Methods 53 (2011) 78–84

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Formation of Tat–TAR containing ribonucleoprotein complexes for biochemical and structural analyses

Janna M. Bigalke [a], Nadine Czudnochowski [a], Friederike Vollmuth [a], Karin Vogel-Bachmayr [a], Kanchan Anand [b], Matthias Geyer [a,*]

[a] Max-Planck-Institut für molekulare Physiologie, Abteilung Physikalische Biochemie, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany
[b] EMBL Heidelberg, Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany

ARTICLE INFO

Article history:
Accepted 6 April 2010
Available online 10 April 2010

Keywords:
HIV-1 Tat
TAR
P-TEFb
Cyclin T1
Transcription elongation

ABSTRACT

Viruses manipulate multiple processes of the host cell machinery in order to replicate successfully in the infected cell. Among these, stimulation of transcription of the viral genes is crucial for lentiviruses such as HIV for increased protein expression levels and generation of escape mutants. The transactivation response (TAR) element at the 5′-end of HIV, SIV, BIV, EIAV or JDV retroviruses forms a unique RNA based promoter element that together with the transcription activator protein Tat stimulates viral gene expression at the level of transcription elongation. TAR is a double stranded non-coding RNA of typically 24–40 nucleotides length. Together with Tat it interacts with the Cyclin T subunit of the positive transcription elongation factor P-TEFb to recruit Cyclin T and its corresponding Cyclin-dependent kinase Cdk9 to the RNA polymerase II. *In vitro* formations of these Tat–TAR containing ribonucleoprotein complexes are a key requisite for biochemical characterizations and interaction studies that eventually will allow structural analyses. Here, we describe purification methods of the different factors employed and chromatography techniques that yield highly specific complex assemblies suitable for crystallization.

1. Introduction

RNA–protein interactions are vital for many cellular regulation processes, including the life cycle of RNA viruses in an infected cell. Human immunodeficiency virus (HIV) encodes a transcriptional activator protein, Tat, that interacts with the viral RNA to stimulate transcription. Tat is expressed early in the viral life cycle and is essential for viral gene expression, replication and pathogenesis (reviewed in Refs. [1,2]). Tat enhances the processivity of RNA polymerase II (RNA pol II) elongation complexes that initiate transcription in the HIV-1 long terminal repeat (LTR) region. To accomplish this, Tat binds the transactivation response (TAR) RNA element that is located at the 5′-end of all nascent HIV-1 transcripts [3]. The Tat–TAR interaction provokes the recruitment of the positive transcription elongation factor (P-TEFb) to the paused RNA pol II. Tat directly interacts with the Cyclin T subunit of P-TEFb [4], which consists of Cyclin T1 (CycT1), or its minor forms T2 or K, and a Cyclin-dependent kinase, Cdk9 [5,6]. P-TEFb phosphorylates the carboxy-terminal domain (CTD) of RNA pol II at serine 2 of the heptad-peptide repeats, which marks the transition from transcription initiation to transcription elongation [7]. A model of Tat–TAR mediated transactivation of viral genes by the recruitment of P-TEFb to RNA pol II is shown in Fig. 1.

Despite its small size, several functional regions have been assigned to Tat that are conserved throughout lentiviruses from various species [1,8]. A variable N-terminal section of 20–50 residues is followed by a cysteine-rich region that in HIV-1 contains a dense accumulation of 7 cysteines within 16 residues (Fig. 2A). This segment ranging from residues 22–37 is thought to form an intermolecular zinc finger with cysteine 261 of human CycT1. While such intermolecular zinc finger would require three cysteines (or two cysteines and one histidine) of Tat, the remaining four cysteines may compose a second intramolecular zinc finger within Tat [9]. Next to the cysteine-rich sequence follows the core region that is best conserved in Tat proteins. A consensus motif of five residues (LGIxY) appears in all lentiviruses. The succeeding arginine-rich motif (ARM, residues 49–57 in HIV-1 Tat) constitutes the most peculiar region in Tat that is essential for TAR RNA binding. This region is also known as cell penetrating peptide, which can serve as a vehicle for the intracellular delivery of macromolecules by crossing the plasma membrane [10–12]. The highly basic ARM region encompasses at least nine residues as in HIV and EIAV but may be longer with up to 14 residues as in BIV. In contrast, the spacing between the core and the ARM region appears more variable in composition and length, which varies from 1 to 5 residues

