109] Indeed, mutational analysis aiming to neutralize the electrical charge of lysine and block acetylation by substitution with the alanine residue demonstrated that acetylation of K50 exclusively regulates Tat transactivation activity (Fig. 2).[~~107~~110]

Proteomic analysis and in vivo experiments showed that Tat acetylated at K50 and K51 residues preferentially binds p32, an inhibitor of splicing factor ASF/SF-2, and mediates its transport to the viral promoter, thus regulating the splicing pattern of HIV-1.[~~108~~-111] Acetylation also decreases cellular uptake of Tat-derived peptides acetylated at either K50 or K51 positions.[~~119~~112] Dysregulation of expression of C5, APBA1, BDNF, and CRLF2 genes associated with inflammation and damage by the K50A mutant has been recently identified in human macrophages.[~~110~~113]

Formatted: Header
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Not Italic, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted: Font: 11 pt, Russian

Acetylation of Tat can be reversed by sirtuin 1 (SIRT1) class III deacetylase. Acetylation and deacetylation cycles are believed to be necessary for the pursuit of ~~the~~ complete HIV transcription (Fig. 2).[~~98~~101] In particular, acetylation results in the release of Tat from TAR leading to translocation of Tat and chromatin-modifying transcriptional coactivators to elongating RNA polymerase and recruitment of chromatin-remodeling complexes while deacetylation by SIRT1 restores Tat basic domain to its initial form so that Tat can interact with P-TEFb and bind TAR RNA, leading to the new transcriptional cycle.[~~98,111~~101,114]

7.2 | METHYLATION

In addition to acetylation, Tat basic domain ~~has been shown to undergo~~undergoes methylation, a process in which methyl groups from S-adenosylmethionine are transferred to proteins regulating many protein functions. Unlike acetylation, methylation does not result in neutralization of residual electrical charge. Because methylation primarily affects arginines and lysines, Tat basic domain serves a prime substrate for post-translational modifications by different methyltransferases.

Methylation of position 50 and 51 lysines of the Tat basic domain can be generally accomplished by the action of SETDB1 and Set7/9-KMT7 methyltransferases (Fig. 3). Both are the members of a broad SET protein family that have been originally shown to specifically methylate lysines of histone H3 tail.[~~112,113~~115,116] Depending on the position of histone H3 lysine, methylation can prompt either transcriptionally active or transcriptionally repressed state of chromatin.[~~114-116~~117-

Formatted: Header

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted

Formatted

Formatted

[119] Later, nonhistone proteins methylated by SETDB1 and Set7/9-KMT7 were discovered,[~~112,117-126~~115,120-129] one of them is Tat protein.[~~127,128~~130,131]

In vitro experiments, using SETDB1, wild-type Tat, and variation of Tat-derived mutant peptides have determined that both K50 and K51 can be methylated by SETDB1, whereas K51 showed an increased amount of methyl groups ~~[128].~~.[131] SETDB1 knockdown resulted in the increase of viral transcription, implying that methylation of K50 and/or K51 can attenuate ~~the~~ HIV-1 ~~transcription~~[~~128~~]transcription[131] similarly to the general assumption that methylation of histone H3 lysine 9 corresponds to the assembly of transcriptionally-silent heterochromatin. ~~The~~ Set7/9-KMT7 ~~lysine methyltransferase,~~ has been also found to monomethylate K51 residue both in vitro and in vivo.[~~127~~130] However, contrary to SETDB1, this monomethylation increased Tat affinity to TAR and Tat transcriptional activity.[~~127~~130] Interestingly, Set7/9 has been identified as one of the methyltransferases that drives monomethylation of histone H3 lysine 4, initiating formation of transcriptionally-active euchromatin.[~~129-131~~132-134] Such strikingly different outcomes in methylation of Tat lysines could favor the notion that methylation can influence HIV-1 ~~biology~~[~~132~~]biology[135] similar to histone methylation that is linked to both transcriptional activation and repression (Fig. 3).[~~133~~136]

Lysines can be demethylated by LSD1 ~~demethylase~~[~~134~~]demethylase[137] which in complexes with different cofactors ~~demethylate~~demethylates histone H3K4/9 [~~133,135~~ 136,138] and nonhistone proteins such as p53,[~~136~~139] E2F1,[~~137~~140] and Tat.[~~138~~141] In vitro experiments demonstrated that Tat K51 residue ~~has been~~was specifically targeted by LSD1 ~~demethyltransferase~~, resulting in HIV-1 transactivation.[~~138~~141]

