Volume 282, number 1, 157–160 FEBS 09643 April 1991

ADONIS 001457939100350M

Production of a full length Tat protein in *E. coli* and its purification

Jean Armengaud, Lidia de Nuova Perez*, Piere Lemay and Jean-Michel Masson

Institut National des Sciences Appliquées, Centre de Transfert en Biotechnologie et Microbiologie, Laboratoire de génie Biochimique, UA 544 du CNRS, Toulouse, France

Received 17 January 1991

A full length *tat* gene was constructed by a combination of polymerase chain reaction (PCR) for the first exon and chemical synthesis for the second exon. This gene was expressed in *E. coli* under the control of the strongly regulated *araB* promoter, either directly or fused to a secretion signal encoding sequence. We then defined a rapid, three-step procedure for the purification of the Tat protein

HIV; Tat; Regulatory protein; Gene expression; Purification

1. INTRODUCTION

Amongst the known retroviruses, HIV-1, the causative agent of AIDS, is one of the most complex with regard to its genetic organization (reviewed in [1]). At least 6 gene products are devoted to the regulatory pathway of the virus expression. One of these proteins, Tat, is a powerful transactivator which is highly conserved between various HIV-1 isolates [2–4].

The Tat protein is 86 amino acids in length and is encoded by two exons. Two Tat monomers assemble into a dimer upon captation of Zn^{2+} or Cd^{2+} ions to yield a functional product [5,6]. It has been shown that only the portion of Tat encoded by the first exon, which consists of 4 different short domains, is required for transactivation [7,8]. Thus, many studies dealing with Tat have been done with this 72 amino acids-long moiety of the Tat molecule [e.g. 9,10] and sometimes with even smaller fragments [11,12]. Only few studies on the gene expression or chemical synthesis of the 86 amino acids-long Tat protein have been reported [11,13–15]. The acquisition of such a full-length Tat product is of prime importance for physico-chemical analysis such as crystallography.

In this study, we report the construction, the expression and the characterization of a semi-synthetic full-length gene coding for the Tat protein in *E. coli* as well as a simple three-step method for the acquisition of pure Tat protein.

**Present address:* Centro de Ingeneria Genetica y Biotenologica, Havana, Cuba

Correspondence address: J.M. Masson, INSA, Centre de Transfert en Biotechnologie et Microbiologie, Complexe Scientifique de Rangueil, 31077 Toulouse Cedex, France. Fax: (33) (61) 55 95 00.

2. MATERIALS AND METHODS

2.1. *Bacterial strains and plasmids*

All construction steps were in the *E. coli* TG1 strain (*Δlac-pro supE thi* F' [*traD36lacI^q proAB lacZ* ΔM15]) [16]. For gene expression the vectors were transferred into the *E. coli* MC1061 strain (*araD139* Δ(*ara-leu*)$_{7696}$ *lacX74 galV galK hsr hsm rpsL*) [17]. Cells were grown in LB medium at 37°C.

For Tat production, cells were grown at 37°C, on LB medium with ampicillin, for one hour before induction (OD_{600} 0.2–0.3) with 0.2% arabinose (w/v). They were then harvested by centrifugation at the end of the exponential growth phase (after 7.5 h of culture) at an OD_{600} value of approximately 2.3.

The expression plasmids used in this work will be described in detail elsewhere (C. Cagnon, V Valverde and J.-M. Masson, in preparation). Vector pARA6 harbours both the *araC* repressor gene and the *araB* promotor from pING1 [18] (Ingene, Santa Monica, CA), fused to a synthetic sequence coding for an idealized *E. coli* signal peptide, upstream from a polylinker that allows for in-phase fusion of the foreign gene. Vector pARA13 harbours the same construct without the signal peptide encoding sequence. In this latter vector, the *araB* promoter has been mutated, in order to increase the expression level without losing the tight control by the Ara C repressor.