* Corresponding author. Address: Max Planck Institute of Molecular Physiology, Department of Physical Biochemistry, Otto-Hahn-Strasse 11, D-44227 Dortmund, Germany. Fax: +49 231 133 2399.
E-mail address: matthias.geyer@mpi-dortmund.mpg.de (M. Geyer).

1046-2023/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2010.04.001

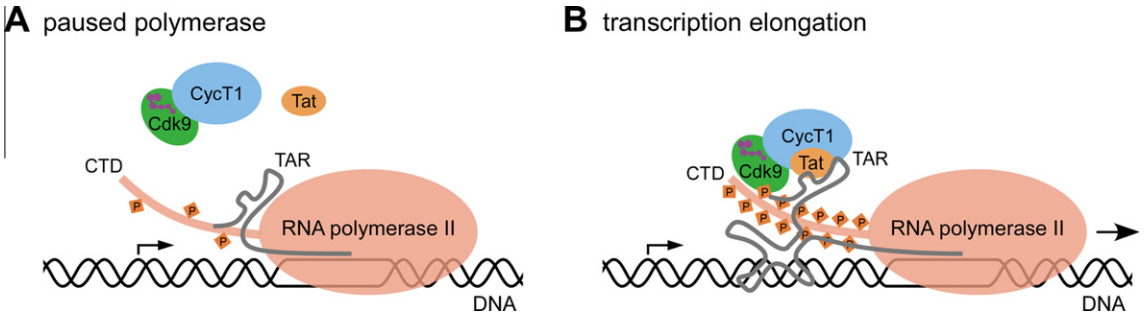

Fig. 1. Model of the Tat–TAR mediated transcription activation complex. (A) Cyclin T1 and Cdk9 (P-TEFb) are recruited by HIV-1 Tat to the paused RNA polymerase II. The nascent pre-mRNA element encodes the transactivation response (TAR) stem loop structure at the 5′-end of the viral transcript. (B) Cdk9 of active P-TEFb hyperphosphorylates the CTD of RNA pol II at position Ser2, which induces transcription elongation of the viral transcript.