Besides lysine, Tat arginine residues can be methylated by arginine methyltransferase PRMT6. Early experiments by Boulanger and colleagues demonstrated that HA-tagged Tat expressed in HEK293T cell line was subjected to methylation by endogenous PRMT6; cotransfection with PRMT6 increased the level of Tat methylation.[142] Knockdown of PRMT6 led to an ~~increased level of~~increase in HIV-1 production, demonstrating that methylation of the basic domain ~~exerts~~exerted a negative effect on Tat transactivation function.[~~139~~142] Further experiments ~~by~~ determined position 52 and 53 arginine residues (R52 and R53) to be specifically methylated by PRMT6.[~~140~~143] Consistent with the previous observation, *in vitro* methylation of R52 and R53 triggered a decrease in Tat interaction with TAR RNA and complex formation with ~~CycT1~~cyclinT1, thus affecting Tat function, whereas *in vivo* experiments augmented ~~Tat's~~the role of Tat in transactivation by downregulating PRMT6.[~~140~~143] PRMT6 has been proposed to be a restriction factor of HIV-~~1~~[~~139-141~~]1[142-144] as an innate cellular response to the viral replication.[~~140~~143] Yet, this restrictive effect is counterbalanced by recently observed downregulation of PRMT6 expression in CD4+ T cell of HIV-infected patients, suggesting that ~~indeed~~the virus can indeed control expression of cellular genes to benefit its replication (Fig. 3).[~~142~~145]

While methylation of Tat is generally linked to the attenuation of Tat transactivation activity, the fate of the methylated Tat remains largely obscure. Compelling results came from the study where instead of ~~downregulation of~~downregulating Tat transactivation function ~~the~~, overexpression of PRMT6 led to increased Tat

Formatted: Header
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt, Russian
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted: Font: 11 pt, Not Italic, Russian
Formatted
Formatted
Formatted: Font: 11 pt, Not Italic, Russian
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Right: 0 cm
Formatted: Font: 11 pt, Russian
Formatted: Indent: Left: 0 cm, Right: 0 cm
Formatted
Formatted
Formatted

stability by protection from proteasome-dependent degradation.[~~143~~146] This strikingly different outcome has been explained as a way to fulfill multifunctional role of Tat apart from ~~the~~its transactivation function, while methylation serves as a molecular switch between Tat functions.[~~143~~146] Further investigation by the same group demonstrated that methylation of R52 and R53 residues of Tat basic domain ~~results~~resulted in the exclusion of Tat-GFP fusion protein from the nucleolus of COS cells; thus ~~can~~it could also modulate Tat localization.[~~144~~147]

Methylation is a complex post-translational modification with pleiotropic effects on protein functions. Arginine contains three nitrogen atoms (one ε and two η) in its side chain each of which can be monomethylated ~~(MMA),~~, symmetrically dimethylated ~~(SDMA)~~ or asymmetrically dimethylated ~~(ADMA).~~. Depending on the state of methylation, thermodynamics of Tat-derived peptides binding capacity to TAR RNA has been recently assayed. Monomethylation of arginine R52 or R53 at ε-nitrogen atom enhanced binding affinity whereas monomethylation or asymmetric dimethylation at η-nitrogen resulted in reduced binding capacity.[~~145~~148] Methylation ~~has been~~was further suggested to influence Tat-TAR RNA interaction in a position- and state-dependent manner. Asymmetric dimethylation of R52 or R53 severely affected Tat-TAR RNA binding while dimethylation of flanking arginines, such as R49 or R57, slightly increased Tat-TAR RNA affinity.[~~146~~149]

8 | CONCLUSIONS

Formatted: Header
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted
Formatted: Font: 11 pt
Formatted
Formatted: Font: 11 pt
Formatted
Formatted
Formatted
Formatted
Formatted: Font: 11 pt, Russian
Formatted: Font: 11 pt
Formatted: Font: 11 pt, Russian
Formatted
Formatted: Font: 11 pt
Formatted
Formatted
Formatted
Formatted: Font: 11 pt, Not Bold, Russian
Formatted: Font: Not Bold
Formatted: Indent: Left: 0 cm, Right: 0 cm
Formatted: Font: 11 pt, Not Bold, Russian

Formatted: Header

Tat is a regulatory protein encoded by the HIV-1 viral genome; it plays a crucial role in regulation of viral and host gene expression. Tat can exhibit multiple functions which are required for viral pathogenesis. It can also enter uninfected ~~cell~~cells and modulate cellular gene expression according to the viral needs, thus leading to oncogenesis or cellular death through apoptosis. Many functions of Tat are attributed to its basic domain (also designated as arginine-rich motif~~,~~ RNA-binding domain, nuclear~~/~~ localization signal, nucleolar localization signal, and protein transduction domain), which is highly conserved among different Tat variants though Tat itself is prone to mutations. This multifunctionality of the Tat basic domain is linked to its high charge and flexible structure. These can lead to interactions with many physiological partners including glycoproteins, proteins or protein/RNA complexes, chromatin of both viral and cellular origin, thus allowing Tat to accomplish various tasks. Tat basic domain is also subjected to post-translational modifications which may ~~expands~~expand and modify its functionality. All these features make the basic domain the key component of Tat protein. ~~We~~In this review, we have summarized the current knowledge on Tat basic domain and its role in Tat functions, but most probably new functions of this viral "Swiss army knife" will be discovered in the near future.