2.2. *Cloning and sequencing*

All cloning steps and restriction analysis were done according to Maniatis et al. [19]. DNA Sequencing was performed as described by Sanger et al. [20]. Oligonucleotides were synthesized on an Applied Biosystem 380A DNA synthesizer. They were purified on NENsorb columns (DuPont deNemours, Boston, MA) according to the manufacturer instructions and phosphorylated as described by Normanly et al. [21]. The polymerase chain reaction used the Klenow fragment of the DNA polymerase of *E. coli* [22].

2.3. *Analysis of protein samples*

Whole cells from 1 ml of culture (OD_{600} = 2.0) were washed once in TN buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl), then disrupted by boiling for 5 min in 50 μl of denaturation buffer (75 mM Tris-HCl, pH 6.8, 600 mM urea, 4% SDS, 10% β-mercaptoethanol, 0.01% Bromophenol blue) before being loaded on the electrophoresis gels.

In the purification steps, the amount of proteins in the samples were measured with the microBradford Kit from Bio-Rad (Richmond, CA). For electrophoresis, the samples were adjusted to 1 mg/ml of protein and diluted with an equal volume of denaturation buffer before boiling.

Samples were analyzed either on 8–25% gradient Phastgels (Pharmacia LKB Biotechnology, Uppsala, Sweden) and silver stained according to the manufacturer specifications, or on 15% polyacrylamide gels following the modification of Caillet-Boudin and Lemay [23] of the Laemmli procedure [24] and stained with Coomassie blue.

The amino-terminal sequence of the *tat* gene product was determined on electrophoretic transfers [25] of tricine-buffered acrylamide slab gels [26] on an Applied Biosystem 470A Protein sequencer coupled to an Applied 120A PTH Analyzer.

3. RESULTS AND DISCUSSION

3.1. *Gene synthesis and assembly*

The first exon of the *tat* gene was synthesized by 12 cycles of PCR, with the TAT1 and TAT2 oligonucleotides as primers and a cloned *Bgl*II fragment of the BH10 isolate as template. The resulting 232-bp DNA fragment is characterized by a *Nco*I site at the 5′ end and a *Sal*I site followed by a *Bgl*II site at the 3′ end (Fig. 1). It was cloned between the *Nco*I and the *Bgl*II sites of the expression vector pARA6.

The second exon was obtained through the annealing of oligonucleotides TAT3 and TAT4, and cloned between the *Sal*I site and the *Kpn*I site of the pARA6 vector already harbouring the first exon. The resulting vector harboured a full length *tat* gene between the *Nco*I and the *Kpn*I restriction sites, the *Sal*I site being destroyed in the process. As the pARA6 vector was designed for periplasmic expression of foreign proteins in *E. coli*, it has a small sequence, 5′ to the *Nco*I site that codes for an idealized signal peptide, in phase with the sequence cloned at this restriction site. In order to have also a cytoplasmic expression system, the full *tat* gene was cloned as a *Nco*I/*Kpn*I fragment into the pARA13 vector, which has the same features as the