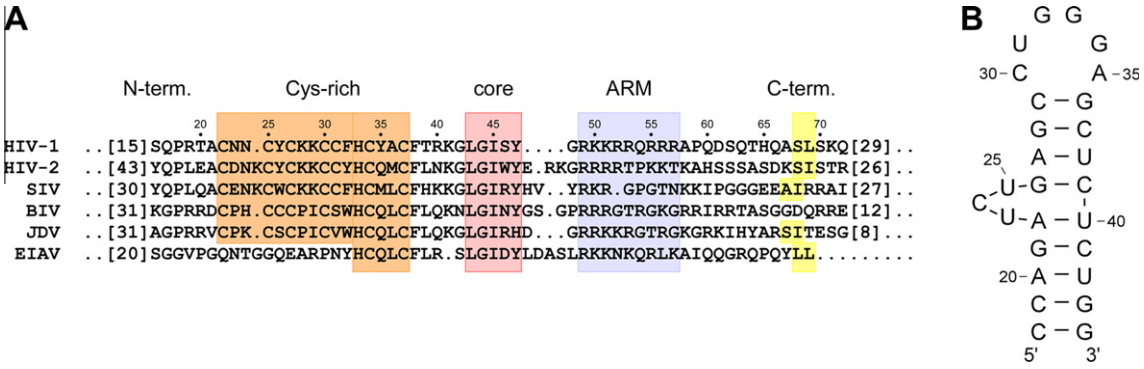

Fig. 2. Tat–TAR responsive elements in lentiviruses. (A) Sequence alignment of Tat proteins from HIV-1, HIV-2, SIV, BIV, JDV and EIAV virus alleles. The Tat protein sequences displayed correspond to HIV-1 SF2 (GenBank Accession No. AAB59879), HIV-2 ST (AAB01356), SIV mandrill (AAA47725), BIV isolate HXB3 (AAA91273), JDV (AAA64392) and EIAV Wyoming (AAG02703). The multiple sequence alignment was performed with the program "Multalign" and further hand-edited. The sequence numbering at the top corresponds to HIV-1 Tat. Conserved elements in the sequence composition of Tat are highlighted. (B) RNA hairpin loop structure of HIV-1 TAR (18–44). The pyrimidine-rich bulge at the 5′ strand is 4 bp apart from the loop structure.

in Tat proteins from different viruses (Fig. 2A). The sequence C-terminal to the ARM is generally more variable and contains up to 50 additional residues in the longest splicing variant of HIV-1. The shortest Tat variant, however, is represented by EIAV, whose additional 12 residues following the ARM region end with a C-terminal dileucine motif that represents another conserved unit among Tat proteins [13]. The length of this Tat ortholog is reminiscent to the shortest fully functional splicing variant of HIV-1 Tat, which encompasses residues 1–72 [8]. Previous studies of Tat alone showed that the protein is largely flexible in the absence of CycT1 or TAR [14,15].

TAR RNA was originally localized to nucleotides 1–80 in the 5′-end of the HIV-1 LTR, but subsequent deletion studies showed the region from nucleotide +19 to +43 to be the minimal responsive element that is both necessary and sufficient for Tat binding *in vivo* and *in vitro* [1–3]. HIV-1 TAR RNA contains a six-nucleotide hairpin loop and a three-nucleotide pyrimidine-rich bulge (UCU) that interrupts the continuity in the stem region of complementary base pairs (Fig. 2B). The bulge is located in the 5′ strand four nucleotide base pairs apart from the apical hairpin loop. The composition of the bulge as well as of the nucleotide pairs that form the RNA double strand, however, varies substantially in the LTR of different lentiviruses. While HIV-1 contains the beforehand mentioned UCU bulge, HIV-2 and JDV contain a two-nucleotide bulge (UU and AU, respectively), BIV has two single U inserts 4 and 5 bp apart from the hairpin loop and EIAV, finally, is a continuous double stranded hairpin loop without any bulge. All TAR RNAs appear to have a rather stable double strand stem structure, while the loop appears highly dynamic in the absence of a ligand [16,17].

Understanding the molecular and structural basis of the ribonucleoprotein complex formation between the viral Tat–TAR assembly and the host cell target protein CycT1 of P-TEFb is a key component not only for the analysis of viral transcription activation but also for the regulation of transcription elongation. Here, we describe molecular techniques that yield highly purified and homogeneous protein and RNA products suitable for functional studies *in vitro* and structure determination.

2. Protein and RNA purifications

2.1. Tat

The coding region of HIV-1 Tat (1–86) or HIV-1 Tat (1–72) was amplified by PCR with primers containing BamHI and EcoRI restriction sites at the 5′- and 3′-end, respectively. PCR products were cloned into the prokaryotic expression vector pGEX-4T-1 (GE Healthcare), modified with a TEV cleavage site to yield a fusion protein with an N-terminal glutathione-*S*-transferase (GST) tag. The resulting plasmid was transformed into *Escherichia coli* BL21(DE)3 cells (Novagen). When using an affinity tag, we found that GST is superior for Tat expression over hexahistidine, since the latter requires nickel ions for binding to the gel matrix. These divalent ions may interfere with the zinc clusters formed by cysteines of Tat. For protein expression cells were grown in LB-medium containing 100 μg/ml Ampicillin at 37 °C. The temperature was lowered to 30 °C upon reaching an optical density of 0.4 at 600 nm and cells were induced with 0.1 mM IPTG at an optical density of 0.6–0.8 at 600 nm.

After 4 h growth cells were harvested, washed in PBS and resuspended in buffer A (20 mM Tris/HCl, pH 7.4, 500 mM NaCl, 10% glycerol, 5 mM β-mercaptoethanol) supplemented with 1 mM