Formatted: Font: 11 pt, Russian

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

Formatted: Font: 11 pt, Russian

ACKNOWLEDGEMENTS

Formatted: Font: 11 pt, Not Bold, Russian

The work was supported by the Russian Science Foundation (grant 17-75-20199) to Y.R.M., by the Russian Foundation for Basic Research (18-54-16002) to E.V.S. and by the Plan Cancer (ENVIBURKITT), ANRS and la Ligue Contre le Cancer to Y.~~V. The work of Y.R.M. was partially conducted in the frame of the Koltzov Institute of~~

Formatted: Font: 11 pt, Russian

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted

Formatted: Font: 11 pt, Russian

~~Developmental Biology government program of basic research № 0108-2018-0004.~~ S.V.

~~CONFLICT OF~~

FINANCIAL AND COMPETING INTEREST DISCLOSURE

~~The authors declare no conflict of interest.~~None declared.

ORCID

Eugene V. Sheval http://orcid.org/0000-0003-1687-1321

Yegor S. Vassetzky https://orcid.org/0000-0003-3101-7043

Formatted: Header

Formatted: Font: 11 pt, Russian, Not Highlight

Formatted: Font: 11 pt, Not Bold, Russian, Pattern: Clear (White), Not Highlight

Formatted: Font: 11 pt, Not Bold, Russian, Pattern: Clear (White), Not Highlight

Formatted: Indent: Left: 0 cm, Right: 0 cm

Formatted: Font: 11 pt, English (U.S.)

Figure legends

~~Figure~~FIGURE 1~~.~~ Structure of the Tat protein. Tat contains the following domains: N-~~terminus,~~terminal acidic domain (1-21 aa), cysteine-rich~~,~~ domain (22-37 aa), hydrophobic core~~,~~ domain (38-47 aa), basic~~,~~ domain (48-59 aa), glutamine-rich~~,~~ ~~RGD motifs~~ domain (60-72), and domain encoded by the second exon.

~~Figure~~FIGURE 2~~.~~ Post-translational modifications of the Tat basic domain by different acetyltransferases. Acetylation of K50 by P300 and pGCN5 leads to association with RNA polymerase II ~~(RNAPII),~~ bromodomain of PCAF, and with SWI/SNF chromatin remodelling complex resulting in liberation of Tat protein from TAR and its subsequent translocation along with the chromatin-remodelling complexes to elongating RNA polymerase II. Deacetylation by SIRTI enables Tat to return to its non-acetylated form and begin a new transactivation cycle.

~~Figure~~FIGURE 3~~.~~ Methylation of Tat basic domain. SETDBI methyltransferase methylates K50 and K51 residues leading to a decrease in transactivation of transcription. Contrary to SETDBI, methylation of the same residues by SET7/9 methyltransferase prompts an increased affinity to TAR RNA resulting in upregulation of transactivation of transcription. PRMT6 (an important host factor) methylates R52 and R53 residues liberating Tat basic domain from the TAR RNA and leading to attenuation of Tat transcriptional activity. PRMT6 modification also allows Tat protein to escape from proteasome-dependent degradation and increases Tat stability.

Formatted: Header

Formatted: English (U.S.)

Formatted: Font: 11 pt, Not Bold, English (U.K.)

Formatted: Indent: Left: 0 cm, First line: 0 cm, Right: 0 cm, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No

Formatted: Font: 11 pt, Not Bold, English (U.S.)

Formatted: Font: 11 pt, English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: Font: 11 pt, English (U.S.)

Formatted: Font: 11 pt, Not Bold, English (U.S.)

Formatted: Font: 11 pt, English (U.S.)

Formatted: Font: 11 pt, English (U.S.)

Formatted: Font: 11 pt, Not Bold, English (U.S.)

Formatted: Font: 11 pt, English (U.S.)

Formatted: English (U.S.)

Formatted: Indent: Left: 0 cm, Right: 0 cm, Border: Top: (No border), Bottom: (No border), Left: (No border), Right: (No border), Between : (No border)