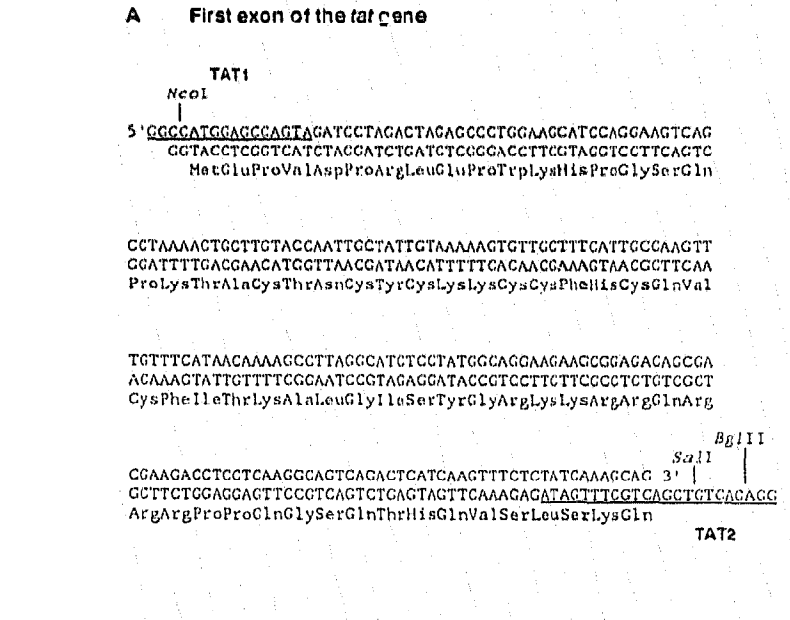

Fig. 1. Sequence of the full-length *tat* gene. Panel A displays the sequence of the first exon of the *tat* gene. The sequences of the oligonucleotides TAT1 and TAT2 used for the PCR amplification of the gene are underlined. The restriction sites introduced in the sequence for cloning purpose are indicated. Panel B presents the sequences of the two synthetic oligonucleotides TAT3 and TAT4 which were annealed to make the second exon of the *tat* gene. The assembly of the two gene fragments is described in the text.

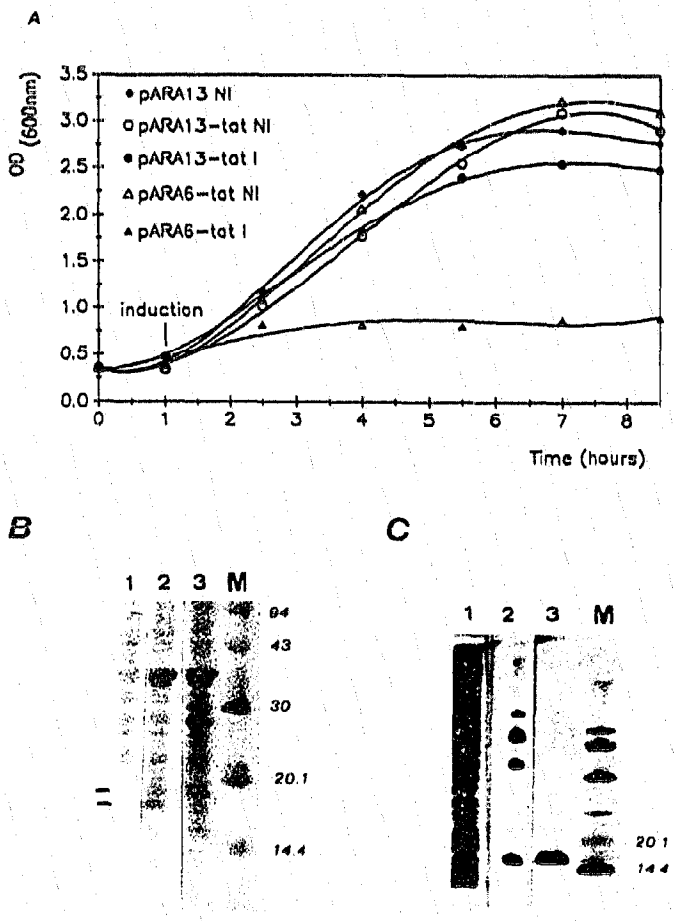

Fig. 2. Expression of the *tat* gene in *E. coli*. Panel A: kinetic ananlysis of the growth of induced (I) or non-induced (NI) MC 1061 cells transformed by pARA6-tat, pARA13-tat or pARA13. Panel B represents a 15% polyacrylamide gel electrophoresis analysis of whole bacterial cell extracts. Lane 1, *E. coli* MC1061[pARA13]; lane 2, *E. coli* MC1061[pARA13-tat]; lane 3, *E. coli* MC1061[pARA6-tat]. M: molecular mass markers (values expressed in kDa). The two bars indicate the difference of apparent molecular mass between the products from pARA13-tat (lane 2) and pARA6-tat (lane 3). Panel C: analysis of the purification steps of the tat gene product by electrophoresis on 8–25% SDS-gradient gels. Lane 1, whole cell extract from induced *E. coli* MC1061 [pARA13-tat]; lane 2; high salt extract (according to Aldovini et al. [13]); lane 3, pure Tat protein obtained after Superose 12 gel chromatography. M: molecular mass markers (values expressed in kDa).

pARA6 vector, save for the signal peptide coding sequence.

3.2. *Expression of the* tat *gene in* E. coli

Preliminary experiments performed in various bacterial strains demonstrated that the *E. coli* MC1061 strain was the most suitable with regard to the level of expression of the cloned gene (not shown). Growth curves of induced or uninduced MC1061 harbouring either pARA6-tat, pARA13-tat or pARA13 alone (as a control) are shown in Fig. 2A. It can be seen that the induction of the expression of *tat* from pARA6 leads to atypical growth of the cells thus suggesting that the production of Tat fused to a signal peptide is somewhat toxic. This is not the case for the cytoplasmic production obtained from pARA13.

Polyacrylamide slab gel analysis of whole bacterial cell extracts showed that the highest level of expression of Tat was obtained with the pARA13-tat construct (Fig. 2B, lane 2), with the induction protocol described in section 2. Moreover, the product expressed by the pARA6-tat plasmid had a higher apparent molecular mass than the Tat protein encoded by the pARA13 construct (Fig. 2B, lane 3). The analysis of various fractions originating from osmotic shocks of MC1061[pARA6-tat] induced cultures showed that (i) no Tat protein was recovered at the periplasmic level and (ii) the signal peptide is not processed, thus accounting for the higher apparent molecular mass. Therefore, further production of Tat was done with MC1061-[pARA13-tat].

Microsequencing of the amino terminus of the Tat protein produced with MC1061[pARA13-tat] and transferred onto immobilon membranes from the polyacrylamide gels, yielded the sequence Met-Pro-Glu-Val-Asp, indicating that the N-terminal methionine is not processed in *E. coli*.

3.3. *Purification of the Tat protein*

Our main goal was to set up a purification protocol that could be scaled-up easily in order to obtain the amount of Tat protein necessary for structural studies. The cells were grown and induced as described in Section 2. The bacterial cells were then treated either as described by Aldovini et al. [13], except that the centrifugation steps were carried out at 4400 × *g*; or as described by Frankel et al. [5]. In our hands, the first method yielded the best results, with Tat being the main protein in the 30% ammonium sulfate pellet (Fig. 2C, lane 2). The yield is very good, as 80-90% of the Tat protein present in the crude extract was recovered after this step (as estimated by densitometry scanning of the electrophoresis gels). This precipitate was then resuspended in water (final concentration 1.4 mg/ml of proteins) and directly applied to an FPLC Superose 12 column. The elution buffer was 20 mM Tris-HCl pH 7.5, 50 mM KCl, 2 mM EDTA. Tat eluted at 2.9 V_0, and was over 95% pure, as can be seen from Fig. 2C, lane 3.

3.4. *Discussion*

By a combination of PCR for the first exon and gene synthesis for the second exon, a sequence coding for a full-length Tat protein was easily assembled and expressed from a specialised vector in *E. coli*. The 86-residue protein is produced in the bacterial cytoplasm in a soluble form, while attempts at producing it in the periplasm failed, as the attached signal sequence was not processed and the protein was not exported. The protein was purified to almost homogeneity in a three-step procedure that is simple enough to allow for easy scale-up. Further biochemical and biological characterization of the protein is in progress.

Acknowledgements: This work was supported by grants from the Région Midi-Pyrénées and the Agence Nationale pour la Recherche contre le SIDA (ANRS). L. de N.P. was the recipient of a fellowship from the Ministère des Affaires Etrangères. We thank Dr. W. Wickner for the gift of plasmid pING1. The authors are indebted to Mrs. M. Segui for her help in protein sequencing.

REFERENCES

[1] Cullen, B.R. and Greene, W.C. (1990) Virology 178, 1-5.
[2] Arya, S.K., Guo, C. Joseph, S.F. and Wong-Staal, F. (1985) Science 229, 69-73.
[3] Fisher, A.G., Feinberg, M.B., Josephs, S.F., Harper, M.E., Marselle, L.M., Reyes, G., Gonda, M.A., Aldovini, A., Debouck, C., Gallo, R.C. and Wong-Staal, F. (1986) Nature 320, 367-371.
[4] Laspia, M.F., Rice, A.P. and Mathews, M.B. (1989) Cell 59, 283-292.
[5] Frankel, A.D., Bredt, D.S. and Pabo, C.O. (1988) Science 240, 70-73.
[6] Frankel, A.D., Chen, L. Cotter, R.J. and Pabo, C.O. (1988) Proc. Natl. Acad. Sci. USA 85, 6297-6300.
[7] Garcia, J.A., Harrich, D., Pearson, L., Misuyasu, R. and Gaynor, R.B. (1988) EMBO J. 7, 3143-3147.
[8] Seigel, L.J., Ratner, L., Derse, J.D., Feinberg, G.R., Feinberg, G.R., Obriez, S.J. and Wong-Staal, F. (1986) Virology 148, 226-231.
[9] Green, M., Ishino, M. and Lowenstein, P.M. (1989) Cell 58, 215-223.
[10] Jeang, K.-T., Shank, P.R., Rabson, A.B. and Kumar, A. (1988) J. Virol. 62, 3874-3878.
[11] Jeyapaul, J., Reddy, M.R. and Khan, S.A. (1990). Proc. Natl. Acad. Sci. USA 87, 7030-7034.
[12] Frankel, A.D., Bianclana, S. and Hudson, D. (1989) Proc. Natl. Acad. Sci. USA 86, 7397-7401.
[13] Aldovini, A., Debouck, C., Feinberg, M.B., Rosenberg, M., Arya, S.K. and Wong-Staal, F. (1986) Proc. Natl. Acad. Sci. USA 83, 6672-6676.
[14] Jakobovits, A., Smith, D.H., Jakobovits, E.B. and Capon, J.D. (1988) Mol. Cell. Biol. 8, 2555-2561.
[15] Chun, R., Glabe, C.G. and Fan, H. (1990) J. Virol. 64, 3074-3077.
[16] Gibson, T.J. (1984) Ph.D. Thesis, University of Cambridge, UK.
[17] Casadaban, M. and Cohen, S. (1980) J. Mol. Biol. 138, 179-207.
[18] Johnston, S., Lee, J.-H. and Ray, D.S. (1985) Gene 34, 137-145.

[19] Maniatis, T., Fritsh, E.F. and Sambrook, J. (1982) In Molecular Cloning. A laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[20] Sanger, F. Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[21] Normanly, J., Masson, J.-M., Kleina, L.G., Abelson, J. and Miller, J.H. (1986) Proc. Natl. Acad. Sci. (USA) 83, 6548–6552.
[22] Mullis, K.B., Faloona, F., Scharf, S.J., Saiki, R.K., Horn, G.T. and Herlich, H.A. (1986) Cold Spring Harbor Symp. Quant. Biol. 51, 263–273.
[23] Caillet-Boudin, M.-L. and Lemay, P. (1986) Electrophoresis 7, 309–316.
[24] Laemmli, U.K. (1970) Nature 227, 680–685.
[25] Plough, M., Jensen, A. and Barkholt, V. (1989) Anal. Biochem 181, 33–39.
[26] Schägger, H. and von Jagow, G. (1987) Anal. Biochem. 166, 368–379.